Molecular Biology Reports (2023) 50:3259–3269
https://doi.org/10.1007/s11033-022-08083-9

ORIGINAL ARTICLE

Development of EST-SSR markers based on transcriptome sequencing for germplasm evaluation of 65 lilies (*Lilium*)

Minmin Chen[1] · Gongping Nie[1] · Xin Li[1] · Liuyan Yang[1] · Youming Cai[1] · Yongchun Zhang[1]

Received: 16 April 2022 / Accepted: 1 November 2022 / Published online: 29 January 2023

Abstract

Background *Lilium* genus consists of approximately 100 species and numerous varieties, many of which are interspecific hybrids, which result in a complicated genetic background. The germplasm identification, genetic relationship analysis, and breeding of *Lilium* rely on exploiting genetic information among different accessions. Hence, an attempt was made to develop new EST-SSR markers and study the molecular divergence among 65 genotypes of *Lilium*.

Methods and results A total of 5509 EST-SSRs were identified from the high-throughput sequencing database of *L.* 'Elodie'. After primer screening, six primer pairs with the most abundant polymorphic bands were selected from 100 primer pairs. Combined with the other 10 reported SSR primers, a total of 16 pairs detected genetic information with an average PIC value of 0.7583. The number of alleles per locus varied from four to 33, the expected heterozygosity varied from 0.3289 to 0.9231, and the observed heterozygosity varied from 0.2857 to 0.5122. Based on the phylogenic results, 22 Asiatic hybrids (A), seven *Longiflorum* × Asiatic hybrids (LA), as well as two native species were grouped. Eighteen Oriental hybrids (O) and nine Oriental × Trumpet (OT) hybrids, four native species, one LO, and one LL (*L. pardalinum* × *L. longiflorum*) variety were grouped.

Conclusions Two major clusters were reported and a large number of genotypes were grouped based on UPGMA and STRU CTURE analysis methods. The PIC value as well as other parameters revealed that the EST-SSR markers selected were informative. In addition, the clustering pattern displayed better agreement with the cultivar's pedigree. The newly identified SSRs in this study provides molecular markers for germplasm characterization and genetic diversity for *Lilium*.

Keywords *Lilium* · EST-SSR · Genetic diversity · Polymorphism · Marker development

Introduction

Lilium spp., perennial ornamental plants belonging to the *Liliaceae* family [1], are widely distributed in the cold and temperate regions of the Northern Hemisphere [2]. In addition, *Lilium* are the second most popular cut flower in the world, more than 10,000 cultivars have been registered by worldwide breeders which prove its popularity. *Lilium* is taxonomically classified into seven sections: *Martagon*, *Pseudolirium* (2a, 2b, 2c, 2d), *Archelirion*, *Liriotypus*, *Sinomartagon* (5a, 5b, 5c), *Leucolirion* (6a, 6b) and *Daurolirion* [3]. While the *Lilium* cultivars are mainly separated into four groups: Asiatic (A), Oriental (O), Longiflorum (L), and Trumpet (T) [4], and they cross with each other to establish different hybrid groups (LA, OT, LO, OA, et al.). China is the central distribution area of *Lilium* with approximately 55 native species [5, 6] distributed in 29 provinces [1]. In addition to the ornamental value, bulbs of some *Lilium* species are rich in nutrients such as carbohydrates, proteins, polysaccharides, and low-fat content which can be used as food [7]. Modern studies also found that lily bulbs contain steroidal saponins [8], alkaloids, and flavonoids [9]. Three *Lilium* bulbs including *L. lancifolium*, *L. pumilum* and *L. brownii* var. *viridulum* have been recorded in the Chinese Pharmacopeia with the function of curing bronchitis, pneumonia, and cough [7].

Molecular marker-assisted breeding has been considered a novel way to shorten the breeding cycle of plants, but it requires a great deal of molecular marker

✉ Yongchun Zhang
saasflower@163.com

[1] Shanghai Key Laboratory of Protected Horticultural Technology, Forestry and Pomology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China

information. Thus the development of molecular markers can help to characterize and identify lines for further use inefficient breeding programs of *Lilium*. In addition, although the two systematic taxonomic classification systems have been widely accepted until now [3, 4], there are still some uncertain nodes in the phylogeny of *Lilium*, which may be due to the lack of appropriate markers. With the development of molecular biology techniques, SSR (Simple Sequence Repeats) markers have become widely employed in plant genetic research [10–12] as well as identification of hybrid authenticity due to the advantages of better stability, high polymorphism, codominance, and easy to use [13, 14]. The expressed sequence tag-simple sequence repeat (EST-SSR) markers, which were more accessible to the studies based on plant functional traits than other genomic SSR types [15], can be developed from the transcriptome data. One hundred and twelve EST-SSR markers had been developed in Oriental *L.* 'Sorbonne', of which 70 were selected to assess the genetic diversity in *Lilium* [16]. Twenty-six EST-SSRs had been developed in five *Lilium* species, in which 19 EST-SSRs were identified with higher polymorphisms among the Asiatic and Oriental cultivars [17]. One hundred and seventy-two EST-SSRs had been developed in *L. regale* by Yuan et al. [18]. One hundred and thirteen SNP markers and 292 common EST-SSR had been developed in four *Lilium* cultivars and tulips [19]. However, because of the complex genetic background of *Lilium* species, more molecular markers need to be developed.

The double-flowered lily cultivars are preferred by many people because of their showy appearances and the absence of pollen which can prevent pollen contamination. 'Elodie' was a special *Lilium* cultivar that had different petaloid organ types according to the petaloidy degree of whorl 3 when planted under the same conditions [20]. The unstable flowering phenotype may result in lower commercial value and ornamental quality, however, it also was a good cultivar to explore functional markers or key genes related to the degree of petaloidy of the stamens. Until now, 4099 EST-SSRs of *Lilium* can be found in the NCBI databases, however, compared with the enormous genome size, the molecular markers, genome, as well as specific functional genes in *Lilium* were still not enough. Thus we conducted the RNA-seq analysis by using the whole three organs as a sample [20]. In addition, developing SSR markers from the transcriptome database, analyzing their distribution and composition characteristics, and evaluating the polymorphism will contribute to the analysis of the genetic diversity of *Lilium*. In this study, we developed novel polymorphic EST-SSR markers from an Illumina transcriptome sequencing data set of *L.* 'Elodie' and tested the genetic diversity of 65 germplasm resources. Results can provide a theoretical basis for *Lilium* germplasm identification and genetic relationship analysis of *Lilium*.

