De novo transcriptome assembly and SSR marker development in apricot (Prunus armeniaca)

Muhammed Ali KÖSE1,2, Necati ÇETİNSAĞ1,2, Kahraman GÜRÇAN1,2,*
1Genome and Stem Cell Research Center, Erciyes University, Kayseri, Turkey
2Department of Agricultural Biotechnology, Erciyes University, Kayseri, Turkey

Abstract: Apricot (Prunus armeniaca) is an important fruit crop worldwide. We have performed a de novo transcriptome assembly for 7 apricot accessions ('Stark Early Orange' (SEO), 'Hacıhaliloğlu' (HH), 'Perfection', 'Iğdır', 'Roxana', 'Esen1', and 'Esen2'), which yielded a total number of transcripts ranging from 30,363 for 'SEO' to 59,751 for 'Iğdır'. The pool of the reads produced from 7 accessions were assembled into 85,766 transcripts, with an average of 1165.69 nt. Functional annotation (Gene Ontology- GO and Kyoto Encyclopedia of Genes and Genomes- KEGG) was performed successfully for the transcripts. Simple sequence repeats (SSRs) were searched in the transcript pool and 14,722 di-, tri- tetra-, penta-, and hexanucleotide motif loci with a minimum of 5 repetitions for all motifs were identified. Primers were designed for 206 loci, and 72 of them were found to be polymorphic by amplifying diverse 24 apricot accesses, including 7 Plum Pox Virus (PPV)-resistant and 17 PPV-susceptible accessions. In order to test the amplification success of publicly available genomic SSRs (gSSRs) for diverse apricot accesses, an additional 88 published Prunus gSSRs were characterized amplifying the same 24 apricots and only 54 (62%) produced polymorphic bands. The new EST-SSRs could be a reliable source of primers for characterization and mapping studies of apricots, especially because they mostly flank easily scorable tri- and tetranucleotide repeats.

Key words: Microsatellite, molecular markers, functional annotation

1. Introduction
The common apricot (Prunus armeniaca L.) is diploid, with eight pairs of chromosomes (2n = 16) and an estimated genome size of 240 million nucleotides (nt) according to the Genome Database for Rosaceae (https://www.rosaceae.org/). Apricot belongs to the family Rosaceae, subfamily Prunoideae, and it is native to China and Central Asia, which are two primary genetic diversity centers for the species. The near-eastern group, including Turkey, Iran, and the Caucasus, is considered as a secondary center of diversity (Vavilov, 1951; Ercisli, 2009; Halasz et al., 2010; Hegedus et al., 2010).

High-throughput next-generation sequencing (NGS) technologies produce large amounts of data and are thus widely used for transcriptome analysis, allowing quantification of RNA transcripts, discovery of new genes, and a vast amount of polymorphic loci. The potential of NGS in apricot science was reviewed by Martínez-Gómez et al. (2011). NGS technologies have been applied in apricot species to facilitate the transcriptome analysis of several biological and agronomical aspects: seasonal bud dormancy (Zhong et al., 2013) and self- and cross-pollinated pistils (Habu et al., 2014) in Japanese apricot, global gene profiling and the search for potential SSR markers (Dong et al., 2014), oil dynamic accumulation in developing seed kernels for the development of woody biodiesel (Niu et al., 2015) in Siberian apricot (Prunus sibirica L.), Plum Pox Virus (PPV) (Sharka) susceptibility/resistance (Rubio et al., 2015), single nucleotide polymorphism (SNP) discovery (Salazar et al., 2015), the study of the development of embryos (Bai et al., 2016), and SNP discovery and genetic characterization via genotyping by sequencing in common apricot (Gürçan et al., 2016).

Simple sequence repeats (SSRs) have been highly preferred due to their high variability, codominant inheritance, suitability for sharing among laboratories, and cross-species transferability (İpek et al., 2016; Sakar and Ünver, 2016; Sorkheh and Khaleghi, 2016). A variety of SSR markers have been developed for Prunus. Examples include almond (Prunus dulcis) (Testolin et al., 2004; Messina et al., 2004), apricot and Japanese apricot (Prunus mume) (Lopes et al., 2002; Decrooq et al., 2003; Vilanova et al., 2006; Li et al., 2010; Wang et al., 2014), cherries (Clarke and Tobutt, 2003; Sorkheh et al., 2016),
and peach (Aranzana et al., 2002; Yamamoto et al., 2002; Howad et al., 2005; Chen et al., 2014; Dettori et al., 2015). SSR markers along with other marker systems have been used in the construction of apricot genetic maps, genetic diversity assessments, and characterization of apricot germplasm collections (Lambert et al., 2007; Lalli et al., 2008; Dondini et al., 2011; Soriano et al., 2012; Rubio et al., 2014, Decrooq et al., 2014, 2016; Gürçan et al., 2015).

