Songklanakarin J. Sci. Technol.
35 (2), 149-158, Mar. - Apr. 2013

SJST SONGKLANAKARIN JOURNAL OF SCIENCE AND TECHNOLOGY
http://www.sjst.psu.ac.th

Original Article

Development of EST-derived markers in *Dendrobium* from EST of related taxa

Narisa Juejun[1, 2], Chataporn Chunwongse[1], and Julapark Chunwongse[1, 2, 3]*

[1] *Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, 73140 Thailand.*

[2] *Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok, 10900 Thailand.*

[3] *Department of Horticulture, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, 73140 Thailand.*

Received 28 June 2012; Accepted 3 February 2013

Abstract

Public databases are useful for molecular marker development. The major aim of this study was to develop expressed sequence tag (EST)-derived markers in *Dendrobium* from available ESTs of *Phalaenopsis* and *Dendrobium.* A total of 6063 sequences were screened for simple sequence repeats (SSRs) and introns. Primers flanking these regions were generated and tested on genomic DNAs of *Phalaenopsis* and *Dendrobium*. Twenty-three percent of amplifiable *Phalaenopsis* EST-derived markers were cross-genera transferable to *Dendrobium*. Forty-one markers from both *Phalaenopsis* and *Dendrobium* that amplified in *Dendrobium* were assessed on six commercial cultivars and six wild accessions. All of them were transferable among *Dendrobium* species. High polymorphism and heterozygosity were observed within wild accessions. Sixteen polymorphic markers were evaluated for linkage analysis on an F_1 segregating population. Seven markers were mapped into three linkage groups, two of which showed syntenic relationship between dendrobium and rice. This relationship will facilitate further quantitative trait loci (QTL) mapping and comparative genomic studies of *Dendrobium*. Our results indicate that *Phalaenopsis* EST-derived markers are valuable tools for genetic research and breeding applications in *Dendrobium*.

Keywords: EST-derived marker, EST data analysis, *Dendrobium*, *Phalaenopsis*, synteny

1. Introduction

Dendrobium, one of the largest genera in the family Orchidaceae, comprises more than 1000 species (Kamemoto *et al.*, 1999). Dendrobium flowers vary a great deal in form, size, color and fragrance. Because of their attractive characteristics, dendrobium orchids are one of the most popular ornamental plants worldwide. Large-scale cultivation of dendrobium orchids for distributions as cut-flowers and potted plants is commonly practised in many tropical countries (Hew and Yong, 2004). Considerable efforts have been put into breeding new and improved commercial cultivars. However, the relatively long life cycles and the lack of knowledge regarding the genetics of most characters have hampered the progress of dendrobium breeding programs. Molecular technology has proven to be a useful tool for both researchers and breeders (Agarwal *et al.*, 2008, Collard and Mackill, 2008, Gupta *et al.*, 1999). Molecular markers for important traits, combined with genetic linkage maps, would allow effective marker-assisted selection (MAS) which would help increase the speed and accuracy of dendrobium breeding programs. To be practically useful in a breeding program, markers should be inexpensive, reproducible, transferable, and predictive of phenotype.

Various types of dominant molecular markers such as random amplified polymorphic DNA (RAPD) (Wang *et al.*, 2006), amplified fragment length polymorphism (AFLP) (Xiang *et al.*, 2003) and inter-simple sequence repeat (ISSR)

* Corresponding author.
Email address: julapark.c@ku.ac.th

(Shen *et al.*, 2006) have been used for dendrobium germplasm identifications and genetic diversity analyses. In recent years, a more informative co-dominant type of marker, i.e., simple sequence repeat (SSR) has been developed in dendrobium (Yue *et al.*, 2006, Gu *et al.*, 2007, Fan *et al.*, 2008). SSR markers are highly polymorphic, highly reproducible, and readily transferable to other populations within the same species (Gonzola *et al.*, 2005) and to related species and genus (Dong *et al.*, 2011). Despite their benefits, only a limited number of dendrobium SSR markers have been published so far. The drawbacks of developing SSR markers are the cost, time and labor involved in cloning and sequencing. Another type of marker, i.e. expressed sequence tag (EST)-derived marker, can be developed using the existing EST databases, hence they are less expensive. This type of marker is currently the most promising technology for marker development in *Dendrobium*.

Over the years, exploitation of the increasing availability of public databases has become an efficient way to generate new polymerase chain reaction (PCR)-based markers. To develop EST-derived markers, specific primer pairs are designed to flank the SSR regions (EST-SSR markers) or the intronic regions (intron-flanking EST markers) of the ESTs. The developments of EST-derived markers have been reported in many plants including potato (Milbourne *et al.*, 1998), grape (Scott *et al.*, 2000), barley (Thiel *et at.*, 2003), and cassava (Zou *et al.*, 2011). In addition to genomic SSRs, EST-derived markers have been used for a variety of applications in molecular genetics, including phylogenetic analysis, parentage identification, and linkage map construction. Since EST-derived markers are from the gene-coding regions, they are highly conserved and highly transferable across taxa (Ellis and Burke, 2007) such as Ericaceae family (Rowland *et al.*, 2003), Prunus family (Decroocq *et al.*, 2003), grass species (Saha *et al.*, 2004). The transferability across taxa of EST-derived marker is also significantly higher in comparison to genomic SSRs. (Cordeiro *et al.*, 2001; Liewlaksaneeyanawin *et al.*, 2004). Therefore, in case of an orphan plant species, researchers are able to take advantage of the EST data belonging to a well-studied species from related taxa in generating a substantial amount of the markers needed for analyzing a less-defined one (Cato *et al.*, 2001; Ellis and Burke, 2007). In addition, the fact that EST-derived markers are gene-specific also makes them useful for QTL mapping (Wei *et al.*, 2005)

In the Orchidaceae family, the genus *Phalaenopsis* has been extensively investigated. In contrast to hundreds of *Dendrobium* ESTs, tens of thousands of *Phalaenopsis* ESTs are available to the public (Fu *et al.*, 2011). Nevertheless, EST-derived markers from these databases have not yet been reported. Thus, the objectives of this study were to: (1) develop EST-derived markers in dendrobium from ESTs of *Phalaenopsis* and *Dendrobium*; (2) assess transferability, polymorphism and heterozygosity of the EST-derived markers; and (3) evaluate the markers for genetic linkage analysis.