Materials and methods

Transcriptome sequencing

Whole three of the petaloid stamen, partially petaloid stamen, and normal stamens of *L.* 'Elodie' were sampled for the transcriptome sequencing. RNA was extracted using the Trizol reagent (Invitrogen, CA, USA), and an RNA-Seq data set containing 1078,058,364 raw reads was generated by Hangzhou Lianchuan Biotechnology Co., Ltd using an Illumina Hiseq 4000 platform. Reads of the three separated libraries were mixed together to assemble and finally normalized to obtain unigenes. The transcriptome data have been submitted to the NCBI Sequence Read Archive database under the accession number PRJNA680629 [20] and the data were analyzed as background data.

Materials

A total of 65 germplasm resources were selected for genetic diversity analysis including 22 Asiatic varieties (A), 18 Oriental varieties (O), seven LA varieties, nine OT varieties, one LL variety, one LO variety, six native species and one variety (Table 1, Supplementary file 1). Six cultivars ('Matrix', 'Roselily Elena', 'Heartstrings', 'Zambesi', 'Fusion', and 'S-1') were selected for primer screening.

Total genomic DNA was extracted from leaves with a plant genomic DNA extraction kit (Tolo Biotech., Shanghai) according to the manufacturer's protocol. The DNA preparations were quantified on 0.8% agarose gels and the DNA concentration was examined using a Nano-Drop Lite spectrophotometer (Thermo Scientific, Massachusetts, USA).

SSR locus identification and primer design

The SSR screening was performed using the MicroSatellite (MISA, http://pgrc.ipk-gatersleben.de/misa/misa.html) [21] within the identified unigenes in the transcriptome, search with two, three, four, five, and six nucleotide repetitive unigenes. The microsatellite characteristics were analyzed statistically by Excel 2007. Primer 3.0 [22] was used for the obtained SSR sequences to design primers batch, and ensure the high amplification efficiency of primers. Primer selection criteria: the repeat unit was two, three, four, and five bases and only one repeat type, the fragment length was greater than 150 bp and less than 300 bp. 20–23 bp, TM value of 60 degrees, the number of base repeats was less than or equal to four. There were no two consecutive A/T bases at the 5′ and 3′ ends, and no repeated sequence in the primer.

Table 1 The germplasm collections of *Lilium* used in this study (A germplasm resource information is missing in this table. I will send you the complete table by email separately)

Sample	Ploidy	Source	Sample	Ploidy	Source
Trendy Savannah (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Entertainer (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Fata Morgana (A)	2n = 2x = 24	Hongyue Horticultural Corporation	After Eight (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Matrix (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Sunny Martinique (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Trendy Havana (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Sunny Keys (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Golden Matrix (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Marlon (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tribal Dance 600 (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Roselily Monica (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Trendy Dakota (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Roselily Natalia (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tiny Padhye (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Reeleeze (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tiny Bee (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Roselily Samantha (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tiny Rocket (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Souvenir (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tiny Invader (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Sorbonne (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Tiny Double You (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Roselily Thalita (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Pink Giant (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Viviana (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Elodie (A)	2n = 3x = 36	Hongyue Horticultural Corporation	Siberia (O)	2n = 2x = 24	Hongyue Horticultural Corporation
Levi (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Love Story (O)	/	Hongyue Horticultural Corporation
Lancifolium Flore Pleno (A)	2n = 3x = 36	Hongyue Horticultural Corporation	Black Beauty (OT)	2n = 2x = 24	Hongyue Horticultural Corporation
Tresor (A)	2n = 4x = 48	Hongyue Horticultural Corporation	Zambesi (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Pink Flight (A)	2n = 4x = 48	Hongyue Horticultural Corporation	Touchstone (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Pearl Frances (A)	2n = 4x = 48	Hongyue Horticultural Corporation	Palazzo (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Heartstrings (A)	2n = 2x = 24	Hongyue Horticultural Corporation	Robina (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Tribal Kiss (A)	–	Hongyue Horticultural Corporation	Manissa (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Morpho Pink (A)	–	Hongyue Horticultural Corporation	Conca D'Or (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Fusion (LL)	2n = 2x = 24	Hongyue Horticultural Corporation	Profundo (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Golden Stone (LA)	2n = 2x = 24	Hongyue Horticultural Corporation	Competition (OT)	2n = 3x = 36	Hongyue Horticultural Corporation
Bach (LA)	2n = 3x = 36	Hongyue Horticultural Corporation	White Triumph (LO)	2n = 3x = 36	Hongyue Horticultural Corporation
Albufeira (LA)	2n = 3x = 36	Hongyue Horticultural Corporation	S-1	2n = 3x = 36	Zhejiang Province
Dynamix (LA)	2n = 3x = 36	Hongyue Horticultural Corporation	*L.* wenshanense	2n = 2x = 24	Wenshan Baihe

Table 1 (continued)

Sample	Ploidy	Source	Sample	Ploidy	Source
Armandale (LA)	2n = 3x = 36	Hongyue Horticultural Corporation	*L.* sargentiae Wils	2n = 2x = 24	Tongjiang Baihe
Methone (LA)	–	Hongyue Horticultural Corporation	*L.* henryi Baker	2n = 2x = 24	Hubei Baihe
Beau Soleil (LA)	–	Hongyue Horticultural Corporation	*L.* leucanthum (Baker) Baker	2n = 2x = 24	Yichang Baihe
Roselily Isabella (O)	2n = 2x = 24	Hongyue Horticultural Corporation	*L.* leucanthum var. centifolium	2n = 2x = 24	Ziji Baihe
Roselily Elena (O)	2n = 3x = 36	Hongyue Horticultural Corporation	*L.* lancifolium Thunb.	2n = 3x = 36	Juandan Baihe
Sunny Azores (O)	2n = 2x = 24	Hongyue Horticultural Corporation			

