Development of RNA-Seq SSR Markers and Application to Genetic Relationship Analysis among Sea Buckthorn Germplasm

He Li
School of Life Science, Nanjing University, Nanjing 210023, China; and Institute of Plant Resources, Dalian Minzu University, Dalian 116600, China

Cheng-Jiang Ruan1, Li Wang, and Jian Ding
Institute of Plant Resources, Dalian Minzu University, Dalian 116600, China

Xing-Jun Tian1
School of Life Science, Nanjing University, Nanjing 210023, China; and Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China

ADDITIONAL INDEX WORDS. Hippophae rhamnoides, transcriptome, unigenes, polymorphism, cluster, assisted breeding

ABSTRACT. Sea buckthorn (Hippophae rhamnoides) is an ecologically and economically valuable species that has been widely cultivated as a new berry crop rich in nutritional and medicinal compounds. RNA Sequencing (RNA-Seq) simple sequence repeat (SSR) markers were developed to evaluate the genetic relationships among 91 plants of 31 cultivars from two subspecies, mongolica and sinensis, as well as intraspecific hybrids between them. A total of 7540 RNA-Seq SSRs were identified as potential molecular markers, in which AG/CT (27.57%) was the most abundant unit type. AT/AT (9.93%), and AAG/CTT (11.95%) are the other main repeat motifs. A total of 110 primer pairs were randomly selected for validation of amplification. Seventeen SSR loci, located in genes encoding metabolic processes and cellulose synthases, were identified to be polymorphic among different sea buckthorn cultivars. These SSR loci generated 48 alleles, ranging from 2 to 5 per locus. Cluster analysis based on the proportion of shared alleles and unweighted pair group method with arithmetic average (UPGMA) algorithm divided all the genotypes into two main groups, with all of the ssp. sinensis cultivars (native to China) and hybrids in one group and ssp. mongolica cultivars (introduced from Russia) in the other group, which was in good agreement with their taxonomic classification. The RNA-Seq SSRs developed in this study have a potential use in the conservation of sea buckthorn germplasm and marker-assisted breeding (MAB).

Plants, particularly horticulture section, are raw material and used by people for food, either as edible products or for culinary ingredients, for medicinal use or ornamental and aesthetic purposes. They are genetically a very diverse group and play a major role in modern society and economy. Fruit and vegetables are an important component of traditional food but are also central to healthy diets of modern urban populations. Their roles in soil and water conservation and in the improvement of soil fertility by efficient nitrogen fixation allow sea buckthorn to be widely used in vegetation construction and the restoration of degenerated ecosystems (Ruan et al., 2007). Native cultivars (e.g., Zhongguoshaji) belonging to ssp. sinensis are well adapted, but have low economic value because of their small fruit and many thorns thus preventing large-scale cultivation and market development. By contrast, cultivars belonging to ssp. mongolica, which occurs naturally in Russia, Mongolia, and other countries, have large fruit and few thorns. However, many ssp. mongolica cultivars showed poor adaptability when introduced into China. Hybrids between the two subspecies have the potential to overcome these disadvantages, e.g., hybridization between ssp. sinensis and ssp. mongolica (introduced from Russia and other countries) (Li et al., 2009; Ruan et al., 2007). Thus, increasing numbers of new cultivars have been introduced from different countries, cultivated, and selected, but the origin and pedigree information are unavailable or are poorly documented for some cultivars, leading to some possible problems with homonyms and synonyms. In cross experiments, a few cultivars with desirable traits of interest (e.g., large fruit, few thorns) were identified from F1 individuals and planted on a wide scale. More optimal parental combinations need to be selected for cross breeding, and a key step is to clarify the genetic relationships among cultivars (Becelaere et al., 2005).
Molecular markers are widely used as a tool for the analysis of genetic diversity, including clarification of the taxonomic and geographic origin of cultivars, sex determination, and population genetic structure of woods. In fact, various types of molecular markers have been widely used to distinguish among sea buckthorn cultivars, e.g., amplified fragment length polymorphism, random amplified polymorphic DNA, intersimple sequence repeat, and sequence-related amplified polymorphism (Bartish et al., 2001; Li et al., 2009, 2010; Ruan, 2006). Compared with the molecular markers mentioned above, SSR (microsatellite) markers, which are 1 to 6 bp DNA regions repeated in tandem, have desirable advantages, such as codominance, random distribution throughout the genome, a high level of polymorphisms, high reproducibility, and technical simplicity (Kalai et al., 2011). SSR can be divided into genomic SSRs and genic SSRs because of the resource of sequences used for SSR identification. Genic SSRs derived from transcriptome or expressed sequence tag sequences are located in expressed genes. These markers can be linked with important phenotypic characteristics through quantitative trait loci analysis (Ramchary et al., 2011; Zhang et al., 2015). After the development of high-throughput sequencing, transcriptome datasets have become valuable resources for SSR discovery, more and more RNA-Seq SSRs are developed to investigate the genetic relationships among species and subspecies, as well as the genetic structure of populations (Liu et al., 2015; Zhang et al., 2015).

RNA-Seq SSRs, as gene-based SSRs, are located in the coding region of the genome and have some intrinsic advantages compared with genomic SSRs (Chen et al., 2016). For an example, they are more conserved and highly transferable to related taxa (Bhandawat et al., 2016). Because of these advantages, RNA-Seq SSRs have been developed and used in many plant species (Bhandawat et al., 2016; Sharma et al., 2011). For sea buckthorn, a number of SSR markers have been developed in several species and subspecies, including nine genomic SSRs (Wang et al., 2008) and 22 genic SSRs (Jain et al., 2010, 2014). However, the quantity and efficiency of SSR markers so far are limited, and there is no report on genetic relationship analysis among sea buckthorn germplasm using RNA-Seq SSR markers.

In this study, a large number of sea buckthorn transcriptome sequences containing SSR loci were generated using the HiSeq sequencing platform (Illumina, San Diego, CA). The distribution of SSR motifs in the sequences generated was characterized, and a set of RNA-Seq SSR markers with polymorphism and functional significance were developed and used firstly to evaluate genetic relationships among 31 cultivars grown in northern China.

**Materials and Methods**

**Plant materials.** A total of 91 individuals from 31 cultivars of sea buckthorn grown in northern China were selected for genetic relationship analysis (Table 1; Supplemental Table 1). They were provided by the Institute of Sea Buckthorn Selection and Breeding in Fuxin, the Research Institute of Berry in Suling, the Jiuchenggong Breeding Base of Sea Buckthorn in Dongsheng, and cultivated lands in Qinghe and Datong. These areas with various geographical and weather conditions ranged between latitudes 36°93′N–47°14′N, longitudes 90°22′E–127°06′E, and altitudes 202–2280 m (Table 2; Supplemental Fig. 1). One cultivar, Shiyou-1, which grows in the experimental field of our university, was used for RNA-Seq. Roots, stems, leaves, and berries were collected from a single plant.

