

RESEARCH ARTICLE

# RNA-Seq SSRs of Moth Orchid and Screening for Molecular Markers across Genus *Phalaenopsis* (Orchidaceae)

Chi-Chu Tsai<sup>1,2\*</sup>, Huei-Chuan Shih<sup>3\*</sup>, Hao-Ven Wang<sup>4\*</sup>, Yu-Shium Lin<sup>1</sup>, Chia-Hung Chang<sup>1</sup>, Yu-Chung Chiang<sup>5,6\*</sup>, Chang-Hung Chou<sup>7\*</sup>

**1** Kaohsiung District Agricultural Research and Extension Station, Pingtung 900, Taiwan, **2** National Pingtung University of Science and Technology, Pingtung 912, Taiwan, **3** Department of Nursing, Meiho University, Pingtung 912, Taiwan, **4** Department of Life Sciences, National Cheng Kung University, Tainan 701, Taiwan, **5** Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung 804, Taiwan, **6** Department of Biomedical Science and Environment Biology, Kaohsiung Medical University, Kaohsiung 807, Taiwan, **7** Research Center for Biodiversity, China Medical University, Taichung 404, Taiwan

\* These authors contributed equally to this work.

\* [tsaicc@mail.kdais.gov.tw](mailto:tsaicc@mail.kdais.gov.tw) (CCT); [yuchung@mail.nsysu.edu.tw](mailto:yuchung@mail.nsysu.edu.tw) (YCC); [choumasa@mail.cmu.edu.tw](mailto:choumasa@mail.cmu.edu.tw) (CHC)



**OPEN ACCESS**

**Citation:** Tsai C-C, Shih H-C, Wang H-V, Lin Y-S, Chang C-H, Chiang Y-C, et al. (2015) RNA-Seq SSRs of Moth Orchid and Screening for Molecular Markers across Genus *Phalaenopsis* (Orchidaceae). PLoS ONE 10(11): e0141761. doi:10.1371/journal.pone.0141761

**Editor:** Xiaoming Pang, Beijing Forestry University, CHINA

**Received:** June 17, 2015

**Accepted:** October 13, 2015

**Published:** November 2, 2015

**Copyright:** © 2015 Tsai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. These short sequence reads have been deposited at NCBI as SRA accession number SRX1253908 and SRX1253909.

**Funding:** This research was supported by funding from the Ministry of Science and Technology, Taiwan, MOST 103-2321-B-067E-001 to CCT and partial from MOST 103-2621-B-110 -001 to YCC.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

### Background

The moth orchid (*Phalaenopsis* species) is an ornamental crop that is highly commercialized worldwide. Over 30,000 cultivars of moth orchids have been registered at the Royal Horticultural Society (RHS). These cultivars were obtained by artificial pollination of inter-specific hybridization. Therefore, the identification of different cultivars is highly important in the worldwide market.

### Methods/Results

We used Illumina sequencing technology to analyze an important species for breeding, *Phalaenopsis aphrodite* subsp. *formosana* and develop the expressed sequence tag (EST)-simple sequence repeat (SSR) markers. After *de novo* assembly, the obtained sequence covered 29.1 Mb, approximately 2.2% of the *P. aphrodite* subsp. *formosana* genome (1,300 Mb), and a total of 1,439 EST-SSR loci were detected. SSR occurs in the exon region, including the 5' untranslated region (UTR), coding region (CDS), and 3'UTR, on average every 20.22 kb. The di- and tri-nucleotide motifs (51.49% and 35.23%, respectively) were the two most frequent motifs in the *P. aphrodite* subsp. *formosana*. To validate the developed EST-SSR loci and to evaluate the transferability to the genus *Phalaenopsis*, thirty tri-nucleotide motifs of the EST-SSR loci were randomly selected to design EST-SSR primers and to evaluate the polymorphism and transferability across 22 native *Phalaenopsis* species that are usually used as parents for moth orchid breeding. Of the 30 EST-SSR loci, ten polymorphic and transferable SSR loci across the 22 native taxa can be obtained. The validated EST-SSR markers were further proven to discriminate 12 closely related *Phalaenopsis* cultivars. The results show that it is not difficult to obtain universal SSR markers by transcriptome deep sequencing in *Phalaenopsis* species.

## Conclusions

This study supported that transcriptome analysis based on deep sequencing is a powerful tool to develop SSR loci in non-model species. A large number of EST-SSR loci can be isolated, and about 33.33% EST-SSR loci are universal markers across the *Phalaenopsis* breeding germplasm after preliminary validation. The potential universal EST-SSR markers are highly valuable for identifying all of *Phalaenopsis* cultivars.

## Introduction

Moth orchids (*Phalaenopsis* spp.) are among the most graceful and popular plants. They consist of approximately 66 natural species worldwide, fifty-six of which are extant [1]. Based on the classification of Christenson [1], *Phalaenopsis* is divided into five subgenera, *Proboscioides*, *Aphyllae*, *Parishianae*, *Polychilos*, and *Phalaenopsis* that are determined mainly by plant size and floral morphology (including callus, lip structure, pollinium number, and other characters). The subgenus *Polychilos* was further subdivided into four sections, including *Polychilos*, *Fuscatae*, *Amboinenses*, and *Zebrinae*. Additionally, the subgenus *Phalaenopsis* was also subdivided into four sections, *Phalaenopsis*, *Deliciosae*, *Esmeralda*, and *Stauroglottis*. *Phalaenopsis* species are found throughout tropical Asia and the larger islands of the Pacific Ocean. All *Phalaenopsis* species, excluding the natural tetraploid species *P. buyssoniana* Rchb.f., have 38 ( $2n = 38$ ) chromosomes [1,2]. Recently, the plastid genome of *P. aphrodite* has been completely sequenced [3], and molecular phylogenies of *Phalaenopsis* species also have been constructed based on the internal transcribed spacer (ITS) of the ribosomal (rDNA) and plastid DNA [4,5,6,7]. Additionally, molecular data were used to determine the inheritance of the natural hybrid, *P. x intermedia*, showing that *P. aphrodite* was the maternal parent and *P. equestris* was the paternal parent [8]. More recently, complete genome sequencing has been conducted in *P. equestris* [9].

Random amplified polymorphic DNA (RAPD) has been conducted to reveal the phylogenetic relationship of 16 *Phalaenopsis* species [10]. Three-hundred-eighty-one RAPD makers derived from 20 primers were obtained. Chuang [11] examined several accessions of *Phalaenopsis aphrodite* subsp. *formosana* and several related *Phalaenopsis* species from the Philippines based on RAPD and inter-simple sequence repeat (ISSR) molecular markers. The results showed that these two molecular techniques could offer informative markers to separate those from samples that are closely related. Another RAPD analysis was conducted by Goh *et al.* [12]. They examined 149 accessions representing 46 species of genus *Phalaenopsis*, and four *Paraphalaenopsis* species were used as outgroups. Six out of twenty random primers were selected for analysis and 123 polymorphic bands have been obtained. Cluster analysis derived from the RAPD molecular markers showed that *Phalaenopsis* form seven groups and are basically congruent with previous studies derived from morphological characters

Generally, the high repeated motifs of microsatellites are prone to mutation through slipped-strand mispairing [13]. The relatively rapid mutation rate, and high frequency in genome have made SSRs to be popular markers for population genetics [14,15,16], hybrid detection [17], linkage mapping, genetic fingerprinting [18,19], evolutionary history [20,21], and taxonomy [22,23]. Young [24] examined DNA fingerprinting of 89 accessions of *Phalaenopsis amabilis* based on microsatellite DNA (simple sequence repeats, SSRs). Three SSR loci were cloned and evaluated from *P. amabilis* accessions. The results indicated that these loci are good molecular markers to identify intraspecific variation of *Phalaenopsis*. EST-SSRs separately

developed from the *Phalaenopsis* ESTs database have obtained 42 [25] and 261 EST-SSR loci [26]. Nine-hundred-fifty potential SSRs in *Phalaenopsis equestris* were discovered by large-scale BAC end sequencing [27]. Deep sequencing technologies offer the possibility of generating numerous SSR markers much faster and at a lower cost compared to library-based methods [28,29,30,31,32].

Here, we performed *de novo* transcriptome deep sequencing of *P. aphrodite* subsp. *formosana* to analyze EST-SSR, develop molecular markers, and test the transferability between most members of *Phalaenopsis* that are used as parents for moth orchid breeding. To our knowledge, this is the first study to develop EST-SSRs by deep sequencing of transcriptomes in *Phalaenopsis* species. Furthermore, the developed EST-SSR markers in the present study can be applied for genetic diversity analysis, gene mapping, linkage map development, marker-assisted selection breeding, and cultivar identification in *Phalaenopsis* species/cultivars.

## Results

### Sequencing and *de novo* assembly

A total of 21,396,423 (30–76 base, 46.5% GC) high-quality PE reads were generated from Sanger/Illumina 1.9 sequencing, approximately 4Gb of sequence data was obtained from leaves of *P. aphrodite* subsp. *formosana*. These short sequence reads have been deposited at NCBI as SRA accession number SRX1253908 and SRX1253909. The reads with high quality bases above Q20 were more than 90% indicated a very good quality calls bases (Q20 means 1 error per 100 sequenced bases) and with per sequence quality score above 38 (if the most frequently observed mean quality below 27 equates to a 0.2% error rate) mean a good quality. The high-quality PE reads were used for *de novo* assembly to join into scaffolds step-by-step, based on paired-end information. Finally, 22,598 unigenes ( $\geq 100$  bp) were generated, with a final unigene N50 length of 2,047 bp and a total length of 29,062,410 bp (Table 1) (Fig 1).