2. Materials and Methods

2.1 SSR mining and primer design

A total of 6063 EST sequences, 5908 *Phalaenopsis* ESTs and 155 *Dendrobium* ESTs, were downloaded from GenBank (http://www.ncbi.nlm.nih.gov) in FASTA-formatted files. Local BLAST was performed on the acquired ESTs using BioEdit Sequence Alignment Editor (Hall, 1999) to minimize redundancy. All unique sequences were screened for SSRs using WebSat server (http://wsmartins.net/websat/). Criteria for SSR search were stretches of di-nucleotide, tri-nucleotide and tetra-nucleotide perfect repeats with at least 5 perfect core motifs. PCR primers were designed to flank SSR regions of the selected ESTs using BatchPrimer3 (You *et al.*, 2008). The parameters for primer design were primer length of 18-25 nucleotides, GC contents of 40-60%, optimum annealing temperature at least 50°C and PCR product size 150-500 base pairs (bps).

2.2 Intron identification and primer design

The orchid EST sequences were aligned against rice complementary DNAs (cDNAs) in the TIGR rice genome annotation database (Ouyang *et al.*, 2007) using NCBI BLASTN program (Altschul *et al.*, 1990).The sequences matched with rice cDNA were used for the prediction of intronic regions. The alignment between orchid ESTs, rice cDNA and rice genomic DNA from GenBank database were performed using GeneDoc software (Nicholas *et al.*, 1997). EST sequences that contain at least one intron were selected. PCR primers were designed to flank intronic regions using BatchPrimer3 with the same parameters as in EST-SSRs primer design. The potential intron polymorphism (PIP) marker server (http://ibi.zju.edu.cn/pgl/pip) (Yang *et al.*, 2007) was also used to design intron-flanking EST markers. In order to identify putative function of the genes, selected ESTs were compared to the non-redundant sequence data from NCBI using BLASTX (Altschul *et al.*, 1997) at http://www.ncbi.nlm.nih.gov /BLAST/Blast.cgi with an E-value < 1e-20.

2.3 Plant material and transferability

All primer pairs were screened for PCR-amplification and cross-genera transferability using DNAs from a *Phalaenopsis* sp. and two *Dendrobium* cultivars, *'Caesar' (2N) (C2) and (D. 'Sri-Racha' x D. 'Snowfire') x D. bigibbum* (CP). The primer pairs that amplified clear PCR products in dendrobiums were tested for cross-species amplification in six *Dendrobium* commercial hybrid cultivars (C2, CP, *D. 'Mayuree White', D. ' Pompadour', D. Sonia 'Dang Piriya', and D. 'Caesar' (4N))* and six *Dendrobium* wild accessions (*D. superbiens*, *D. hercoglossum*, *D. signatum*, *D. discolor*, *D. stratiotes*, *and D. crumenatum*). This same set of primers was subsequently used for linkage analysis. The mapping population consisted of 224 progenies from a cross between

C2 and CP. Genomic DNA was extracted from young leaves following the protocol described by Fulton *et al.* (1995).

2.4 PCR amplification and electrophoresis

PCR was performed in a 20 µl reaction containing 25 ng of genomic DNA, 0.1 mM of dNTP, 1.5 mM of $MgCl_2$, 0.25 µM of each primer, 1x PCR buffer and 1 units of *Taq* polymerase (Fermentas Inc., MD, USA), Amplification was carried out on a PTC-100 Thermal-Cycler (MJ research, MA, USA) by 'Touch Down PCR' (Don *et al.*, 1991) with some modifications as follows: pre-denaturation at 94°C for 2 min; followed by 10 cycles of denaturation at 94°C for 30 sec, annealing at 60 or 65°C for 30 sec and extension at 72°C for 1 min, with the gradual decrease of annealing temperature by 1°C per cycle; followed by 24 cycles at 94°C for 30 sec, 50 or 55°C for 30 sec, and 72°C for 1 min; and the final extension at 72°C for 5 min. PCR products were separated on 4.5% polyacrylamide gel in 1X TBE and visualized by silver staining (Bassam *et al.*, 1991). PhiX174/*Hinf* I was used as a molecular weight standard marker.

2.5 Polymorphism and heterozygosity

To assess the level of polymorphism and heterozygosity, the number of alleles (N_a), the observed and expected heterozygosities (H_o and H_e) (Nei, 1978), and the polymorphism information content (PIC) value (Botstein *et al.*, 1980) for each marker loci were estimated, using data from the six commercial cultivars and the six wild accessions. Mean values of the three parameters for each group of markers (*Phalaenopsis* EST-SSRs, *Phalaenopsis* intron-flanking markers, *Dendrobium* EST-SSRs, and *Dendrobium* intron-flanking markers) were determined. PowerMarker V3.25 (Liu and Muse, 2005) was used for the calculation of N_a, H_o, H_e, and PIC values.

2.6 Linkage analysis

PCR products that showed monomorphism between C2 and CP were digested with 2 units of six restriction endonucleases: *Hinf*I, *Alu*I, *Taq*I, *Hae*III, *Rsa*I and *Bst*NI, at optimum temperature for 3 hours to detect restriction-site polymorphism. All polymorphic markers between the two parents were subsequently used to assess the entire C2xCP mapping population. Linkage analysis was carried out using a cross-pollinators (CP) algorithm in JoinMap 3.0 (Van Ooijen and Voorrips, 2001). A minimum logarithm of odds (LOD) score of 3.0 and the Haldane's mapping function were used.