**RNA isolation and transcriptomes sequencing.** Total RNA was isolated from the roots, stems, leaves, and berries of a single plant of cultivar Shiyou-1 using the Spin Column Plant Total RNA Purification Kit (Shanghai Sangon Biotech Co., Shanghai, China). The RNA quality was confirmed using RNase-free agarose gel electrophoresis. The concentration of total DNA was determined using a spectrophotometer (Qubit 3.0, Life Technologies, CA), and the intactness of the RNA was also confirmed using RNA ISOLATION AND TRANSCRIPTOMES SEQUENCING.

| Cultivar       | Code | Trees (no.) | Origin | Subspecies* | Growth site in China |
|----------------|------|-------------|--------|-------------|---------------------|
| Zhuangyanhuang | ZYH  | 5           | Russia | mongolica   | Fuxin, Liaoning      |
| Wucifeng       | WCF  | 5           | Russia | mongolica   | Fuxin, Liaoning      |
| Liusha-1       | LS1  | 5           | Russia | mongolica   | Fuxin, Liaoning      |
| Siberia rumianes | SR  | 4           | Russia | mongolica   | Fuxin, Liaoning      |
| Fangxiang      | FX   | 2           | Russia | mongolica   | Fuxin, Liaoning      |
| Yalishanda-12  | YLSD12 | 4       | Russia | mongolica   | Fuxin, Liaoning      |
| Jiuyuexiang    | JYH  | 2           | Russia | mongolica   | Fuxin, Liaoning      |
| Nanren         | NR   | 1           | Russia | mongolica   | Fuxin, Liaoning      |
| MZ14           | MZ14 | 3           | Russia | mongolica   | Suiling, Heilongjiang |
| Shoudu         | SD   | 3           | Russia | mongolica   | Suiling, Heilongjiang |
| Xin’e-2        | XE2  | 3           | Russia | mongolica   | Suiling, Heilongjiang |
| Xin’e-3        | XE3  | 3           | Russia | mongolica   | Suiling, Heilongjiang |
| Fenlan         | FL   | 3           | Russia | mongolica   | Suiling, Heilongjiang |
| Quyisike        | QYSK | 1           | Russia | mongolica   | Qinghe, Xinjiang     |
| Zhongguoshai   | ZGSJ | 3           | China  | sinesis     | Suiling, Heilongjiang |
| Juda           | JD   | 3           | China  | sinesis     | Dongsheng, Inner Mongolia |
| Jianpingdahuang| JPDH | 3           | China  | sinesis     | Dongsheng, Inner Mongolia |
| Manhanci       | MHC  | 3           | China  | sinesis     | Dongsheng, Inner Mongolia |
| Zhongxiongyou  | ZXY  | 3           | China  | sinesis     | Dongsheng, Inner Mongolia |
| Zhongguoshai** | ZGSJ** | 1       | China  | sinesis     | Datong, Qinghai      |
| Zaijiao-1      | ZJ1  | 1           | China  | M × S       | Fuxin, Liaoning      |
| Z24            | Z24  | 3           | China  | M × S       | Suiling, Heilongjiang |
| HS-56          | HS6  | 3           | China  | M × S       | Suiling, Heilongjiang |
| Liaofoza       | LFZ  | 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Zaciyou –1     | ZCY1 | 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Zaciyou-10     | ZCY10| 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Zaciyou-12     | ZCY12| 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Xinzaci-26     | XZC26| 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Shiciyou-2     | SCY-2| 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Shiciyou-5     | SCY-5| 3           | China  | M × S       | Dongsheng, Inner Mongolia |
| Shiciyou-30    | SCY30| 3           | China  | M × S       | Dongsheng, Inner Mongolia |

*M × S = ssp. mongolica (♂) × ssp. sinesis (♀).
RNA was identified using a spectrophotometer (NanoDrop 2000; Thermo Fisher Scientific, Waltham, MA). Transcriptome sequencing was performed by BGI Tech Solutions Co. (Shenzhen, China).

DE NOVO SEQUENCE ASSEMBLY AND DEVELOPMENT OF SSR MARKERS. A total of 5.17 G clean reads were used to de novo assemble transcriptomes using the Trinity software and then assembled into 62,166 unigenes. The MIcroSAtellite [MISA (Leibniz-Institut für Pflanzenbiochemie, 2002)] was used to identify SSR motifs with criteria of 12, 6, 5, 5, 4, and 4 minimum repeat units for mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats, respectively. Based on MISA results, primer pairs were designed using Primer 3 (Rozen and Skaletsky, 2000) with the size of the polymerase chain reaction (PCR) products set up to the range from 100 to 200 bp. Parameters were set up including 50 to 60 °C annealing temperature, 40% to 60% GC content, 18 to 24 bp primer length, and 100 to 200 bp amplicon size. All candidate SSR primer pairs were synthesized by Shanghai Sangon Biotech Co.

MARKER VALIDATION. A total of 110 candidate SSR primer pairs were screened for amplification and transferability using three samples from cultivars YLSD12, ZGSL, and LFZ, which represented different genetic backgrounds (ssp. mongolica, ssp. sinensis, hybrid). Among SSR loci with expected amplicons, 40 SSR loci were selected according to the amplification quality and validated using 91 individuals of 31 cultivars. Genomic DNA of each sample was extracted from young leaves by the hexadecyl trimethyl ammonium bromide method (Ruan et al., 2004). The quality and quantity of DNA were measured using a spectrophotometer (NanoDrop 2000) and visualized on a 1.0% agarose gel. PCR amplification was performed in 20 μL volumes containing 40 ng of DNA template, 1× PCR buffer, 1.5 mM MgCl2, 0.15 mM of each dNTP (Takara, Dalian, China), 1.5 U of Taq polymerase (Takara), and 0.5 μM of each primer. The PCR conditions included an initial denaturation at 94 °C for 2 min, 35 cycles of 30 s at 94 °C for denaturation, 30 s at 54–60 °C for annealing, and 45 s at 72 °C for extension, with a final extension 7 min at 72 °C, using a thermal cycler (C1000 Touch; Bio-Rad, Hercules, CA). PCR products were electrophoresed on 8% nondenaturing polyacrylamide gels using a standard dual-cooled vertical unit (SE 600 Ruby; GE Healthcare Life Sciences, Pittsburgh, PA) and visualized by silver staining. A 20 bp DNA ladder (Takara) was used to estimate the SSR allele size.