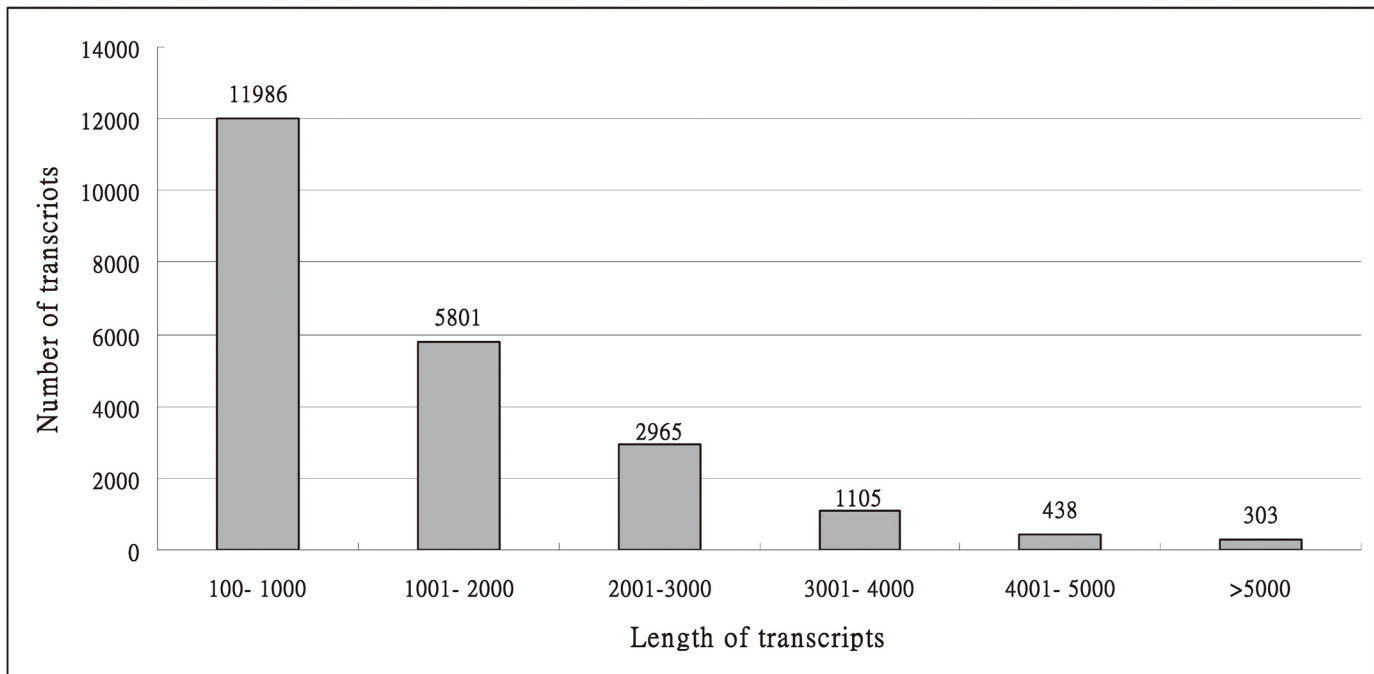
### Frequency and distribution of different types of EST-SSR loci

The 22,598 unigenes generated in this study were used to search potential microsatellites that were defined as perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 9, 6, 5, 5, and 4 repeats, respectively. There are potentially 1,439 EST-SSRs that can be found after SSR mining. The potential 1,439 EST-SSRs were further classified and shown in Table 2, the di-nucleotide repeats were shown to be the most abundant (741, 51.49%), followed by tri- (507, 35.23%), hexa- (121, 8.41%), tetra- (59, 4.10%), and penta- nucleotide repeats (11, 0.76%). Of di-nucleotide repeat motifs, the AG/CT di-nucleotide repeat was the most abundant motif (422, 29.33%), followed by TC/GA (205, 14.25%), and GC/GC was the rarest motif (0, 0%). Of tri-nucleotide repeat motifs, AGA/TCT was the most abundant motif (73, 5.07%), followed by GAA/TTC (59, 4.10%), AAG/CTT (56, 3.89%), and both ACG/CGT and TAC/GTA were the rarest motif (1, 0.07%) (Fig 2).

**Table 1. Summary of sequence assembly after Illumina sequence.**

Total number of unigenes	22,598
Sum of scaffolds	29,062,410 bp
Max scaffolds size	16,771 bp
Min scaffolds size	113 bp
Average scaffolds size	1,286 bp
N50	2,047

doi:10.1371/journal.pone.0141761.t001



**Fig 1. Summary distribution of the lengths of the 22,598 unigenes from raw reads (>100 bp, mean length = 1,286 bp, N50 = 2,047 bp, Min = 113 bp, Max = 16,771 bp).**

doi:10.1371/journal.pone.0141761.g001

### Preliminary validation for the developed EST-SSR loci

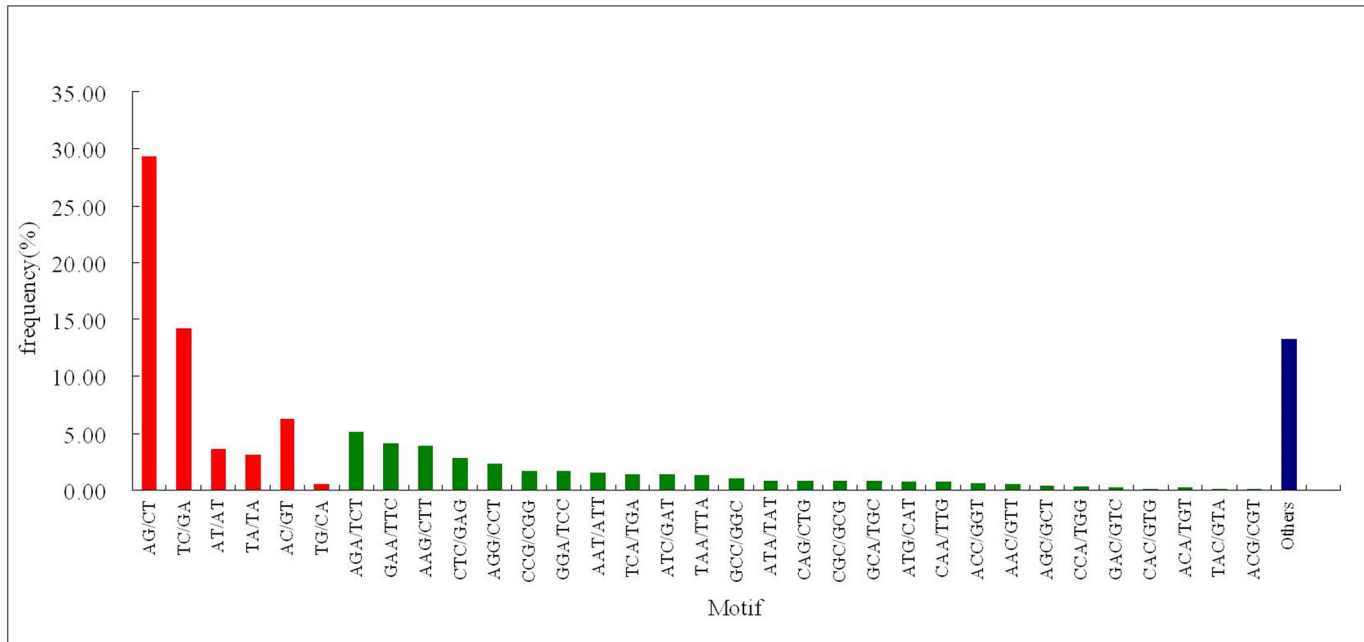
Most EST-SSRs are either di- (51.49%) or tri-nucleotide motifs (35.23%). The average number of potential EST-SSRs per unigene is 0.064. Of the detected EST-SSRs, 1,051 EST-SSRs were obtained for suitable primer designation by BatchPrimer3 (Table 2). The information of EST-SSR primers in this study is shown in S1 Table. Of the 421 primer pairs for tri-nucleotide motifs, 30 EST-SSR loci were randomly selected to evaluate the polymorphism and transferability across 22 native *Phalaenopsis* species that are representative germplasms for breeding most *Phalaenopsis* cultivars. Of these, ten EST-SSR loci were stably amplified, polymorphic, and transferrable SSR loci across 22 native *Phalaenopsis* species (S1 Fig). In total, 70 amplifying

**Table 2. Numbers of EST-SSRs identified and primer designation in *Phalaenopsis aphrodite* subsp. *formosana* in the study.**

	Numbers of SSRs
Total number of unigenes	22598
Total number of identified EST-SSRs	1,439 (1,051)*
Dinucleotide	741 (507)
Trinucleotide	507 (421)
Tetranucleotide	59 (29)
Pentanucleotide	11 (6)
Hexanucleotide	121 (88)

\*The number of suitable primer designation for PCR is shown in parentheses.

doi:10.1371/journal.pone.0141761.t002



**Fig 2. Summary distribution of EST-SSR loci repeat motifs.** The number of EST-SSR loci repeat motifs derived from Solexa transcriptome *de novo* sequencing. Di-, tri-nucleotide motifs, and others were highlighted in red, green and blue boxes, respectively.

doi:10.1371/journal.pone.0141761.g002

bands were detected by 10 primer pairs across 22 native *Phalaenopsis* species, and the number of amplifying bands per primer pairs ranged from 3 to 16, with an average of 7. The polymorphism information content (PIC) value across 22 native *Phalaenopsis* species ranged from 0.163 to 0.889, with an average of 0.588 (Table 3). The amplified products derived from EST-SSR PCR across 22 native *Phalaenopsis* species are shown to be one or two bands for most of SSR loci, such as the loci Pap-3222 (Fig 3a) and Pap-4825 (Fig 3b). Genetic similarity between 22 native moth orchids was evaluated by principal coordinate analysis (PCoA), and the three-dimensional representation provided by the plot shows a certain degree of separation between different species (Fig 4a). The resolution of the first, second and third axes show 24.87%, 20.27%, and 16.76% of the variance, respectively. Compared to 22 native taxa (Fig 5) by 10 polymorphic EST-SSR loci, genetic compositions among different species are obviously scattered between taxa but can be grouped at the three axes on Sections *Zebrinae*, *Phalaenopsis*, *Deliciosae*, and *Stauroglottis* (Fig 4a).

