Development and Identification of Expressed Sequence Tag-based SSR Markers Associated with the Heterostyly Trait in Primula forbesii

Cunquan Yuan¹, Zhiyi Qu¹, Huitang Pan, Tangren Cheng, Jia Wang, and Qixiang Zhang²

Beijing Key Laboratory of Ornamental Plants Germplasm Innovation & Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, School of Landscape Architecture, Beijing Forestry University, Beijing 100083, P.R. China

ADDITIONAL INDEX WORDS. bulked segregant analysis, BSA, simple sequence repeats, EST-SSR, pin style, S-locus, thrum style

ABSTRACT. Heterostylyous Primula forbesii is an important ornamental flower in China because of its long-lasting flowers and winter bloom. This study aimed to develop markers of expressed sequence tag–simple sequence repeats (EST-SSRs) that are associated with heterostyly and that can be used for molecular-assisted selective breeding in P. forbesii. We investigated 114,474 unigenes and identified 25,095 SSRs in P. forbesii. Dinucleotide repeats (46.14%), mononucleotide repeats (44.65%), and trinucleotide repeats (8.27%) were the most abundant SSRs. Among the 25,095 SSRs, 10,645 SSR primer pairs were successfully designed, of which 130 primer pairs were randomly selected for further amplification validation using eight accessions of P. forbesii; 98 pairs produced clear and stable polymerase chain reaction (PCR) products, and 28 pairs showed polymorphism. Bulked segregant analysis (BSA) was conducted for the F₁ population with respect to thrum style and pin style by scanning 28 polymorphic SSR primer combinations. One SSR marker, c64326, linked to the heterostyly trait at a genetic distance of ≈3.70 cM was identified. The marker c64326 was further validated in two populations with an accuracy of 97.92% and 90.63%. The novel and linked EST-SSR markers can be valuable resources for genetic diversity analysis, mapping, and marker-assisted breeding in P. forbesii.

Plants in the genus Primula have attracted the attention and love of the world’s flower gardening enthusiasts for their rich flower colors and distinctive flower patterns. Primula forbesii, which is characterized by the production of abundant attractive flowers and stable florescences lasting for 3 months in winter when other species remain dormant, is one of the most important ornamental floriculture crops in China. Primula forms either pin or thrum flowers. Pin flowers have a long style with the stigma at the corolla mouth and the anthers attached midway down the corolla tube, whereas thrum flowers have anthers that are positioned at the corolla mouth and a short style with the stigma attached at the midpoint of the corolla tube (Kappel et al., 2017). Heterostyly is an important characteristic to promote cross-pollination through the natural selection of reciprocal positions of the anthers and stigma, providing a basis for genetic variability. Heterostyly is controlled by the S-locus. The S-locus consists of a number of genes that control floral traits, consisting of at least three pairs of closely linked genes, G, P, and A, in which G determines the height of the style, A determines the height of the anther, and P determines the pollen size (Lewis and Jones, 1992) and the relative position of pollen (Pamela and Dowrick, 1956). Several S-locus–related genes, such as those indirectly involved in floral heteromorphy, have been identified from mutants and through phenotypic variations, including sepaloid (Li et al., 2008; Webster and Gilmartin, 2003) and Hose in Hose (Li et al., 2010). There are additional S-locus–related genes, such as Is/I, which controls pistil biochemical characteristics; Ip/ip, which controls pollen biochemical characteristics; Pp/pp, which controls pollen size; Mpp/mpp, which controls the pollen size dominant factor; Il, which is a lethal gene; G/g, which controls style length factor A; and Gm/gm, which controls style length factor B (Kappel et al., 2017; Lewis and Jones, 1992). Many studies have investigated heterostyly in Primula (Barrett and Shore, 2008; Kappel et al., 2017; Yoshida et al., 2011). Molecular markers and genes linked with the S-locus have been identified in Primula (Kappel et al., 2017). Manfield et al. (2005) obtained the first molecular marker, Primula vulgaris S-locus polymorphism 1 (PvSLP1) (8.8-kb-long DNA sequence), of the Primula S-locus. According to morphological differences in different growth stages, Webster and Gilmartin (2006) found that the gene expression of the S-locus was significantly different at different developmental stages, such as the expression of the G gene in 5-mm-long buds and the expression of the A gene in 11-mm-long buds. Li et al. (2010) obtained two RFLP markers, PvSL1 and PvSL2, which are closely linked to the P and A genes in the S-locus, via cloning complementary DNA and Southern blot analysis. The authors found that the genes that make up S-loci, such as P and A, may have multiple alleles. They also determined the genetic distance between G, P, and A. Li et al. (2008) found marker PvGlo linked to the S-locus and with a function similar to those of the B-type gene in the flower development model. Yoshida et al. (2011) found four markers linked to the Primula sieboldii S-locus via a quantitative trait loci (QTL) analysis and made the first attempt to predict and direct flower production based on
molecular markers. Nowak et al. (2015) sequenced ≈63% of the genome of *Primula veris*, detecting 19,507 genes and four sites linked with the S-locus, and found a primrose-specific gene, *PveGLO2*. Cuong Nguyen et al. (2016) identified the CYP734A50 gene as being exclusively expressed in the styles of S-morph plants. To date, great progress has been made in understanding the molecular-genetic basis of heterostyly; however, the intrinsic molecular mechanism is still unknown. In addition, the heterostyly trait is determined by the S-locus as a supergene, meaning that several individual genes reside at the S-locus; therefore, the exploration of more molecular markers linked to the S-locus in different species will help to identify the causal heterostyly genes.