Although a variety of SSRs have been identified in *Prunus* species including apricot, more reliable SSR markers are needed for apricot molecular breeding, particularly for anchoring the parental maps and construction of saturated maps of apricot. Available primers occasionally fail during polymerase chain reaction (PCR), probably due to their having not been tested on a wide set of apricots representing overall apricot genetic diversity. Furthermore, the useful number of SSRs lowers drastically while mapping them to biparental segregation populations since SSRs do not always produce polymorphic alleles for parental accessions. Additionally, the majority of available primers were usually developed from SSR-enriched genomic libraries, derived primarily from intergenic DNA regions. In contrast, EST-SSRs are specifically developed from transcribed regions of the genome and present high potential for linkage to loci of interest. Thus, polymorphic EST-SSRs are valuable in constructing linkage maps, presenting considerable utility for MAS. Here, we report the NGS sequencing and transcriptome profiling of 7 apricot accessions and the development of 72 polymorphic EST-SSR loci obtaining allele sizes of 24 diverse accessions, including 7 PPV-resistant and 16 PPV-susceptible accessions. Additionally, we have studied 88 previously published gSSRs amplifying the same diverse 24 apricots in order to exhibit how many of the published *Prunus* primers are useable in apricot breeding programs.

2. Materials and methods

2.1. Plant material

We used 7 accessions including international, national, and local accessions for the transcriptome analysis: 'SEO', 'Hachhaliloğlu' (HH), 'Perfection', 'İğdır', 'Roxana', 'Esen1', and 'Esen2'. 'SEO' is PPV-resistant and the most commonly used donor for PPV resistance breeding programs in Europe. ‘HH’ is susceptible to PPV and accounts for most dried apricots (about 70%) in Turkey, and thereby is also the main cultivar for apricot studies and breeding programs in Turkey. 'İğdır', 'Perfection', and ‘Roxana’ are important apricots for fresh consumption in Turkey, the United States, and Europe, respectively. 'Esen1' and 'Esen2' are local accessions preferred in the Kayseri region and are highly susceptible to PPV. For characterization of SSR loci, 24 accessions representing genetic diversity in apricot were used including 7 PPV-resistant ('SEO', 'Harlayne', 'Zard', 'M2243', 'M2244', '2_7', 'No 8 Zerdalı') and 17 susceptible apricots ('1297', '2254', 'Adilcevaz 2', 'Gü 2', 'Güz Erği', 'Hachhaliloğlu', 'Harcoat', 'Levent', 'Markuleştir', 'Mektep', 'No 8 Mektep', 'Polumella', 'Roxana', 'Şekerpare Benzeri', 'Y5', 'Y6')

2.2. RNA isolation, library construction, and Illumina sequencing

Fresh leaf tissue (100 mg) was used for total RNA extraction via a commercial plant RNA isolation kit. Integrity, quantity, and quality of the RNA of the 7 samples were determined using a spectrophotometer and a 2100 Bioanalyzer RNA nano chip device (Agilent, Santa Clara, CA, USA). RNA samples were used for sequencing when the RNA Integrity Number (RIN) was >7. The cDNA libraries were constructed and sequenced using an Illumina HiSeq 2000, located at the Genetic Engineering and Biotechnology Institute (GEBI) of TÜBİTAK, Turkey.

2.3. De novo transcriptome assembly and functional and structural annotation

Initially, clean reads were obtained using a FASTX-Toolkit by removing adaptor sequences, discarding the low-quality reads (Phred quality score less than 20, Q ≥ 20) and ambiguous nucleotides ('N') at the end of reads, and finally removing the short reads (length below 50 nt). The high-quality reads were then assembled to reconstruct transcripts for each sample using Trinity software (Haas et al., 2013). All reads of seven accessions were also combined and used to assemble transcripts, which is called the ‘pool’ of transcripts. Functional annotation (Gene Ontology-KEGG) was performed by means of sequence similarity (e-value cutoff of 10⁻⁵) using Trapid (Van Bel et al., 2013), and SSRs were identified using GMATo (Wang et al., 2013) and SSRIT (http://www.gramene.org/db/markers/ssrtool). Transcripts were aligned to Japanese apricot (*Prunus mume*) and peach (*Prunus persica*) genomes using Blast2go software (Conesa et al., 2005).