The best fit numbers of grouping is inferred as two by the  $\Delta K$  evaluation ( $\Delta K = 64.57$  when  $K = 2$ ) in the Bayesian clustering analysis. Two genetic components were estimated using assignment test and each taxon was either high percentage of component 1 (Blue color on Fig 4b) or component 2 (Orange color on Fig 4b), except three taxa including *Phalaenopsis amabilis* (Taxon abb. 15) belong to Section *Phalaenopsis*, *Phalaenopsis equestris* (Taxon abb. 21), and *Phalaenopsis lindenii* (Taxon abb. 22) belong to Section *Stauroglottis*, revealed an admixture genetic components. Based on the best fit number of grouping, 22 taxa of *Phalaenopsis* were divided into two groups (Fig 4b). The first group with high percentage of component 1 are included 20 taxa of *Phalaenopsis*, except 2 taxa, *Phalaenopsis equestris* and *P. lindenii* of Section *Stauroglottis* were grouped into the second group.

**Table 3. The evaluation of the polymorphism and transferability of 10 EST-SSR primers across 22 native *Phalaenopsis* species.**

SSR loci	Motif	Primer sequence (5'→3')	Size	No. of sample	No. of alleles	Heterozygosity (H)	PIC	Ta (°C)
Pap-1059	(CGC) <sub>8</sub>	F:AGAAGTTCGATTCTGCTATGA R:GGGAAGGAAAGAGAGATGTAA	150–135	22	3	0.091	0.163	55
Pap-1358	(TCA) <sub>8</sub>	F:CTGACGGAAGATTGAAAATTA R:TGGTCTTCGGTAAGAAGTATG	180–135	22	9	0.682	0.818	55
Pap-1520	(GCT) <sub>8</sub>	F:ATCAGCCTTCATGATCTTCTT R:AACTCTACCACCATCAGCAG	152–128	22	4	1.000	0.491	55
Pap-1904	(GCG) <sub>8</sub>	F:GGTGCATTTGAACTTGAATA R:CCCCAATTCTCAAATTTCTAT	183–180	22	3	0.000	0.163	55
Pap-3222	(GAG) <sub>8</sub>	F:GAGTATTGAATCCCCAAGTTT R:TTCAGAATCATCTTTCTCCTG	180–126	22	16	0.546	0.889	55
Pap-3268	(AAC) <sub>9</sub>	F:TAACTCGCCTTCTCGTCTTA R:TTTTTCCATTACTGTTTGATGA	160–145	22	6	0.182	0.601	55
Pap-3754	(TCC) <sub>8</sub>	F:AGTCTGAAGCTTCTTCTTGCT R:CAATATAGAGGAGGAGCAGGT	151–136	22	6	0.182	0.597	55
Pap-4282	(AGA) <sub>8</sub>	F:CTATGCTTCCCACAGAAACC R:CTGTGATCCACCATCCTTAC	213–185	22	11	0.818	0.862	55
Pap-4356	(AAG) <sub>8</sub>	F:CTATTGTGAAGAAGGAGGTGA R:CTGTTACTAACCTGCGTTGAT	158–310	22	4	0.182	0.496	55
Pap-4825	(CTC) <sub>8</sub>	F:ACCAGCTTCTACATTTCCAAT R:AAGATCTTCATTGATCCTTTTG	162–141	22	8	0.318	0.803	55
			Average		7	0.400	0.588	

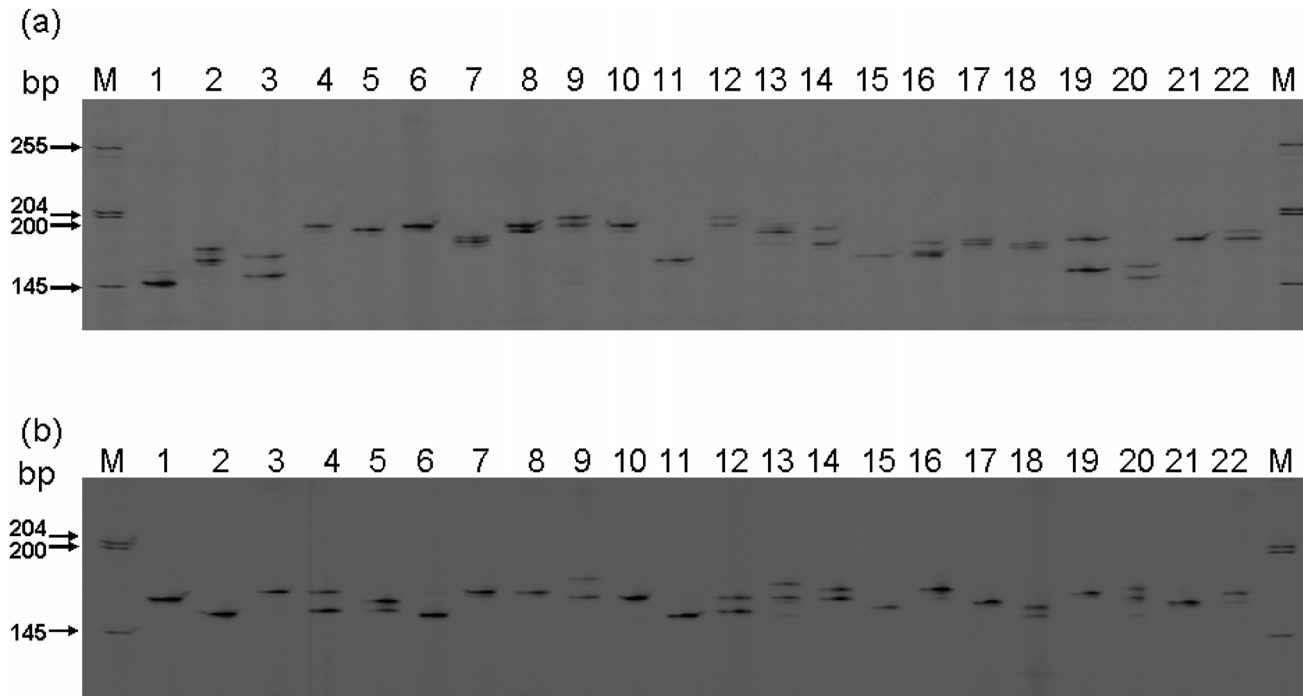
doi:10.1371/journal.pone.0141761.t003

### Application of validated EST-SSR loci for *Phalaenopsis* cultivars identification

The EST-SSRs studied were further used to identify 12 commercialized *Phalaenopsis* cultivars, including white, red and yellow floral color groups (Table 4). The morphological characters of the same floral color of plant materials studied are very similar. It is not easy to identify them based on either vegetative or reproductive characters (such as floral color, size, and morphology). Three validated polymorphic and transferable primer pairs (i.e., SSR loci Pap-3222, Pap-4825, and Pap-4282) for EST-SSR were selected to discriminate 12 commercialized *Phalaenopsis* cultivars. According to the amplified PCR products, more than two bands can be found within an individual (Fig 6). In white floral color group (Fig 7a–7d), each of cultivars can be identified according to both SSR loci Pap-3222 and Pap-4282 (Fig 6a). In red floral color group (Fig 7e–7h), each of cultivars can be identified by using SSR locus Pap-4825 (Fig 6b). In yellow floral color group (Fig 7i–7l), each of cultivars can be identified by using SSR locus Pap-3222 (Fig 6c). Using the aforementioned three EST-SSR markers, each of the 12 commercialized *Phalaenopsis* cultivars can be discriminated.

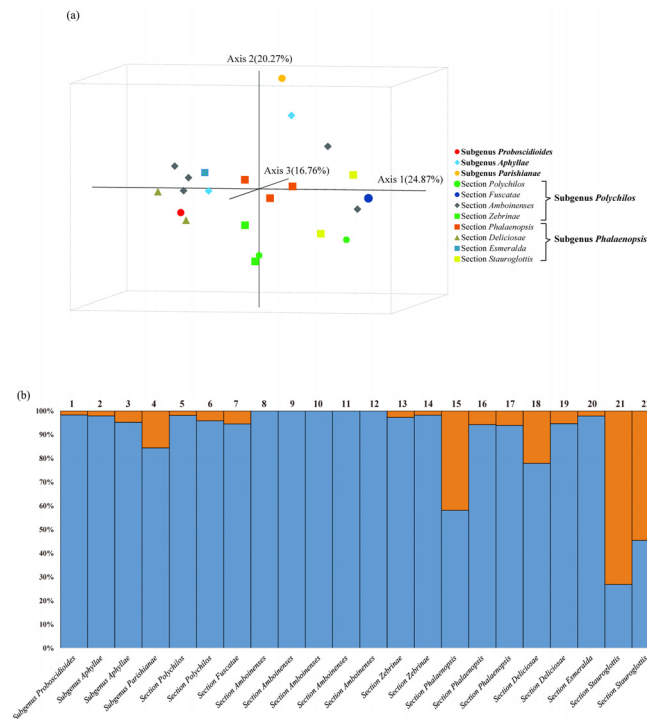
### Discussion

The genome size of *Phalaenopsis aphrodite* subsp. *formosana* was estimated to be approximately 1,300 Mb (2.81 pg/diploid genome), which is relatively small compared to other *Phalaenopsis* species [33,34]. After *de novo* assembly of deep sequencing data, the obtained sequence covered 29.1 Mb, or approximately 2.2% of the *P. aphrodite* subsp. *formosana* genome.