3. Results

3.1 Frequency and distribution of SSRs in ESTs

From 4534 non-redundant *Phalaenopsis* ESTs, 466 ESTs were found containing SSR motifs (Table 1). The length of SSRs ranged from 10-82 bps and 81% of SSRs were less than 21 bps. The most abundant sequence was di-nucleotide repeat, followed by tri- and tetra-nucleotide repeat. $(AG)_n$ and $(CCG)_n$ were the most frequent motifs for di- and tri-nucleotide repeats, respectively. Two tetra-nucleotide repeats were TATT and CTCG. From 145 non-redundant *Dendrobium* ESTs, 29 ESTs were found containing SSRs. The motifs found are di- and tri-nucleotide repeats with the length ranged from 10-58 bps and 85% of SSRs were less than 18 bps. All SSRs found were perfect repeats. $(AG)_n$ was also the most frequent motifs and $(CAG)_8$ was the only tri-nucleotide repeat found.

3.2 Frequency and distribution of introns in ESTs

Analysis of the tentative intron/exon junctions in *Phalaenopsis* revealed 1182 ESTs contained at least one intron with a maximum of ten introns observed. The majority of intron-containing ESTs (42.6%) contained one intron. In *Dendrobium* ESTs, 59 contained at least one intron. As in *Phalaenopsis*, ESTs with one intron were also the majority (76.3%), but the number of introns ranged from one to five introns.

3.3 Primer design, amplification and transferability

A total of 288 primer pairs (263 from *Phalaenopsis* and 19 from *Dendrobium*) were designed. The remaining sequences were not used because the DNA sequences flank-

Table 1. Numbers of ESTs containing SSRs/introns and EST-derived primers from 4534 non-redundant *Phalaenopsis* ESTs and 145 non-redundant *Dendrobium* ESTs

Item	PS	PI	DS	DI
EST containing SSR/intron	466	1182	29	59
primer designed	177	92	5	14
primer, amplifiable with *Phalaenopsis* DNA	76	54	0	0
primer, amplifiable with *Dendrobium* DNA	5	25	3	8

PS, *Phalaenopsis* EST-SSR; PI, *Phalaenopsis* intron-flanking; DS, *Dendrobium* EST-SSR; DI, *Dendrobium* intron-flanking.

ing the SSR and the intron regions were either too short or were not suitable for primer design, e.g., low GC content. One hundred and thirty primer pairs derived from *Phalaenopsis* ESTs were able to amplify genomic DNA from a *Phalaenopsis* sp. but only 30 primers from this group gave clear and stable amplifications in two dendrobium cultivars used as testers. Eleven primer pairs from *Dendrobium* ESTs amplified only in dendrobiums. The details of 41 amplifiable primer sequences and their related information are described in Table 2. Putative gene functions were identified for 40 sequences.

3.4 Polymorphism and heterozygosity

The N_a, *He*, and PIC values for each marker are shown in Table 3. Among the four groups of EST-derived markers, EST-SSRs from *Dendrobium* exhibited the highest polymorphism in both cultivated and wild species. GQ250049DS showed eight different alleles among 12 *Dendrobium* species (Figure 1). For intron-flanking markers from both genera, the heterozygosities detected were higher in wild species than in cultivated species. Intron-flanking markers from *Dendrobium* also revealed higher polymorphism than intron-flanking markers from *Phalaenopsis*.

3.5 Linkage analysis

Of the 41 primer pairs, five showed length polymorphism between two parents of the mapping population and 11 showed restriction-site polymorphism, and thus were used as cleaved amplified polymorphic sequence (CAPS) markers (Table 2). The low level of polymorphism between the two cultivars (39%) for such an outcrossing species like *Dendrobium* was due to the fact that both C2 and CP were hybrids from similar background. Sixteen polymorphic markers were used in linkage analysis. Seven markers were mapped to three linkage groups (LGs) (Figure 2). Two markers from LG1, HA642879DI and CB032915PI were related to rice genes, Os09g36730 and Os09g07510, on rice chromosome 9. Two markers on LG3, CB034027PS and CB032380PI, were also related to Os04g0467100 and Os04g0448800 on chromosome 4 in rice.

4. Discussion

The percentages of SSRs in *Phalaenopsis* ESTs (10.3%) and *Dendrobium* ESTs (20%) were high in comparison to barley (3.4%), wheat (3.2%), maize (1.5%), sorghum (3.6%), rice (4.7%) (Kantety *et al.*, 2002). However, they are approximately in the same range as in coffee (13.5%) (Aggarwal *et al.*, 2007). The highest proportion of SSRs in *Phalaenopsis* ESTs was di-nucleotide repeat, followed by tri-nucleotide repeat, similar findings have been reported in rubber tree (Feng *et al.*, 2009) and walnut (Zhang *et al.*, 2011). These results are in contrast to earlier studies in several plants including cereal crops (Varshney *et al.*, 2002), tall fescue (Saha *et al.*, 2004), barrel medic (Eujayl *et al.*, 2004) and cultivated flax (Soto-Cerda *et al.*, 2011), where the most abundant classes of SSRs found were tri-nucleotide repeats. Similar to grape (Cordeiro *et al.*, 2001), barley (Thiel *et al.*, 2003), and rubber tree (Feng *et al.*, 2009), AG/TC was the most common di-nucleotide motif in *Phalaenopsis* and *Dendrobium* ESTs. The predominance of CCG repeats among tri-nucleotide motifs in this study are in agreement with those found in several cereal grains (Kantety *et al.*, 2002) and grass species (Saha *et al.*, 2004). As suggested by Varshney *et al.* (2005), the differences in EST-SSR frequencies and repeats

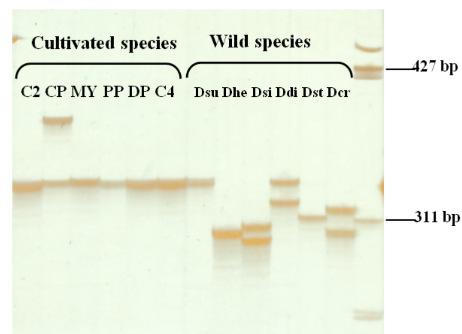

Figure 1. PCR amplification of EST-derived marker GQ250049DS showing a total of eight alleles in 12 *Dendrobium* accessions. C2 = *D.* 'Caesar' (2N), CP = (*Dendrobium* 'Sri-Racha' x *D.* 'Snowfire') x *D. bigibbum*, MY = *D.* 'Mayuree White', PP = *D.* 'Madam Pompadour', DP = *D.* 'Dang Piriya', C4 = *D.* 'Caesar' (4N), Dsu = *D. superbiens*, Dhe = *D. hercoglossum*, Dsi = *D. signatum*, Ddi = *D. discolor*, Dst = *D. stratiotes*, Dcr = *D. crumenatum*. The standard marker on the right of the gel was PhiX174 digested with *Hinf*I.