2.4. Development and characterization of EST-SSR primers and assessment of published gSSR primers

After determining SSR loci in transcripts, primers were designed for 206 randomly selected loci with long internal repeats (≥20 nt) and flanking sequences sufficient for primer design. The PRIMER3 program (Rozen et al., 2000) was used for designing primers with the following criteria: 22–24 bp length, 30%–55% GC content, and 60 °C annealing temperature. Initial screening of markers in order to identify polymorphic loci was performed by amplifying DNA of 24 apricot genotypes and running agarose gels. The PCRs were conducted in a volume of 15 µL containing water, 12 pM each of forward and reverse primer.
primers, 1X reaction buffer, 35 μM MgCl₂, 27 μM each of dNTPs, ~25 ng of template DNA, and 0.25 U of DNA polymerase. The thermal cycler was programmed for denaturation at 94 °C for 5 min followed by 35 cycles of 94 °C for 40 s, 60 °C for 30 s for primer annealing, and 72 °C for 40 s with a final 7-min extension step at 72 °C. The products were run on 2% agarose gels. Fluorescently labeled SSR genotyping and PCR was performed according to the method described by Schuelke (2000) for allele sizing and characterization of the loci that appeared polymorphic in the initial screening on agarose gels. PCR reactions were performed including M13 primer fluorescently labeled with 6-FAM, NED, PET, or VIC amplifying the same 24 apricots. For multiplexing, 1 µL of labeled products of each of four primers was combined with distilled water to a final volume of 200 µL. A 1-µL aliquot was loaded into an ABI 3500 capillary electrophoresis instrument (Applied Biosystems, Foster City, CA, USA). For marker characterization, a genetic similarity matrix based on the proportion of shared alleles was generated and the expected heterozygosity (He), observed heterozygosity (Ho), and polymorphism information content (PIC) were calculated using PowerMarker V3.025 software (Liu and Muse, 2005). In order to increase reliable SSR loci and moreover to exhibit how many of the existing SSRs are indeed usable in apricot breeding programs, a total of 88 previously published gSSRs were also amplified using the same 24 accessions. They included 20 primers developed in almond (Testolin et al., 2004), 50 in apricot (Lopes et al., 2002; Messina et al., 2004; Vilanova et al., 2006), 8 in tetraploid cherry (Cantini et al., 2001), and 10 in peach (Sosinski et al., 2000).

### 3. Results

#### 3.1. De novo transcriptome assembly, functional annotation, and classification

After removing low-quality sequences and trimming adapter and ambiguous nucleotides ('N') at the end of reads, de novo assembly of all clean reads was obtained for the seven accessions (Table 1). De novo assembly produced an average of 43,023 sequences with an average N50 of 1762 nt. Among the seven accessions, apricot ‘Iğdır’ produced the highest number of transcripts (59,751), and total length (76,625,300 nt), as well as average length (1282.41 nt) and maximum length (15,046 nt). The average GC content ranged from 41.75% (‘Iğdır’) to 43.67% (‘SEO’). After combining the quality reads from all 7 samples, the assembly of reads generated a total of 85,766 pool transcripts. The maximum and average length of the pool transcripts were 16,371 and 1165, respectively. Among the pool transcripts, the majority (35,306) ranged between 201 and 500 nt, followed by 15,163 transcripts in the size range of 501–1000 nt. The number of transcripts for the pool with a size over 3000 nt became 6359 and N50 was 2006 nt.

We performed annotation for the transcripts of 7 accessions and the pool. GO assignments were used to classify the functions of the sequences. The transcripts of the pool (85,766 fragments) were assigned to 158,816 GO terms as depicted in the Figure. Of these GO terms, 65,119 (41%), 46,225 (29.11%), and 47,477 (29.89%) were classified into the three GO main categories of “biological process”, “cellular component”, and “molecular function”, respectively. The transcripts were assigned to 20 functional groups in the biological process category and “metabolic

### Table 1. Summary of outputs of sequence analysis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total number of sequences</th>
<th>Total length (nt)</th>
<th>Average sequence length (nt)</th>
<th>Maximum sequence length (nt)</th>
<th>N25</th>
<th>N50</th>
<th>N75</th>
<th>Total GC count (nt)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEO</td>
<td>30,363</td>
<td>28,532,152</td>
<td>939.70</td>
<td>11,477</td>
<td>2325</td>
<td>1529</td>
<td>818</td>
<td>12,461,102</td>
<td>43.67</td>
</tr>
<tr>
<td>Esen1</td>
<td>41,385</td>
<td>35,313,704</td>
<td>853.30</td>
<td>11,364</td>
<td>2058</td>
<td>1331</td>
<td>713</td>
<td>15,297,593</td>
<td>43.32</td>
</tr>
<tr>
<td>Esen2</td>
<td>31,002</td>
<td>35,721,316</td>
<td>1152.23</td>
<td>11,464</td>
<td>2626</td>
<td>1777</td>
<td>1090</td>
<td>15,531,837</td>
<td>43.48</td>
</tr>
<tr>
<td>Perfection</td>
<td>53,641</td>
<td>62,461,631</td>
<td>1164.44</td>
<td>14,376</td>
<td>2911</td>
<td>1937</td>
<td>1132</td>
<td>26,421,993</td>
<td>42.30</td>
</tr>
<tr>
<td>Iğdır</td>
<td>59,751</td>
<td>76,625,300</td>
<td>1282.41</td>
<td>15,046</td>
<td>3096</td>
<td>2066</td>
<td>1268</td>
<td>31,992,294</td>
<td>41.75</td>
</tr>
<tr>
<td>Roxana</td>
<td>45,247</td>
<td>51,677,243</td>
<td>1142.11</td>
<td>14,899</td>
<td>2860</td>
<td>1871</td>
<td>1099</td>
<td>22,020,860</td>
<td>42.61</td>
</tr>
<tr>
<td>HH</td>
<td>39,778</td>
<td>45,176,279</td>
<td>1135.71</td>
<td>12,972</td>
<td>2756</td>
<td>1824</td>
<td>1084</td>
<td>19,421,704</td>
<td>42.99</td>
</tr>
<tr>
<td>Average</td>
<td>43,023</td>
<td>47,929,618</td>
<td>1165.69</td>
<td>16,371</td>
<td>3077</td>
<td>2006</td>
<td>1149</td>
<td>41,314,448</td>
<td>41.32</td>
</tr>
</tbody>
</table>

NA, Not applicable.
process” (18,125) and “cellular process” (17,691) dominated, followed by “single-organism process” (9104).

