Identification of Expressed Sequence Tag-simple Sequence Repeat Markers from the De Novo Transcriptome Sequence of Red Raspberry (Rubus idaeus L.)

Jianfeng Liu, Bowen Yang, Yue tong Ming, Yuchu Zhang, and Yunqing Cheng

Jilin Provincial Key Laboratory of Plant Resource Science and Green Production, Jilin Normal University, Siping, Jilin 136000, People's Republic of China

Additional index words. Rubus idaeus, EST-SSR markers, de novo transcriptome sequence

Abstract. Rubus idaeus has remarkable economic and cultural value. Developing efficient simple sequence repeat (SSR) markers is necessary for the molecular breeding of red raspberry. In this study, SSR mining was performed using the de novo transcriptome sequence of R. idaeus. In total, 14,210 SSR sequences were identified from 11,158 SSR-containing unigenes. In all the SSR sequences, mononucleotide, dinucleotide, and trinucleotide repeats were the most common, and their number and percentage were 1323 (9.31%), 6752 (47.52%), and 4897 (34.46%), respectively. Of the mononucleotide and dinucleotide repeats, A/T, AG/CT, AT/AT, and AC/GT were more abundant and accounted for 9.09%, 37.82%, 6.51%, and 3.14% of the total repeat number, respectively. In the trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats, the nucleotide (NT) patterns AAG/CTT, AAAG/CTTT, AAAAG/CTTTT, and AAGAGG/CCTCTT were the most frequent, and accounted for 14.11%, 0.38%, 0.57%, and 0.23% of the total SSRs, respectively. Of the 480 SSR-containing unigenes with gene ontology (GO) annotation, the classification results showed that they were mainly involved in binding, catalytic, and transporter molecular functions. Most of the 3441 SSR-containing unigenes with the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation were involved in the following top five pathways: metabolic, RNA transport, spliceosome, protein processing in the endoplasmic reticulum, and mRNA surveillance. Thirty pairs of primers derived from the red raspberry transcriptome were randomly selected to assess their polymorphism by using 15 red raspberry germplasms, in which the polymorphism information content (PIC) values ranged from 0.50 to 0.86, with a mean of 0.73, thereby indicating a high level of polymorphism. The unweighted pair group method with arithmetic mean clustering results indicated that the thirty pairs of primers could precisely distinguish the germplasms. This study reveals the SSR distribution characteristics of red raspberry and provides a scientific basis for further genetic diversity studies and genetic linkage map construction for this species.

Rubus idaeus is a red-fruited species of Rubus L. and has remarkable economic and cultural value. This genus is distributed over both the hemispheres and is commonly cultivated in temperate regions (Somsteby and Heide, 2009). Its berry has important health benefits; therefore, its breeding is now increasing (Chen et al., 2011; Jimenez-Garcia et al., 2012; Kafkas et al., 2008). The identification of reliable molecular markers associated with disease resistance and flower characteristics and genes responsible for the morphological characteristics of plants is necessary for the molecular breeding of R. idaeus. Simple sequence repeats are tandem repeats of 1–6 bp sequence motifs. They are considered the most suitable because of their desirable genetic attributes, including relative abundance, high polymorphism, codominant inheritance, transferability to related species and genera, and good genome coverage (Guo et al., 2016). The expressed sequence tag (EST)-SSRs are widely used in the genetic linkage map and fingerprint construction, molecular marker-assisted breeding, and evolutionary research (Mei et al., 2012). Because of the high cost of developing SSR markers, their usage has previously been limited to a certain extent. With the rapid development of high-throughput sequencing technology, large amounts of EST sequence data have been deposited in several public databases in recent years. In silico mining of SSR markers from the vast deposited sequences provides a time- and cost-effective alternative choice. Although the genome information of R. idaeus is available, its EST resource in the National Center for Bioinformatics Information (NCBI) is still scarce. On 17 May 2016, only 1532 EST sequences and 4820 unigenes of R. idaeus L. were found in NCBI.