PCR reaction system and electrophoresis detection

Then 100 primer pairs (Supplementary file 2) were selected for the initial screening to ensure the polymorphism as well as the optimal annealing temperature of each pair. The PCR reactions were performed in Thermo Cycler ABI2720 (Applied Biosystems, Foster City, California, USA) following the standard protocol of the 2 × Master mix (Tsingke Biotechnology Co., Ltd.) in a final volume of 20 µL, which containing 1 µL of genomic DNA, 0.15 µL forward primer (10 µM), 1.2 µL reverse primer (10 µM), 10 µL of 2 × Tsingke Master mix, 6.45 µL of deionized water and 1.2 µL Premix Taq™ Hot Start Version (TaKaRa, China). The PCR procedure was as follows: 5 min of initial denaturation at 95 °C, 35 cycles of 94 °C for 30 s, annealing temperature from 54 °C to 60 °C for 30 s, and 72 °C for 30 s; followed by a final 5 min extension step at 72 °C and a 4 °C holding temperature. The PCR products were separated by 3% agarose gel electrophoresis. Primer pairs that amplified strong bands within the expected size were selected for the second polymorphic screening of the PCR fluorescent labeling technique. The amplification system and procedure were the same as the first time. Then, the PCR products with distinct bands were run on an ABI 3730 Genetic Analyzer, and data was obtained using GeneMapper 4.1 software (Applied Biosystems, Foster City, CA, USA).

EST-SSR analysis, cluster analysis, and population structure analysis

Screened primers with better polymorphism and higher specificity were used for the diversity analysis in the 65 varieties combined with the other 10 previously reported primers [18, 23] (Table 2). Gene Mapper 4.1 software was used to conduct the fragment analysis. The size of the acquired SSR marker was converted by Convert 1.31 [24]. POPGEN version 1:32 program [25] was conducted to generate the observed number of alleles (Na), effective number of alleles (Ne), observed heterozygosity value (Ho), expected heterozygosity value (He), polymorphism information content (PIC). A dendrogram of an UPGMA (unweighted pair-group method with arithmetic means) cluster analysis was generated in population genetic software of Populations 1.2.30 [26]. ITOL V6.5.2 (http://itol.embl.de/) [27] was used to modify the cluster picture.

The population structure was assessed in a model-based approach using STRUCTURE 2.3.4 [28]. The run length was set to 200,000 MCMC (Markov chain Monte Carlo) using correlated allele frequencies under an admixture model. Structure analysis was carried out by setting the number of sub-populations (k) from k = 1 to k = 10. The K value is the corresponding K value when the likelihood value LnP(D) is maximum. If there is no obvious turning point in LnP(D), the method combining LnP(D) with a specific statistic ΔK is needed to determine the most appropriate K value [29]. When ΔK is the maximum value, the corresponding K value is the best. The results of 10 repeated operations of the best K value were calculated with the online software STRUCTURE HARVESTER [30].

Results

SSR identification and primer polymorphism analysis of the transcriptome data

Totally, 5509 potential EST-SSR motifs were identified and used for the mining of EST-SSRs. The SSR distribution characteristics of the two, three, four, five, and six nucleotide repeat motifs were shown in Supplementary file 3 and Supplementary file 4. The proportion of different

Table 2 Characteristics of 16 polymorphic EST-SSR markers in *Lilium*

Number	Repeat motif	Primer sequence (5′–3′)	Product (bp)	Tm (°C)	Gene bank accession number/References
EST-SSR45	$(GCG)_5$	F: GTCTTGCGAGCGTGTCAATA R: CCATCCCTACATCAAGACCG	204	60	OP573464
EST-SSR59	$(GGC)_6$	F: ACAATCAGTACCTGCGGACC R:CAGCTCCTTGTACCTCAGCC	137	60	OP573465
EST-SSR71	$(CTC)_6$	F: GGCGGAATAAACGAAACTCA R: CTCTCGGAGCTCCTGTATGG	241	60	OP573466
EST-SSR77	$(CGAG)_7$	F: CCCAACTCAAGAAACGAAGC R: AAGATGAGGCGATTGACAGC	235	60	OP573467
EST-SSR84	$(CGC)_5$	F: CGGAAGACGACCTTCAAGAC R: ATTCTCAGCAGCGACGATCT	131	59	OP573468
EST-SSR100	$(CAC)_5$	F:TCTGTGAAGAATGCTGTCGG R:CAGTAACAGGCAGCCAAACA	136	59	OP573469
EL42	$(GCA)_4$	F:AAGCATGCTGAGCTGTTGTCAG R:CTGCTTGAGTTGGTGTTGTTCG	150–180	61	Lee et al. [6]
ZJU04	$(CT)_8$	F:CCGCGTGTCTTCACTCCTAT R:GGGTGTCTTCCATCCTTTCA	150–172	58	Du et al. [23]
IVFLMRE118	$(TCTCCC)_4$	F:CTCCCTTTCCCTCTCCTTG R:GCTCCGACGATAAGTACCG	200–248	56	Yuan et al. [18]
IVFLMRE381	$(GAG\cdots GAA)_8$	F:GCGGGATCTCGCAGTTCT R:GTCGCAGTCTTCAAAGTCGT	272–296	57	Yuan et al. [18]
IVFLMRE100	$(AACCCT)_5$	F:CTCCTTCCCCAGAAAACCA R:TGACTAAAATGAAGAGGACGG	182–200	54	Yuan et al. [18]
IVFLMRE141	$(TTAGGG)_4$	F:AAAACTCAACACGGATCACA R:GCCATCCAATTCTTCTCCTA	143–173	57	Yuan et al. [18]
IVFLMRE107	$(AG)_{12}$	F:TGTAACCCTTGACATAACCAT R:ATAGCCCTATATCCTGTCACT	176–214	55	Yuan et al. [18]
IVFLMRE136	$(ACGCCG)_3$	F:TCTCTTCGTCTTCCATTGTG R:ATCCTTGCTCACCTCCTCTG	184–196	57	Yuan et al. [18]
IVFLMRE252	$(CCG)_6$	F:TAGACCTCGTGCCGTTATCA R:TGGCGTAGCAGTTGTCTTGG	185–260	58	Yuan et al. [18]
IVFLMRE725	$(CCCTCA)_4$	F:TCTCCGGCATACCAAATC R:GCGTACCTGCTCCTGTTC	156–186	56	Yuan et al. [18]