With respect to molecular function, 13 functional groups were detected and the “binding” term was the most highly represented term (20,974 transcripts), followed by “catalytic activity” (17,677) and “transporter activity” (2448) processes. Coming to the cellular component category, “cell part” and “cell” were the major terms and both were represented by 14,001 transcripts, followed by the term “organelle” (6355) among the 16 functional groups. With respect to KEGG analysis, a total of 5612, 4657, 5865, 5885, 5848, 5847, and 5923 transcripts of ‘SEO’, ‘Esen1’, ‘Esen2’, ‘Perfection’, ‘Iğdır’, ‘Roxana’, and ‘HH’ were assigned with 324, 224, 325, 327, 323, 326, and 323 KEGG pathways, respectively.

3.2. SSR mining and characterization of EST-SSR and published gSSRs

To identify the repeat loci proper for SSR marker development, we screened the transcripts of 7 accessions and the pool for di- to hexanucleotide repeat loci with a minimum of 5 repetitions for all motifs. The total number of repeats discovered for each accession ranged from 4356 (‘SEO’) to 11,037 (‘Iğdır’) and was 14,722 for the pool (Table 2). Frequencies of SSR motifs for each accession were

![Figure. Gene ontology (GO) distributions of assembled transcripts of the pool.](image)

**Table 2. Distribution to different repeat type classes.**

<table>
<thead>
<tr>
<th>Motif</th>
<th>SEO</th>
<th>Esen1</th>
<th>Esen2</th>
<th>Perfection</th>
<th>Iğdır</th>
<th>Roxana</th>
<th>HH</th>
<th>Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Di</td>
<td>2664</td>
<td>3191</td>
<td>3646</td>
<td>5946</td>
<td>7405</td>
<td>4898</td>
<td>4484</td>
<td>10,123</td>
</tr>
<tr>
<td>Tri</td>
<td>1616</td>
<td>1814</td>
<td>2027</td>
<td>2866</td>
<td>3386</td>
<td>2367</td>
<td>2420</td>
<td>4263</td>
</tr>
<tr>
<td>Tetra</td>
<td>56</td>
<td>76</td>
<td>84</td>
<td>172</td>
<td>187</td>
<td>120</td>
<td>105</td>
<td>288</td>
</tr>
<tr>
<td>Penta</td>
<td>11</td>
<td>14</td>
<td>26</td>
<td>31</td>
<td>41</td>
<td>25</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>Hexa</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>25</td>
<td>18</td>
<td>11</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>4356</td>
<td>5107</td>
<td>5792</td>
<td>9040</td>
<td>11,037</td>
<td>7421</td>
<td>7047</td>
<td>14,722</td>
</tr>
</tbody>
</table>
determined. For the pool sequences, the most abundant repeat motif was dinucleotide at 10,123 (68.8%), followed by trinucleotide at 4263 (28.95%), tetranucleotide at 288 (1.95%), pentanucleotide at 30 (0.2%), and hexanucleotide at 18 (0.12%). The dinucleotide AG/CT motifs constituted more than three-quarters (8042, 79.44%) of the total number of dinucleotides, followed by AT/AT (1073, 10.6%), AC/GT (1002, 9.9%), and CG/CG (6, 0.05%). Considering trinucleotides, AGG/CTT was the most abundant repeat motif.

Primers were designed for 206 SSR repeats of ≥20 nt of the pool transcripts. In the initial screening, 171 of the 206 primer pairs produced bands, while the remaining 35 failed to produce PCR products. Of these 171, 65 primer pairs produced nonspecific amplification far from SSR loci patterns and 34 produced monomorphic bands and were not investigated further; the remaining 72 generated one or two polymorphic bands in a diploid manner per accession and could be scored with confidence. Of the 72 SSR loci, 33 contained trinucleotide motifs, while 27 and 12 were tetranucleotide and dinucleotide motifs, respectively. The primers AK1, AK164, and AK200 produced alleles 100–200 nt longer than the original locus size determined by the Primer3 program, suggesting the presence of introns on these loci. We obtained PCR products for all accessions at all loci. For the 72 loci, allele sizes were obtained for the 24 accessions in accordance with the SSR locus evolution. A total of 293 alleles were produced at the 72 loci. The number of alleles per locus ranged from 2 to 12, with a mean of 4.07. Mean He, Ho, and PIC were 0.48, 0.66, and 0.43, respectively. PIC values were >0.50 for 25 loci and ranged from 0.50 to 0.90 for these 25 loci (Table 3).