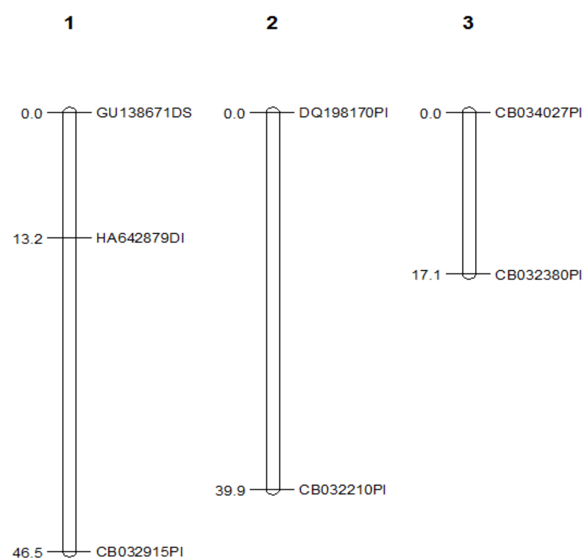

Figure 2. Three linkage groups of 7 EST-derived markers obtained from C2xCP F_1 progeny. The marker names are shown on the right of the bar and the distances between markers in centiMorgan (cM) on the left.

Table 2. Oligonucleotide sequences and related information for the 41 EST-derived primer pairs

Marker	Forward primer (5'-3')	Reverse primer (5'-3')	Assigned function	T_a [1]	Size range,bp	Mapping[2]
AF038840-1DI	GAGAACTGGGGATTCTTCGAG	TGGGAGATTCCGTGGTTCAG	1-aminocyclopropane-1-carboxylate oxidase	65 hold 55	904-954	*C
AF193815DI	TATCTCAAGACTGTGCAGGG	TCCTTGAACTGCTCCTCTGTCA	isocitrate lyase	65 hold 55	154-382	
AF420239DI	TACGAAGTTCCCCTCGGCTA	CATAGCCCTCGAACCCAAT	S-adenosyl-L-methionine decarboxylase	65 hold 55	414-418	
AF485892DI	AGACCAAAAGCTGCTTGCAT	TTCTTTAGGTGCGTGTTCCA	MYB family protein	60 hold 50	673-753	
AF485893DI	TCACAAAGTATGGGCATGGA	TTGTGCTGCAATCTGAGACC	MYB family protein	60 hold 50	382-625	
AY746972DI	GGAGCAACCCACTTGATAAACC	TGAGTGAGCATCCTCACATGC	ethylene receptor	65 hold 55	262	
EF612438DI	TGTTGCCATTCAAGCTGTTC	CGTCACATTTCATGGTGGAG	actin	60 hold 50	504-573	
HA642879DI	AAGGCTCACACCAACAAAGG	TGTCCGTCCTCCCTTGTAAC	MYB family protein	65 hold 55	361-398	*
AF100336DS	TGGTGCTCCTTTACCTATCTGC	GTACACAGGAACTTGCATCGAC	phenylalanine ammonia-lyase	65 hold 55	186-192	
GQ250049DS	CGACTGTGGATGTTTCAGTTGT	CTGTCACCCAAGAAATCAGTTG	Glyceraldehyde-3-phosphate dehydrogenase	60 hold 50	279-373	*
GU138671DS	CTTGTCCCGACTCCTTATTACC	TTTGTGGAAGAGAAGACACAGC	1-aminocyclopropane-1-carboxylate synthase	60 hold 50	408-425	*
AJ563284PI	GTGGACGCAGAATATGTGGCGAA	AGTTCCAAAACGGCAGCGCAA	ethylene response sensor	65 hold 55	276	
AY134752PI	CTCTCTTTATGCCAGTGGTCG	ATTGGGACAGTATGGCTGACAC	actin	65 hold 55	776-808	*C
AY771992PI	GATGCTGGAGAAATACCAGCAG	GTTAAGAGAGTTCAGATCCTCCC	MADS box	65 hold 55	330-356	*
CB032056PI	CTTGGGGTGAAGCAGATGAT	TTGGATCAGGTCGAGAGCTT	elongation factor-1 alpha	65 hold 55	248-362	
CB032073PI	GCTATGAATGCAGCCGAAAT	AATGTTGCTGGAAAGCCAGT	ARF-like small GTPase 1	65 hold 55	898-918	*C
CB032210PI	GGAGCAACAAATAGGCCAAA	GTTGAGCACCTCAGCATCAA	Cell division cycle protein	65 hold 55	406-460	*C
CB032380PI	GCCGTCACACCTCTTGATCT	CCCAGCAATGTCCGAATAAT	phosphate carrier protein	65 hold 55	450-463	*C
CB032610PI	ATATGGGAGGGATGGGTTTC	GCTTCTCCAAGCCATAAGCA	GDP-mannose 3,5-epimerase	65 hold 55	662-763	*C
CB032877PI	CCACCAATCCGAAGCTACC	CACGGAACATCACCAACAAG	Auxin-responsive protein	65 hold 55	935-1,022	
CB032915PI	CGCCTTTGGAGACTTTGTGT	GCATATCAGGGGAACTTGGA	Phosphatase	65 hold 55	946-977	*C
CB033161_1PI	ATCTTCGTCAAAACCCTAACG	AACGAGGTGAAGCGTCGAT	Ubiquitin protein	65 hold 55	205	*C
CB033572_3PI	GAAGGAGCAGTGGAGTCCTG	GTGCGATCTCTGGAACCAAT	ubiquitin-conjugating enzyme	60 hold 50	473-683	
CB033695PI	TTCCTGGAGTTCGAGTTTCG	AGACCCTCAGGATCCTTGCT	mago nashi-like protein	65 hold 55	534	
CB033731PI	GCCGTGCTTTCTCTTTATGC	CGAGCTCCTGCTCATAGTCC	actin	65 hold 55	366-376	
CB033806PI	TGGAGATCATCTTGCCACTG	ACGCACCATTAGAGGTTTGG	protein phosphatase	65 hold 55	406-408	
CB033831PI	GTCATCCGCAAGAATCTCGT	CGACAACCTTCTCCACCTTG	heat shock protein 90	65 hold 55	528	
CB034260PI	TTGATTGTGGAAAGGCAAAA	TGATATGCTTCCAAGCCTGAG	putative DEAD/H box polypeptide 36 protein	65 hold 55	422-554	
CB034284PI	TGGGAAGCCTACAGTTGTTGA	CTCCTGCTCCCACTTTGTGT	thioredoxin-like protein HCF164	65 hold 55	310-314	
CB034567PI	CAGGGGTCGAAGAAGAAGAA	TGTACTCGATGACGCTCCAG	cellulase	65 hold 55	465	
CB034644PI	TTGAAAATATGGTCCGCACA	GCAATATCTTGATTGGCAGGA	dynamin-related protein 1E	65 hold 55	426-450	
CB034671PI	TTCCAGGCTGAGATCAATCA	AGCAGCAAGAGCCTCCATAA	heat shock protein	65 hold 55	313-410	
CK857666_1PI	AGAGGCTTTTCTACAAACTTCTAA	GCAGCCAAGATCTCCAAGTC	NADP-malic enzyme	65 hold 55	324	