Although great progress has been made in identifying molecular markers linked to S-locus in *Primula*, very few studies have focused on *P. forbesii*, and little genomic information is known for *P. forbesii*. In addition, the transferability of SSR markers between *Primula* species is exceptionally low. SSR markers are considered to be effective in genetic diversity analysis, linkage, and QTL mapping; marker-assisted selection; and so on (Rosazlina et al., 2015; Taheri et al., 2018). Expressed sequence tag SSRs have shown advantages because of their inexpensive development, high level of genetic diversity, and high transferability to related taxa (Taheri et al., 2018); however, we previously found that the versatility of SSR markers in *Primula* is poor. Among the 300 SSR primer pairs designed from *Primula maximowiczii* transcriptome data, only four primer pairs have been amplified successfully in *P. forbesii*. Moreover, we selected 20 published SSR primer pairs from *P. sieboldii* from the National Center for Biotechnology Information (NCBI) database (Ueno et al., 2009) to amplify with the *P. forbesii* DNA template; however, only one primer pair was successfully amplified (Z.Y. Qu, C.Q. Yuan, H.T. Pan, T.R. Cheng, J. Wang, and Q.X. Zhang, unpublished data). Wang (2016) performed hybrid identification using 20 SSR primers of *P. sieboldii* and identified four primer pairs that were capable of amplifying specific bands in parents and more than half of the progeny. The results showed a good specific locus with *P. sieboldii* and a poor specific locus with *Primula saxatilis*. These results suggest that the transferability of SSRs in *Primula* is poor, and specific SSR markers must be developed for this genus.

Considering the above, few SSR primers have been developed in *P. forbesii*. This study aimed to 1) develop novel EST-derived SSRs for *P. forbesii* and 2) identify EST-SSR markers that are associated with heterostyly traits in *P. forbesii*. The EST-SSR markers and the linkage identified in this study will serve as a valuable resource for genetic diversity analysis, parentage analysis, mapping, and marker-assisted breeding in *P. forbesii*.

**Materials and Methods**

**Construction of segregating populations and heterostyly investigation.** The heterostylos F1 segregating populations were established by crossing three thrum-style forms (T1/T2/T3) with three pin-style forms (P1/P2/P3) of *P. forbesii* (Table 1). Finally, six F1 segregating populations were obtained and used for heterostyly investigation (Table 1).

To screen the polymorphic SSR markers, eight individuals, including parents P2, T2, three pin-style and three thrum-style F1 individuals, were randomly selected. A segregating population, T2 × P2, comprising 115 F1 individuals, was selected as the screening population. A reciprocal cross population, including P2 × T2, comprising 96 (randomly selected 48 pin-style plants and 48 thrum-style plants) F1 individuals, was selected as the standard population to assess the accuracy and applicability of the developed linked markers.

**DNA extraction and BSA.** Total genomic DNA was extracted from young leaves using a new DNA extraction kit (Tiangen Biotech Co., Beijing, China). After quality examination, the DNA samples were stored at –20 °C until use. For the T2 × P2 population, the DNA was extracted from 115 F1 individuals and their parents. Eight pin-style plants and eight thrum-style plants were randomly selected from the 115 offspring of the F1 population, and their DNA concentrations were adjusted and equilibrated to construct a pin-style gene pool (Bp) and thrum-style gene pool (Bt) for BSA.

| Table 1. Phenotypic traits of three thrum-style and three pin-style parents of *Primula forbesii* that were used for the construction of segregating populations. |
|---------------------------------|---------------------------------|---------------------------------|-----------------|
| **Parent** | **Name** | **Description of characteristics** | **Traits icon** |
| **Thrum-style** | T1 | Style height lower than anthers; PURPLE- | |
| | T2 | VIOLET GROUP N80A; flower diameter 2.7 cm; | |
| | T3 | 6–8 flowers per round; 5–7 round; 30 × 30 cm crown; leaves dark green | |
| **Pin-style** | P1 | Style height higher than anthers PURPLE- | |
| | P2 | VIOLET GROUP N80A; flower diameter 2.7 cm; | |
| | P3 | 6–8 flowers per round; 5–7 round; 30 × 30 cm crown; leaves dark green | |
Table 2. Forward and reverse primer pair sequences, simple sequence repeat (SSR) type, annealing temperature (Tm), and length of the 28 polymorphic expressed sequence tag–SSR markers that were used for bulked segregant analysis in *Primula forbesii*.