Table 3. Primer sequences and characteristics of new EST-SSR loci for apricot.

<table>
<thead>
<tr>
<th>Primer sequence (5'→3')</th>
<th>Size</th>
<th>Motif</th>
<th>n</th>
<th>He</th>
<th>Ho</th>
<th>PIC</th>
<th>Primer sequence (5'→3')</th>
<th>Size</th>
<th>Motif</th>
<th>n</th>
<th>He</th>
<th>Ho</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ak1f accctgacccagttgaatgg</td>
<td>295</td>
<td>aagc</td>
<td>3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>ak105f ccaaggggttactgaagc</td>
<td>156</td>
<td>aca</td>
<td>3</td>
<td>0.5</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>ak1r tatttactgctgtgcaatc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak105r ccacagcttcttctgatctt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak3f gttggtgacaggggtcataagga</td>
<td>151</td>
<td>ca</td>
<td>6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>ak107f tccggagttgcaagttctta</td>
<td>257</td>
<td>tca</td>
<td>6</td>
<td>0.6</td>
<td>0.9</td>
<td>0.5</td>
</tr>
<tr>
<td>ak3r tccacaaatctgtagaatgt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak107r gatgtgtgacaaggagtag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak5fctgctgctctcagaaactct</td>
<td>152</td>
<td>ta</td>
<td>5</td>
<td>0.7</td>
<td>0.9</td>
<td>0.6</td>
<td>ak109f ctgctgctgtaggtagcaggt</td>
<td>132</td>
<td>atg</td>
<td>3</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>ak5r aacggaccaaaactcttcta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak109r ggaggtgctcaacgacag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak7fattggaaggggcaacacaa</td>
<td>191</td>
<td>tg</td>
<td>4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.5</td>
<td>ak110f gctgctgctgtgtgatgtg</td>
<td>140</td>
<td>atg</td>
<td>4</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>ak7cr ttcgggctctggttctggtc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak110r ttccagagcagagcagacag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak12f caccaacctttaaaccct</td>
<td>198</td>
<td>aga</td>
<td>5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.5</td>
<td>ak114f ccaacccctttgtacacga</td>
<td>257</td>
<td>cca</td>
<td>2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>ak12rcgatcgcacatcttcttg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak114rttccagagcttcagacgc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak18f ccaacacgcaacatctctaca</td>
<td>245</td>
<td>ac</td>
<td>8</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>ak118f gtcgaaactgggtgtag</td>
<td>258</td>
<td>tgc</td>
<td>4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>ak18rttcgggctctggtttctggtc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak118r gtcgaaactgggtgcatg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak21faatctgtgatggcaacacac</td>
<td>242</td>
<td>agg</td>
<td>3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>ak122f ctgtgacactcactctcact</td>
<td>254</td>
<td>atg</td>
<td>5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>ak21rttgctacacttcttccttc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak122rttcgagcttcactctcag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak23fctccttccttactctctttct</td>
<td>149</td>
<td>aacc</td>
<td>4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>ak123f ccaacgctcttcctcttta</td>
<td>257</td>
<td>ctt</td>
<td>5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>ak23rcgctgctctggaattcaga</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak123rttcgacatcacttcctctc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak24fctctgctgtctgtgcattgac</td>
<td>313</td>
<td>cac</td>
<td>3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>ak124f tctctctgtgtaacgcctctt</td>
<td>146</td>
<td>cct</td>
<td>7</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>ak24artactctcaccgccagctgag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak124rtttcgttccttgtgggagag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak30f aagccgctacttattcttgca</td>
<td>270</td>
<td>tg</td>
<td>4</td>
<td>0.6</td>
<td>0.3</td>
<td>0.6</td>
<td>ak130f taccgtaaagccacccattt</td>
<td>141</td>
<td>ctc</td>
<td>3</td>
<td>0.5</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>ak30rc aaccaacacttgacccgaatc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak130rc ccaacactctcgaacccact</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak32f tatctgctggccacctctct</td>
<td>192</td>
<td>agtg</td>
<td>3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>ak132f ccccaacattacctggaag</td>
<td>264</td>
<td>cag</td>
<td>7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>ak32rcgctttgtaggacctgtaa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak132rc ccatattacaactgctgca</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak33f ggtacgagcttggaagaaag</td>
<td>206</td>
<td>tcca</td>
<td>3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>ak133f ccctctattcctcatgatc</td>
<td>269</td>
<td>ttct</td>
<td>5</td>
<td>0.7</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>ak33r aacgagaaagtggagagggga</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak133rc tttctctctctctctctct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak36fctcaattgcacagttgagc</td>
<td>178</td>
<td>atta</td>
<td>3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>ak138fgagacagacatcttccttg</td>
<td>270</td>
<td>aat</td>