FUNCTIONAL ANNOTATION OF UNIGENES CONTAINING POLYMORPHIC SSRs. The unigenes with validated genic SSRs were aligned with BLASTX against four protein databases [U.S. National Center for Biotechnology Information (NCBI) nonredundant (Nr) proteins (Pruitt et al., 2007), Swiss-Prot (Bairoch and Apweiler, 1996), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2008) and Clusters of Orthologous Groups (COG) (Tatusov et al., 2000, 2003)] and one nucleotide database [NCBI nucleotide (Nt) sequences (Altschul et al., 1990)] with an E-value threshold of 1.0E-5 (Altschul et al., 1997; Cameron et al., 2004). Using nucleotide-based annotation, Blast2GO (Conesa et al., 2005) software was used to obtain gene ontology (GO) annotation categories defined by molecular function, cellular component and biological process ontologies.
**Data Analysis.** The microsatellites were scored as codominant marker data for genetic diversity analysis. The number of alleles \( (N_a) \), effective number of alleles \( (N_e) \), observed and expected heterozygosity \( (H_o \text{ and } H_e) \), Shannon’s information index, and polymorphic information content (PIC) for each of the genic SSR locus were calculated using POPGENE version 1.32 (Yeh et al., 1999) and PowerMarker version 3.25 (Liu and Muse, 2005) software packages. A genetic similarity matrix based on the proportion of shared alleles was generated, and a UPGMA tree was constructed using PowerMarker. The dendrogram was displayed and labeled using MEGA 6 software (Tamura et al., 2013) to reveal genetic relationships between the 31 sea buckthorn cultivars. Principal coordinates analysis (PCoA) from distance-standardized method was also performed to determine the genetic relatedness among the sampled cultivars using GenAlEx 6.5 (Peakall and Smouse, 2006, 2012).

**Results and Discussion**

**Distribution and frequency of different SSR types.** De novo transcriptional sequencing is a rapid and cost-effective method for developing microsatellite markers and is a promising method for marker analysis in species with no reference genome (Grabherr et al., 2011; Zheng et al., 2013). Here, we showed that transcriptome sequencing is a very useful tool for microsatellite marker development in sea buckthorn. Of the 62,166 unigenes derived from the transcription sequence, 9.89% unigenes showed the presence of SSR, 2.20% higher than that reported by Jain et al. (2014). A total of 6147 unigenes contained SSR, 504 had more than 1 SSR, and 231 had SSRs of different motifs (compound SSRs). The proportion of six SSR repeat motifs was not evenly distributed among 7540 SSR loci (Fig. 1). Dinucleotide repeat motifs were the most abundant [3025 (40.12%)] followed by tri [2423 (32.14%)] of SSR loci. Of all the repeat motif types, the \( (AG/CT)_n \) dinucleotides was the most abundant at frequency of 27.57%, which was observed in many other plant species (Chen et al., 2016; Zhou et al., 2015). The other two main unit types were the \( (AT/AT)_n \) dinucleotide and \( (AAG/CTT)_n \) trinucleotide repeat motifs, occurred at frequencies of 9.93% and 11.95%, respectively. The \( (AG/CT)_n \) dinucleotides (68.73%) was the most abundant in dinucleotide SSR units, followed by \( (AT/AT)_n \) dinucleotides (24.76%) and \( (AC/GT)_n \) dinucleotides (6.41%). However, GC dinucleotide repeats are relatively rare in the present study (only 0.10%), which was very similar to previous reports (Dong et al., 2014; Jain et al., 2014; Zhou et al., 2012). In trinucleotide SSR units, the \( (AAG/CTT)_n \) trinucleotides were the richest SSR loci repeats (37.19%), followed by \( (ATC/ATG)_n \) trinucleotides (14.86%), \( (AAT/ATT)_n \) trinucleotides (12.96%), \( (ACC/GGT)_n \) trinucleotides (7.59%), \( (AAC/GTT)_n \) trinucleotides (7.30%), \( (AGG/CCT)_n \) trinucleotides (7.26%), and \( (AGC/CTG)_n \) trinucleotides (7.18%).

The number of SSR repeats in each different motif length ranged from 4 to 23, with 6 repeats being the most abundant [1475 (19.56%)], followed by 5 [1419 (18.82%)], and 7 repeats [1093 (14.50%)] as the next most abundant (Table 3). The most abundant repeats in each motif type were 12 repeats in mononucleotides [1283 (82.40%)]), 6 repeats in dinucleotides [843 (27.87%)]), 5 repeats in tri- [1287 (53.16%)] and tetranucleotides [114 (75.50%)], and 4 repeats in penta- [126 (94.44%)] and hexanucleotides [240 (100%)]. Considering the length of microsatellite loci, 85.78% SSR repeats were less...
Table 4. Characterization of 17 polymorphic SSR markers including SSR locus, primer sequence, repeat motif, melting temperature (Tm), observed number of alleles (Na), effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), polymorphism information content (PIC), and Shannon’s information index (Is) among sea buckthorn cultivars.