| Code      | Primer sequence | SSR type | Tm (°C) | Length (bp) |
|-----------|-----------------|----------|---------|-------------|
| e68610    | ATAAGGAACCTCATACATCATATTTCCA | (A)10 | 56 | 166 |
| e80741    | ACCTGAGACATCCTGCTTCCA | (A)12 | 56 | 190 |
| e69493    | CTCTCGTTCCTGTGGTGAGG | (CT)8 | 56 | 107 |
| e34551    | ATTATCCACAAAGACGGCCCGG | (TA)8 | 56 | 103 |
| e75176    | AAAACCGTGTGATGTTGGGA | (AG)8 | 56 | 203 |
| e79939    | GCTGTTACAGAGGGAACCCACCA | (TA)10 | 56 | 154 |
| e107997   | TTTGAGATTTGTTTCTTGGAGT | (AG)6 | 59 | 249 |
| e131542   | TCAGTCTTCGTTGTTTCCA | (TA)5 | 56 | 182 |
| e98289    | GATGTTCAGACCTGCTGACT | (GC)6 | 56 | 209 |
| e4763     | ATCTGGAATGCTGTCAATCAT | (CA)5 | 56 | 169 |
| e4137     | GCCATATCGACCTGGAAGT | (CA)5 | 56 | 156 |
| e39084    | GGGCCATCCAGCTAAATCTTTC | (TC)6 | 54 | 150 |
| e89428    | GATTTGGTGATGAACTGAGG | (AC)5 | 56 | 117 |
| e24351    | AGCCACCAATTTCCCTGGTCT | (ATT)7 | 59 | 183 |
| e117883   | CAAGAGAGAGGGCAGGACAC | (AC)6 | 56 | 122 |
| e107377   | TGCTCGAGATGGTCAATCAT | (AC)5 | 56 | 180 |
| e107879   | AAGGGGAATTTTTCCCTTTCG | (AG)6 | 56 | 124 |
| e112336   | TTTGTTGTGATGAGGAGCTT | (TG)7 | 59 | 209 |
| e35718    | GTTACGCGACGCGATAGGTA | (AG)7 | 56 | 168 |
| e36556    | AACCTTGTGCTCCTGGAACA | (AG)6 | 56 | 199 |
| e66809    | CGTCGTGCTGCTGCTGCTAT | (AG)6 | 56 | 243 |
| e42125    | TACGGCAAAACAGAGTCG | (AT)6 | 56 | 150 |
| e46791    | GTTCCTCTCCCCTGTGATGA | (AT)6 | 56 | 187 |
| e93232    | AAGCTGCAAGAGACAGTGC | (AG)7 | 56 | 187 |
| e64326    | GGAATTTAATGTCTCTGATGCC | (CT)6 | 56 | 137 |
| e112228   | TTCTTGTTCTGGTCAAGGTT | (GTG)6 | 56 | 125 |
| e41380    | GCATCTATTGACTGTCTTGG | (CT)6 | 56 | 126 |
| e125436   | TAAACGGCAAGTTGGGACC | (CA)6 | 56 | 193 |
| e3007     | TGCCGCAAATATTGAAATACAAA | (T)6 | 56 | 174 |
| e99307    | GTACAGGTGCTTGGTGCGTCT | (CT)6 | 56 | 233 |
| e94361    | TCTACCCATCTTACCAACTTC | (CT)6 | 56 | 183 |

**Development and screening of polymorphic SSR markers.** Total RNA was extracted from pistils of *P. forbesii*, and then, magnetic beads with Oligo (dT) were used to enrich and isolate messenger RNA. The constructed libraries were sequenced using an Illumina (San Diego, CA) HiSeq 2000 platform, and de novo assembly was performed with the Trinity program (Grabherr et al., 2011). SSR loci were screened using MicroSatellite software (Thiel et al., 2003). The criteria for the SSR search were repeat stretches having a minimum of 10 repeat units for mononucleotide SSRs, 6 repeat units for dinucleotide SSRs, and 5 repeat units in cases of trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs. Primer pairs were designed using Primer Premier software (Primer 5; PREMIER Biosoft, Palo Alto, CA) with the following parameters: 1) a primer length of 10 to 26 base pairs (bp), with an optimum length of 16 bp; 2) an annealing temperature (Tm) of 55 to 65 °C and a Tm difference of less than 5 °C between the upstream primer and the downstream primer; 3) GC content of 40% to 70%; 4) a product size of 100 to 350 bp; and 5) the absence of primer dimers, mismatches, hairpin structures, and so forth. A total of 130 dinucleotide and trinucleotide repeat primer pairs were randomly selected for further verification and polymorphic screening.