<td>3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>ak36rtacctcttgagcctcctc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak138rc ccaacaaacatgcctctctgta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak37ftgctgttgggaagaaatt</td>
<td>174</td>
<td>aatg</td>
<td>5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
<td>ak140f cacaggtctggtgtgtaga</td>
<td>272</td>
<td>ga</td>
<td>6</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>ak37rctgctgtttggagcagaga</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak140rttaggtgtagggataacacag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak38fgagacagaaatctctctcgtg</td>
<td>234</td>
<td>gctg</td>
<td>4</td>
<td>0.5</td>
<td>0.7</td>
<td>0.5</td>
<td>ak143f actgctgctgctctcttct</td>
<td>268</td>
<td>aat</td>
<td>4</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>ak38rgagctgctgctctctctcttcg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ak143rc taggtgctgctgcttacag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primer sequence (5’-3’)</td>
<td>Size</td>
<td>Motif</td>
<td>n</td>
<td>He</td>
<td>Ho</td>
<td>PIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------</td>
<td>-------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak39f:gtgcaaattcacaagc</td>
<td>264</td>
<td>agg</td>
<td>3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak39r:tgatggggccacacag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak44f:aaaggaacctggtggatt</td>
<td>216</td>
<td>ct</td>
<td>7</td>
<td>0.7</td>
<td>0.5</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak44r:agattgctcaacgtaag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak57f:aggggaccaaaacaaataa</td>
<td>216</td>
<td>ttg</td>
<td>7</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak57r:ttgtgctcaacgtaagcatca</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak62f:agcttcttcatatcagg</td>
<td>274</td>
<td>tta</td>
<td>2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak62r:ctttggtctccctccagg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak69f:cgctctccaggatatg</td>
<td>235</td>
<td>gat</td>
<td>3</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak69r:cgctctccaggatatg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak73f:agggtggtggttgttagg</td>
<td>164</td>
<td>aaga</td>
<td>7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak73r:agggtggtggttgttagg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak75f:ctctccactcctcattg</td>
<td>142</td>
<td>ttcc</td>
<td>2</td>
<td>0.5</td>
<td>0.3</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak75r:ctctccactcctcattg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak80f:gcaccaaccaaaaccaagga</td>
<td>235</td>
<td>aga</td>
<td>4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak80r:gcaccaaccaaaaccaagga</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak88f:gacctctgctccctcc</td>
<td>259</td>
<td>ataa</td>
<td>6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak88r:atctctccttctctcct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak90f:caatggtggttgtgacagca</td>
<td>145</td>
<td>tat</td>
<td>5</td>
<td>0.7</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak90r:caatggtggttgtgacagca</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak92f:gcaccaacaccccaacagt</td>
<td>275</td>
<td>ttaa</td>
<td>3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak92r:gcaccaacaccccaacagt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak98f:ccatctctccttcttcta</td>
<td>142</td>
<td>gagg</td>
<td>3</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak98r:ccatctctccttcttcta</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak99f:gcaccaacaccccaacagt</td>
<td>258</td>
<td>cagt</td>
<td>2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak99r:gcaccaacaccccaacagt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak102f:ctctctccttctcttc</td>
<td>155</td>
<td>taaa</td>
<td>2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak102r:ctacagctctctctctct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak103f:atctctccttcttcttc</td>
<td>254</td>
<td>gca</td>
<td>3</td>
<td>0.6</td>
<td>0.2</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak103r:ctctctccttcttcttc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak104f:atctctccttcttcttctttc</td>
<td>236</td>
<td>at</td>
<td>4</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak104r:ctctctccttcttcttcttctttc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak105f:catctctccttcttcttc</td>
<td>229</td>
<td>agaa</td>
<td>2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ak105r:catctctccttcttcttc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n: Number of alleles, He: expected heterozygosity, Ho: observed heterozygosity, PIC: polymorphism information content.
For the assessment analysis of published SSRs for 24 diverse apricots, 88 primers developed in Prunus species were investigated. Of the total 88 primer pairs, 54 produced nice polymorphic bands in apricot and they were further characterized, obtaining allele sizes of 24 apricot accessions. Of the remaining, 15 (17%), and 14 (16%) produced monomorphic and nonspecific amplicons, respectively, while 5 (6%) failed to produce amplicons (Table 4). The 54 polymorphic loci produced 411 alleles and the number of alleles per locus ranged from 2 to 18, with a mean of 7.47. Mean values for He, Ho, and PIC were 0.69, 0.60, and 0.65, respectively (Table 5). PIC values were <0.50 for 8 loci and ranged from 0.50 and 0.90 for the remaining loci, which indicates that they are highly informative.