Table 2. Continued

Marker	Forward primer (5'-3')	Reverse primer (5'-3')	Assigned function	T_a [1]	Size range,bp	Mapping [2]
CK857699PI	CGGACACCATTGATAACGTG	AGGGTGGACTCCTCTGGAT	polyubiquitin	65 hold 55	146	
DQ198170PI	ATGGCAGGGAGTGGGTTCTTT	ACAGAATCGGGGATGACATTG	IAA hydrolase	65 hold 55	283-792	*C
X79905-2PI	AACGCATCACCATTAAGCCGC	TCCACAGCTCCAGCTGAGCA	S-adenosyhomocysteine hydrolase	65 hold 55	268-450	
CB032425PS	ACAAGACCACTTCGAGAACTCC	AGGGTGTTAGCAATTAGGCAGA	Phosphatase	65 hold 55	608-823	*C
CB032617PS	GAGAGCAGGGATTTAGGTGATG	CAACAGATACAAATGCAGGGAA	Stress responsive protein	65 hold 55	354-427	*C
CB033302PS	AATTCAGTACCATTGCGATCC	TCCGTATTTTGAGCCTTCAGTT	vesicle-associated 727-like membrane protein	65 hold 55	560	
CB033889PS	TGTGATTGATTGGAGTTGATGC	GTGGTCGAACCTACGAAGAGAC	Not assigned	60 hold 50	133	
CB034027PS	TCTTCCTCTCCCAAACCAATAA	AGTGGAGCAGTAGCCCTGTAAG	uncharacterized protein	60 hold 50	314-335	*

[1] T_a, annealing temperature, according to 'Touch-down PCR' (Don *et al.*, 1991).
[2] Mapping: '*' indicates that marker was used in linkage analysis, 'C' indicates that marker was used as CAPS marker.

motifs can possibly be from the differences in SSR search criteria used in different studies.

Analysis of intron number per EST revealed that the ESTs containing only one intron were in the majority for both *Phalaenopsis* and *Dendrobium*. The average number of introns per intron-containing EST was 2.1. This result is in the same range as other studies in eukaryotic genes. Deutsch and Long (1999) showed that the number of intron per gene in both plant and animal vary from 2.2 to 4.8 introns. However, most ESTs in this study were not complete sequences and the calculations were based on ESTs that contained intron(s). These results may not represent the intron distribution in orchid ESTs.

Twenty-three percent of EST-derived primers from *Phalaenopsis* could amplify products from *Dendrobium* DNAs but none of *Dendrobium* EST-derived primers amplified *Phalaenopsis* DNA. This result is still inconclusive because only ten primers were obtained from 145 *Dendrobium* ESTs. More *Dendrobium* markers are needed to test the cross-genera transferability from *Dendrobium* to *Phalaenopsis*. The EST-derived markers that are designed from coding regions were reported to have high rates of cross-genera and cross-species amplifications. However in this study, the cross-genera transferability of *Phalaenopsis* EST-derived markers to *Dendrobium* was low compared to the cross-genera transferability of some plant species, e.g., *Vaccinium* sp. to *Rhododendron* sp. (70-90%) (Rowland and Dhanaraj, 2003), erianthus to sorghum (60%) (Cordeiro *et al.*, 2001), barley to rice (40%) (Thiel *et al.*, 2003) and tall fescue to ryegrass, rice and wheat (38-66%) (Saha *et al.*, 2004). The reason for this may possibly be the high level of genetic diversity among Orchidaceae family and the distant relationship between *Phalaenopsis* and *Dendrobium*. According to Liewlaksaneeyanawin *et al.* (2004), the transferability rate decrease as the evolutionary distance between the source and target species increases. Non-amplifiable primer could result from the intron insertion within primer sequences that interrupted primer annealing or the presence of large intron fragment between primer pair that interrupted the PCR extension (Thiel *et al.*, 2003; Varshney *et al.*, 2005; Zou *et al.*, 2011). The amplifications of 41 EST-derived markers in 12 *Dendrobium* accessions show the high level of the transferability across species. These results were expected, as the EST markers were derived from the transcribed region of the DNA that would be more conserved across population and species than the non-transcribed region (Rowland and Dhanaraj, 2003).

According to Holland *et al.* (2001), EST-SSRs are more polymorphic than intron-flanking markers. Similar results were observed in this study. *Dendrobium* EST-SSRs yielded the highest number of alleles, the expected heterozygosities and the PIC values in both cultivated dendrobiums and wild species. The heterozygosities in *Dendrobium* hybrid cultivars were lower than those in the wild species. The differentiation of N_a and H_e between the cultivated and the wild species, especially from intron-flanking markers, reflects the

Table 3. Characteristics of EST-derived markers applied for cultivated, wild, and all species of *Dendrobium* used in this study.

