

RESEARCH ARTICLE

Open Access

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

Xiaobai Li<sup>1\*</sup>, Feng Jin<sup>2</sup>, Liang Jin<sup>1</sup>, Aaron Jackson<sup>3</sup>, Cheng Huang<sup>4</sup>, Kehu Li<sup>5</sup> and Xiaoli Shu<sup>6\*</sup>

## 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.

\* Correspondence: hufanfan1982815@gmail.com; shuxl@zju.edu.cn

<sup>1</sup>Zhejiang Academy of Agricultural Sciences, Shiqiao Road 139, Hangzhou 310021, Peoples Republic of China

<sup>6</sup>State Key Lab of Rice Biology, International Atomic Energy Agency Collaborating Center, Zhejiang University, Hangzhou 310029, Peoples Republic of China

Full list of author information is available at the end of the article

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 standing 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 *C. ensifolium* transcriptome, we obtained 55 new polymorphic micro-satellite 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<sup>®</sup> 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 Li's 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.

**Table 1 Fifty nine cymbidium accessions used for genetic analysis**

Accession	Name	Group <sup>a</sup>	Species
1	Tiegusu	4	<i>C. ensifolium</i>
2	Qingshanyuquan	4	<i>C. ensifolium</i>
3	Jinsimawei	4	<i>C. ensifolium</i>
4	Jinhe	4	<i>C. ensifolium</i>
5	Yinsimawei	4	<i>C. ensifolium</i>
6	Dayibai	4	<i>C. ensifolium</i>
7	Dahongzhusha	2	<i>C. ensifolium</i>
8	QiuHong	4	<i>C. ensifolium</i>
9	Baodao	4	<i>C. ensifolium</i>
10	Jinhe	2	<i>C. ensifolium</i>
11	Tianhe	4	<i>C. ensifolium</i>
12	Shisantaobao	4	<i>C. ensifolium</i>
13	TuerA	2	<i>C. lancifolium</i>
14	TuerB	2	<i>C. lancifolium</i>
17	DuohualanA	5	<i>C. floribundum</i>
18	GuoxianglanA	2	<i>C. suavissimum</i>
19	ShayelanA	1	<i>C. cyperifolium</i>
20	ShayelanB	1	<i>C. cyperifolium</i>
21	ShayelanC	1	<i>C. cyperifolium</i>
22	ShayelanD	1	<i>C. cyperifolium</i>
23	QiubeidonghuiA	2	<i>C. qiubeiense</i>
24	ShayelanE	1	<i>C. cyperifolium</i>
25	LvlanA	1	<i>C. faberi</i>
26	GuoxianglanB	5	<i>C. suavissimum</i>
27	lvlanB	1	<i>C. faberi</i>
28	DuohualanB	5	<i>C. floribundum</i>
29	Yuhudie	2	<i>C. goeringii</i>
30	Yinhe	5	<i>C. goeringii</i>
31	Silan	2	<i>C. goeringii</i>
32	Hexingmei	5	<i>C. goeringii</i>
33	Dasongmei	2	<i>C. goeringii</i>
34	Yipin	2	<i>C. goeringii</i>
35	Huangmei	2	<i>C. goeringii</i>
36	Puchunhong	2	<i>C. goeringii</i>
37	Chunjiansuxin	2	<i>C. goeringii</i>
38	Hongmeigui	2	<i>C. goeringii</i>
39	Wenyi	2	<i>C. goeringii</i>
40	Jiuxianmudan	2	<i>C. goeringii</i>
41	Dayipin	3	<i>C. faberi</i>
42	Ruyisu	2	<i>C. faberi</i>
43	Jiepeimei	3	<i>C. faberi</i>
44	Xinshanghai mei	3	<i>C. faberi</i>
45	Laoranzi	3	<i>C. faberi</i>

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

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

<sup>a</sup>Five groups indicated by population structure analysis.

### 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 (*Fst*), 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

were used for the subsequent assembly, and then resulted in 101,423 unigenes (139,385,689 residues). The length of unigenes averaged 1,374 bp and ranged from 351 bp to 17,260 bp. The data were uploaded to the NCBI (<http://orchidbase.itps.ncku.edu.tw/est/home2012.aspx>) for public use (Accession: SRA098864).

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. cyperifolium*, *C. qiubeiense*, *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 Group5 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. cyperifolium* and *C. goeringii*; Group 2 included *C. ensifolium*, *C. lancifolium*, *C. suavissimum*, *C. qiubeiense*, *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.