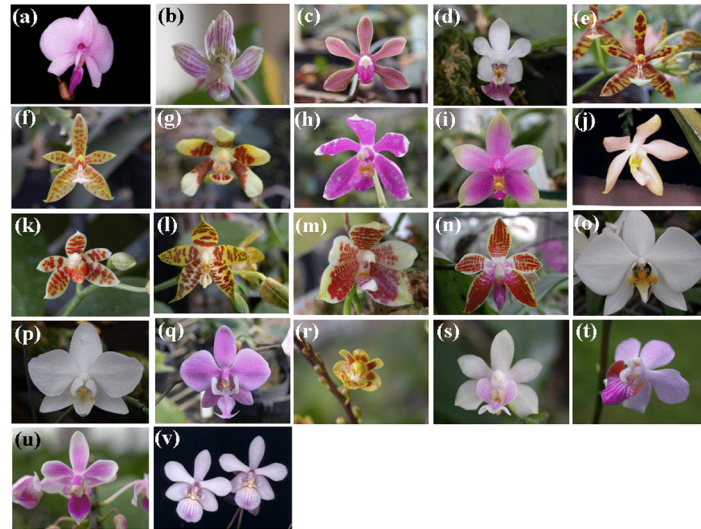
**Fig 3. The polymorphism of 22 *Phalaenopsis* species detect by SSR-PCR analysis.** (a) The polymorphism of *Phalaenopsis* taxa at Pap-3222 SSR locus. Lanes 1~22 represent 22 *Phalaenopsis* species listed in Table 5. (b) The polymorphism of *Phalaenopsis* taxa at Pap-4825 SSR locus. Lanes 1~22 represent 22 *Phalaenopsis* species listed in Table 5.

doi:10.1371/journal.pone.0141761.g003



**Fig 4. Estimated genotypic group structure for 22 moth orchids.** (a) Using the first three axes in principle coordinate analysis (PCoA). (b) Using the assignment test with Bayesian clustering analysis on the best fit numbers ( $K = 2$ ) of grouping based on 10 polymorphic microsatellite loci.

doi:10.1371/journal.pone.0141761.g004



**Fig 5. Images of 22 *Phalaenopsis* species.** Images (a)-(v) represent samples 1–22 shown in Table 5.

doi:10.1371/journal.pone.0141761.g005

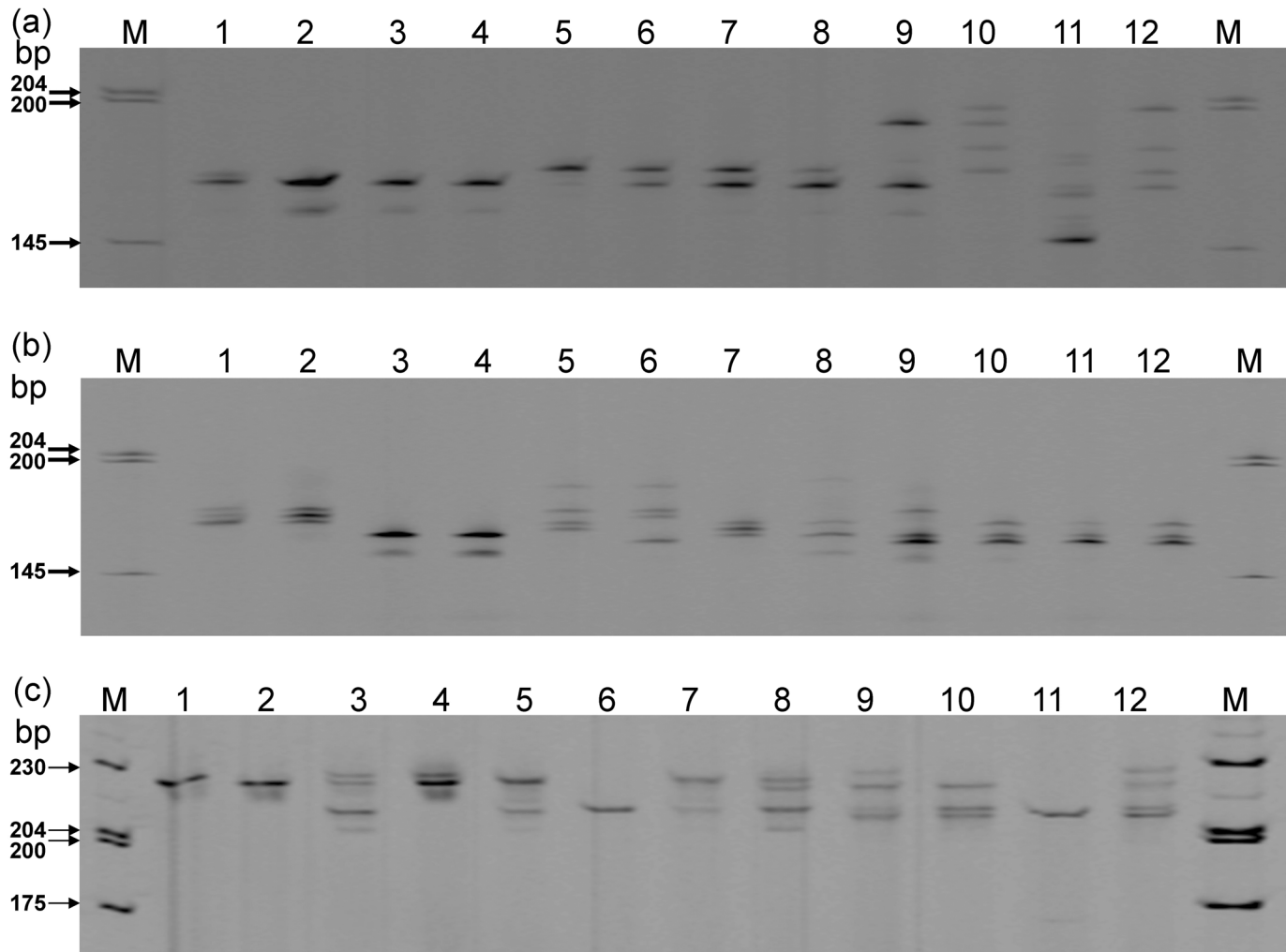
Excluding mono-nucleotide repeats that are not nearly useful for molecular markers [35], a total of 1,439 EST-SSR loci, including di-, tri-, tetra-, penta-, and hexa-nucleotide motifs, were detected across the transcriptome of 29.1 Mb in *P. aphrodite* subsp. *formosana*. Excluding mono-nucleotide repeat motifs, di-nucleotide repeat motifs of EST-SSRs were the most abundant type (51.49%) of microsatellites in the study. This result is consistent with EST-SSR studies in loblolly pine and spruce [36]. *Pinus contorta* [37], blueberry [38], rubber tree [39]. In contrast, tri-nucleotide motifs are also revealed to be the most abundant type of EST-SSRs in several studies [40,41,42,43], which are consistent with the maintenance of the open reading frame (ORF) coding. The untranslated regions (UTRs) are richer in SSRs than coding regions, particularly the 5'-UTRs [43,44,45]. Thus, the most abundant type of di-nucleotide repeats in the *P. ahrodite* subsp. *formosana* and some plants may result from their high proportion of 5'UTR-SSRs.

**Table 4. The twelve commercialized *Phalaenopsis* cultivars studied.**

Abb.	Cultivars name	Floral color	Flower size in diameter (cm)
1	<i>P. Sogo Yukidian</i> 'V3'	white	12–13
2	<i>P. Sogo Musadian</i>	white	12–13
3	<i>P. I-Hsin Diamond</i>	white	12–13
4	<i>P. Chainport Dorothy</i>	white	12–13
5	<i>P. Ruey Lih Beauty</i>	red	8–10
6	<i>P. Shih-Dong Red Rose</i> 'Fantasy Rose'	red	8–10
7	<i>P. Ruey-Lih Red Rose</i>	red	8–10
8	OX 1172	red	8–10
9	<i>P. Sogo Meili</i>	Yellow	6–7.5
10	Sogo F3005	Yellow	6–7.5
11	<i>P. Sogo Shito</i> 'Sogo F2999'	Yellow	6–7.5
12	<i>P. Sogo Sweet</i>	Yellow	6–7.5

doi:10.1371/journal.pone.0141761.t004

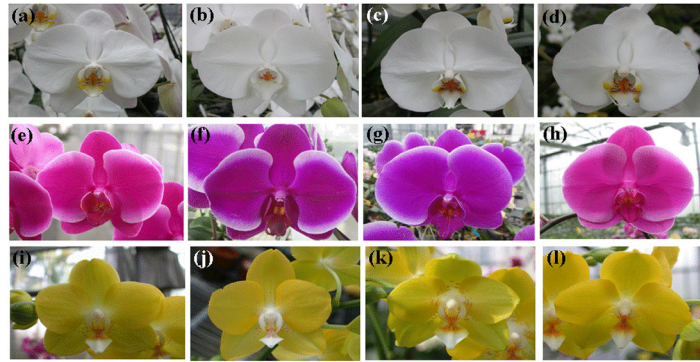




**Fig 6. The polymorphism of 12 *Phalaenopsis* varieties by SSR-PCR analysis.** The polymorphism of 12 *Phalaenopsis* varieties at (a) Pap-3222 SSR, (b) Pap-4825 SSR, and (c) Pap-4282 SSR loci. Lanes 1–12 represent 12 *Phalaenopsis* varieties/lines listed in Table 4. Lanes 1–4 represent four similar commercialized cultivars with white floral color; Lanes 5–8 represent four similar commercialized cultivars with yellow floral color; Lanes 9–12 represent four similar commercialized cultivars with red floral color.