SSR types varied greatly and mainly focused on dinucleotide and trinucleotide repeats. The trinucleotide repeats number was 1446, accounting for 26.25% of the total SSR repeats, followed by dinucleotide repeats (1318), accounting for 23.92% of the total repeats. The number of four, five, and six nucleotide repeats types were 27, 11, and 43 respectively, totally accounting for 1.47% of the total SSR repeats. There were seven dinucleotide dominant motifs and 29 trinucleotide dominant motifs. Among the dinucleotide repeat motifs, the dominant repeat was AG/CT (9.29%) followed by GA/TC (7.01%). The dominant trinucleotide repeat was CCG/CGG (3.49%) (Supplementary file 5).

Development of EST-SSR markers and EST-SSR analysis

A total of 2057 primer pairs were designed from flanking SSR regions by Primer 3 software. One-hundred primers were selected randomly (Supplementary file 2) and screened in eight varieties including 'Matrix' (A), 'Roselily Elena' (O), 'Heartstrings' (LA hybrids), 'Zambesi' (OT hybrids), 'Fusion' (LL hybrid), 'Claude Shride' (Matagon), *L. regale*, and 'S-1' (varieties). Forty-four primer pairs (44%) did not give any amplification products. A total of 56 primer pairs (56%) produced reproducible polymorphic fragments. From these 56 primer pairs, a total of 13 pairs of primers with clear and polymorphic

Table 3 Genetic characterization of 16 polymorphic EST-SSR markers in *Lilium*

Locus	Na	Ne	I	PIC	Ho	He
SSR45	10	7.0392	2.0667	0.8579	0.4510	0.8664
SSR59	4	1.4826	0.6635	0.3255	0.2917	0.3289
SSR71	11	4.2639	1.7255	0.7655	0.3137	0.7731
SSR100	12	4.4219	1.8660	0.7739	0.2885	0.7814
SSR77	8	2.1248	1.1444	0.5294	0.5094	0.5344
SSR84	9	3.7629	1.5142	0.7342	0.4151	0.7412
EL42	8	2.2997	1.2197	0.5652	0.4815	0.5704
ZJU04	17	3.8644	1.9966	0.7412	0.4821	0.7479
IVFLMRE118	10	6.6336	2.0204	0.8493	0.4681	0.8584
IVFLMRE381	13	4.2343	1.8773	0.7638	0.5122	0.7733
IVFLMRE107	23	12.1857	2.7813	0.9179	0.3947	0.9302
IVFLMRE141	18	5.4258	2.0315	0.8157	0.4921	0.8222
IVFLMRE100	20	5.3153	2.1529	0.8119	0.2857	0.8192
IVFLMRE136	27	7.4352	2.5764	0.8655	0.3871	0.8725
IVFLMRE252	27	10.1603	2.7331	0.9016	0.4259	0.9100
IVFLMRE725	33	11.7828	2.9552	0.9151	0.3793	0.9231
Mean	15.6250	5.7770	1.9578	0.7583	0.4111	0.7658
St. Dev	8.3337	3.2646	0.6253	0.1601	0.0808	0.1618

Na observed number of alleles, *Ne* effective number of alleles, *I* Shannon index, *PIC* Polymorphism information content, *Ho* observed heterozygosity value, *He* expected heterozygosity value

bands were screened, in which six pairs of stable polymorphisms primers (SSR45, SSR59, SSR71, SSR77, SSR84, and SSR100) were the best, other seven primers (SSR22, SSR23, SSR43, SSR47, SSR60, SSR63, and SSR81) could not be amplified in partial varieties. Thus six polymorphic primers were selected in the 13 primers to conduct fluorescent labeling PCR combined with the other 10 previously reported primers [18, 23].

Allele richness of SSR loci

Sixteen pairs of primers detected a total of 250 allele loci in 65 sample individuals (Table 3). Among them, the minimum number of alleles was four, the maximum number of alleles was 33, and the average number of alleles per site was 15.6250. The total number of effective alleles was 92.432, and the value range was 1.4826 (SSR59) to 12.1857 (IVFLMRE107). The average number of effective alleles per locus was 5.7770. The Shannon Index (I) ranged from 0.6635 (SSR59) to 2.9552 (IVFLMRE725), with an average value of 1.9578. The value range of polymorphic information content (PIC) was 0.3255 (SSR59) to 0.9179 (IVFLMRE107), and the average value was 0.7583. Among them, 16 pairs of primers had higher polymorphic information (PIC > 0.25). In summary, the 16 EST-SSR primers had higher polymorphisms. The numerical ranges of the observed heterozygosity (Ho) and expected heterozygosity (He) were 0.2917 (SSR59) to 0.5122 (IVFLMRE381) and 0.3289 (SSR59) to 0.9302 (IVFLMRE107), respectively, with the mean values of 0.4111 and 0.7658.