In the present study, transcriptome sequencing of the red raspberry (R. idaeus ‘Sijihong’) was conducted, and 59,173 unigenes were acquired after de novo assembly of the sequences. Thus, our transcriptome sequence data are considerably more representative than the current ESTs and unigenes available in R. idaeus L. The massive transcriptome sequence information might be beneficial for screening effective SSR markers in red raspberry. This study aimed to identify and characterize the EST-SSR distribution characteristics by using our transcriptome sequence data and to validate them. Our findings might provide effective molecular markers for further studies on germplasm diversity evaluation, quantitative trait locus mapping, and genetic mapping of important traits in red raspberry.

Materials and Methods

Transcriptome sequencing and assembly. In 2015, our red raspberry transcriptome sequencing was completed by BGI Co., Ltd. Total RNA was extracted from the roots of the red raspberry cultivar Sijihong by using the RNA Easyspin Isolation System (Aidlab Biotech, Beijing, China), and RNase-free DNase I was used to eliminate the residual genomic DNA from the raw RNA extract according to the manufacturer’s protocol (Promega, Beijing, China). The stem tips of ‘Sijihong’ are biennial; its primocane can grow several meters during their first year, and rooting often occurs at the tip of the primocane during autumn. Therefore, the development is divided into four stages: stem elongation stage, withhold growing stage, adventitious root inducing stage, and adventitious root growth stage. The stem tips were randomly collected from each development stage, and the total RNA was extracted from these samples. Equal amount of RNA from different time points for each stage was pooled for library preparation and sequencing. The cDNA library preparation and Illumina sequencing for transcriptome analysis were performed as previously described (Cheng et al., 2015). After low-quality and adaptor-polluted base (N) reads were removed, Trinity was used to perform de novo assembly by using the clean reads (Grabherr et al., 2011), followed by TGICL clustering of transcripts to unigenes (Pertea et al., 2003). All the raw transcriptome data were deposited in Sequence Read Archive (SRA) (NCBI BioSample Accession Number: SAMN06013479).

Received for publication 27 Dec. 2016. Accepted for publication 21 Feb. 2017.

This work was financially supported by the Natural Science Foundation of Jilin Provincial Department of Education (2015213).

1Corresponding author. E-mail: chengyunqing1977@163.com.
SSR detection. MISA (MicroSatellite) identification tool (http://pgsc.ipk-gatersleben.de/misa) was used to screen SSRs in unigenes, followed by primer designing for each SSR by using Primer3 (Thiel et al., 2003). The minimum number of mononucleotide repeats was 12; 2–6 NT repeats were refined as previously described (Boccacci et al., 2015). For compound repeats, the maximum length of interruption was set as 100 bp. The types and percentage of SSR repeat motif types and their frequency were calculated.

GO and KEGG enrichment analysis of SSR-containing unigenes. The annotations of SSR-containing unigenes were acquired after basic local alignment search tool (BLAST) (Altschul et al., 1990) was used to align them to NT and KEGG. Subsequently, GO annotations were obtained using Blast2GO (B2G) (Conesa et al., 2005) with NR annotations, followed by GO classification analysis. The KEGG pathway functional enrichment analysis was performed using phyper, a function of RSEM software package (Li et al., 2011).

SSR primer synthesis and validation. For each SSR-containing unigene, primers were designed using Primer 3.0 version 2.2.2 (http://bioinfo.ut.ee/primer3; Untergasser et al., 2012) with default parameters. The results of the primer design are provided in Supplemental Table 1. Thirty randomly chosen primers were synthesized by Shanghai Biological Engineering Co., Ltd., and their sequences are listed in Supplemental Table 2. Their availability was validated by isolating the total genomic DNA from 15 raspberry cultivars by using the cetyltrimethylammonium bromide method (Soengas, 2006). The polymerase chain reaction (PCR) amplification was performed in 20-μL reaction mixtures with the following composition: 1.0 μL of template DNA (50 ng/μL), 1.0 μL each of the 10 μM 6-carboxyfluorescein-labeled M13 forward primers and reverse primers, 10×Ex Taq buffer (TAKARA BIO Inc., Kusatsu, Japan), and 7.0 μL deionized water. The PCR was performed using the following parameters: 30 cycles of predegradation for 3.0 min at 94 °C, denaturation for 30 s at 94 °C, annealing for 30 s at 55 °C, and extension for 1 min at 72 °C, followed by a 10-min final extension step at 72 °C. Capillary electrophoresis of PCR products was performed using 3730XL sequencer (Applied Biosystems, Foster City, CA). The PIC value was calculated as follows: PIC = 1 – ∑Pr^2 (i = 1 – k, where k = allele number and Pr = frequency of the i allele). Finally, cluster analysis was performed using unweighted pair-group method with arithmetic average (UPGMA) method by using NTSYS-pc version 2.10b software package (Exeter Software, Setauket, NY).