Marker	Cultivated species				Wild species				All species			
	N_a	H_e	H_o	PIC	N_a	H_e	H_o	PIC	N_a	H_e	H_o	PIC
AF038841-1DI	2	0.44	0.67	0.35	2	0.22	0.25	0.19	3	0.62	0.43	0.55
AF193815DI	1	0	0	0	3	0.64	0	0.56	3	0.43	0	0.39
AF420239DI	1	0	0	0	2	0.5	0	0.38	2	0.47	0	0.36
AF485892DI	1	0	0	0	3	0.57	0.17	0.50	3	0.35	0.08	0.32
AF485893DI	1	0	0	0	4	0.64	0.33	0.57	4	0.45	0.17	0.40
AY746972DI	1	0	0	0	1	0	0	0	1	0	0	0
EF612438DI	1	0	0	0	3	0.50	0	0.45	3	0.29	0	0.27
HA642879DI	3	0.54	0.83	0.46	7	0.84	1.00	0.82	7	0.75	0.91	0.72
Mean DI	**1.4**	**0.12**	**0.19**	**0.10**	**3.1**	**0.49**	**0.22**	**0.43**	**3.3**	**0.42**	**0.20**	**0.38**
AF100336DS	5	0.74	0.50	0.70	4	0.71	0.17	0.65	5	0.77	0.33	0.73
GQ250049DS	2	0.15	0.17	0.14	7	0.82	0.50	0.80	8	0.63	0.33	0.61
GU138671DS	3	0.63	0.83	0.55	4	0.69	0.33	0.64	5	0.77	0.58	0.73
Mean DS	**3.3**	**0.50**	**0.50**	**0.46**	**5**	**0.74**	**0.33**	**0.70**	**6**	**0.72**	**0.42**	**0.69**
Mean D	**1.9**	**0.23**	**0.27**	**0.20**	**3.6**	**0.56**	**0.25**	**0.51**	**4**	**0.50**	**0.26**	**0.46**
AJ563284PI	1	0	0	0	1	0	0	0	1	0	0	0
AY134752PI	3	0.29	0.17	0.27	3	0.46	0.20	0.41	3	0.60	0.18	0.52
AY771992PI	2	0.49	0.83	0.37	4	0.67	0	0.62	6	0.79	0.42	0.76
CB032056PI	1	0	0	0	2	0.50	0	0.38	2	0.38	0	0.30
CB032073PI	2	0.28	0.33	0.24	1	0	0	0	2	0.22	0.25	0.19
CB032210PI	2	0.50	1.00	0.38	3	0.50	0.33	0.45	3	0.54	0.67	0.46
CB032380PI	2	0.49	0.83	0.37	2	0.15	0.17	0.14	2	0.44	0.50	0.35
CB032610PI	1	0	0	0	2	0.50	0	0.38	2	0.40	0	0.32
CB032877PI	2	0.47	0.75	0.36	3	0.53	0.25	0.47	4	0.73	0.50	0.68
CB032915PI	1	0	0	0	3	0.56	0	0.50	3	0.51	0	0.44
CB033161_1PI	1	0	0	0	1	0	0	0	1	0	0	0
CB033572_3PI	3	0.54	0.80	0.47	7	0.83	0.67	0.81	7	0.79	0.73	0.76
CB033695PI	1	0	0	0	1	0	0	0	1	0	0	0
CB033731PI	1	0	0	0	2	0.28	0	0.24	2	0.17	0	0.15
CB033806PI	1	0	0	0	2	0.38	0	0.30	2	0.22	0	0.19
CB033831PI	1	0	0	0	1	0	0	0	1	0	0	0
CB034260PI	1	0	0	0	3	0.59	0.25	0.51	3	0.54	0.17	0.46
CB034284PI	2	0.15	0.17	0.14	1	0	0	0	2	0.12	0.13	0.11
CB034567PI	1	0	0	0	1	0	0	0	1	0	0	0
CB034644PI	1	0	0	0	3	0.63	0	0.55	3	0.37	0	0.34
CB034671PI	1	0	0	0	2	0.18	0.2	0.16	2	0.09	0.09	0.08
CK857666_1PI	1	0	0	0	1	0	0	0	1	0	0	0
CK857699PI	1	0	0	0	1	0	0	0	1	0	0	0
DQ198170PI	2	0.28	0	0.24	5	0.74	0.17	0.69	5	0.66	0.08	0.61
X79905-2PI	1	0	0	0	2	0.28	0	0.24	2	0.17	0	0.15
Mean PI	**1.4**	**0.14**	**0.20**	**0.11**	**2.3**	**0.31**	**0.09**	**0.27**	**2.5**	**0.31**	**0.15**	**0.27**
CB032425PS	2	0.38	0.50	0.30	4	0.72	0.33	0.67	5	0.65	0.42	0.61
CB032617PS	2	0.38	0.17	0.30	2	0.28	0	0.24	3	0.35	0.08	0.32
CB033302PS	1	0	0	0	1	0	0	0	1	0	0	0
CB033889PS	1	0	0	0	1	0	0	0	1	0	0	0
CB034027PS	4	0.65	0.50	0.60	5	0.76	0.33	0.73	7	0.80	0.42	0.78
Mean PS	**2**	**0.28**	**0.23**	**0.24**	**2.6**	**0.35**	**0.13**	**0.33**	**3.4**	**0.36**	**0.18**	**0.34**
Mean P	**1.5**	**0.16**	**0.20**	**0.13**	**2.3**	**0.32**	**0.10**	**0.28**	**2.6**	**0.32**	**0.15**	**0.29**
Mean Total	**1.6**	**0.18**	**0.22**	**0.15**	**2.7**	**0.38**	**0.14**	**0.34**	**3**	**0.37**	**0.18**	**0.33**

N_a, Number of allele per locus; H_e, expected heterozygosity; H_o, observed heterozygosity;
PIC, Polymorphism informa-tion content.

narrow genetic diversity in cultivated dendrobiums. Pedigrees of many cultivated dendrobiums showed the similarities of parental species used in breeding programs. An inbreeding depression caused by accumulation of deleterious alleles could result in the reduction of yield traits such as flower per raceme, raceme yield and so on (Bobisud and Kamemoto, 1982) and could be a problem for genetic improvement of *Dendrobium*. Interspecific hybridization could introduce new gene pool to solve this limited diversity problem. The EST-derived markers could be utilized to assess the genetic variability in wild species that can be introgressed into the cultivated dendrobiums.