Clusters based on the dendrogram

Cluster analysis was performed and the 65 samples could be roughly divided into two groups (Fig. 1). Group 1 included accessions of 18 Oriental varieties, nine OT (Oriental×Trumpet) varieties, one LL (*L. pardalinum* × *L. longiflorum*) variety, and one LO (Longiflorum×Oriental), and four native species. Group 2 included accessions of 22 Asiatic varieties, seven LA (Longiflorum×Asiatic) varieties, and two native species *(L. wenshanense*, *L. lancifolium*), and one unknown variety. The A and LA samples were divided into one group, and there was an obvious genetic distance among partial A and LA varieties, which could be divided into two clusters. The unknown variety S-1 was divided into the LA clusters. O and OT samples were clustered into one group, while the genetic distance between O and OT was not obvious, there was no significant separation. Four native species (*L. sargentiae*, *L. henryi*, *L. leucanthum*, *L. leucanthum* var. *Centifolium*) were clustered together. *L. leucanthum* var. *Centifolium* was the variant type of *L. leucanthum*, our results showed that they were grouped. *L. lancifolium* 'Flore pleno' was the double type of *L. lancifolium*, our results showed that they were grouped. There was more cross-mixing between groups and populations, indicating that there was more genetic variation within the population.

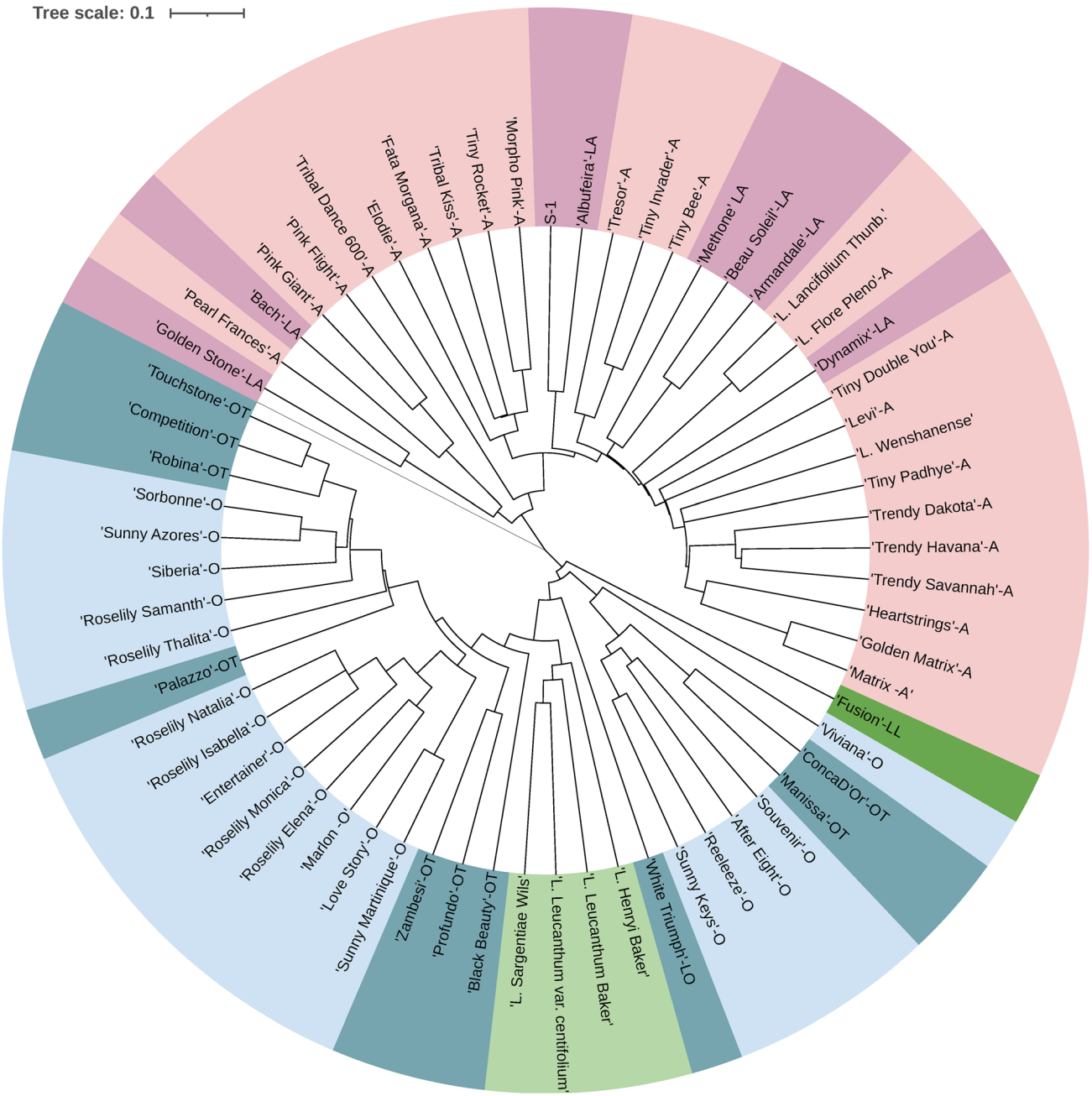

Fig. 1 UPGMA cluster phenogram showing the genetic relationship among 65 *Lilium* accessions

According to the population structure analysis, the optimal delta K value calculated by Evanno et al. [29] in STRUCTURE HARVESTER of the 65 *Lilium* samples was 2 in our research, indicating the existence of two major gene clusters. Different gene clusters were represented by different colors (Fig. 2). The STRUCTURE results grouped the 65 genotypes into two subgroups which had the same results as the UPGMA cluster analysis.

Discussion

A large number of high-quality clean reads, which were obtained from RNA-sequencing data of *L.* 'Elodie', were assembled. A total of 1067,547,290 unigenes were obtained, which provides a substantial basis for EST-SSR identification and characterization [20]. Simple sequence repeats have been applied to plant genetics and breeding owing to their advantages including codominant inheritance, hypervariability, multiallelic nature, reproducibility, relative abundance,