Results

SSR number and proportion in the transcriptome of red raspberry. After the transcriptome sequence assembly, 59,173 unigenes were obtained. The total sequence length was 72,027,384 bp, and the average length was 1217 bp. There were 11,158 SSR-containing sequences in the transcriptome of red raspberry. Among these sequences, more than one SSR was found in 2299 unigenes, and compound SSRs were contained in 1207 unigenes. The total number of SSR was 14,210, and an SSR was found for every 5068-bp sequence (Supplemental Table 3). The longest and shortest SSR sequences were 402 and 12 bp, respectively, and their average length was 24.60 bp.

In the transcriptome sequence data, large differences were observed between different SSR sequences with respect to their quantity and proportion (Fig. 1), which were relatively high for mononucleotide, dinucleotide, and trinucleotide SSRs, measuring 1323 (9.31%), 6752 (47.52%), and 4897 (34.46%), respectively. The sum of mononucleotide, dinucleotide, and trinucleotide SSRs accounted for 91.29% of the total identified SSRs. The number of tetranucleotide, pentanucleotide, and hexanucleotide SSRs was small; their sum was 1238 and accounted for 1.45%, 2.34%, and 4.93% of the total number of SSRs.

SSR motif types and proportions in the transcriptome of red raspberry. In total, 235 SSR motif types were found in the red raspberry transcriptome (Fig. 2). In general, the SSR motif types increased with an increase in the number of NTs. The NT number of SSRs was defined as arguments (x), and SSR motif type was defined as a variable (y). The exponential function was used to simulate the changing trend in the number of motif types along with the number of NTs in an SSR motif. The results indicated that the SSR motif type increased exponentially with an increase in the NT number. Their correlation coefficient was 0.9951, which was highly significant (P < 0.01).

**Fig. 1.** Simple sequence repeat (SSR) number of different nucleotide repeats in the raspberry transcriptome data

**Fig. 2.** Types of simple sequence repeat motifs in every nucleotide repeat in the raspberry transcriptome data
The percentages of repeat motifs less than 0.1% are not listed.

Table 1. Major simple sequence repeat motifs and their percentage in the raspberry transcriptome data.

| Repeat type | Motifs | No.  | Percent |
|-------------|--------|------|---------|
| Mononucleotide | A/T | 1,292 | 9.09 |
| Dinucleotide  | AG/CT | 5,375 | 37.83 |
| Trinucleotide | AAGAGG/CCTCTT | 32 | 0.23 |
| Tetrancleotide | AAAAG/CTTTT | 81 | 0.57 |
| Pentancleotide | AAGG/CTTTT | 32 | 0.23 |
| Hexancleotide | AAAGG/CCCCCTT | 32 | 0.23 |

Different SSR motif proportions were compared (Table 1). The top 10 SSR motifs with the highest proportion were as follows: AG/CT (5375; 37.83%), AAGG/CTTTT (81; 0.57%), AAG/CTTT (55; 0.39%), AG/CT (25; 0.18%), AAT/AT (25; 0.18%), AAGG/CTTT (32; 0.23%), AAGG/CTTTT (32; 0.23%), AAAT/ATTT (19; 0.13%), AACAGG/CCCCCTT (32; 0.23%), and AACAGG/CCCCCTT (32; 0.23%).