All SSR primers were synthesized by Sangon Biotech Co. (Shanghai, China). T2 and P2 and six randomly selected F₁ individuals were used as DNA templates to screen the polymorphic SSR markers. PCR amplifications were performed in a final reaction volume of 20 μL, consisting of 10 μL of 2 × Taq PCR Master Mix (Biomiga, San Diego, CA), 1 μL of each primer (10 μM), 1 μL (20 ng μL⁻¹) of genomic DNA and 7 μL of double-distilled H₂O. The PCR program was as follows: 94 °C for 3 min; 35 cycles of 94 °C for 30 s, annealing temperature for 30 s, 72 °C for 1 min; and a final extension at 72 °C for 10 min, then held at 10 °C. The amplified products were detected on 8% polyacrylamide gel and visualized with the silver staining method. The polymorphism of the primers was further analyzed using an Illumina (San Diego, CA) HiSeq 2000 platform, and de novo assembly was performed with the Trinity program (Grabherr et al., 2011). SSR loci were screened using MicroSatellite software (Thiel et al., 2003). The criteria for the SSR search were repeat stretches having a minimum of 10 repeat units for mononucleotide SSRs, 6 repeat units for dinucleotide SSRs, and 5 repeat units in cases of trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs. Primer pairs were designed using Primer Premier software (Primer 5; PREMIER Biosoft, Palo Alto, CA) with the following parameters: 1) a primer length of 10 to 26 base pairs (bp), with an optimum length of 16 bp; 2) an annealing temperature (Tm) of 55 to 65 °C and a Tm difference of less than 5 °C between the upstream primer and the downstream primer; 3) GC content of 40% to 70%; 4) a product size of 100 to 350 bp; and 5) the absence of primer dimers, mismatches, hairpin structures, and so forth. A total of 130 dinucleotide and trinucleotide repeat primer pairs were randomly selected for further verification and polymorphic screening.

All SSR primers were synthesized by Sangon Biotech Co. (Shanghai, China). T2 and P2 and six randomly selected F₁ individuals were used as DNA templates to screen the polymorphic SSR markers. PCR amplifications were performed in a final reaction volume of 20 μL, consisting of 10 μL of 2 × Taq PCR Master Mix (Biomiga, San Diego, CA), 1 μL of each primer (10 μM), 1 μL (20 ng μL⁻¹) of genomic DNA and 7 μL of double-distilled H₂O. The PCR program was as follows: 94 °C for 3 min; 35 cycles of 94 °C for 30 s, annealing temperature for 30 s, 72 °C for 1 min; and a final extension at 72 °C for 10 min, then held at 10 °C. The amplified products were detected on 8% polyacrylamide gel and visualized with the silver staining method. The polymorphism of the primers was further analyzed using an Illumina (San Diego, CA) HiSeq 2000 platform, and de novo assembly was performed with the Trinity program (Grabherr et al., 2011). SSR loci were screened using MicroSatellite software (Thiel et al., 2003). The criteria for the SSR search were repeat stretches having a minimum of 10 repeat units for mononucleotide SSRs, 6 repeat units for dinucleotide SSRs, and 5 repeat units in cases of trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs. Primer pairs were designed using Primer Premier software (Primer 5; PREMIER Biosoft, Palo Alto, CA) with the following parameters: 1) a primer length of 10 to 26 base pairs (bp), with an optimum length of 16 bp; 2) an annealing temperature (Tm) of 55 to 65 °C and a Tm difference of less than 5 °C between the upstream primer and the downstream primer; 3) GC content of 40% to 70%; 4) a product size of 100 to 350 bp; and 5) the absence of primer dimers, mismatches, hairpin structures, and so forth. A total of 130 dinucleotide and trinucleotide repeat primer pairs were randomly selected for further verification and polymorphic screening.
by nondenaturing polyacrylamide gel electrophoresis. The primers that met the requirements were further analyzed in the offspring of the hybrid progeny.

**SCREENING OF SSR MARKERS LINKED TO THE HETEROSTYLY TRAIT.** BSA with respect to thrum style and pin style for population T2 × P2 (115 offspring) was conducted by scanning 28 polymorphic EST-SSR markers. The primer information is documented in Table 2. Based on the floral form and band data, the recombination rate between the marker and the target gene was calculated using the following formula:

\[
\text{Recombination rate} = \left( \frac{\text{number of recombinant individual}}{\text{total number of offspring}} \right) \times 100
\]

The percentage of the recombination rate was about estimated as the genetic linkage distance in centimorgans.