Xue *et al.* (2010) were the first to report two genetic linkage maps for two *Dendrobium* species using two-way pseudotestcross strategy with an F_1 population. Both maps were based on RAPD and sequence-related amplified polymorphism (SRAP) markers, which are dominant markers and not readily transferable to other populations. The EST-derived markers will bridge information among populations and among laboratories. Since *Dendrobiums* are outcrossing species with inbreeding depression (Kamemoto *et al.*, 1999), advanced generations for mapping are not obtainable. An F_1 population was also used to evaluate our markers for mapping potential. For the mapping calculation, we used CP algorithm of JoinMap 3.0 program which generated a single map for both parents. Unexpectedly, seven out of 16 markers were mapped into three linkage groups. The high percentage of markers mapped could be by chance, but was quite promising. For further QTL mapping, the EST-derived markers are potentially useful because they are gene-specific functional markers. The syntenic relationship between dendrobium and rice suggests the possibility of targeting markers to particular loci using comparative genomics (Tamura *et al.*, 2009).

In this study, we developed 41 PCR-based gene specific markers in *Dendrobium* using EST databases of *Phalaenopsis* and *Dendrobium*. All markers were transferable across genera and species. The level of polymorphisms and heterozygosities are high particularly among wild species. They showed great potential for genetic linkage mapping and comparative genomic study. Considering the large numbers of *Phalaenopsis* ESTs available to date, we are hopeful that EST-derived markers from *Phalaenopsis*, a well-intensively studied genus in *Orchidaceae* family, will advance *Dendrobium* research and breeding programs in the future. Furthermore, research on other less-studied orchid species could also benefit from the same approach using minimal resources.

Acknowledgements

We are grateful to Dr David Hole for his valuable comments. We thank Mr Michael Cooper for proofing the manuscript. This research was supported by Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus and the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education (AG-BIO/PERDO-CHE). The authors would like to declare no conflict of interest.

References

Agarwal, M., Shrivastava, N. and Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports. 27, 617-631.

Aggarwal, R.K., Hendre, P.S., Varshney, R.K., Bhat, P.R., Krishnakumar, V. and Singh, L. 2007. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theoretical and Applied Genetics. 114, 359-372.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215, 403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. 25, 3389-3402.

Bassam, B.J., Anollés, G.C. and Gresshoff, P.M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Analytical Biochemistry. 196, 80–83.

Bobisud, C. A. and Kamemoto, H. 1982. Selection and inbreeding in amphidiploid *Dendrobium* (Orchidaceae). Journal of the American Society for Horticultural Science. 107, 1024-1027.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 32, 314–331.

Cato, S.A., Gardner, R.C., Kent, J. and Richardson, T.E. 2001. A rapid PCR-based method for genetically mapping ESTs. Theoretical and Applied Genetics. 102, 296-306.

Collard, B.C.Y. and Mackill, D.J. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences. 363(1491), 557-572.

Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M. and Henry, R.J. 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Science. 160, 1115-1123.

Decroocq, V., Favé, M.G., Hagen, L., Bordenave, L.and Decroocq, S. 2003. Development and transferability of apricot and grape EST microsatellite markers across taxa. Theoretical and Applied Genetics. 106, 912-922.

Deutsch, M. and Long, M. 1999. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Research. 27, 3219-3228.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. 1991. 'Touchdown' PCR to circumvent spurious

priming during gene amplification. Nucleic Acids Research. 19(14), 4008.

Ellis, J.R. and. Burke, J.M. 2007. Short Review: EST-SSRs as a resource for population genetic analyses. Heredity. 99, 125-132.

Eujayl, I., Sledge, M.K., Wang, L. May, G.D., Chekhovskiy, K., Zwonitzer, J.C. and Mian, M.A.R. 2004. *Medicago truncatula* EST-SSR reveal cross-species genetic markers for *Medicago* spp. Theoretical and Applied Genetics. 108, 414-422.

Fan, W.J., Luo, Y.M., Li, X.X., Gu, S., Xie, M.L, He, J., Cai, T. and Ding, X.Y. 2009. Development of microsatellite markers in *Dendrobium fimbriatum* Hook, an endangered Chinese endemic herb. Molecular Ecology Resources. 9, 373-375.

Feng, S.P., Li, W.G., Huang, H.S., Wang, J.Y. and Wu, Y.T. 2009. Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). Molecular Breeding. 23, 85-97.

Fu, C.H., Chen, Y.W., Hsiao, Y.Y., Pan, Z.J., Liu, Z.J., Huang, Y.M., Tsai, W.C. and Chen, H.H. 2011. OrchidBase: A collection of sequences of the transcriptome derived from orchid. Plant and Cell Physiology. 52, 238-243.

Fulton, T.M., Chunwongse, J. and Tanksley, S.D. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter. 13, 207-209.

Gonzalo, M.J., Oliver, M., Garcia-Mas, J., Monfort, A., Dolcet-Sanjuan, R., Katzir, N., Arús, P. and Monforte, A. J. 2005. Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theoretical and Applied Genetics. 110, 802-811.

Gu, S., Ding, X.Y., Wang, Y., Zhou, Q., Ding, G., Li, X.X. and Qian, L. 2007. Isolation and characterization of microsatellite markers in *Dendrobium officinale*, an endangered herb endemic to China. Molecular Ecology Notes. 7, 1166-1168.

Gupta, P.K., Varshney, R.K., Sharma, P.C. and Ramesh, B. 1999. Molecular markers and their applications in wheat breeding. Plant Breeding. 118, 369-390.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41, 95-98.