**Table 2 List of the 62 *C. ensifolium* genic-SSR primers including their unigenes annotation**

Name	Product size (bp)	SSR	SSR location	Primer	Homologs in non-redundant database (accession in Genebank)	GO annotation	KOG annotation	KEGG annotation	Allele number	PIC
SSR01	400-500	(AC)8	utr5	F: AACGCCATGTCCAATACCC R: GGAGGGCTTATTTGCAGCG	PREDICTED: probable transcription factor KAN2-like (XP_002278005.2)	GO: 0003677	KOG1601	NULL	5	0.552
SSR02	300-400	(AC)8	utr5	F: CTCCTTCAAGCTTCTGCC R: GACCGCAGCGTTAATGACC	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: CTCGGTTCATTTGCAGCCC R: GGGTGGGTATGGCGAAATC	PREDICTED: mitochondrial import receptor subunit TOM20 (XP_002269795.1)	GO: 0045040	NULL	NULL	7	0.690
SSR04	400-500	(AC)8	utr3	F: AGAATCTGCCAACCTTGATAC R: GCAGATGCCAGTTAGAATGGG	NULL	NULL	NULL	NULL	6	0.657
SSR05	1000	(AC)8	utr3	F: AGAACTGCAGGTGTGAAGC R: GGCTTGAAGTGGCGATAACC	PREDICTED: protein CbbY, chromosomal-like isoform 1 (XP_003574671.1)	GO: 0016787	NULL	NULL	3	0.125
SSR06	600	(AC)9	utr3	F: GCGTCTGCTGAAACGATGG R: AAACAGCGCCTGTCATTCC	Putative steroid 22-alpha-hydroxylase (AAN60994.1)	GO: 0016020	KOG0157	K09587	2	0.063
SSR07	300-400	(AC)9	utr3	F: ACGCTGCATCCCATTTAC R: CAGTCTGTTGAGGAAGCCG	PREDICTED: uncharacterized protein LOC100243361 (XP_002276849.2)	GO: 0008987	NULL	K03517	4	0.180
SSR08	100-200	(AC)10	utr3	F: TGCTGGAATACATGCGAGAC R: GTTGCCGAAGCCAGTGC	Predicted protein (XP_002298559.1)	GO: 0023014	KOG0610		14	0.753
SSR11	600	(AG)10	utr3	F: AACTGACAAGCATCTGCAAG R: CTGCTGCATTGGCCTTACC	Uncharacterized protein LOC100273319 precursor (NP_001141232.1)	GO: 0005774	NULL	NULL	6	0.477
SSR12	300	(AG)11	utr5	F: TCAGCCGAGGTTAGTATACGG R: CTTGCCATCTCAGCAGTCG	PREDICTED: phosphatidylinositol-4-phosphate 5-kinase 9-like (XP_002265706.1)	GO: 0016020	KOG0229	K00889	NA	NA
SSR13	400-500	(AG)11	utr5	F: GCTGCTGCTTGGTGGAAAC R: GCGCTCGTTGTATGGCTTG	Predicted protein (XP_002317724.1)	GO: 0005488	NULL	NULL	6	0.343
SSR14	300	(AG)11	utr5	F: CACAGCAGCTCACAATCCTG R: TACAGCCCTGTTTACCGCC	Unnamed protein product (CBI20568.3)	GO: 0006099	KOG1257	K00029	8	0.467
SSR15	100-200	(AG)11	utr3	F: CCTTCTCCCGTACCAG R: CTTGCGTTGGCGTTTAGGG	PREDICTED: uncharacterized protein LOC100825549 (XP_003558805.1)	GO: 0005783	NULL	NULL	4	0.339
SSR16	300-400	(AG)11	utr5	F: GCCCACAGCAATCCATCTG R: GCAGTCGAAGAAACCGTGG	PE repeat family protein (XP_003014087.1)	NULL	NULL	NULL	7	0.348
SSR17	400	(AG)11	utr5	F: GGATCACCAACAGCATGGG R: TCCACCAAGAGCAAGGATG	Transcription factor (ADG57844.1)	GO: 0003677	NULL	K09060	4	0.417