| SSR locus     | Primer sequence (5’ to 3’) | Repeat motif | Tm (°C) | Size (bp) | Na | Ne | Ho   | He    | PIC          | Is         |
|---------------|-----------------------------|--------------|---------|-----------|----|----|------|-------|-------------|------------|
| CL1076.Contig2_HR_301 | F: CATGTAGCCACGATATTGCAA | (TTC)5       | 58      | 157       | 2  | 1.2300 | 0.2088 | 0.1880 | 0.1095 | 0.3346     |
|               | R: AGAGGCAAGAAGGGAAGAAG     | (TTCT)5      | 56      | 159       | 4  | 1.3923 | 0.3297 | 0.2833 | 0.2529 | 0.5247     |
| Unigene28552_HR_2397 | F: ATGTGATGTTGAGCAACCT     | (TTC)6       | 60      | 131       | 3  | 2.5270 | 0.5495 | 0.6076 | 0.5207 | 0.9899     |
|               | R: GCAATGGTTATAGGATAGG     | (TTC)6       | 60      | 125       | 4  | 1.9299 | 0.5934 | 0.4845 | 0.4371 | 0.8845     |
| Unigene14574_HR_2027 | F: GTCAACAGGCTTTAGTTACTG   | (GGACGA)4    | 60      | 125       | 2  | 1.2432 | 0.2198 | 0.1967 | 0.1765 | 0.3463     |
|               | R: CAAAGAACCTAACCACACTC    | (ACA)5       | 56      | 125       | 2  | 1.2432 | 0.2198 | 0.1967 | 0.1765 | 0.3463     |
| CL2151.Contig4_HR_595 | F: AGTTTCGTGGATGTCGTTCATC  | (AC)8        | 54      | 139       | 2  | 1.2969 | 0.2637 | 0.2302 | 0.2027 | 0.3899     |
|               | R: GTTCACGAAATTTCTCCTCT    | (TC)7        | 60      | 155       | 2  | 1.2834 | 0.2527 | 0.2220 | 0.1965 | 0.3794     |
| CL2339.Contig9_HR_634 | F: AGAGATAGTGGAGATAAGTCCCCGC | (GGATG)4     | 60      | 147       | 5  | 3.5374 | 0.5495 | 0.7213 | 0.6667 | 1.3673     |
| CL2394.Contig1_HR_655 | F: CACCATCATATAATCCAAAAATGG | (AC)8        | 54      | 139       | 2  | 1.2969 | 0.2637 | 0.2302 | 0.2027 | 0.3899     |
|               | R: TTGATATTTGATGGATATTTTCT | (TC)7        | 60      | 155       | 2  | 1.2834 | 0.2527 | 0.2220 | 0.1965 | 0.3794     |
| CL2969.Contig3_HR_789 | F: ATAAAACAATGGCAAACATCGT   | (GAA)5       | 60      | 154       | 3  | 1.7297 | 0.5824 | 0.4242 | 0.3457 | 0.6666     |
|               | R: AGAGCACTCGCAGGGATAGGC    | (TTTCC)4     | 60      | 135       | 4  | 2.0822 | 0.6154 | 0.5226 | 0.4321 | 0.8672     |
| CL338.Contig1_HR_138   | F: AAGCTGGAGGAACTAGATGCC    | (AGA)6       | 60      | 157       | 2  | 1.9247 | 0.4286 | 0.4831 | 0.3650 | 0.6735     |
|               | R: TGATATTTTGATCGAGGCCAAG  | (AAG)6       | 60      | 157       | 2  | 1.9247 | 0.4286 | 0.4831 | 0.3650 | 0.6735     |
| CL3511.Contig1_HR_929 | F: AGGTCCGTTACCATTTGAGCA    | (TC)7        | 60      | 154       | 2  | 1.1655 | 0.1538 | 0.1428 | 0.1319 | 0.2712     |
|               | R: TTGTTGTTACTGAGGAGCGAAG  | (TC)7        | 60      | 154       | 2  | 1.1655 | 0.1538 | 0.1428 | 0.1319 | 0.2712     |
| CL3558.Contig1_HR_936 | F: AGGTCTGTACATTGAGGCGAACAGAG | (AGA)6   | 60      | 154       | 2  | 1.1655 | 0.1538 | 0.1428 | 0.1319 | 0.2712     |
|               | R: AACGAGGCGAACCTATACCA    | (TGGTGAATAGG) | 60      | 154       | 2  | 1.1655 | 0.1538 | 0.1428 | 0.1319 | 0.2712     |
| CL3904.Contig3_HR_1006 | F: GTGTTCACTGAGAGGCAAGAG   | (GCT)7       | 60      | 154       | 3  | 1.2209 | 0.1648 | 0.1820 | 0.1695 | 0.3672     |
|               | R: TGGTGAATAGGCTAGGCAGGTA  | (GCT)7       | 60      | 154       | 3  | 1.2209 | 0.1648 | 0.1820 | 0.1695 | 0.3672     |
| CL4371.Contig1_HR_1073 | F: ATATTTATTACCATGCTTACAC  | (GCT)7       | 60      | 154       | 3  | 1.2209 | 0.1648 | 0.1820 | 0.1695 | 0.3672     |
|               | R: TGGTGAATAGGCTAGGCAGGTA  | (GCT)7       | 60      | 154       | 3  | 1.2209 | 0.1648 | 0.1820 | 0.1695 | 0.3672     |
| CL4750.Contig2_HR_1131 | F: ATTATTTGCTTTGTCACGACTA  | (AT)10       | 57      | 157       | 2  | 1.4936 | 0.4176 | 0.3322 | 0.2758 | 0.5124     |
|               | R: TTGTTGTTACTGAGGAGCGAAG  | (TGGTGAATAGG) | 60      | 154       | 2  | 1.1655 | 0.1538 | 0.1428 | 0.1319 | 0.2712     |
than 20 nucleotides long, and no microsatellite with the length higher than 30 bp was observed in this study.

**Polymorphism of the SSR Markers.** One hundred and ten SSR loci were selected to validate their amplification. Of these 110 SSR loci, 67 (60.91%) generated amplicons with expected size, 7 (6.36%) amplified nonspecific products, 36 (32.73%) failed to amplify the DNA products. The deviation from the expected size could be caused by the presence of introns, large insertions, lack of specificity in flanking sequences or even assembly errors (Saha et al., 2004; Varshney et al., 2005). The failure of amplification of primer pairs could be due to the location of the primers across splice sites, invalid primers, or poor-quality sequences (Chen et al., 2016; Li et al., 2015).

Owing to the inconsistency of amplification effects among 67 SSR loci-generated amplicons, 40 SSR markers with clear amplicons were finally selected to validate genetic diversity. Of these 40 SSR loci, 17 (42.50%) SSRs showed polymorphism in 31 sea buckthorn cultivars (Table 4). Amplification profile of “CL2339.Contig9_HR_634” marker locus across all the genotypes was shown in Supplemental Fig. 2. A total of 48 alleles ranging in size from 124 to 169 bp were detected after PCR amplification of all samples (Table 4). The Na varied from two to five, with an average of 2.8235 alleles per locus. Among 48 alleles generated by the 17 SSR loci, 39 were polymorphic alleles, ranging from 1 to 5 per locus. The Ne ranged from 1.2209 to 3.5374, with an average of 1.7402. CL2339.Contig9_HR_634, Unigene14574_HR_2027, CL3904.Contig3_HR_1006, and CL338.Contig1_HR_138 were informative SSR loci, each revealing more than two effective alleles distributed among all of the individuals.

Observed heterzygosity (Ho) and He ranged from 0.2088 and 0.1820 to 0.6154 and 0.7213, respectively. Shannon’s information index (I) averaged 0.6310 and ranged from 0.3346 to 1.3673. The value of polymorphism information content (PIC), regarded as discriminating power, varied from 0.1319 to 0.6667, with an average value of 0.3208. Loci CL2339.Contig9_HR_634 (PIC = 0.6667), Unigene14574_HR_2027 (PIC = 0.5207) and CL3904.Contig3_HR_1006 (PIC = 0.5072) showed higher effectiveness because of their high informativity. The average PIC of RNA-Seq SSR loci was less than that of our previous studies in sea buckthorn (Li et al., 2009, 2010; Ruan, 2004; Ruan et al., 2006). The low PIC value suggested that the genic sequences used for developing these markers were highly conserved in sea buckthorn germplasm used in this study.