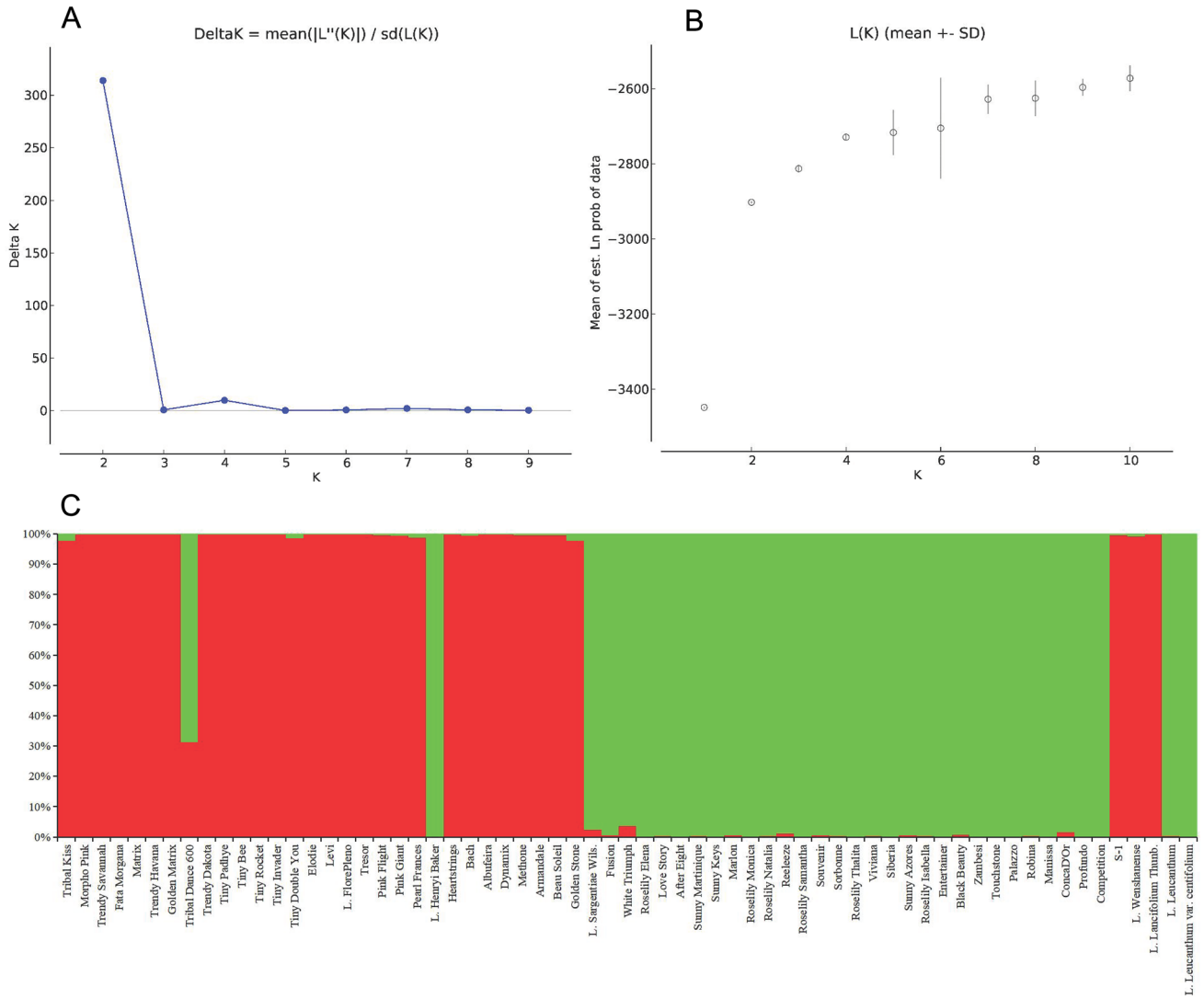

Fig. 2 STRUCTURE analysis of 65 *Lilium* germplasm resources. **A** Line chart of K value with Delta K value; **B** Line chart of K value with Ln probability of data; **C** Model-based structure of *Lilium* germplasm resources (K = 2). **A** Magnitude of Delta K as a function of K; **B** Plot of the Ln probability of data, over 5 runs for each K value; **C** Different colors represent different gene banks

and extensive genome coverage [31]. This study helped to identify 5509 potential SSRs (10.36%) in the 53,182 annotated unigenes which was higher than other reports in *Lilium* [18, 23] but lower than some plants such as *Camellia* (15.5%) [32], *Ricinus communis* (28.4%) [33] and *Coffee* [34]. Du et al. [23] reported that *Lilium* had more trinucleotide motifs than dinucleotide motifs, the same result was noticed in our study with 1,446 trinucleotide motifs and 1,318 dinucleotide motifs. Clear motif type bias reported in *L.* 'Elodie' was AG/CT (9.29%) for dinucleotide, the same result had been described in other *Lilium* species and cultivars [17, 18, 23]. In addition, the most abundant trinucleotide motif type in *L.* 'Elodie' was CCG/CGG (3.49%), which was also reported as the most abundant tri-nucleotide repeats in other monocot plants [21, 35–37].

The development of SSR markers for genetic diversity screening of *Lilium* germplasm has been widely adopted [16, 18, 23]. In the present study, 44 primer pairs in the 100 selected SSR primers did not give any amplification products, the main reason was probably due to the large genome of *Lilium* and the presence of large introns within the primer binding regions which might prevent the primer annealing [38]. A total of 56 primers (56%) produced reproducible polymorphic fragments, from which six primers amplified in the six different samples were selected for the diversity assessment of *Lilium*. Because only one organ including petaloid stamen, partially petaloid stamen, and normal stamen of *L.* 'Elodie' was sampled for transcriptome sequencing, the sequencing data varied a lot and might hinder the possibility to distinguish isoforms, other ten SSR markers developed by

other researchers from different cultivars and tissues [18, 23] were also selected for polymorphism screening and genetic diversity analysis. Totally, the 16 primer pairs detected 250 alleles, with an average of 15.625 alleles per SSR locus. In addition, SSR primers amplified with an average PIC of 0.7583, the results were consistent with the previous reports of 0.76 in 84 *Lilium* germplasm accessions [17]. However, other studies in *Lilium* reported comparatively lower rates of genetic diversity with average PIC of 0.55 [23], 0.53 [16], and 0.493 [18], which might be related to the number of markers and germplasm resources. The number of alleles observed at each locus in the data set was high, ranging from 1.4826 (SSR59) to 12.1857 (IVFLMRE107), indicating the capacity of the SSRs to detect relatively higher polymorphism. The EST-SSR markers developed from the transcriptomic data was conserved across the genomes of the closely related species, thus the polymorphic EST-SSR markers had transferability among species [31]. Our results indicated that the six EST-SSRs developed in *L.* 'Elodie' (LA) had been cross-amplified in other species of *Lilium*, which was consistent with other reports [17]. Most of the SSR markers used in this study had a PIC value ≥ 0.3255, and primer IVFLMRE725 had a PIC value of 0.9151 with the highest number of alleles (33) denotes the strong correlation between PIC and allele richness, the same results were also reported in *Vigna unguiculata* spp. *sesquipedalis* [39]. In addition, ten previously reported primers were mostly developed from different cultivars and organs, which can enrich the limitations of EST-SSR markers development in single tissue in this study. These EST-SSR markers can be used as candidate markers for future genetic map construction and evolutionary analysis in Liliaceae plants.