**Table 2 List of the 62 *C. ensifolium* genic-SSR primers including their unigenes annotation (Continued)**

SSR18	300	(AG)11	utr5	F: TGAACGGTTGGCTCTAGTTC R: AGCAAGCACTGACCTGAAAC	Conserved hypothetical protein (XP_002527260.1)	NULL	NULL	NULL	13	0.519
SSR21	300-500	(GT)8	utr3	F: TGGGCGACAGATCGAGTTC R: ACATGGACCACAGCATTCC	Hypothetical protein OsJ_08996 (EAZ25197.1)	NULL	NULL	NULL	15	0.794
SSR22	200-300	(GT)9	utr3	F: TATGCGTCTCTCCAACCG R: AAGCTAGTGGCCTTTGGTG	14-3-3-like protein B-like (ACQ45020.1)	GO: 0019904	KOG0841	K06630	10	0.572
SSR23	100-200	(GT)10	utr3	F: CGGCGATCGATTTATGAGCC R: CGATACTCCTCAATGTCGTGG	PREDICTED: beta-amylase 1, chloroplastic isoform 1 (XP_002285569.1)	GO: 0005634	NULL	K01177	NA	NA
SSR24	200-300	(GT)11	utr5	F: TCGGTAACCTGTTGCAAGG R: ACCTGTGAAGCTACCAGAC	PREDICTED: flavin-containing monooxygenase YUCCA6-like (XP_003550114.1)	GO: 0050661	NULL	K11816	2	0.063
SSR25	100-250	(GT)11	utr3	F: GAATCTCTCGCACCCGAAG R: TGGACAACATCAAGTGACGC	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B, putative (XP_002528338.1)	GO: 0006536	NULL	K02434	NA	NA
SSR26	100-250	(AAG)7	utr3	F: GCTTTATGCGACATCTGCG R: CGTCGGTTCCATGCACATC	Unnamed protein product (CBI25980.3)	GO: 0005634	KOG1901	NULL	11	0.638
SSR27	500-600	(AGC)5	utr3	F: CTGCCTTACAGCTAATGCC R: GCATGCTTGGACGCTGAAC	Os04g0512400 (NP_001053298.1)	GO: 0046872	NULL	NULL	3	0.313
SSR29	200-300	(AGC)6	utr3	F: AGCAAACGGCAAGTCATGG R: ATTCTGACTACCAGCCGGAC	RING finger protein 113A, putative (XP_002522169.1)	GO: 0016020	NULL	K13127	2	0.033
SSR30	200-300	(AGG)5	utr3	F: AAACGAAGGGCTGGAAGTC R: TTTGACATCGGGAAGTGGC	NULL	NULL	NULL	NULL	9	0.486
SSR31	100-200	(AGG)5	utr5	F: GGGATGCATAGACCTTTCGC R: CAGGTTCAACGGCATCGTG	Protein MSF1, putative (XP_002535293.1)	GO: 0005739	KOG3336	NULL	2	0.033
SSR32	1000-1100	(AGG)5	utr3	F: CTCCGGCCTCTGGTACTC R: AGTGATGAGGCTTGGACCG	PREDICTED: HVA22-like protein j (XP_002281038.1)	NULL	KOG1726	NULL	7	0.601
SSR34	700-900	(AGG)6	utr5	F: GAGAGGGAATTGCAGTGCC R: ACCGAGCTAGCACTTCATC	Hypothetical protein (BAI68347.1)	NULL	NULL	NULL	6	0.696
SSR35	700-900	(ATC)5	utr5	F: AGAGTGATTGTCCAGCTCCG R: TGCCTCTCTGGTGATGTC	PREDICTED: diacylglycerol kinase-like (XP_003534537.1)	GO: 0009395	NULL	K07029	4	0.475
SSR36	400-500	(ATC)5	utr3	F: AGTATTGGACCCTCCAGGC R: AGAGGATCATGGTGTAGGC	NULL	NULL	NULL	NULL	5	0.536
SSR37	200-300	(ATC)5	utr5	F: GGCCTAGCCAGCCCTTC R: ATTTGGATCGACAAGCGG	NULL	NULL	NULL	NULL	3	0.205