Verification and polymorphism analysis of SSR markers. Thirty pairs of SSR primers were randomly chosen and synthesized, and the PCR amplification was performed using genomic DNA samples from 15 raspberry germplasms. Each pair of primers chosen produced at least an amplicon, but only five pairs of primers were found to have high polymorphism characteristics (Supplemental Table 7). In total, 30 polymorphism loci were identified (Table 4). Their PIC values ranged from 0.50 to 0.86, with a mean of 0.73. The unweighted pair group method with arithmetic average (UPGMA) cluster analysis of the 15 raspberry samples was performed (Fig. 4). Eight R. idaeus germplasms were clustered into category I at 0.61, and the other seven germplasms were included in category II at 0.60. In category I, all the eight R. idaeus germplasms were characterized by a hollow fruit. In category II, all the members were Rubus occidentals or their hybrids, and they were characterized by a solid fruit. Except for raspberry cultivars Mac Black and Thornless Red, all the germplasms could be distinguished from each other. Thus, clustering results based on UPGMA method were almost consistent with those of morphological and biological classification.

Discussion

EST-SSRs from Rubus L. are scarce. Overall, 1149 ESTs were obtained from the cDNA library of the red raspberry cultivar Heritage, and 131 SSR-containing sequences were acquired (Bushakra et al., 2015). In this study, in total, 11,158 SSR-containing unigenes and 235 SSR motifs were identified from 59,173 unigenes after transcriptome assembly and SSR search, and 18.67% of the total members were SSR-containing unigenes. Both the cell and cell part subcategories contained 226 members and accounted for 47.1% of the total. The organelle subcategory contained 123 members, accounting for 25.6% of the total. The molecular function category included 10 subcategories. Among them, binding, catalytic, and transporter subcategories contained the largest number of members, and they comprised 233, 222, and 47 SSR-containing unigenes, accounting for 48.5%, 46.3%, and 9.8% of the total members, respectively. The biological process category included 18 subcategories. The proportion of the metabolic process, cellular process, localization establishment, and biological regulation was the highest compared with that of the other subcategories, and they comprised 325, 251, 69, and 59 SSR-containing unigenes, accounting for 52.9%, 52.3%, 14.4%, and 12.3% of the total members, respectively.

Of the 11,158 SSR-containing unigenes, 3441 unigenes had matching KEGG annotations. KEGG pathway enrichment analysis of SSR-containing unigenes was conducted to elucidate their potential metabolic pathways.
nonrepetitive fractions of plant genomes. Our original sequencing data, unigene amount, and the number of developed SSR motif types were considerably more than those included in the current database and showed little duplication of data with the available SSR-containing sequences in NCBI after sequence comparison. The GO classification and KEGG pathway enrichment analysis of SSR-containing unigenes indicated that they were involved in binding, catalytic, and transporter functions, and their biological functions included metabolism, RNA transport, spliceosome, protein processing, and mRNA surveillance, indicating that SSR-containing unigenes might regulate important agricultural traits in raspberry. Thus, these SSR sequences could be widely used in marker-assisted breeding and selection of red raspberry.

The relative abundance of motif types has been attributed to the species-specific SSR distribution in the genome and SSR types (genomic SSR or EST-derived SSR markers; Postolache et al., 2014; Trivedi, 2006; Zhang et al., 2014). The trinucleotide repeat motifs are usually the most abundant in the EST data (Cavagnaro et al., 2011; Zhou et al., 2016); this is consistent with the notion that protein-coding sequences tolerate better frame-shift mutations (InDels) of 3 bp or in multiples of 3 bp than other InDel lengths. However, dinucleotide repeats in red raspberry transcriptome were the most abundant type, which is inconsistent with the results obtained for

Table 2. The frequency of different repeat times for each repeat type of simple sequence repeats in the raspberry transcriptome.