According to the Royal Horticultural Society (RHS), the genus *Lilium* was classified into nine groups based on a particular characteristics or a combination of characteristics [3]. The Asiatic subgroup was composed mainly of AA lines. *L. lancifolium*, one of the breeding parents involved in currently available cultivars of Asiatic hybrids, should be classified into the Asiatic group. Our result was in agreement with the genetic background of *L. lancifolium*, *L. lancifolium* 'Flore pleno' (A), the double flower cultivar of *L. lancifolium*, was grouped together with *L. lancifolium* with a single flower. LA lines were hybrids of *L. longiflorum* and *L. formolongi* [40], which was supported by the clustering pattern. *L. lancifolium* was the partial original parent of LA lines [41], therefore, it had a closer relationship with the AA lines, *L. Longiflorum* × Asiatic hybrids were genetically closer to Asiatic hybrids than to Oriental hybrids. Oriental hybrids were derived from eastern species such as *L. auratum*, *L. japonicum*, *L. rubellum*, *L. speciosum* and their hybrids with *L. henryi*. In the present study, *L. henryi* was classified into the Oriental subgroup, which was in agreement with the genetic background of *L. henryi*. *L. sargentiae* was involved in the currently available cultivars of Trumpet hybrids and Aurelian hybrids [42]. *L. leucanthum* var. *Centifolium* and *L. leucanthum* had the same trumpet flowers and they were classified into the same group with *L. sargentiae*. OT lines were Oriental trumpet hybrids derived from the progenies of Oriental hybrids and Chinese aurelianese [43]. Lee et al. [17] also clustered OO and OT into one group. This finding was also supported by our clustering pattern. Population structure analysis was useful in understanding genetic diversity analysis of *Lilium*, 65 germplasm resources were divided into two major clusters with the same results as UPGMA analysis method.

In our research, 'Fusion' (LL) and 'White Triumph' (LO) were both grouped into the Oriental subgroup which may be because fewer LL and LO sub-genome germplasm resources were collected. The germplasm resources between LA varieties and Asiatic hybrids, OT varieties and Oriental hybrids were still difficult to be distinguished, which had been reported in previous papers on *Lilium* [16, 17], thus the number of EST-SSRs is still needed to be increased in the future. In addition, it is necessary to develop more specific EST-SSR markers to distinguish the four sub-genomes (O, L, A, and T) of *Lilium*.

In general, although 4099 EST-SSRs of *Lilium* have been reported in the NCBI databases, new EST-SSRs are necessary to be developed to supplement the *Lilium* genetic information. In addition, the germplasm diversity evaluation of *Lilium* is also important work to understand the genetic relationship between cultivated species and wild species. In this paper, six new EST-SSRs were developed to enrich the biological data of *Lilium* by transcriptome sequencing. We collected some new varieties which were different from previous papers. Therefore, the genetic relationship analysis based on the resources was conducive to distinguishing the genetic relationship and serving the specific breeding objectives in the future.

Conclusion

In this study, EST-SSRs were analyzed from an Illumina transcriptome sequencing data set of *L.* 'Elodie'. After primer screening, six pairs of polymorphic EST-SSR markers were selected from 100 pairs with the most abundant polymorphic bands. The newly identified SSRs combined with the other 10 reported SSRs were verified in the genetic diversity analysis of 65 *Lilium* germplasm resources. Two major clusters were reported and a large number of genotypes were grouped together based on UPGMA and STRU CTURE analysis methods. This study provides valuable information for *Lilium*, as well as the molecular-marker-assisted breeding in the further.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-022-08083-9.

Author contributions MC and YZ conceived the theme of the study and designed the experiment; MC, GN and XL performed the experiment; MC, LY and YC analyzed the SSR data; MC wrote the manuscript. All co-authors read and approved the final manuscript.

Funding This work was supported by the Shanghai Agriculture Applied Technology Development Program, China (Grant No. X2020-02-08-00-12-F01463).

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval The article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication All the authors have read and consented to submit the manuscript.