4. Discussion
In the last decade, high-throughput sequencing technologies together with an increasing number of reliable bioinformatics tools have accelerated genomic and transcriptomic studies in crop breeding. Transcriptome sequencing has become an efficient method to generate high levels of genomic data, large EST sequences, and molecular markers. In this study, we have performed transcriptome analysis of leaves from 7 apricot accessions and obtained an average of 43,023 transcripts (N50: 1762 nt) for all accessions with a mean length 1095 nt. The results are compatible with recently published transcriptome analysis of apricots: Bai et al. (2016) obtained 59,851, 57,163, and 54,792 contigs from three embryo samples of apricot with average contigs size exceeding 427 nt (N50 of 908 nt). Niu et al. (2015) generated 124,070 contigs (N50: 1603 bp) with the mean length of 829.62 bp in Siberian apricot. Habu et al. (2014) obtained 40,061 contigs ranging from 201 to 7477 bases with an average length of 572 bases by using another type of high-throughput technology, 454 pyrosequencing.

In order to perform functional annotation and classification, the assembled transcripts were subjected to GO annotations and KEGG analysis. The transcripts of the pool (85,766 fragments) were classified into the three GO main categories of biological process, cellular component, and molecular function with numbers 65,119 (41%), 46,225 (29.11%), and 47,477 (29.89%), respectively. The present findings are in agreement with previous results; Dong et al. (2014) reported the highest assignments of transcripts of embryos of Siberian apricot to biological processes (54,667, 40.72%), followed by cellular components (51,551, 38.40%) and molecular functions (28,039, 20.88%). Similarly, GO assignment of transcripts sequenced from buds, leaves, stems, flowers, fruit pulp, and seeds of commercial apricot resulted in maximum hits to biological processes, followed by cellular components and molecular function (Bai et al., 2016). The highest hits were related to the biological process GO category, which may indicate that the analyzed tissues were undergoing extensive metabolic activities. Consistent with that, in the present study and in the other two studies "metabolic processes” and “cellular process” were the most frequent terms in the biological GO category; moreover, KEGG analysis revealed that the highest numbers of transcripts were involved in metabolic pathways. Indeed, we used the freshest growing leaves of shoot tips, which explains the extensive metabolic activities found. The predicted GO terms, together with the KEGG analysis, are useful for future research into gene functions.

Here we identified 14,722 di-, tri-, tetra-, penta-, and hexanucleotide motif loci with a minimum of 5 repetitions for all motifs. The dinucleotide AG/CT motifs constituted more than three-quarters (8042, 79.44%), and CG/CG was

### Table 4. Amplification and polymorphism of published gSSRs in 24 diverse apricots.

<table>
<thead>
<tr>
<th>Source plant</th>
<th>Source article</th>
<th>n</th>
<th>Polymorphic loci (%)</th>
<th>Monomorphic loci (%)</th>
<th>No amplification</th>
<th>Nonspecific amplification (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almond</td>
<td>1</td>
<td>20</td>
<td>9 (45)</td>
<td>2 (10)</td>
<td>3 (15)</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Apricot</td>
<td>2</td>
<td>22</td>
<td>12 (55)</td>
<td>8 (36)</td>
<td>2 (9)</td>
<td>-</td>
</tr>
<tr>
<td>Apricot</td>
<td>3</td>
<td>20</td>
<td>13 (65)</td>
<td>1 (5)</td>
<td>-</td>
<td>6 (30)</td>
</tr>
<tr>
<td>Apricot</td>
<td>4</td>
<td>8</td>
<td>5 (63)</td>
<td>1 (13)</td>
<td>-</td>
<td>2 (25)</td>
</tr>
<tr>
<td>Cherry*</td>
<td>5</td>
<td>8</td>
<td>6 (75)</td>
<td>2 (25)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Peach</td>
<td>6</td>
<td>10</td>
<td>9 (90)</td>
<td>-</td>
<td>1 (10)</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>88</td>
<td>54(62)</td>
<td>14 (16)</td>
<td>5 (6)</td>
<td>15 (17)</td>
</tr>
</tbody>
</table>

*Tetraploid; n, number of SSRs. 1, Testolin et al. (2004); 2, Lopes et al. (2002); 3, Messina et al. (2004); 4, Vilanova et al. (2006); 5, Cantini et al. (2001); 6, Sosinski et al. (2000).
The dinucleotide AG/CT motifs are known to be the most common motifs in plants, and the CG motif is very rare in plants (Varshney et al., 2005). Our findings are consistent with the previous report of Dong et al. (2014), who found that the dinucleotides GA/TC (48.17%) are the most common motif in Siberian apricot. Considering trinucleotides, AGG/CTT was the most abundant repeat motif, which is compatible with previous reports that AAG/CTT was the most common motif in *A. duranensis*, *A. batizocoi*, and *A. hypogaea* (Huang et al., 2016), as well as in *Picea abies* and *Pinus taeda* (Ranade et al., 2015).

SSR developments in *Prunus* have continued since 2000 (Yamamoto et al., 2002; Xu et al., 2004; Xie et al., 2006; Vendramin et al., 2007; Chen et al., 2014; Dettor et al., 2015; Sorkheh et al., 2016). The novel 72 EST-SSR loci will be useful in apricot breeding and especially in mapping and anchoring parental maps. The new 72 EST-SSRs identified in the present study enlarge the number of EST-SSRs identified in apricot species, including the 34

Table 5. Characteristics of published gSSR loci in 24 apricot accessions.