**Table 2 List of the 62 *C. ensifolium* genic-SSR primers including their unigenes annotation (Continued)**

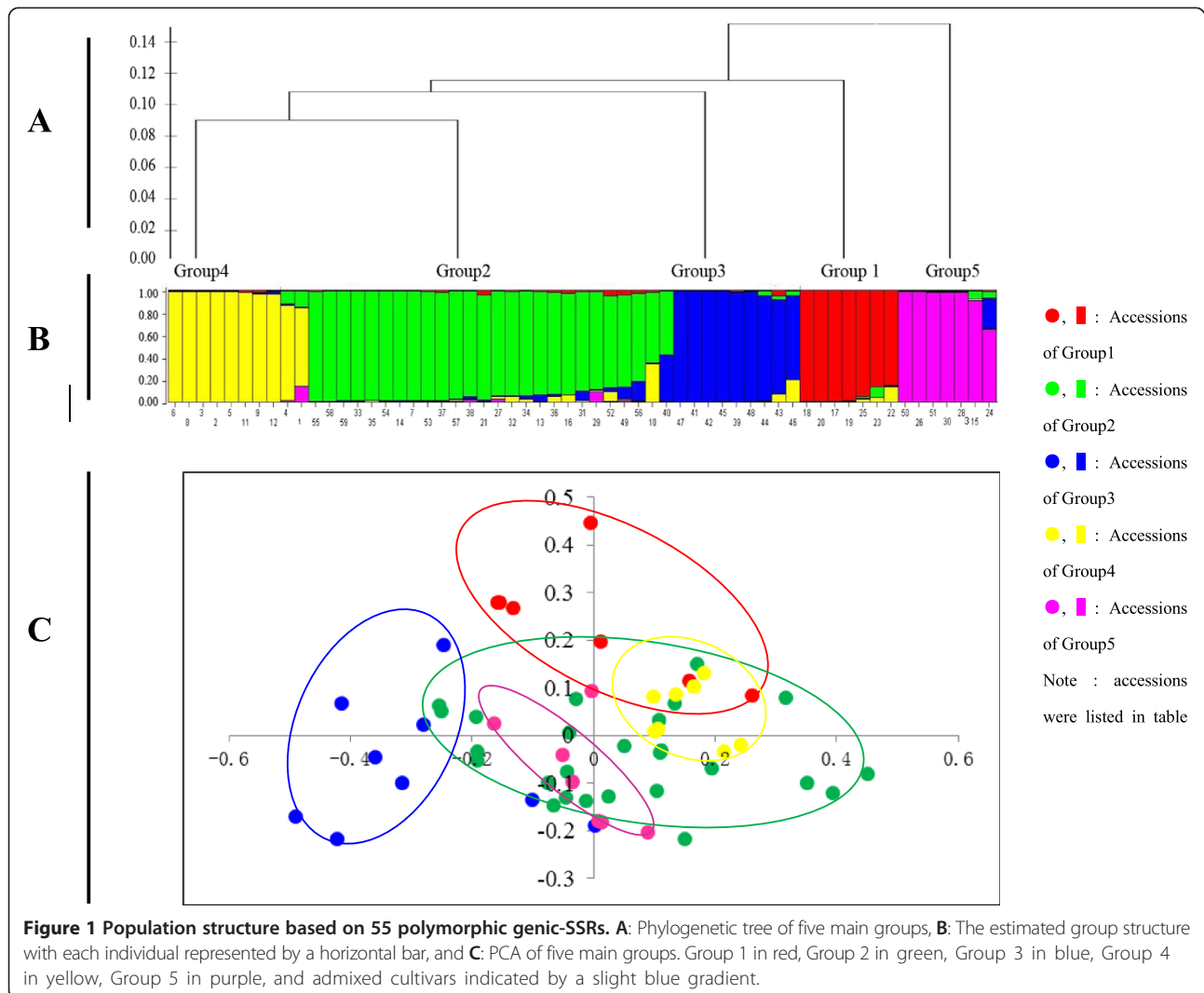
SSR38	200-300	(ATC) <sub>6</sub>	utr3	F: TAGCCCATGCCAGTGTCC R: AACTGCCACAAGAGAAGGC	LOC100285373 (NP_001151738.1)	GO: 0007165	NULL	NULL	3	0.149
SSR39	1000-1100	(ATC) <sub>6</sub>	utr3	F: ACAGACTGCCACTGTTC R: GCCTGCCTTTGCTCCTTG	unnamed protein product (CBI38283.3)	GO: 0008234	KOG1870	K11835	5	0.401
SSR40	400	(ATC) <sub>6</sub>	utr5	F: ACAAGCATCATCCCAAATTCC R: GCAGAAACTGGAGCTTGCC	PREDICTED: probably inactive leucine-rich repeat receptor-like protein kinase At2g25790-like (XP_002267653.1)	GO: 0007165	NULL	NULL	NA	NA
SSR42	200-300	(CCG) <sub>5</sub>	utr5	F: GACGACATATCGCGTTCGG R: CTCAGCCACACCAAGAGG	unnamed protein product (CBI18667.3)	GO: 0003779	KOG0160	NULL	6	0.564
SSR43	500	(CCG) <sub>5</sub>	utr5	F: GGAGCTGCATACGCAAGTG R: AGCTTCTCACTGCCTCCAG	glycinebetaine/proline transporter (BAJ07206.1)	GO: 0015193	NULL	NULL	7	0.572
SSR44	300-400	(CCG) <sub>5</sub>	utr5	F: CGTCGACTCCTCGAGATCC R: GCGTTAGCAGCAGTCTTGG	predicted protein (BAJ93650.1)	GO: 0046872	NULL	NULL	NA	NA
SSR45	400-500	(CCG) <sub>5</sub>	utr5	F: GCCTTACACATCCCTTCCAAC R: TGCCTGCTGATAGTTTGCC	unnamed protein product (CBI33381.3)	GO: 0005515	KOG0550	NULL	5	0.338
SSR46	200-300	(CCG) <sub>6</sub>	utr5	F: CCTTCGTGGACTCAACAGC R: TCTCGTGCAAGGATCGGTC	hypothetical protein SORBIDRAFT_01g031510 (XP_002465065.1)	NULL	NULL	NULL	2	0.063
SSR47	400-500	(CCG) <sub>6</sub>	utr3	F: GCAGGTGCTCCTCATCGGAG R: CTCCGGCTAACTCCATCCC	CONSTANS-like protein (ADN97077.1)	GO: 0005622	KOG1601	NULL	NA	NA
SSR49	300	(CCG) <sub>7</sub>	utr3	F: AGAGGGCCACCTGCTTTC R: GCCAATTGCCAGATGGACG	predicted protein (XP_002312577.1)	NULL	KOG1863	NULL	6	0.549
SSR52	400-500	(CCT) <sub>4</sub>	utr5	F: AAGAGGCACTGCAAGACCC R: CGTTCCAGCAACCCATAGC	hypothetical protein SORBIDRAFT_01g031070 (XP_002465040.1)	NULL	NULL	NULL	8	0.378
SSR53	100-200	(CCT) <sub>4</sub>	utr5	F: GCTGAAGGTTCCGGTCTCTC R: TCCGCCTCTTAAAGCCGAC	PREDICTED: uncharacterized protein LOC100830480 (XP_003580351.1)	NULL	NULL	NULL	8	0.700
SSR54	200-300	(CCT) <sub>4</sub>	utr5	F: ATCTTCCCTCCACATCGGC R: TGGAGAAGAGTCGACCAGC	hypothetical protein MTR_1g083540 (XP_003591171.1)	GO: 0005886	NULL	NULL	5	0.422
SSR55	200-300	(CCT) <sub>4</sub>	utr5	F: TGGAATGGTTCTAGGGCTTC R: CCACTGGTACCCTCCTTGG	hypothetical protein (CCA65980.1)	NULL	NULL	NULL	2	0.323
SSR56	900-1000	(CCT) <sub>5</sub>	utr5	F: TGCTTCATTGTTGAGGCG R: AGTGGACGGAGAGTCAAGC	predicted protein (XP_002324427.1)	GO: 0008643	KOG0254	NULL	5	0.315
SSR59	200-300	(CGG) <sub>5</sub>	utr3	F: GTTCCAACGGTCAGCTCG R: GTGATGTGGTAGCATCGCC	leucine-rich repeat transmembrane protein kinase family protein (NP_177007.1)	GO: 0005524	NULL	NULL	3	0.432