**Gene Functions of Unigenes Containing Polymorphic SSRs.** For annotation of the sequences containing the 17 validated SSRs, relevant unigenes were searched against the five databases (see Materials and Methods), and the results were shown in Supplemental Table 2. A total of 16 unigenes showed significant similarity to known proteins in the Nr sequence database. The results showed that the sequences were similar to hypothetical/predicted proteins mainly from *Prunus persica* (29.4%), *Vitis vinifera* (23.5%), and *Ricinus communis* (23.5%). Based on sequence homology with Nt, Swissprot, KEGG, and COG database, putative functions mainly included 1) multicopper oxidase or L-ascorbate oxidase, participating in oxidation-reduction process, related to plant growth and development, and stress resistance; 2) translation initiation factors, related to regulation of translation; and 3) leucine-rich repeat receptor-like tyrosine-protein kinase and serine/threonine protein kinase, participating in signal transduction in cell. Among predicted proteins, omega-hydroxypalmitate O-feruloyl transferase, participating in suberin biosynthetic process in epidermal tissues, correlated with adaptation to the environmental changes in plants, such as resistance to desiccation and providing defense to many pathogens (Pollard et al., 2008; Schreiber, 2010). Another putative protein, DGAT 1, a key enzyme in lipid metabolism was also found in this study (Fatima et al., 2012). The RNA-Seq SSRs in these genes were potential to become functional markers for MAB when the genes are identified to regulate important agronomical traits (Zhang et al., 2015).
Based on GO, the sequences containing polymorphic SSRs were assigned to three major categories, namely, biological process, molecular function, and cellular component. Among biological process, the functions focused on metabolic processes, cellular organization, and signal transduction. The main molecular functions of these sequences were categorized for binding and catalytic activity. Under cellular component, organelles involved in vesicular transport and energy metabolism constituted the major proportion.

**Genetic relationship analysis.** The 17 polymorphic genic SSRs developed in this study were used to assess the genetic relationships between 31 sea buckthorn cultivars. A UPGMA dendrogram based on the proportion of shared alleles was constructed (Fig. 2) and revealed two major groups. The cultivars of *V. mongolica* clustered into one large group without exception, indicating independent evolution, whereas those of *V. sinensis* and hybrids between *V. mongolica* and *V. sinensis* were distributed in the other group. This result was in good agreement with the available pedigree and genetic background information (Supplemental Table 1). The cultivars of *V. mongolica* were introduced as seeds or asexual reproduction materials from Russia. Oppositely, the cultivars of *V. sinensis* were native to China, and some have close relationships with the male parent of the hybrids selected in this study and thus cluster together with the hybrids. Thus, the Russian cultivars growing in China showed a long genetic distance from the Chinese native cultivars and hybrids in the present.

The first group in the dendrogram was composed of Cluster I and II, which included all the cultivars of *V. mongolica* grown in Fuxin and Suiling. According to the information provided by documents of cultivation places and previous literatures (Jin et al., 1998), the origins of these cultivars were seeds or asexual reproduction materials mainly collected from research institutes in Moscow (lat. 55°45′N, long. 37°34′E) and Novosibirsk (lat. 55°10′N, long. 82°00′E). Owing to the geographic difference, Russian cultivars growing in the above regions showed obvious difference in phenotypic characters and resistance (Jin et al., 1998). Interestingly, SD and FL, originated from Moscow, was in Cluster I, and QYSK, LS1, SR, and FX, originated from Novosibirsk, were all clustered into the other one. Genetic differences between Moscow and Novosibirsk germplasm could be due to random sampling of genes that have no bearing on fitness. These results indicated sea buckthorn cultivars of *V. mongolica* generated different branches in phylogenesis because of various environmental factors, although we were not able to identify it among Russian cultivars because of the lack of pedigree information of them.

The second group contained 17 cultivars and was divided into three subgroups, namely Cluster III, IV, and V. The Cluster IV consisted of all the cultivars of *V. sinensis*. In this subgroup, JD and ZGSJ, both from Fengning population of *V. sinensis*, were clustered closely. ZXY, JPDH, and Manhanci, respectively from Yibei population, Jianping population, and Manhanshan population, clustered together. ZGSJ, a cultivar domesticated from wild plants (ssp. *sinensis*), was close to ZXY. The hybrids from different origins were classified into Cluster III and V. Cluster III consisted of 4 cultivars, namely SCY2, SCY5, SCY30, and ZCY10. Among them, SCY2, SCY5, and SCY30, with similar male parent from Manhanshan population, clustered together. Probably because of poorly documented, cultivar ZCY10, with male parent of Fengning population, also was classified in this subgroup. Cluster V contained seven cultivars whose male parents all stemmed from Fengning population. In this subgroup, ZJ1, LFZ, Z24, and HS56 closely located in the tree, and ZXY1, ZXY12, and XZC26 clustered together. ZJ1 and LFZ, with similar parentage, had the closest relationship.

A PCoA was performed as the further illustration of the relationships between the cultivars (Fig. 3). The percentages of total variation explained by the first two axes were 44.07% and 15.10%, respectively. Similar to cluster analysis, PCoA also did not group the cultivars based on the location specific, and intermixing of one group of hybrids across the coordinates was observed, which was in accordance with previous studies (Anandan et al., 2016). All of the cultivars of *V. mongolica* distributed in the quadrant I and IV with two groups, which was similar to the results using the UPGMA method. The cultivars of *V. sinensis* composed another group that stayed on the quadrant III. The hybrids mainly concentrated in the quadrant II whereas four cultivars spread in the quadrant II and IV. These results of genetic relatedness were basically consistent with the UPGMA cluster analysis, and the major groups were related to known pedigree relationships.

Germplasm characterization is necessary to provide information on gene sources for future use and to prevent their disappearance. Our results represent one of the first attempts at detailed sea buckthorn genotyping by means of microsatellites and thus provide meaningful data for breeding programs in China. Indeed, this report is the first to analyze genetic relationships of sea buckthorn germplasm using RNA-Seq SSRs. The newly developed SSRs would not only enrich molecular markers, but also facilitate genetic researches in sea buckthorn.
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.

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

Anandan, A., M. Anumalla, S.K. Pradhan, and J. Ali. 2016. Population structure, diversity and trait association analysis in rice (Oryza sativa L.) germplasm for early seedling vigor (ESV) using trait linked SSR markers. PLoS One 11:e0152406.

Bairoch, A. and R. Apweiler. 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res. 24:21–29.

Bal, M.L., V. Meda, N.S. Naik, and S. Satya. 2011. Sea buckthorn: A potential source of valuable nutrients for nutraceuticals and cosmeceuticals. Food Res. Int. 44:1718–1727.

Bartish, I.V., N. Jeppsson, U. Swenson, and H. Nybom. 2002. Phylogeny of Hippophae (Elaeagnaceae) inferred from parsimony analysis of chloroplast DNA and morphology. Syst. Bot. 27:41–54.

Becelaere, G.V., L.L. Edward, A.H. Paterson, and P.W. Chee. 2005. Pedigree-vs. DNA marker-based genetic similarity estimates in crop. Crop Sci. 45:2281–2287.

Bhandawat, A., G. Singh, S.A. Raina, J. Kaur, and K.R. Sharma. 2016. Development of genomic SSR marker resource from RNA-Seq data in Dendrocalamus latiflorus. J. Plant Biochem. Biotechnol. 25:179–190.

Cameron, M., H.E. Williams, and A. Cannane. 2004. Improved gapped alignment in BLAST. IEEE/ACM Trans. Computational Biol. Informatics 1:116–129.