References

1. Long YY, Zhang JZ, Zhang LN (1999) Lily-the king of corm flowers. Jindun Press, Beijing, pp 1–2
2. Van Tuyl JM, Arens P, Shahin A, Marasek-Ciołakowska A, Barba-Gonzalez R, Kim HT, Lim KB (2018) *Lilium*. Ornamental crops. Handbook Plant Breeding 11:481–512
3. Comber HF (1949) A new classification of the genus *Lilium.* Lily Yearbook. R Hort Soc 13:86–105
4. Shahin A, Smulders MJM, van Tuyl JM, Arens P, Bakker FT (2014) Using multi-locus allelic sequence data to estimate genetic divergence among four *Lilium* (Liliaceae) cultivars. Front Plant Sci 5:567
5. Kong Y, Bai J, Lang L, Bao F, Dou X, Wang H, Shang H (2017) Floral scents produced by *Lilium* and *Cardiocrinum* species native to China. Biochem Syst Ecol 70:222–229
6. Lee CS, Kim SC, Yeau SH, Lee NS (2011) Major lineages of the genus *Lilium* (Liliaceae) based on nrDNA ITS sequences, with special emphasis on the Korean species. J Plant Biol 54:159–171
7. Kong Y, Wang H, Lang L, Dou X, Bai J (2021) Metabolome-based discrimination analysis of five *Lilium* bulbs associated with differences in secondary metabolites. Molecules 26:1340
8. Munafo JP, Gianfagna TJ (2011) Quantitative analysis of steroidal glycosides in different organs of Easter lily (*Lilium longiflorum* Thunb.) by LC-MS/MS. J Agric Food Chem 59:995–1004
9. Francis JA, Rumbeiha W, Nair MG (2004) Constituents in Easter lily flowers with medicinal activity. Life Sci 76(6):671–683
10. Kumar SPJ, Susmita C, Sripathy KV, Agarwal DK, Pal G, Singh AN, Kumar S, Rai AK, Simal–Gandara J (2022) Molecular characterization and genetic diversity studies of Indian soybean (*Glycine max* (L.) Merr.) cultivars using SSR markers. Mol Biol Rep 49:2129–2140
11. Ma L, Wang X, Yan M, Liu F, Zhang S, Wang X (2022) Genome survey sequencing of common vetch (*Vicia sativa* L.) and genetic diversity analysis of Chinese germplasm with genomic SSR markers. Mol Biol Rep 49:313–320
12. Nyabera LA, Nzuki IW, Runo SM, Amwayi PW (2021) Assessment of genetic diversity of pumpkins (*Cucurbita* spp.) from western Kenya using SSR molecular markers. Mol Biol Rep 48:2253–2260
13. Liu C, Dou Y, Guan X, Fu Q, Zhang Z, Hu Z, Zheng J, Lu Y, Li W (2017) De novo transcriptomic analysis and development of EST-SSRs for *Sorbus pohuashanensis* (Hance) Hedl. PLoS ONE 12(6):e0179219
14. Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR) anchored polymerase chain reaction amplification. Genomics 20(2):176–183
15. Ukoskit K, Posudsavang G, Pongsiripat N, Chatwachirawong P, Klomsaard P, Poomipant P, Tragoonrung S (2018) Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. Genomics 111:1–9
16. Du F, Jiang J, Jia H, Zhao X, Wang W, Gao Q, Mao W, Wu Y, Zhang L, Grierson D, Xia Y, Gao Z (2015) Selection of generally applicable SSR markers for evaluation of genetic diversity and identity in *Lilium*. Biochem Syst Ecol 61:278–285
17. Lee SI, Park KC, Song YS, Son JH, Kwon SJ, Na JK, Kim JH, Kim NS (2011) Development of expressed sequence tag derived-simple sequence repeats in the genus *Lilium*. Genes Genom 33:727–733
18. Yuan S, Ge L, Liu C, Ming J (2013) The development of EST-SSR markers in *Lilium regale* and their cross-amplification in related species. Euphytica 189:393–419
19. Shahin A, van Kaauwen M, Esselink D, Bargsten JW, van Tuyl JM, Visser RG, Arens P (2012) Generation and analysis of expressed sequence tags in the extreme large genomes *Lilium* and *Tulipa*. BMC Genomics 13:640
20. Chen M, Nie G, Yang L, Zhang Y, Cai Y (2021) Homeotic transformation from stamen to petal in *Lilium* is associated with *MADS*-box genes and hormone signal transduction. Plant Growth Regul 95:49–64
21. 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
22. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386
23. Du F, Wu Y, Zhang L, Li XW, Zhao XY, Wang WH, Gao ZS, Xia YP (2015) De novo assembled transcriptome analysis and SSR marker development of a mixture of six tissues from *Lilium* Oriental hybrid 'Sorbonne'. Plant Mol Biol Rep 33:281–293
24. Glaubitz JC (2004) CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Mol Ecol Notes 4:309–310
25. Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157–163
26. Langella O (2007) Populations 1.2.30: population genetic software (individuals or populations distances, phylogenetic trees). France. http://bioinformatics.org/~tryphon/populations/. Accessed 14 Sept 2012
27. Letunic I, Bork P (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acid Res 47:W256–W259
28. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
29. 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)
30. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

31. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334
32. Ma JQ, Ma CL, Yao MZ, Jin JQ, Wang ZL, Wang XC, Chen L (2012) Microsatellite markers from tea plant expressed sequence tags (ESTs) and their applicability for cross-species/genera amplification and genetic mapping. Sci Hortic 134:167–175
33. 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 10:278–287
34. Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372
35. Baisakh N, Subudhi PK, Arumuganathan K, Parco AP, Harrison SA, Knott CA, Materne MD (2009) Development and interspecific transferability of genic microsatellite markers in *Spartina* spp. with different genome size. Aquat Bot 91:262–266
36. Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546
37. Wang YW, Samuels TD, Wu YQ (2011) Development of 1030 genomic SSR markers in switchgrass. Theor Appl Genet 122:677–686
38. Wang Z, Yu G, Shi B, Wang X, Qiang H, Gao H (2014) Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and *in silico* mapping onto the *M. truncatula* genome. PLoS ONE 9(3):e92029
39. Tantasawat P, Trongchuen J, Prajongjai T, Seehalak W, Jittayasothorn Y (2011) Variety identification and comparative analysis of genetic diversity in yardlong bean (*Vigna unguiculata* spp. *sesquipedalis*) using morphological characters, SSR and ISSR analysis. Sci Hort 124:204–216
40. Lim KB, Barba-Gonzalez R, Zhou S, Ramanna MS, van Tuyl JM (2008) Interspecific hybridization in lily (*Lilium*): taxonomy and commercial aspects of using species hybrids in breeding. Floriculture, ornamental and plant biotechnology, vol 5. Global Science Books, Japan, pp 146–151
41. McRae EA (1998) Lilies: a guide for growers and collectors. Timber Press, Portland
42. Lim KB, Van Tuyl JM (2006) *Lily*, *Lilium* hybrids. Flower Breed Genet 19:517–537
43. Asano Y (1989) *Lilium* L. In: Tsukumoto Y (ed) The great dictionary of horticulture, vol 5. Sykakukum, Tokyo, pp 198–209