**SEQUENCE ANALYSIS OF LINKAGE MARKERS.** The products of the linkage markers were recovered and sequenced from the bands obtained with nondenaturing polyacrylamide gel electrophoresis, and three biological replicates were carried out. The sequence was analyzed using Basic Local Alignment Search Tool (Altschul et al., 1990) with the NCBI database.

**Results**

**CHARACTERIZATION OF EST-SSRS.** A total of 114,474 unigenes were screened for SSRs. The SSR screening resulted in the identification of 20,142 unigenes containing 25,095 SSRs using MISA and 3939 unigenes containing more than one SSR. Of the 25,095 SSRs, 11,580 (46.14%) were dinucleotide repeats, 11,204 (44.65%) were mononucleotide repeats, and 2075 (8.27%) were trinucleotide repeats. The three repeats represented most (24,859, 99.06%) of the EST-SSRs in *Primula forbesii*. A total of 55 repeat motifs were found among the 25,095 SSRs. Among the mononucleotide repeats, A/T was the most common (99.43%) motif. Among the dinucleotide repeats, AG/CT was the most common (74.46%) motif. Among the trinucleotide repeats, AAT/ATT was the most common (21.97%) motif (Table 3). The number of repeats of the different motifs varied from 5 to 24, although the vast majority of them were from 5 to 12. The maximum frequency of repeats was 10 (6007, 23.94%), followed by 7 (2825, 11.26%), 11 (2370, 9.44%), 8 (2256, 8.99%), 9 (2246, 8.95%), 5 (1385, 5.52%), 12 (1116, 4.45%), and >12 (2953, 11.77%) (Table 4). Most of the SSR region length was from 10 to 20 bp, with a maximum frequency (66.77%) of 10 to 15 bp, followed by 16 to 20 bp (30.20%), and with a length of more than 20 bp representing only 3.03% (Fig. 1).

**SCREENING OF THE POLYMORPHIC EST-SSRS.** Among the 25,095 SSRs, 10,645 SSR primer pairs were successfully

### Table 4. Repeat times of each repeat motif in simple sequence repeat loci among the transcriptome of *Primula forbesii.*

| Repeat type | Repeats of each repeat motif (no.) |
|-------------|-----------------------------------|
|             | 5       | 6       | 7       | 8       | 9       | 10      | 11      | 12      | 13–24   |
| Mononucleotide | —       | —       | —       | —       | —       | —       | 4964    | 2177    | 1113    | 2950    |
| Dinucleotide  | —       | 3373    | 2496    | 2230    | 2244    | 1043    | 192     | 2       | 0       |
| Trinucleotide | 1173    | 549     | 325     | 25      | 0       | 0       | 0       | 1       | 2       |
| Tetranucleotide | 191    | 11      | 4       | 0       | 0       | 0       | 0       | 1       | 1       |
| Pentanucleotide | 15      | 2       | —       | —       | —       | —       | —       | —       | —       |
| Hexanucleotide | 6       | 2       | —       | 1       | 1       | —       | —       | —       | —       |
| **Total**    | 1385    | 3937    | 2825    | 2256    | 2246    | 6007    | 2370    | 1116    | 2953    |
| **Ratio (%)**| 5.52    | 15.69   | 11.26   | 8.99    | 8.95    | 23.94   | 9.44    | 4.45    | 11.77   |

### Table 5. Construction of segregating populations and the \( \chi^2 \) test of the segregation ratio of pin-/thrum-style trait in six \( F_1 \) segregating populations of *Primula forbesii.*

| Population | Different types of offspring | Theoretical ratio | \( \chi^2 \) | \( P \) |
|------------|------------------------------|-------------------|-----------|-------|
| T1 × P1    | 156 84 72                   | 1:1               | 0.923     | 0.337 |
| T2 × P2    | 115 54 61                   | 1:1               | 0.426     | 0.514 |
| T3 × P3    | 100 47 53                   | 1:1               | 0.360     | 0.549 |
| P1 × T1    | 87 46 41                    | 1:1               | 0.281     | 0.592 |
| P2 × T2    | 192 101 91                  | 1:1               | 0.521     | 0.470 |
| P3 × T3    | 177 87 90                   | 1:1               | 0.051     | 0.822 |

\( \chi^2 \) = pin-style genotype of \( F_1 \) individual, \( T \) = thrum-style genotype of \( F_1 \) individual.

Fig. 1. Length distribution of simple sequence repeat loci among the transcriptome of *Primula forbesii*. bp = base pairs.
designed using Primer 5 software, of which 130 were randomly selected for further validation using eight accessions of *P. forbesii*. The results showed that 98 primer pairs produced clear and stable PCR products, with an amplification efficiency rate of 75.38%. Twenty-eight of the 103 primer pairs that produced PCR showed polymorphism (Table 2).