Chen, C., X.M. Xu, Y. Chen, M.Y. Yu, F.Y. Wen, and H. Zhang. 2013. Identification, quantification and antioxidant activity of acylated flavonol glycosides from sea buckthorn (Hippophae rhamnoides ssp. sinesis). Food Chem. 141:1573–1579.

Chen, H.L., X. Chen, J. Tian, Y. Yang, Z.X. Liu, X.Y. Hao, L.X. Wang, S.H. Wang, J. Liang, L.Y. Zhang, F.X. Yin, and X.Z. Cheng. 2016. Development of gene-based SSR markers in rice bean (Vigna umbellata L) based on transcriptome data. PLoS One 11:e0151040.

Conesa, A., S. Götz, J.M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Dong, S.B., Y.L. Liu, J. Niu, Y. Ning, S.Z. Lin, and Z.X. Zhang. 2014. Identification, quantification and antioxidant activity of acylated flavonol glycosides from sea buckthorn (Hippophae rhamnoides ssp. sinesis). Food Chem. 141:1573–1579.

Fatima, T., C.L. Snyder, W.R. Schroeder, D. Cram, R. Dutla, D. Wishart, R.J. Weselake, and P. Krishna. 2012. Fatty acid composition of developing sea buckthorn (Hippophae rhamnoides L.) berry and the transcriptome of the mature seed. PLoS One 7:e34099.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Fatima, T., C.L. Snyder, W.R. Schroeder, D. Cram, R. Datla, D. Jain, A., S. Chaudhary, and P.C. Sharma. 2014. Mining of microsatellites using next generation sequencing of seabuckthorn (Hippophae rhamnoides L.) transcriptome. Physiol. Mol. Biol. Plants 20:115–123.

Jin, Z.P., D.M. Lan, and S.Q. Zhou. 1998. Studies on selection of Chinese elite seabuckthorn cultivars and introduction of Russian large berry species. Hippophae 11(4):10–16 (in Chinese with an English abstract).

Kalia, R.K., M.K. Rai, S. Kalia, R. Singh, and A.K. Dhawan. 2011. Microsatellite markers: An overview of the recent progress in plants. Euphytica 177:309–334.

Kallio, H., M. Lassila, E. Järvenpää, G.G. Haraldsson, S. Jonsdottir, and B.R. Yang. 2009. Inositol and methylinositois in sea buckthorn (Hippophae rhamnoides) berries. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 877:1426–1432.

Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamashita. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484.

Leibniz-Institut für Pflanzenbiochemie, and Kulturpflanzenforschung. 2002. The MicroSatellite identification tool (MISA). 5 Jan. 2015. <http://pgrc.ipk-gatersleben.de/misa/>.

Li, C.I., Q.F. Ling, C. Ge, Z.Q. Ye, and X.F. Han. 2015. Transcriptome characterization and SSR discovery in large-scale loach Paramisgurnus dabyranus (Cobitidae, Cypriniformes). Gene 557:201–208.

Li, H., C.J. Ruan, and J.A. Teixeira da Silva. 2009. Identification and genetic relationship based on ISSR analysis in a germplasm collection of sea buckthorn (Hippophae L.) from China and other countries. Sci. Hort. 123:263–271.

Li, H., C.J. Ruan, J.A. Teixeira da Silva, and B.Q. Liu. 2010. Associations of SSR markers with dried-shrink disease resistance in a germplasm collection of sea buckthorn (Hippophae L.). Genome 53:447–457.

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

Liu, Y.L., P.F. Zhang, M.L. Song, J.L. Hou, M. Qing, W.Q. Wang, and C.S. Liu. 2015. Transcriptome analysis and development of SSR molecular markers in Glycrrhiza uralensis Fisch. PLoS One 10:e0143017.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295.

Peakall, R. and P.E. Smouse. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research - An update. Bioinformatics 28:2537–2539.

Pollard, M., F. Beisson, Y. Li, and J.B. Ohlrogge. 2008. Building lipid barriers: Biosynthesis of cutin and suberin. Trends Plant Sci. 13:236–246.

Prupp, K.D., T. Tatusova, and D.R. Maglott. 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35:D61–D65.

Pollard, M., F. Beisson, Y. Li, and J.B. Ohlrogge. 2008. Building lipid barriers: Biosynthesis of cutin and suberin. Trends Plant Sci. 13:236–246.

Pruss, K.D., T. Tatusova, and D.R. Maglott. 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35:D61–D65.

Raffo, A., F. Paoletti, and M. Antonelli. 2004. Changes in sugar, organic acid, flavonol and carotenoid composition during ripening of berries of three seabuckthorn (Hippophae rhamnoides L.) cultivars. Eur. Food Res. Technol. 9:360–368.

Ramchari, N., V.D. Nguyen, X. Li, C.P. Hong, V. Dhandapani, S.R. Choi, G. Yu, Z. Piao, and Y. Lim. 2011. Genetic microsatellite markers in Brassica rapa: Development, characterization, mapping, and their utility in other cultivated and wild Brassica relatives. DNA Res. 18:305–320.

Rozen, S. and H. Skaltsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132:365–386.

Ruan, C.J. 2006. Genetic relationships among some sea buckthorn cultivars from China, Russia and Mongolia using AFLP markers. J. Hort. Sci. Biotech. 81:409–414.

Ruan, C.J., P. Qin, J.W. Zheng, and Z.X. He. 2004. Genetic relationships among some cultivars of sea buckthorn from China, Russia and Mongolia based on RAPD analysis. Sci. Hort. 101:417–426.