doi:10.1371/journal.pone.0141761.g006

SSR occurs in the exon region, including a 5' UTR, CDS, and 3'UTR, on average every 20.22 kb in *P. aphrodite* subsp. *formosana*. The average density of EST-SSRs was likely shown to be relatively low frequency in *P. aphrodite* subsp. *formosana* compared to most other species; such as one EST-SSR is found every 14 kb in *Arabidopsis* [46], every 19 kb in rice [47], every 19.4 kb averaged across maize, rice, soybean, and wheat [48], and every 1.77 kb in castor bean [42]. Although the frequencies of EST-SSRs were shown to vary in different species, these results may not be accurate because of the different strategies that were used to mine the EST-SSRs. In this study, the EST-SSR mining parameters were set to search perfect di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of 9, 6, 5, 5, and 4 repeats, respectively. The parameters are relatively stringent compared to other studies; for example, the EST-SSR mining parameter in castor bean was set to identify perfect mono-, di-, tri-, tetra-, penta- and hexa-nucleotide motifs with a minimum of 10, 5, 4, 4, 4, and 4 repeat subunits,



**Fig 7. Images of 12 *Phalaenopsis* varieties.** Images (a)–(l) represent samples 1–12 shown in [Table 4](#).

doi:10.1371/journal.pone.0141761.g007

respectively. The study included mono-nucleotide repeats and a parameter with low repeat number (a cut-off value of 5) on di-nucleotide SSR mining resulting of the high average density of EST-SSRs (an EST-SSR every 1.77 kb) in castor bean [42]. Using the same SSR mining parameters, the frequency of SSR loci in *P. aphrodite* subsp. *formosana* is higher than that of loblolly pine and spruce, which on average have an EST-SSR every 49.8 kb [36]. Therefore, the average density of EST-SSRs cannot be compared among different species if the mining parameters are not identical. Because both di-nucleotide and tri-nucleotide repeats are the two most abundant types of microsatellites [35], the average density of SSRs is highly dependent on the parameter for the minimum repeat units of di- and tri-nucleotide repeat types of microsatellites.

Di-nucleotide repeat units in *P. aphrodite* subsp. *formosana*, AG motifs were the most frequent SSR motifs, about 29.33% of the total isolated loci and the lowest frequency (0%) in GC repeat di-nucleotide repeat units in the study. Similar results also can be found in SSR loci of other plants derived from either EST-SSR [39,42,43] or genome-wide SSR [42,43,45,49,50,51]. In tri-nucleotide repeat units, AGA, GAA, and AAG are the three highest microsatellite frequencies in *P. aphrodite* subsp. *formosana*. These results are consistent with those found in SSR loci of other plants derived from either EST-SSRs [39,42,43] or genome-wide SSRs [42,43,45,47,50,51]. According to this study and previous studies, both AG and AGA/GAA/AAG repeat units are shown to be high frequency SSR motifs in most of the clarified plants. AG or AGA/GAA/AAG repeat motifs in the 5'UTR upstream region of genes were thought to play significant roles in regulating gene expression and translation in *Arabidopsis* [52,53], and positive selection of AG and AGA/GAA/AAG repeat motifs respectively can be found in the 5'UTR and 5' coding region of *Arabidopsis* [43].

GC repeat units in low frequency of SSR motifs are shown in most SSR transcriptome-wide studies as previously described. GC-rich regions might be relatively stable, resulting in less replication slippage [54,55]. Furthermore, GC-rich and AT-rich motifs are respectively found in exon and intron regions for the splice site recognition in plant genes [56,57]. Additionally, the coding region of di-nucleotide SSR motif repeats occurs less frequently because of functional constraints, therefore di-nucleotide SSRs were preferentially concentrated in 5'- and 3' untranslated regions (UTR) [39,58,59], as well as both 5' and 3'UTR regions usually show AT-rich motifs, which is implicated in mediating RNA stability [60]. Overall, these patterns might lead to the low efficiency of GC repeat unit SSR motifs in plants. The low frequency of GC repeat

units of EST-SSR motifs has been revealed in various species, from yeast, plants, and vertebrates [35].

Deep sequencing clearly offers a rapid strategy of acquiring the sequences required to discover SSRs and to design specific primers to obtain useful SSR markers. Additionally, EST-SSR markers usually have a higher amplification efficiency, and are more likely to be transferable across species than SSR markers derived from non-coding regions of the genome [61,62,63,64].

To the present, there is no universal SSR markers across genus *Phalaenopsis*, even though several studies have focused on SSR mining in *Phalaenopsis* with transferable evaluation across part germplasm of the genus [24,25,27], or with transferable evaluation by using several cultivars [26]. To develop universal SSR markers for all commercialized *Phalaenopsis* cultivars, the transferability and polymorphisms of SSR loci cloned from EST-SSRs require validation. According to the molecular phylogeny of *Phalaenopsis* [6] and the orchid hybrid database at the Royal Horticultural Society (RHS), twenty-two native *Phalaenopsis* species were selected from the representatives of all subgenera and sections, which include most of the breeding parents for historical *Phalaenopsis* breeding programs. To develop universal SSR markers for all commercialized *Phalaenopsis* cultivars, EST-SSR markers were isolated and screened, resulting from having a higher amplification efficiency and are more likely to be transferable across species than those derived from non-coding regions of the genome [62].

Di-nucleotide repeat unit microsatellites always have larger repeat numbers and high levels of polymorphism in diverse plants [51,65,66,67]. Although the performance of higher polymorphism derived from di-nucleotide repeat units implied that these markers could be used efficiently, the higher efficiency of SSR-PCR stutter products will be easily found in di-nucleotide repeat units with larger repeat numbers [68,69]. Thus, tri-nucleotide, tetra-nucleotide or penta-nucleotide repeats usually amplify more faithfully by PCR than di-nucleotide repeats [70]. The stutter products result from the slipped-strand mispairings as a natural process of SSR mutation *in vivo* [71]. SSR markers will interfere with the high ratio of stutter products, especially in polyploidy plants [72]. Because *Phalaenopsis* cultivars have different ploids, including diploids, triploids and tetraploids [73], di-nucleotide repeat units of SSR motifs might not be suitable for plant identification. In a study of transcriptome-wide EST-SSR searching, 4–6 nucleotide repeat units of SSR motifs are at too low of a frequency to develop transferable and polymorphic SSR markers across the diversified breeding germplasm of *Phalaenopsis* cultivars; thus, tri-nucleotide repeat unit microsatellites are considered to be suitable SSR markers to identify *Phalaenopsis* cultivars.

Of 507 potential tri-nucleotide microsatellites, 421 were suitable for primer designation. Of these, 30 tri-nucleotide microsatellites were randomly selected to evaluate the transferability and polymorphism across the 22 native *Phalaenopsis* species which are usually used as parents for moth orchid breeding. The result showed that the transferability of the EST-SSRs across the 22 native *Phalaenopsis* species is approximately 33.33% (10/30). According to the PCoA results, the resolution of the first three axes explain 61.91% of the variation between species EST-SSR multilocus genotypes. The distribution of results in the three dimensional plot were clearly scattered between taxa, but can be grouped on cross related species of Sections *Zebrinae*, *Phalaenopsis*, *Deliciosae*, and *Stauroglottis*. Based on the Bayesian clustering analysis, 22 moth orchid divided into two groups and the Section *Stauroglottis* was separated out of others. This provides evidence of genetically distinct units in native moth orchids and potential molecular tools to identify commercialized cultivars/lines. The data were sampled from across five subgenera of the genus *Phalaenopsis* [1]. Previous systematic studies have indicated that this genus included members of two genera, *Phalaenopsis* and *Doritis* [74]. In the Orchidaceae, hybrids derived not only from different species but also different genera are continually crossed, and F1 hybrids from more advanced generations have been produced on a large scale [75]. In moth

orchids, there are over 30,000 *Phalaenopsis* cultivars registered in the RHS orchid hybrid database. The results show that the crossing barriers within orchidaceous plant genera are relatively low. Additionally, the amplified products derived from EST-SSR PCR for 22 native *Phalaenopsis* species are shown to be one or two bands, respectively, showing homozygosity and heterozygosity as described by several studies [76,77].

In the analysis of 12 commercialized *Phalaenopsis* cultivars, more than two bands may exist for an individual, as shown on Fig 6. The plant materials could be explained as polyploids as described by Diwan et al. [78]. This result is consistent with the chromosome karyotype of commercialized *Phalaenopsis* cultivars that are often triploid or tetraploid [79]. Additionally, 12 *Phalaenopsis* cultivars can be differentiated from one another based on the analysis of three SSR loci (Fig 6). Thus, the EST-SSR markers developed in this study could be efficiently used to differentiate closely related *Phalaenopsis* cultivars.

## Conclusions

The study shows that transcriptome analysis based on deep sequencing is a powerful tool to develop EST-SSR loci in non-model species. A total of 1,439 EST-SSR loci from *Phalaenopsis* species were obtained. After the preliminary validation of EST-SSR loci, about 33.33% EST-SSR markers are transferable across the *Phalaenopsis* breeding germplasm. These characterized and uncharacterized universal EST-SSR markers can be potentially applied to identify all of *Phalaenopsis* cultivars in the future.

## Materials and Methods

### Plant materials

Twenty-two *Phalaenopsis* taxa were obtained from wild populations and cultivated in the greenhouse at Kaohsiung District Agricultural Research and Extension Station in Taiwan by C. C. Tsai. Voucher specimens were deposited at the herbarium of the National Museum of Natural Science, Taiwan (TNM) and are listed in Table 5 and Fig 5. To test the transferability for commercialized *Phalaenopsis* cultivars, twelve commercialized varieties were collected for study and are listed in Table 4 and Fig 7.