| Repeat no. | Mononucleotide | Dinucleotide | Trinucleotide | Tetranucleotide | Pentanucleotide | Hexanucleotide |
|------------|----------------|--------------|---------------|-----------------|-----------------|----------------|
| 4          | 0              | 0            | 0             | 0               | 258             | 580            |
| 5          | 0              | 0            | 2,379         | 141             | 66              | 42             |
| 6          | 0              | 1,659        | 1,122         | 46              | 6               | 49             |
| 7          | 0              | 1,061        | 618           | 4               | 2               | 8              |
| 8          | 0              | 850          | 409           | 6               | 0               | 4              |
| 9          | 0              | 709          | 56            | 1               | 0               | 7              |
| 10         | 0              | 526          | 86            | 0               | 2               | 2              |
| 11         | 0              | 423          | 56            | 3               | 0               | 1              |
| 12         | 479            | 254          | 39            | 4               | 0               | 1              |
| 13         | 263            | 58           | 53            | 0               | 0               | 3              |
| 14         | 154            | 85           | 24            | 0               | 0               | 2              |
| 15         | 99             | 91           | 17            | 0               | 0               | 1              |
| 16         | 67             | 108          | 8             | 0               | 0               | 0              |
| 17         | 43             | 80           | 8             | 0               | 0               | 0              |
| 18         | 32             | 93           | 6             | 0               | 0               | 0              |
| 19         | 17             | 97           | 11            | 0               | 0               | 0              |
| 20         | 15             | 85           | 0             | 0               | 0               | 0              |
| 21         | 5              | 99           | 2             | 1               | 0               | 0              |
| 22         | 14             | 82           | 0             | 0               | 0               | 0              |
| 23         | 73             | 84           | 1             | 0               | 0               | 0              |
| 24         | 1              | 44           | 0             | 0               | 0               | 0              |
| 25         | 2              | 57           | 0             | 0               | 0               | 0              |
| 26         | 1              | 52           | 0             | 0               | 0               | 0              |
| 27         | 2              | 47           | 0             | 0               | 0               | 0              |
| 28         | 4              | 34           | 0             | 0               | 0               | 0              |
| 29         | 5              | 31           | 0             | 0               | 0               | 0              |
| 30         | 3              | 11           | 2             | 0               | 0               | 0              |
| 31         | 4              | 7            | 0             | 0               | 0               | 0              |
| 32         | 7              | 5            | 0             | 0               | 0               | 0              |
| 33         | 0              | 9            | 0             | 0               | 0               | 0              |
| 34         | 8              | 2            | 0             | 0               | 0               | 0              |
| 35         | 1              | 2            | 0             | 0               | 0               | 0              |
| 36         | 0              | 1            | 0             | 0               | 0               | 0              |
| 37         | 3              | 0            | 0             | 0               | 0               | 0              |
| 38         | 2              | 1            | 0             | 0               | 0               | 0              |
| 39         | 2              | 2            | 0             | 0               | 0               | 0              |
| 40         | 1              | 2            | 0             | 0               | 0               | 0              |
| 41         | 1              | 1            | 0             | 0               | 0               | 0              |
| 41–73      | 15             | 0            | 0             | 0               | 0               | 0              |
| Total      | 1,323          | 6,752        | 4,897         | 206             | 332             | 700            |

Fig. 3. The gene ontology functional classification of simple sequence repeat-containing unigenes. Note: Unigenes were annotated in three categories: cellular components, molecular functions, and biological processes.
Table 3. Metabolic pathway enrichment analysis of simple sequence repeat (SSR)-containing unigenes.

| No. | Pathway                | SSR-containing unigenes with pathway annotation (3441) | Percent of SSR-containing unigenes (%) | Pathway ID | Level 1                  | Level 2                  |
|-----|------------------------|--------------------------------------------------------|----------------------------------------|------------|--------------------------|--------------------------|
| 1   | Metabolic pathways     | 775                                                    | 22.52                                  | ko01100    | Metabolism               | Global map               |
| 2   | RNA transport          | 363                                                    | 10.55                                  | ko03013    | Genetic information processing | Translation             |
| 3   | Spliceosome            | 261                                                    | 7.59                                   | ko03040    | Genetic information processing | Transcription           |
| 4   | Protein processing in  | 206                                                    | 5.99                                   | ko04141    | Genetic information processing | Folding, sorting and degradation |
|     | endoplasmic reticulum  |                                                        |                                        |            |                          |                          |
| 5   | mRNA surveillance      | 187                                                    | 5.43                                   | ko03015    | Genetic information processing | Translation             |
| 6   | Endocytosis            | 169                                                    | 4.91                                   | ko04144    | Cellular processes        | Transport and catabolism |
| 7   | Purine metabolism      | 135                                                    | 3.92                                   | ko06230    | Metabolism                | Nucleotide metabolism    |
| 8   | Ubiquitin mediated     | 125                                                    | 3.63                                   | ko04120    | Genetic information processing | Folding, sorting and degradation |
|     | proteolysis            |                                                        |                                        |            |                          |                          |
| 9   | Pyrimidine metabolism  | 112                                                    | 3.25                                   | ko00240    | Metabolism                | Nucleotide metabolism    |
| 10  | Regulation of actin    | 110                                                    | 3.20                                   | ko04810    | Cellular processes        | Cell motility             |
|     | cytoskeleton           |                                                        |                                        |            |                          |                          |