**Identification of SSRs linked with the heterostyly trait.** To further identify the EST-SSR markers linked with the pin-/thrum-style trait in *P. forbesii*, six F₁ segregating populations were constructed by crossing three pin-style (P1, P2, P3) with three thrum-style (T1, T2, T3) *P. forbesii*. The pin-/thrum-style traits of the F₁ population were investigated, and the observed segregation ratios and the theoretical segregation ratios were analyzed. The results showed that after the $\chi^2$ test, the segregation ratio of the probability value of each population was above 0.05 (Table 5), which meant that the segregation ratio was in line with the expected value of 1:1, indicating that the trait of heterostyly is indeed controlled by an S-locus supergenome. This phenomenon is consistent with previous studies showing that long-column flower-type plants (pin-style) are controlled by a homozygous genotype (s/s), whereas the short-column flower-type plants (thrum-style) are the result of a heterozygous genotype (S/s) (Lewis and Jones, 1992; Pamela and Dowrick, 1956). The results indicated that these groups are in line with the requirements of the segregating population.

Eight pin-style and eight thrum-style offspring were randomly selected from the T2 × P2 population, and the genomic DNA was extracted individually and mixed equally with the same style to separately construct the Bₚ and Bₜ. The Bₜ and Bₚ pools, combined with their parents, T2 and P2, were used for BSA. After screening with the 28 polymorphic primer pairs that we obtained previously, a pair of SSR primers (C64326) showed stable polymorphism between parents (T2 and P2) and gene pools (Bₜ and Bₚ) (Fig. 2). The obtained marker c64326 was further validated in the F₁ population of T2 × P2 (115 offspring), and the genetic linkage distance between the obtained primers and the control gene S-locus was calculated. The results showed that among the 115 offspring that we detected, three individuals did not amplify successfully. Among the 112 successful amplifications, the band type was consistent with the phenotype in 108 progeny amplifications; the band type was inconsistent with the phenotype in four progeny amplifications, suggesting that genetic recombination may have occurred among these four progeny. According to the statistical results, the interconnection distance (recombination rate) between the SSR molecular marker and the S-locus of heterostyly was calculated to be $\approx 3.70$ cM (Fig. 3).

**Verification of the application and the accuracy of the SSR marker C64326.** To verify and assess the applicability and accuracy of the SSR-linked marker c64326, segregating populations P2 × T2 (96 individuals) and T1 × P1 (96 individuals) were used. Forty-eight thrum-style and 48 pin-style offspring were randomly selected from each population separately. The results showed that the developed marker c64326 was applicable to the P2 × T2 and T1 × P1 populations, and the accuracy rates were 97.92% and 90.63%, respectively.

**Sequence analysis of the SSR-linked marker.** The product of the linkage marker was recovered and sequenced from the bands obtained through nondenaturing polyacrylamide gel electrophoresis, and three biological replicates were carried out. The sequencing results showed that the sequences
of three biological replicates were the same. The length of product that was amplified with pin-style DNA templates was 139 bp, whereas the lengths of products that were amplified with thrum-style DNA templates were 139 and 143 bp. In addition, the sequence comparison analysis results showed that the sequence of the amplified product that was 139 bp in both pin-style and thrum-style was identical (Fig. 4, Table 6). Two types of sequences were further compared, and sequence analysis results showed the following: 1) The amplified fragments had GA/CT two-base repeats; the fragment that was 143 bp in length had eight GA/CT repeats, and the fragment that was 139 bp in length had six GA/CT repeats. 2) The similarity of these two fragments was 97%, and the difference between these two fragments was 4 bp (Table 6). Previous studies have shown that pin-style plants are controlled by a homozygous genotype (s/s) and the thrum-style plants are the result of a heterozygous genotype (S/s) (Lewis and Jones, 1992; Pamela and Dowrick, 1956). Accordingly, we speculated that the 139-bp fragment in the SSR marker represented the homozygous loci and is closely linked to the recessive allele and that the 143-bp fragment represented the heterozygous loci and is closely linked to the dominant allele.

Discussion

In this study, 10,645 EST-SSR primer pairs were successfully designed from 25,095 SSR loci that were identified by screening 114,474 unigenes in *P. forbesii*. Dinucleotide repeats (46.14%), mononucleotide repeats (44.65%), and trinucleotide repeats (8.27%) were the most abundant SSRs. Then, we randomly selected 130 primer primers for further validation, and 28 primer primers showed polymorphism. Next, we used these 28 polymorphic SSR primer combinations to screen the F1 population with a BSA strategy. One SSR marker, c64326, which was linked to the heterostyly trait at a genetic distance of \(\sim 3.70 \text{ cM}\), was identified.