### RNA extraction, cDNA library construction, sequencing, Data filtering, *de novo* assembly

For Illumina transcriptome deep sequencing, total RNA was extracted from the fresh leaves of *P. aphrodite* subsp. *formosana* using the RNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's protocol. RNA quantity and quality was verified using NanoDrop ND 1000 (Thermo Scientific, Hudson, NH, USA) and 2100 Bioanalyzer (Agilent Technologies), respectively. The cDNA library was constructed according to the manufacturer's instructions for the mRNA-Seq Sample Preparation Kit (Illumina Inc., San Diego, CA). The constructed paired-end library was prepared by using the Genomic Sample Preparation Kit (Illumina) according to the manufacturer's instructions. After validation on an Agilent Technologies 2100 Bioanalyzer, the library was sequenced using Illumina HiSeq™ 2000 (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. *De novo* transcriptome assembly for the high quality reads ( $Q < 20$ ) was performed using Trinity software [79].

### Isolation of EST-SSR loci and Primer Designation

SSR loci were isolated in all of the unigenes from *P. aphrodite* subsp. *formosana* with SciRoKo 3.4 software [80]. The searching parameters of SSR loci were set to identify perfect di-, tri-,

**Table 5. Names of the specimens, geographical distributions, and sources for the plant material used in the study.**

Taxon abb.	Taxa	Systematics <sup>a</sup>	Geographical distribution	Source <sup>b</sup>
1	<i>Phalaenopsis lowii</i> Rchb.f. <sup>b</sup>	Subgenus <i>Proboscidioides</i>	Myanmar and adjacent western Thailand	KDAIS KC-88
2	<i>Phalaenopsis minus</i> (Seidenf.) E. A. Christ.	Subgenus <i>Aphyllae</i>	Endemic to Thailand	KDAIS KC-227
3	<i>Phalaenopsis braceana</i> (J. D. Hook.) E. A. Christ.	Subgenus <i>Aphyllae</i>	Bhutan and China	KDAIS KC-289
4	<i>Phalaenopsis parishii</i> Rchb.f.	Subgenus <i>Parishianae</i>	Eastern Himalayas, India, Myanmar, and Thailand	KDAIS KC-316
5	<i>Phalaenopsis manni</i> Rchb.f.	Subgenus <i>Polychilos</i> Section <i>Polychilos</i>	Northeast India, Nepal, and China to Vietnam	KDAIS KC-22
6	<i>Phalaenopsis cornu-cervi</i> (Breda) Bl. & Rchb.f.	Section <i>Polychilos</i>	Northeast India and the Nicobar Islands to Java and Borneo	KDAIS KC-23
7	<i>Phalaenopsis fuscata</i> Rchb.f.	Section <i>Fuscatae</i>	Malaysia (Malay Peninsula), Borneo (West Koetai)	KDAIS KC-115
8	<i>Phalaenopsis pulchra</i> (Rchb. f.) Sweet	Section <i>Amboinenses</i>	Endemic to the Philippines (Luzon and Leyte)	KDAIS KC-17
9	<i>Phalaenopsis violacea</i> Witte	Section <i>Amboinenses</i>	Indonesia (Sumatra) and Malaysia (Malay Peninsula)	KDAIS KC-153
10	<i>Phalaenopsis micholitzii</i> Rolfe	Section <i>Amboinenses</i>	Philippines (Mindanao)	KDAIS KC-382
11	<i>Phalaenopsis maculata</i> Rchb. f.	Section <i>Amboinenses</i>	Malaysia (Pahang), East Malaysia (Sabah and Sarawak), and Indonesia (Kalimantan Timur)	KDAIS KC-49
12	<i>Phalaenopsis amboinensis</i> J. J. Sm.	Section <i>Amboinenses</i>	Indonesia (Molucca Archipelago and Sulawesi)	KDAIS KC-157
13	<i>Phalaenopsis inscriptiosinensis</i> Fowlie	Section <i>Zebrinae</i>	Endemic to Indonesia (Sumatra)	KDAIS KC-298
14	<i>Phalaenopsis comingiana</i> Rchb.f.	Section <i>Zebrinae</i>	Borneo (Sarawak and elsewhere on the island)	KDAIS KC-346
15	<i>Phalaenopsis amabilis</i> (L.) Blume	Subgenus <i>Phalaenopsis</i> Section <i>Phalaenopsis</i>	Widespread from Sumatra and Java to the southern Philippines, and east to New Guinea and Queensland, Australia	KDAIS KC-96
16	<i>Phalaenopsis aphrodite</i> subsp. <i>formosana</i>	Section <i>Phalaenopsis</i>	Taiwan	
17	<i>Phalaenopsis schilleriana</i> Rchb.f.	Section <i>Phalaenopsis</i>	Endemic to the Philippines	KDAIS KC-429
18	<i>Phalaenopsis chibae</i> Yukawa	Section <i>Deliciosae</i>	Endemic to Vietnam	KDAIS KC-488
19	<i>Phalaenopsis deliciosa</i> Rchb. f.	Section <i>Deliciosae</i>	Widespread from Sri Lanka and India to the Philippines and Sulawesi	KDAIS KC-255
20	<i>Phalaenopsis pulcherrima</i> (Lindl.) J. J. Sm.	Section <i>Esmeralda</i>	Widespread from northeast India and southern China throughout Indochina to Malaysia (Malay Peninsula), Indonesia (Sumatra), and East Malaysia (Sabah)	KDAIS KC-256
21	<i>Phalaenopsis equestris</i> (Schauer) Rchb.f.	Section <i>Stauroglottis</i>	Philippines and Taiwan	KDAIS KC-203
22	<i>Phalaenopsis lindenii</i> Loher	Section <i>Stauroglottis</i>	Endemic to the Philippines	KDAIS KC-119

<sup>a</sup> The systematic characterizations of *Phalaenopsis* are based on Christenson (2001).

<sup>b</sup> Plant materials were cultivated at the Kaohsiung District Agricultural Improvement Station (KDAIS), Taiwan; their voucher specimens were deposited at the herbarium of the National Museum of Natural Science, Taiwan (TNM).

doi:10.1371/journal.pone.0141761.t005

tetra-, penta-, and hexa-nucleotide motifs with a minimum of 9, 6, 5, 5, and 4 repeats, respectively. SSR motifs and their complementary SSR motifs were considered the same type of SSR motifs, such as those that were subsequently classified according to theoretically possible combinations, such as an AAG equivalent to CTT on a complementary strand. The specific primers for each of EST-SSR loci were separately designed using BatchPrimer3 developed by You et al. [81]

## DNA extraction and EST-SSRs PCR amplification

To validate the polymorphism and transferability of the EST-SSR markers derived from transcriptome deep sequencing, 22 native *Phalaenopsis* species and 12 commercialized cultivars were the plant materials respectively examined. One hundred milligrams of fresh leaves was ground in liquid nitrogen, and genomic DNA was extracted using the CTAB method [82]. For economic validation purposes, the designed forward primers for each of the SSR loci were elongated from the M13 (-21) 18 bp sequence (5' -TGTAACAACGACGGCCAGT-3' ) to inexpensively label PCR products as described by Schuelke [83]. PCR conditions and IRDye label procedure were referenced from Tsai et al [19]. The labeled PCR products were denatured in loading dye (10 mg/ml blue dextran in formamide), and separated by 6.5% polyacrylamide gel (19:1, 7 M urea) electrophoresis using LI-COR 4300 DNA analyzer (LI-COR, Lincoln, Nebraska USA). Allele sizes were determined using IRDye 700 size standards (50–350 bp, LI-COR). The experiments were repeated three times and only the target bands consisted with three separate experiment were used for genotyping.

## Statistical analysis

The degree of polymorphism, including the number of amplifying bands per primer pairs with an average and the polymorphism information content (PIC) value were calculated using PowerMarker version 3.25 [84]. A principle coordinate analysis (PCoA) was performed to evaluate the degree of separation between different species. The PCoA was conducted with GenAlEx ver. 6.4 [85]. To evaluate the assistance of genotyping group information, the genetic clustering algorithms based on Bayesian-clustering approach were using by using the program STRUCTURE ver. 2.3.4 [86]. The posterior probability of the genetic groups from 1 to 22 was estimated using the Markov chain Monte Carlo (MCMC) approach based on the admixture model with 20 separate runs for each possible group to estimate the stability. Each run contained of 1,000,000 burn-in steps followed by 10,000,000MCMCsteps. To evaluate the best fit number of grouping, the  $\Delta K$  method [87] by STRUCTURE HARVESTER v. 0.6.8 [88] was using.

## Supporting Information

**S1 Fig. The polymorphism of 22 *Phalaenopsis* species at ten characterized EST-SSR loci in the study.** Lanes 1~22 represent 22 *Phalaenopsis* species listed in Table 4.

(PDF)

**S1 Table. The information of EST-SSR primers of the study.**

(PDF)

## Acknowledgments

This research was supported by funding from the Ministry of Science and Technology, Taiwan, MOST 103-2321-B-067E-001 to CCT, partial from MOST 103-2621-B-110-001 to YCC, and from the China Medical University, Taichung, Taiwan to CHC.

## Author Contributions

Conceived and designed the experiments: CCT HCS HVW YCC C. Chou. Performed the experiments: CCT YCC C. Chou. Analyzed the data: CCT YCC C. Chou HVW. Contributed reagents/materials/analysis tools: YSL HCS C. Chang. Wrote the paper: CCT HCS HVW YCC C. Chou. Conceived of the study, edited the manuscript, and approved the final manuscript: CCT HCS HVW YSL C. Chang YCC C. Chou.