Table 4. Sequence of simple sequence repeat (SSR) primers developed from the raspberry transcriptome.

| Primer no. | Primer sequence (5'-3') | Repeat motif | Expected size/bp | No. loci | No. polymorphic loci | Polymorphism information content |
|------------|--------------------------|--------------|------------------|----------|----------------------|----------------------------------|
| SSR5       | F: TGTAAAACGACGGCCAGTGAGAAAACC CATTTCACCAGTTA R: TGTTCAGATAATGAAGTGGACCC | AT           | 121              | 6        | 6                    | 0.81                             |
| SSR10      | F: TGTAAAACGACGGCCAGTGGGTCTCTTA R: CCAGTAGATCCGAGATCGAAAAT | TCT          | 105              | 8        | 8                    | 0.86                             |
| SSR13      | F: TGTAAAACGACGGCCAGTGTGAAAGGCT TCTATTTGCACAC R: CATCAAAAACTGGAACCCAAAAC | GAAA         | 128              | 5        | 5                    | 0.73                             |
| SSR20      | F: TGTAAAACGACGGCCAGTGGGATGGTGAAGAGGCT TCTATTTGCACAC R: CATCAAAAACTGGAACCCAAAAC | AGAAA        | 148              | 5        | 5                    | 0.50                             |
| SSR30      | F: TGTAAAACGACGGCCAGTGGGATGGTGAAGAGGCT TCTATTTGCACAC R: CATCAAAAACTGGAACCCAAAAC | CACGAGGCTATG | 138              | 6        | 6                    | 0.75                             |

Fig. 4. Dendrogram of the 15 germplasms produced using the unweighted pair group method with arithmetic mean clustering method derived from five simple sequence repeat markers in raspberry. Note: 1: Huangshu; 2: Heritage; 3: Meercr; 4: Australian Red; 5: Fengmanhong; 6: Autumn Bliss; 7: Bristol; 8: Rubus niveus; 9: Boysen; 10: Mac Black; 11: Thornless Red; 12: M22; 13:Shownee; 14: Black Butte; and 15: Rubus coreorpholius.

alternative splicing, and other forms of posttranscriptional modifications. The frameshift mutations of 2 bp or in multiples of 2 bp might be beneficial for regulating post-transcriptional RNA cleavage and improving the adaptation flexibility to environmental changes, growth, and virulence.

The PIC of a given genetic marker provides the probability that the marker might be inherited by the offspring of one heterozygous parent. PIC has been widely used to determine the polymorphism detection capability of a molecular marker. In Rosales, the PIC of the polymorphic SSRs markers developed from Siberian apricot (Prunus sibirica L.) genomic library ranged from 0.11 to 0.90, with an average of 0.62 (Wang et al., 2014b). In the present study, 30 polymorphism loci were obtained using five pairs of primers with high polymorphism characteristics, and their PIC values varied from 0.50 to 0.86, with an average of 0.73. Thus, the SSRs developed in this study, based on the red raspberry transcriptome, were about as good as those previously reported (Wang et al., 2014b). In addition, the cluster analysis results of different red raspberry cultivars performed using our developed SSRs were relatively consistent with those of previous morphological and biological classifications, indicating that our developed SSRs are valuable makers and can be used in further genetic diversity and genetic linkage map construction studies.