**Table 2 List of the 62 *C. ensifolium* genic-SSR primers including their unigenes annotation (Continued)**

SSR60	200-300	(CGG)5	utr5	F: TACGGTTTCGACCAGCCTC R: CCATGCAGATCGGGCAAAG	Unnamed protein product (CBI41056.3)	GO: 0005634	KOG0265	K10143	4	0.203
SSR62	300-400	(CGG)6	utr5	F: GGTGGTTAGACCAGCTCC R: TCCTCAAGGCAAAGCTCCC	Hypothetical protein OsI_29809 (EAZ07555.1)	GO: 0005634	NULL	NULL	6	0.570
SSR63	100-200	(CGG)6	utr5	F: CTTCTCCACCTGGATCGC R: CTGCCGATCAATCCGAGAC	Uncharacterized protein LOC100277474(NP_001144494.1)	GO: 0008270	NULL	NULL	4	0.308
SSR64	400-500	(CGG)6	utr3	F: CGCTCAAAGAGATGGCACG R: TAGTACGGCGCTGCTTGAG	Os01g0226200 (NP_001042462.1)	NULL	NULL	NULL	11	0.627
SSR66	300-400	(CGG)7	utr3	F: CATCTTCTTGCCCGATGC R: CCCGCCAAATTCGAGACC	PREDICTED: pentatricopeptide repeat-containing protein At5g42310, mitochondrial (XP_002272226.1)	NULL	KOG4197	NULL	4	0.126
SSR68	100-200	(GAT)5	utr3	F: CCAGATCGAATGGCTACGC R: CAAGGAGCTCGTCAAGG	Hypothetical protein VITISV_010525 (CAN79523.1)	GO: 0003723	NULL	NULL	4	0.211
SSR69	200-300	(GAT)5	utr5	F: GTTTAGGCTAGCAGTGCGG R: TGAGAACGTAGTGAAGTTGCC	NULL	NULL	NULL	NULL	3	0.149
SSR70	200-300	(GAT)7	utr3	F: CCCAACGCAGAACGATAGC R: CGGTGGCACAAATGGAACG	NULL	NULL	NULL	NULL	5	0.529
SSR71	400-500	(GGT)5	utr5	F: GCATCGAAACCACTGTCGC R: CCCTAGCCGGAGTCTCAAC	Hypothetical protein SORBIDRAFT_09g018170 (XP_002439663.1)	NULL	NULL	NULL	2	0.262
SSR73	100-200	(GGT)5	utr3	F: GGACACAATGGAGACGAAGG R: TGCATGAAACCACATGGC	T4.15 (CCH50976.1)	GO: 0044238	NULL	NULL	16	0.863
SSR75	400-500	(GTT)6	utr3	F: GCCTTTGACCATCCGTGC R: GGCCGCCATGAGTAAGAAC	Mitogen-activated protein kinase 1 (AEQ28763.1)	GO: 0043622	KOG0660	K04371	2	0.118
SSR76	500-600	(GTT)6	utr5	F: AGACAGAGAGTCCCTAAAGGC R: CAGGGATGTTAAGTGGGCTG	NULL	NULL	NULL	NULL	7	0.519
SSR77	300-400	(GTT)6	utr3	F: TTTGTGGCAGTGGAAAGCG R: TGATACCAATGGCAAGGCG	NULL	NULL	NULL	NULL	5	0.470
SSR79	200-300	(GTT)6	utr5	F: AGGATTCATGTAGCCGACCTC R: TCCCTGAAGGAGGCAAACC	Hypothetical protein OsI_35425 (EEC67831.1)	NULL	NULL	K10728	2	0.207
SSR80	400-500	(GTT)7	utr3	F: GCACCCAGCTTGTGAGG R: CCCATACATTACAGGCAAGC	NULL	NULL	NULL	NULL	8	0.626

Note: A total of 62 genic-SSR markers successfully amplified were listed, however 55 polymorphic markers were used in subsequent population analysis or cross species comparison. NULL: no annotation. NA: monomorphic marker.