Ruan, C.J., J.A. Teixeira da Silva, H. Jin, H. Li, and D.Q. Li. 2007. Research and biotechnology in sea buckthorn (Hippophae spp.). Med. Aromat. Plant Sci. Biotechnol. 1:47–60.
Saha, M., M. Mian, I. Eujayl, J. Zwonitzer, L. Wang, and G. May. 2004. Tall fescue EST–SSR markers with transferability across several grass species. Theor. Appl. Genet. 109:783–791.
Sakar, E., H. Unver, M. Bakir, M. Ulas, and Z.M. Sakar. 2016. Genetic relationships among olive (Olea europaea L.) cultivars native to Turkey. Biochem. Genet. 54:348–359.
Saridas, M.A., N.E. Kafkas, M. Zarifikhosroshahi, O. Bozhaydar, and S.P. Kargi. 2016. Quality traits of green plums (Prunus cerasifera Ehrh.) at different maturity stages. Turk. J. Agr. For. 40:655–663.
Schreiber, L. 2010. Transport barriers made of cutin, suberin and associated waxes. Trends Plant Sci. 15:546–553.
Sharma, H., R. Kumar, V. Sharma, V. Kumar, P. Bhardwaj, P.S. Ahuja, and R.K. Sharma. 2011. Identification and cross-species transferability of 112 novel unigene-derived microsatellite markers in tea (Camellia sinensis). Amer. J. Bot. 98:e133–e138.
Solmaz, I., Y.A. Kacar, O. Simsek, and N. Sari. 2016. Genetic characterization of Turkish snake melon (Cucumis melo L. subsp. melo flexuosus Group) accessions revealed by SSR markers. Biochem. Genet. 54:534–543.
Sun, K., X.L. Chen, R.J. Ma, C.B. Li, Q. Wang, and S. Ge. 2002. Molecular phylogenetics of Hippophae L. (Elaeagnaceae) based on the internal transcribed spacer (ITS) sequences of nrDNA. Plant Syst. Evol. 235:121–134.
Swenson, U. and I.V. Bartish. 2002. Taxonomic synopsis of Hippophae (Elaeagnaceae). Nord. J. Bot. 22:369–374.
Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.
Tatusov, R.L., N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhodov, A.N. Nikolskaya, B.S. Rao, S. Smimov, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, and D.A. Natale. 2003. The COG database: An updated version includes eukaryotes. BMC Bioinformatics 4:41.
Tatusov, R.L., M.Y. Galperin, D.A. Natale, and E.V. Koonin. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28:33–36.
Varshney, R.K., A. Graner, and M.E. Sorrells. 2005. Genic microsatellite markers in plants: Features and applications. Trends Biotechnol. 23:48–55.
Wang, A., Q. Zhang, D. Wan, Y. Yang, and J. Liu. 2008. Nine microsatellite DNA primers for Hippophae rhamnoides ssp. sinensis (Elaeagnaceae). Conserv. Genet. 9:969–971.
Yeh, F.C., R.C. Yang, and T. Boyle. 1999. POPGENE version 1.32. 1 Dec. 2016. <https://sites.ualberta.ca/fyeh/popgene_download.html>.
Zhang, L.W., Y.R. Li, A.F. Tao, P.P. Fang, and J.M. Qi. 2015. Development and characterization of 1,906 EST-SSR markers from unigenes in jute (Corchorus spp.). PLoS One 10:e0140861.
Zheng, X., C. Pan, Y. Diao, Y. You, C. Yang, and Z. Hu. 2013. Development of microsatellite markers by transcriptome sequencing in two species of Amorphophallus (Araceae). BMC Genomics 14:490.
Zhou, Y., F. Gao, R. Liu, J. Feng, and H. Li. 2012. De novo sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in Amorppiptanthus mongolicus. BMC Genomics 13:266.
| Cultivar           | Pedigree/background information                                                                 |
|-------------------|--------------------------------------------------------------------------------------------------|
| Zhuangyuanhuang   | selected from the seedlings of mixed seeds from Moscow or Novosibirsk, Russia                    |
| Yaliishanda-12    | selected from the seedlings of varieties from Moscow or Novosibirsk, Russia                        |
| Juuyuehuang       | selected from the seedlings of varieties from Moscow or Novosibirsk, Russia                        |
| Nanren            | selected from the germchits of varieties from Moscow or Novosibirsk, Russia                       |
| Wucifeng          | selected from the seedlings of mixed seeds from Moscow or Novosibirsk, Russia                    |
| Liusha-1          | selected from the seedlings of mixed seeds from Novosibirsk, Russia                              |
| Siberia rumianes  | selected from the asexual reproduction materials from Novosibirsk, Russia                        |
| Fangxiang         | selected from the seedlings of varieties from Novosibirsk, Russia                                |
| Shoudu            | selected from the seedlings of varieties from Moscow, Russia                                      |
| Xin’e-2           | selected from the seedlings of mixed seeds collected from Russia                                  |
| Xin’e-3           | selected from the seedlings of mixed seeds collected from Russia                                  |
| Fenlan            | selected from the seedlings of mixed seeds collected from Moscow, Russia                          |
| MZ14              | selected from the seedlings of varieties from Russia                                              |
| Qiuyisike         | selected from the seedlings of varieties from Novosibirsk, Russia                                |
| Zhongguoshaji     | selected from the seedlings of ZGSJ* (Fengning population) growing in China                      |
| Juda              | selected from the seedlings of ZGSJ* (Fengning population) growing in China                      |
| Jianpingdahuang   | selected from the seedlings of ZGSJ* (Jianping population) growing in China                       |
| Manhanci          | selected from the seedlings of ZGSJ* (Manhanshan population) growing in China                    |
| Zhongxiongyou     | selected from the seedlings of ZGSJ* (Yibei population) growing in China                         |
| Zhongguoshaji*wild| domesticated from wild plants of ZGSJ*                                                         |
| Zajiao-1          | progeny of Chuyi (introduced from Russia) and ZGSJ* (Fengning population)                        |
| Z24               | progeny of Alieyi (introduced from Russia) and ZGSJ* (Fengning population)                        |
| HS-56             | progeny of HS-22 (seedling of ‘Wulangemu’ introduced from Mongolia) and ZGSJ* (Fengning population) |
| Liaofuza          | progeny of Chuyi (introduced from Russia) and ZGSJ* (Fengning population)                        |
| Zaciyou –1        | F1 hybrid of Wulanshalin (seedling of ‘Wulangemu’ introduced from Mongolia) and ZGSJ* (Fengning population) |
| Zaciyou-10        | F1 hybrid of Wulanshalin (seedling of ‘Wulangemu’ introduced from Mongolia) and ZGSJ* (Fengning population) |
| Zaciyou-12        | F1 hybrid of Wulanshalin (seedling of ‘Wulangemu’ introduced from Mongolia) and ZGSJ* (Fengning population) |
| Xinzaci-26        | progeny of Chuyi (introduced from Russia) and ZGSJ* (Fengning population)                        |
| Shicyou-2         | progeny of Taiyang (introduced from Russia) and ZGSJ* (Manhanshan population)                    |
| Shicyou-5         | progeny of Taiyang (introduced from Russia) and ZGSJ* (Manhanshan population)                    |
| Shicyou-30        | progeny of Yousheng (introduced from Russia) and ZGSJ* (Manhanshan population)                   |

*ZGSJ* = native Zhongguoshaji, representing all populations of *H. sinensis* in China.
Supplemental Fig. 1. Growth sites of sampled sea buckthorn cultivars in northern China.
Supplemental Fig. 2. Electrophoresis profile of CL2339.Contig9_HR_634 marker for 91 plants of 31 sea buckthorn accessions. M: 20 bp DNA Ladder (Takara, Dalian, China). 1–91: number of plants.
## Supplemental Table 2. Gene annotation of the unigene sequences containing polymorphic genic SSRs developed in this study using the following databases: NCBI non-redundant proteins (Nr), Swiss-Prot, Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), gene ontology (GO), and Nonredundant Nucleotide (Nt).