The versatility of SSR markers differs among families. Zhang (2010) asserted that the versatility of SSR markers among different species mainly depends on the conservatism of flanking sequences and the stability of sequence evolution. For species with good conservation, the interspecies versatility is better and vice versa. It was found that the versatility of different families, genera, and species was different, as seen in the high versatility of *Medicago*, *Helianthus*, and *Roegneria*; a medium degree of versatility of *Paederia*, *Oryza*, and *Zea mays*; and the low versatility of *Gossypium* species and *Lagenaria*. Our previous study showed that the versatility of SSR markers of *Primula* was poor. Among the 300 SSR primer pairs designed from *P. maximowiczii* transcriptome data, only four primer pairs were amplified successfully in *P. forbesii*. Moreover, we selected 20 published SSR primer pairs from *P. sieboldii* from the NCBI database (Ueno et al., 2009) to amplify with the *P. forbesii* DNA template; however, only one primer pair was successfully amplified (Z.Y. Qu, C.Q. Yuan, H.T. Pan, T.R. Cheng, J. Wang, and Q.X. Zhang, unpublished data). Wang (2016) performed hybrid identification using 20 SSR primers of *P. sieboldii* and identified four primer pairs that were capable of amplifying specific bands in parents and more than half of the progeny. The results showed a good specific locus with *P. sieboldii* and a poor specific locus with *P. saxatilis*. 

Fig. 3. Partial image of the polymerase chain reaction amplification products of simple sequence repeat primer c64326 in F1 offspring, thrum-style gene pool, pin-style gene pool, and their parents of *Primula forbesii*. P2 = pin-style-type parent, T2 = thrum-style-type parent, M = marker, B2 = thrum-style gene pool, BP = pin-style gene pool. Each gene pool was mixed with eight pin-style and eight thrum-style offspring of F1 individuals from the T2 x P2 population. “P” represents pin-style; “T” represents thrum-style. “ ▲ ” performed differently from the bands of parents. bp = base pairs.

Fig. 4. Polymerase chain reaction amplification products of the simple sequence repeat primer c64326 that was used for sequencing in *Primula forbesii*. Lanes 1 to 3 refer to the amplification products in three randomly selected pin-style individuals of the F1 population; Lanes 4 to 6 refer to the amplification products in three randomly selected thrum-style individuals of the F1 population. Lane M is the DNA size marker. Each band was recovered and sequenced. bp = base pairs.
Table 6. Sequence of the polymerase chain reaction amplification products of c64326 in pin-style and thrum-style genotypes of *Primula forbesii.*

| Template | Sequence | Length (bp) |
|----------|----------|-------------|
| Pin-style | TATTTGAGAGCTTTTGGGGGTTCTTTTTGTGCCTTTGCAACTCCACTATGAGTTGTCCTATTTCTGA | 139 |
| Thrum-style | TATTTGAGAGCTTTTGGGGGTTCTTTTTGTGCCTTTGCAACTCCACTATGAGTTGTCCTATTTCTGA | 139 |

bp = base pairs.

These results suggest that the transferability of SSR in *Primula* species is poor. Although great progress has been made in identifying of molecular markers linked to S-locus in *Primula,* very few studies have focused on *P. forbesii,* and few SSR markers have been developed to date. In this study, we screened and developed EST-SSRs derived from the expressed portion of the genome. These markers can be assayed as gene-based functional markers for diversity assessment, gene mapping, and marker-assisted selection. This report is the first comprehensive study on the development and analysis of a large set of SSR markers in *P. forbesii.*

*P. forbesii* presents a typical type of heterostyly. Our previous study found that bumble bee (*Bombus* sp.) pollination is much more efficient than artificial hybridization in *P. forbesii.* However, when a bee population is used, the ratio of thrum-style and pin-style plants in the parent population largely affects the efficiency of hybrid seed production and seed quality (Qu et al., 2017). Therefore, predicting the heterostyly using linkage markers at an early stage will be of great benefit for *P. forbesii* hybrid seed production. BSA generally detects a single multi-allelic locus; therefore, through the construction of gene pools, it can effectively eliminate the effects of the environment. For the quality traits that are usually controlled by one or a few major genes, such as those governing heterostyly, the use of SSR-BSA to carry out linkage markers can be done effectively and quickly through the prescreening of the gene pool, thereby reducing the amount of DNA and saving research time. In the present study, screening 130 SSR markers revealed 28 pairs of polymorphic SSR primers from the transcriptome of *P. forbesii.* By using 28 developed polymorphic SSR primers with BSA methods, we found an SSR marker, c64326, which is linked to the heterostyly trait; the genetic distance is ≈3.70 cM. The linkage marker developed here serves the purpose of predicting the column type at the seedling stage of *P. forbesii* and guiding the arrangement of the production according to the seedling scheme. This approach will ultimately shorten the production time and improve the efficiency of seed production. Moreover, it will provide a solid basis for cloning heterostyly-related genes.