## References

1. Christenson EA (2001) *Phalaenopsis*. Timber Press, Portland, OR, USA, 330 pp.
2. Tanaka R, Kamemoto K (1984) Chromosome in orchids: counting and numbers. In: Arditti J. (Ed.), *Orchid Biology: Reviews and Perspectives*. Cornell University Press, Ithaca, New York.
3. Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, et al. (2006) The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* 23: 279–291. PMID: [16207935](#)
4. Tsai CC, Huang SC, Chou CH (2006a) Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcribed spacer of the nuclear ribosomal DNA. *Plant Syst Evol* 256: 1–16.
5. Tsai CC, Huang SC, Chou CH (2009) Phylogenetics, biogeography, and evolutionary trends of the *Phalaenopsis sumatrana* complex inferred from nuclear DNA and chloroplast DNA. *Biochem Syst Ecol* 37: 633–639.
6. Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH (2010a) Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the plastid and nuclear DNA. *Plant Syst Evol* 288: 77–98.
7. Tsai CC, Sheue CR, Chen CH, Chou CH (2010b) Phylogenetics and biogeography of the *Phalaenopsis violacea* (Orchidaceae) species complex based on nuclear and plastid DNA. *J Plant Biol* 53: 453–460.
8. Tsai CC, Huang SC, Huang PL, Chen FY, Su YT, Chou CH (2006b) Molecular evidences of a natural hybrid origin of *Phalaenopsis × intermedia* Lindl. *J Hort Sci Biotechnol* 81: 691–699.
9. Cai J, Liu X, Vanneste K, Proost S, Tsai WC, Liu KW, et al. (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet* 47: 65–72. doi: [10.1038/ng.3149](#) PMID: [25420146](#)
10. Fu YM, Chen WH, Tsai WT, Lin YS, Chou MS, Chen YH (1997) Phylogenetic studies of taxonomy and evolution among wild species of *Phalaenopsis* by random amplified polymorphic DNA markers. *Rept Taiwan Sugar Res Inst* 157: 27–42. (Chinese with English abstract)
11. Chuang HT (2002) Identification of some species in the genus *Phalaenopsis* to Taiwan and Philippine by using RAPD and ISSR molecular markers. *Mater Thesis, Graduate Institute of Agriculture, National Chiayi University, Chiayi, Taiwan*. (Chinese with English abstract)
12. Goh MWK, Kumar PP, Lim SH, Tan HTW (2005) Random amplified polymorphic DNA analysis of the moth orchids, *Phalaenopsis* (Epidendroideae: Orchidaceae). *Euphytica* 141: 11–22.
13. Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203–221. PMID: [3328815](#)
14. Hsu TW, Shih HC, Kuo CC, Chiang TY, and Chiang YC (2013) Characterization of 42 microsatellite markers from poison ivy, *Toxicodendron radicans* (Anacardiaceae). *Int J Mol Sci* 14: 20414–20426. doi: [10.3390/ijms141020414](#) PMID: [24129176](#)
15. Huang CL, Ho CW, Chiang YC, Shigemoto Y, Hsu TW, Hwang CC, et al. (2014) Adaptive divergence with gene flow in incipient speciation of *Miscanthus floridulus/sinensis* complex (Poaceae). *Plant J* 80: 834–847. doi: [10.1111/tpj.12676](#) PMID: [25237766](#)
16. Tsai CC, Wu PY, Kuo CC, Huang MC, Yu SK, Hsu TW, et al. (2014) Analysis of microsatellites in the vulnerable orchid *Gastrodia flavilabella*: the development of microsatellite markers and cross-species amplification in *Gastrodia*. *Bot Stud* 55:72
17. Liao PC, Tsai CC, Chou CH, Chiang YC (2012) Introgression between cultivars and wild populations of *Momordica charantia* L. (Cucurbitaceae) in Taiwan. *Int J Mol Sci* 13: 6469–6491. doi: [10.3390/ijms13056469](#) PMID: [22754378](#)
18. Chiou CY, Chiang YC, Chen CH, Yen CR, Lee SR, Lin YS, et al. (2012) Development and characterization of 38 polymorphic microsatellite markers from an economical fruit tree, the Indian jujube. *Am J Bot* 99: e199–e202. doi: [10.3732/ajb.1100500](#) PMID: [22539510](#)
19. Tsai CC, Chen YKH, Chen CH, Weng IS, Tsai CM, Lee SR, et al. (2013) Cultivar identification and genetic relationship of mango (*Mangifera indica*) in Taiwan using 37 SSR markers. *Sci Hort* 164:196–201

20. Ge XJ, Hsu TW, Hung KH, Lin CJ, Huang CC, Huang CC, et al. (2012) Inferring multiple refugia and phylogeographical patterns in *Pinus massoniana* based on nucleotide sequence variation and fingerprinting. *Plos One* 7: e43717. doi: [10.1371/journal.pone.0043717](https://doi.org/10.1371/journal.pone.0043717) PMID: [22952747](https://pubmed.ncbi.nlm.nih.gov/22952747/)
21. Ge XJ, Hung KH, Ko YZ, Hsu TW, Gong X, Chiang TY, et al. (2015). Genetic divergence and biogeographical patterns in *Amentotaxus argotaenia* species complex. *Plant Mol Biol Rep* 33: 264–280.
22. Hite JM, Eckert KA, Cheng KC (1996) Factors affecting fidelity of DNA synthesis during PCR amplification of d(C-A) n-d(G-T) n microsatellite repeats. *Nucleic Acids Res* 24: 2429–2434. PMID: [8710517](https://pubmed.ncbi.nlm.nih.gov/8710517/)
23. Ho CS, Shih HC, Liu HY, Chiu ST, Chen MH, Ju LP, et al. (2014) Development and characterization of 16 polymorphic microsatellite markers from Taiwan cow-tail fir, *Keteleeria davidiana* var. *formosana* (Pinaceae) and cross-species amplification in other *Keteleeria* taxa. *BMC Res Notes* 7:255. doi: [10.1186/1756-0500-7-255](https://doi.org/10.1186/1756-0500-7-255) PMID: [24755442](https://pubmed.ncbi.nlm.nih.gov/24755442/)
24. Young L G (2004) The study of microsatellite DNA identification in *Phalaenopsis*. Master Thesis, Institute of Forensic Science, Central Police University, Taoyuan Hsien, Taiwan. (Chinese with English abstract)
25. Han SY (2005) Molecular cloning and characterization of cDNA-SSRs in *Phalaenopsis*. Master Thesis, Department of Biology, National Cheng Kung University, Tainan, Taiwan. (Chinese with English abstract)
26. Zhang SM, Chen C, Chen FF, Wang T (2013) Analysis on Genetic Diversity of 16 *Phalaenopsis* Cultivars Using EST-SSR Markers. *J Plant Genet Resour* 14: 560–564. (Chinese with English abstract)
27. Hsu CC, Chung YL, Chen TC, Lee YL, Kuo YT, Tsai WC, et al. (2011) An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biol* 11: 3–13. doi: [10.1186/1471-2229-11-3](https://doi.org/10.1186/1471-2229-11-3) PMID: [21208460](https://pubmed.ncbi.nlm.nih.gov/21208460/)
28. Abdelkrim J, Robertson BC, Stanton JAL, Gemmell NJ (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46: 185–192. doi: [10.2144/000113084](https://doi.org/10.2144/000113084) PMID: [19317661](https://pubmed.ncbi.nlm.nih.gov/19317661/)
29. Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G, Wingfield M, et al. (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques* 46: 217–223. doi: [10.2144/000113085](https://doi.org/10.2144/000113085) PMID: [19317665](https://pubmed.ncbi.nlm.nih.gov/19317665/)
30. Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, et al. (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* 11: 569–586. doi: [10.1186/1471-2164-11-569](https://doi.org/10.1186/1471-2164-11-569) PMID: [20950470](https://pubmed.ncbi.nlm.nih.gov/20950470/)
31. Csencsics D, Brodbeck S, Holderegger R (2010) Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *J. Hered* 101: 789–793. doi: [10.1093/jhered/esq069](https://doi.org/10.1093/jhered/esq069) PMID: [20562212](https://pubmed.ncbi.nlm.nih.gov/20562212/)
32. Zalapa JE, Simon PW, Hummer KE, Bassil NV, Senalik DA, Zhu H, et al. (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor Appl Genet* 124: 87–96. doi: [10.1007/s00122-011-1689-2](https://doi.org/10.1007/s00122-011-1689-2) PMID: [21904845](https://pubmed.ncbi.nlm.nih.gov/21904845/)
33. Hsiao YY, Chen YW, Huang SC, Pan ZJ, Fu CH, Chen WH, et al. (2011) Gene discovery using next-generation pyrosequencing to develop ESTs for *Phalaenopsis* orchids. *BMC Genomics* 12: 360. doi: [10.1186/1471-2164-12-360](https://doi.org/10.1186/1471-2164-12-360) PMID: [21749684](https://pubmed.ncbi.nlm.nih.gov/21749684/)
34. Chen WH, Lin TY, Tsai CC, Tang CY, Kao YL (2013) Estimating nuclear DNA content within 50 species of the genus *Phalaenopsis* Blume (Orchidaceae). *Sci Hort* 161: 70–75.
35. Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10: 967–981. PMID: [10899146](https://pubmed.ncbi.nlm.nih.gov/10899146/)
36. Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007) Characterization of EST-SSRs in loblolly pine and spruce. *Tree Genet Genom* 3: 251–259.
37. Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11:180. doi: [10.1186/1471-2164-11-180](https://doi.org/10.1186/1471-2164-11-180) PMID: [20233449](https://pubmed.ncbi.nlm.nih.gov/20233449/)
38. Rowland LJ, Alkharouf N, Darwish O, Ogden EL, Polashock JJ, Bassil NV, et al. (2012) Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. *BMC Plant Biol* 12:46. doi: [10.1186/1471-2229-12-46](https://doi.org/10.1186/1471-2229-12-46) PMID: [22471859](https://pubmed.ncbi.nlm.nih.gov/22471859/)
39. Li D, Deng Z, Qin B, Liu X, Men Z (2012) *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13: 192. doi: [10.1186/1471-2164-13-192](https://doi.org/10.1186/1471-2164-13-192) PMID: [22607098](https://pubmed.ncbi.nlm.nih.gov/22607098/)
40. Blanca J, Canizares J, Roig C, Ziarsolo P, Nuez F, Pico B (2011) Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12:104. doi: [10.1186/1471-2164-12-104](https://doi.org/10.1186/1471-2164-12-104) PMID: [21310031](https://pubmed.ncbi.nlm.nih.gov/21310031/)