## Discussion

### Diversity

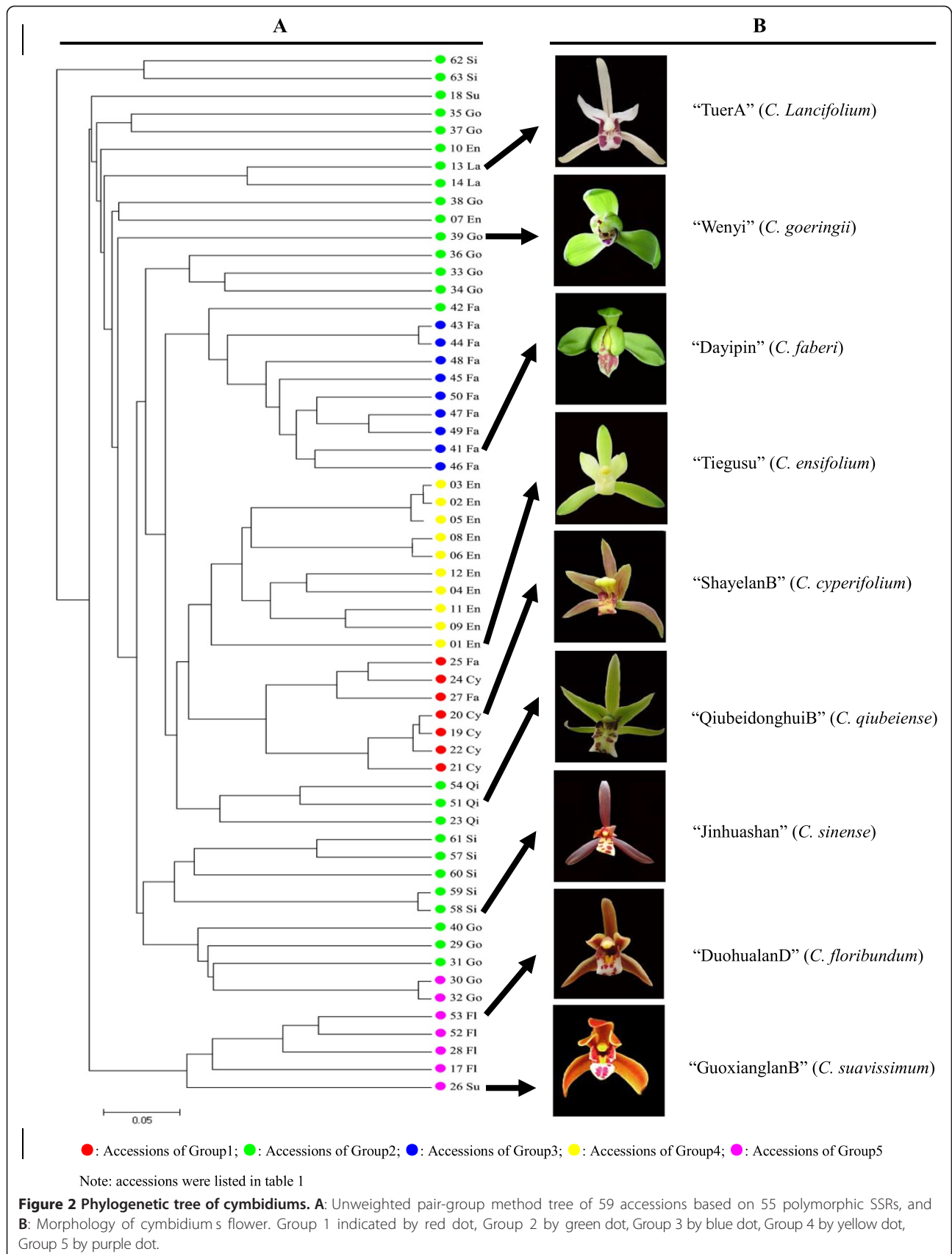
Because genic-SSR markers are derived from transcribed regions of DNA, they are expected to be more conserved and have a higher rate of transferability than anonymous SSR markers [38]. Herein, 55 *C. ensifolium* polymorphic genic-SSR markers exhibited 100% transferability across the 59 accessions of the 9 *Cymbidium* species tested. It