Hew, C.S. and Yong, J.W.H. 2004. The physiology of tropical orchids in relation to the industry, 2nd Ed. World Scientific Publishing Co. Pte. Ltd., Singapore.

Holland, J.B., Helland, S.J., Shaorpova, N. and Rhyne, D.C. 2001. Polymorphism of PCR-based markers targeting exon, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. Genome. 44, 1065-1076.

Kamemoto, H., Amore, T.D. and Kuehnle, A.K. 1999. Breeding *Dendrobium* orchid in Hawaii. University of Hawaii Press, Honolulu.

Kantety, R.V., La Rota, M., Matthews, D. E. and Sorrells, M. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology. 48, 501-510.

Liewlaksaneeyanawin, C., Ritland, C.E., El-Kassaby, Y.A. and Ritland, K. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theoretical and Applied Genetics. 109, 361-369.

Liu, K. and Muse, S.V. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics. 21, 2128-2129.

Milbourne, D., Meyer, R.C., Collins, A.J., Ramsay, L.D., Gebhardt, C. and Waugh, R.1998. Isolation, characterisation and mapping of simple sequence repeat loci in potato. Molecular and General Genetics. 259, 233-245.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89, 583-590.

Nicholas, K.B., Nicholas, H.B., Jr, and Deerfield, D.W., II. 1997. GeneDoc: Analysis and visualization of genetic variation. EMBNEW.NEWS. 4, 14.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J. and Buell, C.R. 2007. The TIGR rice genome annotation resource: Improvements and new features. Nucleic Acids Research. 35, 883-887.

Rowland, L.J., Dhanaraj, A.L., Polashock, J. and Arora, R. 2003. Utility of blueberry-derived EST-SSR primers in related Ericaceae species. HortScience. 38, 1428-1432.

Saha, M.C., Mian, M.A.R., Eujayl, I., Zwonitzer, J.C., Wang, L.J. and May, G.D. 2004. Tall fescue EST-SSR markers with transferability across several grass species. Theoretical and Applied Genetics. 109, 783-791.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S. and. Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theoretical and Applied Genetics. 100, 723-726.

Shen, J., Ding, X., Liu, D., Ding, G., He, J., Li, X., Tang, F. and Chu, B. 2006. Intersimple sequence repeats (ISSR) molecular fingerprinting markers for authenticating populations of *Dendrobium officinale* Kimura *et* Migo. Biological and Pharmaceutical Bulletin. 29, 420-422.

Soto-Cerda, B.J., Carrasco, R.A., Aravena, G.A., Urbina, H.A. and Navarro, C.S. 2011. Identifying novel polymorphic microsatellites from cultivated flax (*Linum usitatissimum* L.). Plant Molecular Biology Reporter. 29, 753-759.

Tamura, K., Yonemaru, J., Hisano, H., Kanamori, H., King, J., King, I.P., Tase, K., Sanada, Y., Komatsu, T. and Yamada, T. 2009. Development of intron-flanking EST markers for the *Lolium/Festuca* complex using rice genomic information. Theoretical and Applied Genetics. 118, 1549-1560.

Thiel, T., Michalek, W., Varshney, R. K. and Graner, A. 2003. Exploiting EST-database for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics. 106, 411-422.

Van Ooijen, J.W. and Voorrips, R.E. 2001. JoinMap® 3.0, Software for the Calculation of Genetic Linkage maps. Plant Research International. Wageningen, The Netherlands.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P. and Graner, A. 2002. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular and Molecular Biology Letters. 7, 537-546.

Varshney, R.K., Sigmund, R., Börner, A., Korzun, V., Stein, N., Sorrells, M.E., Langridge, P. and Graner, A. 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science. 168, 195-202.

Wang, H.Z., Lu, J.J.,Shi, N.N. and Ying, Q.C. 2006. Analysis of genetic diversity and affinity relationships among 13 species of *Dendrobium* Sw. by RAPD. Chinese Traditional and Herbal Drugs. 37, 558-592.

Wei, H., Fu, Y. and Arora, R. 2005. Intron-flanking EST–PCR markers: From genetic marker development to gene structure analysis in *Rhododendron*. Theoretical and Applied Genetics. 111, 1347-1356.

Xiang, N., Hong, Y. and Lam-Chan, L.T. 2003. Genetic analysis of tropical orchid hybrids (*Dendrobium*) with fluorescence amplified fragment-length polymorphism (AFLP). Journal of American Society for Horticultural Science. 128, 731-735.

Xue, D., Feng, S., Zhao, H., Jiang, H., Shen, B., Shi, N., Lu, J., Liu, J. and Wang, H. 2010. The linkage maps of *Dendrobium* species based on RAPD and SRAP markers. Journal of Genetics and Genomics 37, 197-204.

Yang, L., Jin, G., Zhao, X., Zheng, Y., Xu, Z. and Wu, W. 2007. PIP: A database of potential intron polymorphism markers. Bioinformatics. 23, 2174-2177.

You, F.M., Huo, N., Gu, Y., Q. Luo, M.C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J. and Anderson, O.D. 2008. BatchPrimer3: A high throughput web application for PCR and sequencing primer design. Bioinformatics. 9, 253.

Yue, G.H., Lam-Chan, L.T.T. and Hong, Y. 2006. Development of simple sequence repeat (SSR) markers and their use in identification of *Dendrobium* varieties. Molecular Ecology Notes. 6, 832-834.

Zhang, R., Zhu, A.D., Wang, X.J., Yu, J., Zhang, H.R., Gao, J.S., Cheng, Y.J. and Deng, X.X. 2010. Development of *Juglans regia* SSR markers by data mining of the EST database. Plant Molecular Biology Reporter. 28, 646-653.

Zou, M., Xia, Z., Ling, P., Zhang, Y., Chen, X., Wei, Z., Bo, W. and Wang, W.2011. Mining EST-derived SSR markers to assess genetic diversity in cassava (*Manihot esculenta* Crantz). Plant Molecular Biology Reporter. 29, 961-971.