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.

Bocacci, P., C. Beltramo, M.A.S. Prando, A. Lembo, C. Santor, S.A. Mehlenbacher, R. Botta, and T.D. Marinoni. 2015. In silico mining, characterization and cross-species transferability of EST-SSR markers for European hazelnut (Corylus avellana L.). Mol. Breed. 35:1–14.

Bushakra, J.M., K.S. Lewers, M.E. Staton, T. Zhebentyayeva, and C.A. Sasaki. 2015. Developing expressed sequence tag libraries and the discovery of simple sequence repeat markers for two species of raspberry (Rubus L.). BMC Plant Biol. 15:258.

Cavagnaro, P.F., S.M. Chung, S. Manin, M. Yildiz, A. Ali, M.S. Alessandro, M. Iorizzo, D.A. Senalik, and P.W. Simon. 2011. Microsatellite isolation and marker development in carrot - genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across apiaceae. BMC Genomics 12:386.

Chen, J., Z. Shen, and J.R. Liu. 2011. Effects of raspberry phytochemical extract on cell proliferation, apoptosis, and serum proteomics in a rat model. J. Food Sci. 76:192–198.

Cheng, Y., J. Liu, H. Zhang, J. Wang, Y. Zhao, and W. Geng. 2015. Transcriptome analysis and gene expression profiling of abortive and

Fig. 4. Dendrogram of the 15 germplasms produced using the unweighted pair group method with arithmetic mean clustering method derived from five simple sequence repeat markers in raspberry. Note: 1: Huangshumei; 2: Heritage; 3: Meercr; 4: Australian Red; 5: Fengmanhong; 6: Autumn Bliss; 7: Bristol; 8: Rubus niveus; 9: Boysen; 10: Mac Black; 11: Thornless Red; 12: M22; 13:Shownee; 14: Black Butte; and 15: Rubus coreorpholius.

carrot and Siberian wildrye (Cavagnaro et al., 2011; Zhou et al., 2016). Being a widely spread species showing strong adaptation to various environments, raspberry needs to be controlled in many cases owing to its asexual and aggressive propagation. The complexity of the eukaryotic transcriptome is generated by the interplay of transcription initiation, termination, alternative splicing, and other forms of posttranscriptional modifications. The frameshift mutations of 2 bp or in multiples of 2 bp might be beneficial for regulating post-transcriptional RNA cleavage and improving the adaptation flexibility to environmental changes, growth, and virulence.

The PIC of a given genetic marker provides the probability that the marker might be inherited by the offspring of one heterozygous parent. PIC has been widely used to determine the polymorphism detection capability of a molecular marker. In Rosales, the PIC of the polymorphic SSRs markers developed from Siberian apricot (Prunus sibirica L.) genomic library ranged from 0.11 to 0.90, with an average of 0.62 (Wang et al., 2014b). In the present study, 30 polymorphism loci were obtained using five pairs of primers with high polymorphism characteristics, and their PIC values varied from 0.50 to 0.86, with an average of 0.73. Thus, the SSRs developed in this study, based on the red raspberry transcriptome, were about as good as those previously reported (Wang et al., 2014b). In addition, the cluster analysis results of different red raspberry cultivars performed using our developed SSRs were relatively consistent with those of previous morphological and biological classifications, indicating that our developed SSRs are valuable makers and can be used in further genetic diversity and genetic linkage map construction studies.

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.

Bocacci, P., C. Beltramo, M.A.S. Prando, A. Lembo, C. Santor, S.A. Mehlenbacher, R. Botta, and T.D. Marinoni. 2015. In silico mining, characterization and cross-species transferability of EST-SSR markers for European hazelnut (Corylus avellana L.). Mol. Breed. 35:1–14.

Bushakra, J.M., K.S. Lewers, M.E. Staton, T. Zhebentyayeva, and C.A. Sasaki. 2015. Developing expressed sequence tag libraries and the discovery of simple sequence repeat markers for two species of raspberry (Rubus L.). BMC Plant Biol. 15:258.