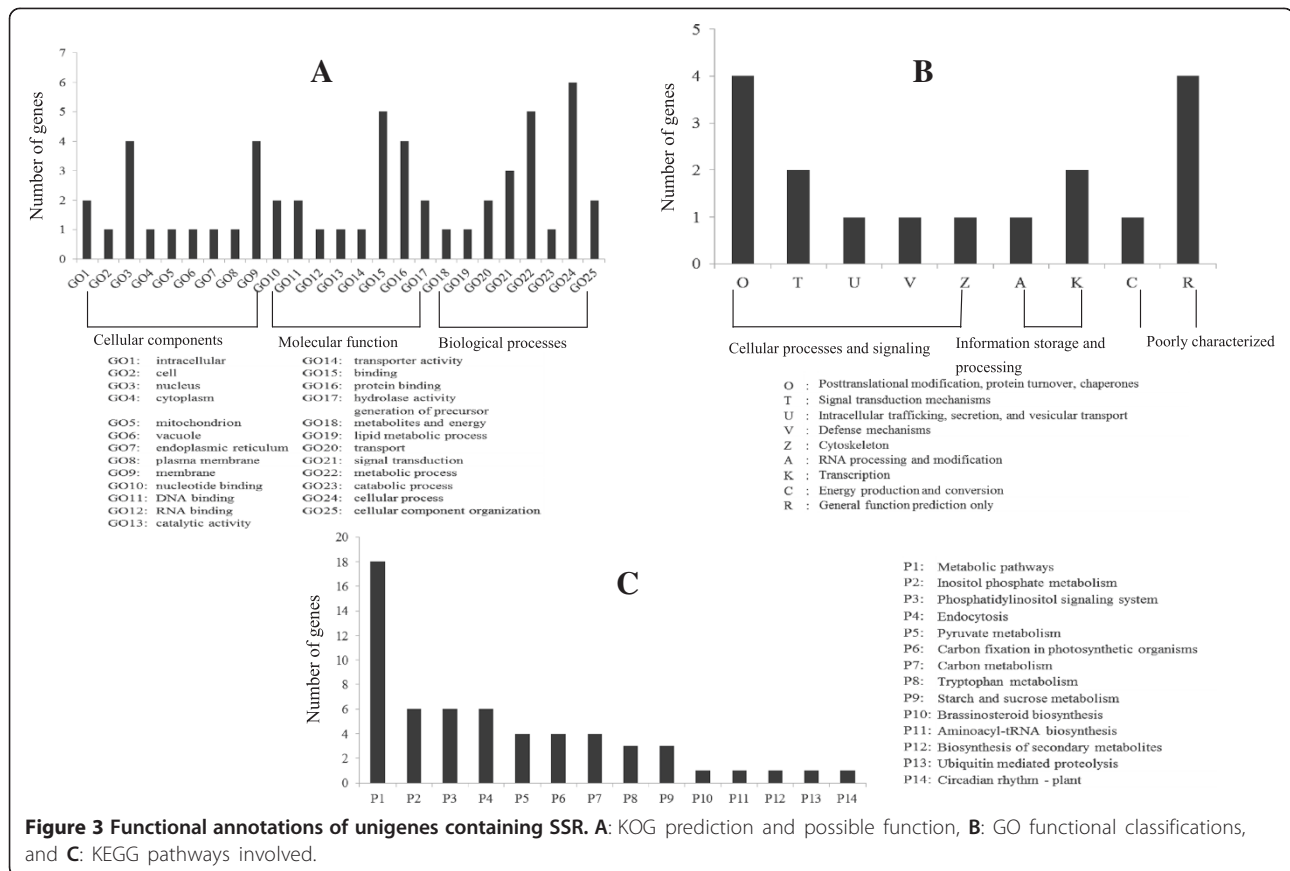
is common that genic-SSRs possess a high potential for inter-specific transferability [39,40]. Other markers such as RAPDs, ISSRs and non-genic SSRs have also been used with success among *C. ensifolium* and the *Cymbidium* species reflecting the genetic similarity among many members of the genus [8,11,15].

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





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. suavisimum* 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

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 *Maxillarianthe*; *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 gene-rich 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 experiment and performed the statistical analysis. XL, and XS conceived 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 Dianxing Wu for materials supply and technical support, and Chongbo 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 M521203), the Special Postdoctoral Science Foundation of China (No. 2013 T60607), and the Foundation for Selected Postdoctoral project of Zhejiang (Bsh1201032), the Qianjiang talents project (No. 2013R10081), and Scientific and technical innovation promotion project of ZAAS (2012R05Y01E04).

### Author details

<sup>1</sup>Zhejiang Academy of Agricultural Sciences, Shiqiao Road 139, Hangzhou 310021, People's Republic of China. <sup>2</sup>Hubei University, College of Life Sciences, Wuhan 430062, People's Republic of China. <sup>3</sup>USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas 72160, USA. <sup>4</sup>Agricultural Technology Extension Stations, Shaoxing County Agricultural Bureau, Shaoxing 312000, Peoples Republic of China. <sup>5</sup>School of Biological Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, SAR, People's Republic of China. <sup>6</sup>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 Yang Institute Technol* 2005, **2**:1-14.
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 Hortic* 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 YH, Chung JW, Cho YI, Park P, Park HS, Chae SC, Park YJ: Development of SSR markers to study diversity in the genus *Cymbidium*. *Biochem Syst Ecol* 2010, **38**:585-594.