41. Niu SH, Li ZX, Yuan HW, Chen XY, Li Y, Li W (2013) Transcriptome characterization of *Pinus tabuliformis* and evolution of genes in the Pinus phylogeny. *BMC Genomics* 14:263. doi: [10.1186/1471-2164-14-263](https://doi.org/10.1186/1471-2164-14-263) PMID: [23597112](https://pubmed.ncbi.nlm.nih.gov/23597112/)
42. Qiu L, Yang C, Tian B, Yang JB, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol* 10: 278. doi: [10.1186/1471-2229-10-278](https://doi.org/10.1186/1471-2229-10-278) PMID: [21162723](https://pubmed.ncbi.nlm.nih.gov/21162723/)
43. Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, et al. (2004) Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. *Bioinformatics* 20: 1081–1086. PMID: [14764542](https://pubmed.ncbi.nlm.nih.gov/14764542/)
44. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. PMID: [12589540](https://pubmed.ncbi.nlm.nih.gov/12589540/)
45. Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30: 194–200. PMID: [11799393](https://pubmed.ncbi.nlm.nih.gov/11799393/)
46. Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156: 847–854. PMID: [11014830](https://pubmed.ncbi.nlm.nih.gov/11014830/)
47. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11: 1441–1452. PMID: [11483586](https://pubmed.ncbi.nlm.nih.gov/11483586/)
48. Gao L, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breed* 12: 245–261.
49. Temnykh S, Park WD, Ayers N, Cartinhour S, Hauck N, Lipovich L, et al. (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor Appl Genet* 100: 697–712.
50. Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman N, et al. (2011) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor Appl Genet* 124: 87–96. doi: [10.1007/s00122-011-1689-2](https://doi.org/10.1007/s00122-011-1689-2) PMID: [21904845](https://pubmed.ncbi.nlm.nih.gov/21904845/)
51. Zhang S, Tang C, Zhao Q, Li J, Yang L, Qie L, et al. (2014) Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in Foxtail millet [*Setaria italica* (L.) P. Beauv]. *BMC Genomics* 15:78. doi: [10.1186/1471-2164-15-78](https://doi.org/10.1186/1471-2164-15-78) PMID: [24472631](https://pubmed.ncbi.nlm.nih.gov/24472631/)
52. Martienssen RA, Colot V (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 293: 1070–1074. PMID: [11498574](https://pubmed.ncbi.nlm.nih.gov/11498574/)
53. Hulzink RJ, de Groot PF, Croes AF, Quaevlied W, Twell D, Wullems GJ, et al. (2002) The 5-untranslated region of the ntp303 gene strongly enhances translation during pollen tube growth, but not during pollen maturation. *Plant Physiol* 129: 342–353. PMID: [12011364](https://pubmed.ncbi.nlm.nih.gov/12011364/)
54. Hentschel CC (1982) Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. *Nature* 295: 714–716. PMID: [6276782](https://pubmed.ncbi.nlm.nih.gov/6276782/)
55. Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20: 211–215. PMID: [1741246](https://pubmed.ncbi.nlm.nih.gov/1741246/)
56. Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58: 473–483. PMID: [2758463](https://pubmed.ncbi.nlm.nih.gov/2758463/)
57. Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, et al. (2012) Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep* 1: 543–556. doi: [10.1016/j.celrep.2012.03.013](https://doi.org/10.1016/j.celrep.2012.03.013) PMID: [22832277](https://pubmed.ncbi.nlm.nih.gov/22832277/)
58. Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res* 10: 72–80. PMID: [10645952](https://pubmed.ncbi.nlm.nih.gov/10645952/)
59. Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, et al. (2008) Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC Genomics* 9: 287. doi: [10.1186/1471-2164-9-287](https://doi.org/10.1186/1471-2164-9-287) PMID: [18558001](https://pubmed.ncbi.nlm.nih.gov/18558001/)
60. Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol Cell Biol* 8: 113–126. PMID: [17245413](https://pubmed.ncbi.nlm.nih.gov/17245413/)
61. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genom* 270: 315–323.
62. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23: 48–55. PMID: [15629858](https://pubmed.ncbi.nlm.nih.gov/15629858/)
63. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, et al. (2010) *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Pomoea batatas*). *BMC Genomics* 11: 726. doi: [10.1186/1471-2164-11-726](https://doi.org/10.1186/1471-2164-11-726) PMID: [21182800](https://pubmed.ncbi.nlm.nih.gov/21182800/)

64. Garcia RAV, Rangel PN, Brondani C, Martins WS, Melo LC, Carneiro MS, et al. (2011) The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genetics* 12: 41–54. doi: [10.1186/1471-2156-12-41](https://doi.org/10.1186/1471-2156-12-41) PMID: [21554695](https://pubmed.ncbi.nlm.nih.gov/21554695/)
65. Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JC, et al. (2005) An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169: 1617–1630. PMID: [15654118](https://pubmed.ncbi.nlm.nih.gov/15654118/)
66. Zhang Z, Deng Y, Tan J, Hu S, Yu J, Xue Q (2007) A genome-wide microsatellite polymorphism database for the Indica and Japonica rice. *DNA Res* 14: 37–45. PMID: [17452422](https://pubmed.ncbi.nlm.nih.gov/17452422/)
67. Jia X, Zhang Z, Liu Y, Zhang C, Shi Y, Song Y, et al. (2009) Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Theor Appl Genet* 118: 821–829. doi: [10.1007/s00122-008-0942-9](https://doi.org/10.1007/s00122-008-0942-9) PMID: [19139840](https://pubmed.ncbi.nlm.nih.gov/19139840/)
68. Kunkel TA, Patel SS, Johnson KA (1994) Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc Natl Acad Sci USA* 91: 6830–6834. PMID: [8041704](https://pubmed.ncbi.nlm.nih.gov/8041704/)
69. Shinde D, Lai Y, Sun F, Arnheim N (2003) *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)<sub>n</sub> and (A/T)<sub>n</sub> microsatellites. *Nucleic Acids Res* 31: 974–980. PMID: [12560493](https://pubmed.ncbi.nlm.nih.gov/12560493/)
70. Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Amer J Hum Genet* 49:746–756. PMID: [1897522](https://pubmed.ncbi.nlm.nih.gov/1897522/)
71. Clarke LA, Rebelo CS, Gonçalves J, Boavida MG, Jordan P (2001) PCR amplification introduces errors into mononucleotide and dinucleotide sequences. *Mol. Pathol.* 54: 351–353. PMID: [11577179](https://pubmed.ncbi.nlm.nih.gov/11577179/)
72. Pfeiffer T, Roschanski AM, Pannell JR, Korbecka G, Schnittler M. (2011) Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). *J Hered* 102: 479–88. doi: [10.1093/jhered/esr024](https://doi.org/10.1093/jhered/esr024) PMID: [21576288](https://pubmed.ncbi.nlm.nih.gov/21576288/)
73. Griesbach RJ (1985) Polyploidy in *Phalaenopsis* orchid improvement. *The Journal of Heredity* 76: 74–75.
74. Sweet HR (1980) The genus *Phalaenopsis*. *The Orchid Digest*, Pomona, CA.
75. Neysa M, Stort S (1984) Sterility barriers of some artificial F1 orchid hybrids: male sterility. *Amer J Bot* 71: 309–318.
76. Cregan PB, Akkaya MS, Bhagwat AA, Lavi U, Jiang R (1994) Length polymorphism of simple sequence repeat (SSR) DNA as molecular markers in plants. In: Gresshoff PM (ed) *Plant genome analysis*. CRC Press, Boca Raton, 43–49.
77. Jiao Y, Jia HM, Li XW, Chai ML, Jia HJ, Chen Z, et al. (2012) Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201. doi: [10.1186/1471-2164-13-201](https://doi.org/10.1186/1471-2164-13-201) PMID: [22621340](https://pubmed.ncbi.nlm.nih.gov/22621340/)
78. Diwan N, Bouton JH, Kochert G, Cregan PB (2000) Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. *Theor Appl Genet* 101: 165–172.
79. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
80. Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23: 1683–1685. PMID: [17463017](https://pubmed.ncbi.nlm.nih.gov/17463017/)
81. You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinform* 9:253.
82. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.
83. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments—a poor man's approach to genotyping for research and high-throughput diagnostics. *Nat Biotechnol* 18: 233–234.
84. Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129. PMID: [15705655](https://pubmed.ncbi.nlm.nih.gov/15705655/)
85. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.
86. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578. PMID: [18784791](https://pubmed.ncbi.nlm.nih.gov/18784791/)
87. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14: 2611–2620. PMID: [15969739](https://pubmed.ncbi.nlm.nih.gov/15969739/)
88. Earl DA, von Holdt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conser Genet Resour* 4: 359–361.