Cavagnaro, P.F., S.M. Chung, S. Manin, M. Yildiz, A. Ali, M.S. Alessandro, M. Iorizzo, D.A. Senalik, and P.W. Simon. 2011. Microsatellite isolation and marker development in carrot - genomic distribution, linkage mapping, genetic diversity analysis and marker transferability across apiaceae. BMC Genomics 12:386.

Chen, J., Z. Shen, and J.R. Liu. 2011. Effects of raspberry phytochemical extract on cell proliferation, apoptosis, and serum proteomics in a rat model. J. Food Sci. 76:192–198.

Cheng, Y., J. Liu, H. Zhang, J. Wang, Y. Zhao, and W. Geng. 2015. Transcriptome analysis and gene expression profiling of abortive and
developing ovules during fruit development in hazelnut. PLoS One 10:e0122072.

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.

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

Guo, Z.H., K.X. Fu, X.Q. Zhang, C.L. Zhang, M. Sun, T. Huang, Y. Peng, L.K. Huang, Y.H. Yan, and X. Ma. 2016. SSRs transferability and genetic diversity of three allogamous ryegrass species. C. R. Biol. 339:60–67.

Jimenez-Garcia, S.N., R.G. Guevara-Gonzalez, R. Miranda-Lopez, A.A. Feregrino-Perezc, I. Torres-Pachecoa, and M.A. Vazquez-Cruz. 2012. Functional properties and quality characteristics of bioactive compounds in berries: Biochemistry, biotechnology, and genomics. Food Res. Intl. 54:1195–1207.

Kafkas, E., M. Özgen, Y. Ozoğüi, and N. Türemiş. 2008. Phytochemical and fatty acid profile of selected red raspberry cultivars: A comparative study. J. Food Qual. 31:67–78.

Li, B., C.N. Dewey, B. Li, and C.N. Dewey. 2011. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:93–99.

Mei, Y., Y. Han, R. Vanburen, R. Ming, L.M. Xu, Y.P. Han, and Y.L. Liu. 2012. Genetic linkage maps for Asian and American lotus constructed using novel SSR markers derived from the genome of sequenced cultivar. BMC Genomics 13:653.

Morgante, M., M. Hanafey, and W. Powell. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. 30:194–200.

Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karayameva, Y.D. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai, and J. Quackenbush. 2003. TIGR gene indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics 19:651–652.

Postolache, D., C. Leonarduzzi, A. Piotti, I. Spanu, A. Roig, B. Fady, A. Roschanski, S. Liepelt, and G.G. Vendramin. 2014. Transcriptome versus genomic microsatellite markers: Highly informative multiplexes for genotyping Abies alba Mill. and congeneric species. Plant Mol. Biol. Rpt. 32:750–760.

Soengas, P., P. Velasco, G. Padilla, A. Ordás, and M.E. Cartea. 2006. Genetic relationships among Brassica napus crops based on SSR markers. HortScience 41:1195–1199.

Sonstebey, A. and O.M. Heide. 2009. Effects of photoperiod and temperature on growth and flowering in the annual (primocane) fruiting raspberry (Rubus idaeus L.) cultivar ‘Polka’. J. Hort. Sci. Biotechnol. 84:439–446.

Thiel, T., W. Michalek, R. Varshney, and A. Graner. 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.

Trivedi, S. 2006. Comparison of simple sequence repeats in 19 Arabica. Genet. Mol. Res. 5:741–772.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40:e115.

Wang, W., Q. Yi, L. Ma, X. Zhou, H.T. Zhao, X.B. Wang, J. Qi, H.Y. Yu, Z.G. Wang, and Q.Q. Zhang. 2014a. Sequencing and characterization of the transcriptome of half-smooth tongue sole (Cynoglossus semilaevis). BMC Genomics 15:470.

Wang, Z., H. Liu, J. Liu, Y.Y. Li, R.L. Wu, and X.M. Pang. 2014b. Mining new microsatellite markers for Siberian apricot (Prunus sibirica L.) from SSR-enriched genomic library. Sci. Hortic. 166:65–69.

Zhang, M., W.H. Mao, G.P. Zhang, and F.B. Wu. 2014. Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. PLoS One 9:e94881.