| Gene ID  | Nr-annotation           | SwissProt-annotation | COG-annotation | GO-biological process                                                                 | GO-molecular function                                                                 | GO-cellular component                     |
|----------|-------------------------|----------------------|----------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|------------------------------------------|
| CL1076.  | hypothetical protein    |                      |                |                                                                                       |                                                                                       |                                          |
| Contig2_HR | PRUPE_ppa011920mg       | Vitis vinifera       |                | GO:0009616/virus induced gene silencing; GO:001050/vegetative phase change              |                                                                                       |                                          |
| Unigene28552_HR | hypomorph                | LOC100256881, transcript variant 2, mRNA |                |                                                                                       |                                                                                       |                                          |
| Unigene14374_HR | hypomorph                | LOC10263682, mRNA |                |                                                                                       |                                                                                       |                                          |
| CL2151.  | unannotated protein product | Vitis vinifera clone |                |                                                                                       |                                                                                       |                                          |
| Contig4_HR |                      |                      |                |                                                                                       |                                                                                       |                                          |
| CL2197.  | eukaryotic translation initiation factor 4E-like [Goneja mar] | Vitis vinifera clone SSOAER237K16 |                | GO:0006413/translation initiation factor NCBP; GO:0006417/translation initiation factor 4E | GO:0003743/translation initiation factor activity | GO:0005737/cytoskeleton                |
| Contig6_HR | multicyclic oxidase, putative [Goneja mar] > (g2)2235533131 | Frangula vesca subsp. vesca | L-ascorbate oxidase homolog | LOC101929598, mRNA | GO:00064664/platelet-type cell wall organization; GO:000832/platelet-type cell wall biogenesis; GO:0055114/oxidation-reduction process | GO:0008447/L-ascorbate oxidase activity; GO:0030596/pectinesterase activity; GO:0009505/plant-cell wall binding; GO:0009506/plastid-oxidation; GO:0016020/membrane; GO:0048464/apoplast |
| CL2319.  | multicopper oxidase, putative [Goneja mar] > (g2)223552817 | L-ascorbate oxidase homolog | LOC101929598, mRNA | Putative multicopper oxidase | LOC101929598, mRNA | LOC101929598, mRNA | LOC101929598, mRNA |
| Contig9_HR | prolyl-tRNA synthetase, putative [Goneja mar] > (g2)223547870 | Prolyl-tRNA synthetase [EC:6.1.1.15] | LOC100262997, mRNA | Prolyl-tRNA synthetase | LOC100262997, mRNA | Prolyl-tRNA synthetase | LOC100262997, mRNA |
| CL2394.  | prolyl-tRNA synthetase, putative [Goneja mar] > (g2)223543870 | Prolyl-tRNA synthetase [EC:6.1.1.15] | LOC100262997, mRNA | Prolyl-tRNA synthetase | LOC100262997, mRNA | Prolyl-tRNA synthetase | LOC100262997, mRNA |
| CL2969.  | DNA binding protein, putative [Goneja mar] > (g2)223800581 | DNA binding protein, putative [Goneja mar] > (g2)223800581 | DNA binding protein, putative [Goneja mar] > (g2)223800581 | DNA binding protein, putative [Goneja mar] > (g2)223800581 | DNA binding protein, putative [Goneja mar] > (g2)223800581 | DNA binding protein, putative [Goneja mar] > (g2)223800581 |
| CL309.   | hypomorph                |                      |                |                                                                                       |                                                                                       |                                          |
| C13347.  | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223543870 | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |
| Contig1_HR | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223548570 | Prolyl-tRNA synthetase-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |
| CL309.   |          |                      |                |                                                                                       |                                                                                       |                                          |
| CL3347.  | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223548570 | Prolyl-tRNA synthetase-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |
| Contig1_HR | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223548570 | Prolyl-tRNA synthetase-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |
| CL3347.  | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223548570 | Prolyl-tRNA synthetase-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |
| Contig1_HR | prolyl-tRNA synthetase-like tyrosine-protein kinase [Goneja mar] > (g2)223548570 | Prolyl-tRNA synthetase-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA | Leucine-rich repeat receptor-like tyrosine-protein kinase [EC:2.7.11.1] | LOC100258928, mRNA |

*Continued next page*
| Gene ID | N-annotation | SwissProt-annotation | GO-biological process | GO-molecular function | GO-cellular component |
|--------|--------------|----------------------|-----------------------|----------------------|----------------------|
| CL338. | PREDICTED: omega-hydroxypalmitate O-feruloyltransferase-like [Vitis vinifera] | PREDICTED: omega-hydroxypalmitate O-feruloyltransferase [EC:2.3.1.188] | GO:0010345/suberin biosynthetic process; GO:0052325/cell wall pectin biosynthetic process | GO:0005773/hydroxycinnamoyltransferase activity; GO:0047672 | — |
| Contig1 | COG-annotation | KO_definition |               |                     |                     |
| 1_HR   |               |                      |                     |                     |                     |
|        | Nt-annotation |                      |                     |                     |                     |
|        |                |                       |                     |                     |                     |
|        | SWISSPROT-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL3541. | hypothetical protein | PREDICTED: Fragaria vesca subsp. vesca LOC101309253, mRNA | — | — | — |
| Contig1 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL3558. | PREDICTED: mediator of RNA polymerase II transcription subunit 7 [Vitis vinifera] | Mediator of RNA polymerase II transcription subunit 7a [LOC100257245], mRNA | GO:0006357/regulation of transcription from RNA polymerase II promoter | GO:0001104/RNA polymerase II transcription cofactor activity | GO:0016592/mediator complex |
| Contig1 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL3904. | hypothetical protein | PREDICTED: Vitis vinifera uncharacterized LOC100255639, mRNA | GO:0006629/lipid metabolic process | GO:0008374/O-acyltransferase activity; GO:0046027/phospholipid:diacylglycerol acyltransferase activity | GO:0005773/vacuole; GO:0005783/ endoplasmic reticulum |
| Contig3 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL4371. | PREDICTED: phospholipid:diacylglycerol acyltransferase 1 [Vitis vinifera] | Phospholipid:diacylglycerol acyltransferase 1 [EC:2.3.1.158] | GO:0008374/O-acyltransferase activity; GO:0046027/phospholipid:diacylglycerol acyltransferase activity | GO:0005773/vacuole; GO:0005783/ endoplasmic reticulum | — |
| Contig1 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL4750. | hypothetical protein | PREDICTED: Vitis vinifera uncharacterized LOC100255639, mRNA | — | — | — |
| Contig2 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
| CL4888. | predicted protein [Populus trichocarpa] | PREDICTED: Vitis vinifera Ocs element-binding factor 1-like [LOC100255634], mRNA | — | — | — |
| Contig1 | N-annotation |                      |                     |                     |                     |
| 1_HR   |                |                       |                     |                     |                     |
|        | SwissProt-     |                       |                     |                     |                     |
|        | annotation     |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |
|        |                 |                       |                     |                     |                     |

*gene ID = gene identification.
*KO = KEGG Orthology.
*No result.