**Literature Cited**

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Barrett, S.C.H. and J.S. Shore. 2008. New insights on heterostyly: Comparative biology, ecology and genetics, p. 3–32. In: V.E. Franklin-Tong (ed.). Self-incompatibility in flowering plants: Evolution, diversity, and mechanisms. Springer, Heidelberg/Berlin, Germany.

Cuong Nguyen, H., C. Kappel, B. Keller, A. Sicard, Y. Takebayashi, H. Breuninger, M.D. Nowak, I. Baehrle, A. Himmelbach, M. Burkart, T. Ebbing-Lohaus, H. Sakakibara, L. Altschmied, E. Conti, and M. Lenhard. 2016. Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. eLife 5:e17956.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Haohen, A. Girirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Kappel, C., H. Cuong Nguyen, and M. Lenhard. 2017. A short story gets longer: Recent insights into the molecular basis of heterostyly. J. Expt. Bot. 68:5719–5730.

Lewis, D. and D.A. Jones. 1992. The genetics of heterostyly, p. 129–150. In: S.C.H. Barrett (ed.). Evolution and function of heterostyly. Springer, Heidelberg/Berlin, Germany.

Li, J., B. Dudas, M.A. Webster, H.E. Cook, B.H. Davies, and P.M. Gilmartin. 2010. *Hose in Hose,* an S locus-linked mutant of *Primula vulgaris,* is caused by an unstable mutation at the *Globosa* locus. Proc. Natl. Acad. Sci. USA 107:5664–5668.

Li, J., M. Webster, B. Dudas, H. Cook, I. Manfield, B. Davies, and P.M. Gilmartin. 2008. The S locus-linked *Primula* homeotic mutant *seploid* shows characteristics of a B-function mutant but does not result from mutation in a B-function gene. Plant J. 56:1–12.

Manfield, I.W., V.K. Pavlov, J.H. Li, H.E. Cook, F. Hummel, and P.M. Gilmartin. 2005. Molecular characterization of DNA sequences from the *Primula vulgaris* S-locus. J. Expt. Bot. 56:1177–1188.

Nowak, M.D., G. Russo, R. Schlabach, H. Cuong Nguyen, M. Lenhard, and E. Conti. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. Genome Biol. 16:12.

Pamela, V. and J. Dowrick. 1956. Heterostyly and homostyly in *Primula obconica.* Heredity 10:219–236.

Qu, Z.Y., C.Q. Yuan, H.T. Pan, and Q.X. Zhang. 2017. Seed producing technology of heterostyous *Primula forbesii,* p. 680–685. In: Q.X. Zhang (ed.). Advances in ornamental horticulture of China. China For. Publ. House, Beijing, China.

Rahzilha, N., R. Jacobsen, M. Orgaard, and A.S. Othman. 2015. Utilizing next generation sequencing to characterize microsatellite loci in a tropical aquatic plant species *Cryptocoryne cordata var. cordata* (Araceae). Biochem. Syst. Ecol. 61:385–389.

Taheri, S., T.L. Abdullah, M.R. Hanafi, M. Sahebi, P. Azizi, and R.R. Shamshiri. 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Molecules 23:E399.

Thiel, T., W. Michalek, R.K. Varshney, and A. Granner. 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.

Ueno, S., Y. Yoshida, Y. Taguchi, M. Honjo, N. Kitamoto, I. Washitani, R. Ohsawa, and Y. Tsumura. 2009. Development of
120 microsatellite markers for *Primula sieboldii* E. Morren for linkage mapping. Conserv. Genet. 10:1945–1952.

Wang, Y. 2016. Backcross breeding of F₁ offspring between *Primula saxatilis* and *Primula sieboldii*. Beijing For. Univ., Beijing, China.

Webster, M.A. and P.M. Gilmartin. 2003. A comparison of early floral ontogeny in wild-type and floral homeotic mutant phenotypes of *Primula*. Planta 216:903–917.

Webster, M.A. and P.M. Gilmartin. 2006. Analysis of late stage flower development in *Primula vulgaris* reveals novel differences in cell morphology and temporal aspects of floral heteromorphy. New Phytol. 171:591–603.

Yoshida, Y., S. Ueno, M. Honjo, N. Kitamoto, M. Nagai, I. Washitani, Y. Tsumura, Y. Yasui, and R. Ohsawa. 2011. QTL analysis of heterostyly in *Primula sieboldii* and its application for morph identification in wild populations. Ann. Bot. 108:133–142.

Zhang, L.D. 2010. Regulation and polymorphism analysis on gene-associated SSRs in plants. Shanghai Jiao Tong Univ., Shanghai, China.