12. Venkateswarlu M, Raje Urs S, Surendra Nath B, Shashidhar HE, Maheswaran M, Veeraiiah TM, Sabitha MG: **A first genetic linkage map of mulberry (*Morus* spp.) using RAPD, ISSR, and SSR markers and pseudotestcross 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, **15**:205–219.
14. Ri R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Besnard G, Bervill A, Martin A, Baldoni L: **A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers.** *Theor Appl Genet* 2003, **106**:1273–1282.
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, **5**:35.
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, Yan T, Xiang L, Jin F, Qin D, Sun C, Xie M: **Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome.** *PLoS One* 2013, **8**(12):e85480.
18. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muccicelli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol* 2011, **29**:644–652.
19. Faircloth BC: **MsaCommander: detection of microsatellite repeat arrays and automated, locus-specific primer design.** *Mol Ecol Resour* 2008, **8**:92–94.
20. Rozen S, Skaletsky H: **From methods in molecular Biology: Primer3 on the www for general users and for biologist programmers.** In *Bioinformatics Methods and Protocols volume 132*. Edited by Misener S, Krawetz SA. New Jersey: Humana Press; 2000:365–385.
21. 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.
22. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M: **KEGG for integration and interpretation of large-scale molecular datasets.** *Nucleic Acids Res* 2012, **40**:109–114.
23. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 2000, **28**:27–30.
24. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes.** *Genome Biol* 2004, **5**:R7.
25. Li XB, Zhang ML, Cui HR: **Data mining for SSRs in ESTs and development of EST-SSR marker in oilseed rape.** *J Mol Cell Biol* 2007, **40**:137–144.
26. Li X, 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 PI96354 conditioning soybean resistance to southern root-knot nematode.** *Theor Appl Genet* 2001, **103**:1167–1173.
28. Nei M, Takezaki N: **Estimation of genetic distances and phylogenetic trees from DNA analysis.** In *Proc 5th World Cong Genet Appl Livstock Prod*. 1983:405–412.
29. Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis.** *Bioinformatics* 2005, **21**:2128–2129.
30. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596–1599.
31. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945–959.
32. Li X, Yan W, Agrama H, Hu B, Jia L, Jia M, Jackson A, Moldenhauer K, McClung A, Wu D: **Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection.** *Genetica* 2010, **138**:1221–1230.
33. Rohlf FJ: *NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1*. 2000.
34. Li X, Yan W, Agrama H, Hu B, Jia L, Shen X, Jackson A, Moldenhauer K, Yeater K, McClung A, Wu D: **Mapping QTLs for improving grain yield using the USDA rice mini-core collection.** *Planta* 2011, **234**:347–361.
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. Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y: **Simple sequence repeats in *Escherichia coli*: abundance distribution, composition, and polymorphism.** *Genome Res* 2000, **10**:62–71.
37. Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P: **Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences.** *Am J Bot* 2012, **99**(2):193–208.
38. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ: **Analysis of SSRs derived from grape ESTs.** *TAG Theor Appl Genet* 2000, **100**(5):723–726.
39. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**:411–422.
40. Gupta PK, 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.
41. Huang Y, Li F, Chen K: **Analysis of diversity and relationships among Chinese orchid cultivars using EST-SSR markers.** *Biochem Syst Ecol* 2010, **38**:93–102.
42. Wang HZ, Wu ZX, Lu JJ, Shi NN, Zhao Y, Zhang ZT, Liu JJ: **Molecular diversity and relationships among *Cymbidium goeringii* cultivars based on inter-simple sequence repeat (ISSR) markers.** *Genetica* 2009, **136**:391–399.
43. Wang HZ, Wang YD, Zhou XY, Ying QC, Zheng KL: **Analysis of genetic diversity of 14 species of *Cymbidium* based on RAPDs and AFLPs.** *Acta Biologicae Exp Sinica* 2004, **37**:482–486.
44. Zhang JX, Li ZL, Fan CM, Cheng SL, Zhao MF, He YQ: **Phylogenetic analysis of the Chinese orchids in Yunnan province using AFLP technique.** *Acta Horticulturae Sinica* 2006, **33**:1141–1144.
45. Woodhead M, Russell J, Squirrel J: **Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions.** *Mol Ecol* 2005, **14**:1681–1695.
46. Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejarival A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferreira S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M: **Inferring non-neutral evolution from human-chimp-mouse orthologous gene trios.** *Science* 2003, **302**:1960–1963.
47. Tiffin P, Hahn MW: **Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. pekinensis.** *J Mol Evol* 2002, **54**:746–753.
48. Kimura M: **The neutral theory of molecular evolution: a review of recent evidence.** *Jpn J Genet* 1991, **66**(4):367–386.
49. Nei M: **Selectionism and neutralism in molecular evolution.** *Mol Biol Evol* 2005, **22**(12):2318–2342.
50. Griffiths S, Dunford RP, Coupland G, Laurie DA: **The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis.** *Plant Physiol* 2003, **131**(4):1855–1867.
51. Fujiwara T, Mitsuya S, Miyake H, Hattori T, Takabe T: **Characterization of a novel glycinebetaine/proline transporter gene expressed in the mesocotyle sheath and lateral root cap cells in barley.** *Planta* 2010, **232**(1):133–143.
52. Grallath S, Weimar T, Meyer A, Gumy C, Suter-Grottemeyer M, Neuhaus JM, Rentsch D: **The AtProT family. Compatible solute transporters with similar substrate specificity but differential expression patterns.** *Plant Physiol* 2005, **137**:117–126.
53. Schwacke R, Grallath S, Breitzkreuz KE, Stransky E, Stransky H, Frommer WB, Rentsch D: **LeProT1, a transporter for proline, glycine betaine, and gamma-amino butyric acid in tomato pollen.** *Plant Cell* 1999, **11**:377–392.
54. Igarashi Y, Yoshida Y, Takeshita T, Nomura S, Otomo J, Yamaguchi-Shinozaki K, Shinozaki K: **Molecular cloning and characterization of a cDNA encoding proline transporter in rice.** *Plant Cell Physiol* 2000, **41**(6):750–756.
55. Ueda A, Shi W, Sanmiya K, Shono M, Takabe T: **Functional analysis of salt-inducible proline transporter of barley roots.** *Plant Cell Physiol* 2001, **42**(11):1282–1289.



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.

doi:10.1186/s12863-014-0124-5

**Cite this article as:** Li *et al.*: Development of *Cymbidium ensifolium* genic-SSR markers and their utility in genetic diversity and population structure analysis in cymbidiums. *BMC Genetics* 2014 **15**:124.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