<table>
<thead>
<tr>
<th>Accession no.</th>
<th>Source species</th>
<th>Size range (nt)</th>
<th>n</th>
<th>He</th>
<th>Ho</th>
<th>PIC</th>
<th>Accession no.</th>
<th>Source species</th>
<th>Size range (nt)</th>
<th>n</th>
<th>He</th>
<th>Ho</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssrPaCITA2</td>
<td>Apricot</td>
<td>240–270</td>
<td>8</td>
<td>0.8</td>
<td>0.1</td>
<td>0.7</td>
<td>UDA-002</td>
<td>Almond</td>
<td>148–184</td>
<td>9</td>
<td>0.9</td>
<td>0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>ssrPaCITA6</td>
<td>Apricot</td>
<td>154–238</td>
<td>6</td>
<td>0.6</td>
<td>0.0</td>
<td>0.5</td>
<td>UDA-004</td>
<td>Almond</td>
<td>172–174</td>
<td>2</td>
<td>0.3</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>ssrPaCITA7</td>
<td>Apricot</td>
<td>204–232</td>
<td>9</td>
<td>0.7</td>
<td>0.5</td>
<td>0.7</td>
<td>UDA-005</td>
<td>Almond</td>
<td>166–192</td>
<td>7</td>
<td>0.6</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>ssrPaCITA10</td>
<td>Apricot</td>
<td>166–192</td>
<td>10</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td>UDA-013</td>
<td>Almond</td>
<td>187–195</td>
<td>5</td>
<td>0.6</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>ssrPaCITA11</td>
<td>Apricot</td>
<td>143–157</td>
<td>5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>UDA-015</td>
<td>Almond</td>
<td>118–142</td>
<td>10</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
</tr>
<tr>
<td>ssrPaCITA12</td>
<td>Apricot</td>
<td>154–174</td>
<td>9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>UDA-020</td>
<td>Almond</td>
<td>154–188</td>
<td>11</td>
<td>0.9</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>ssrPaCITA14A</td>
<td>Apricot</td>
<td>158–202</td>
<td>8</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
<td>UDA-021</td>
<td>Almond</td>
<td>107–137</td>
<td>11</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>ssrPaCITA18</td>
<td>Apricot</td>
<td>142–180</td>
<td>7</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
<td>UDA-022</td>
<td>Almond</td>
<td>178–190</td>
<td>5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>ssrPaCITA20</td>
<td>Apricot</td>
<td>216–219</td>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>UDA-025</td>
<td>Almond</td>
<td>130–136</td>
<td>4</td>
<td>0.7</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>ssrPaCITA22</td>
<td>Apricot</td>
<td>182–185</td>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>aprigms3</td>
<td>Apricot</td>
<td>268–280</td>
<td>7</td>
<td>0.8</td>
<td>0.4</td>
<td>0.8</td>
</tr>
<tr>
<td>ssrPaCITA23</td>
<td>Apricot</td>
<td>157–173</td>
<td>7</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
<td>aprigms8</td>
<td>Apricot</td>
<td>191–223</td>
<td>9</td>
<td>0.8</td>
<td>0.3</td>
<td>0.8</td>
</tr>
<tr>
<td>ssrPaCITA27</td>
<td>Apricot</td>
<td>240–280</td>
<td>7</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
<td>aprigms10</td>
<td>Apricot</td>
<td>292–304</td>
<td>8</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>PMS2</td>
<td>Cherry*</td>
<td>187–227</td>
<td>7</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
<td>aprigms11</td>
<td>Apricot</td>
<td>178–184</td>
<td>4</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>PMS30</td>
<td>Cherry</td>
<td>110–170</td>
<td>10</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>aprigms16</td>
<td>Apricot</td>
<td>225–257</td>
<td>9</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>PceGA25</td>
<td>Cherry</td>
<td>183–207</td>
<td>6</td>
<td>0.7</td>
<td>0.5</td>
<td>0.7</td>
<td>UDAp-046</td>
<td>Apricot</td>
<td>125–161</td>
<td>11</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>PMS40</td>
<td>Cherry</td>
<td>118–132</td>
<td>3</td>
<td>0.4</td>
<td>0.0</td>
<td>0.3</td>
<td>UDAp-047</td>
<td>Apricot</td>
<td>191–223</td>
<td>6</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>PceGA59</td>
<td>Cherry</td>
<td>183–195</td>
<td>5</td>
<td>0.7</td>
<td>0.9</td>
<td>0.7</td>
<td>UDAp-048</td>
<td>Apricot</td>
<td>168–172</td>
<td>4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>PMS67</td>
<td>Cherry</td>
<td>151–181</td>
<td>8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
<td>UDAp-410</td>
<td>Apricot</td>
<td>137–169</td>
<td>12</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms1</td>
<td>Peach</td>
<td>179–195</td>
<td>7</td>
<td>0.7</td>
<td>0.4</td>
<td>0.7</td>
<td>UDAp-411</td>
<td>Apricot</td>
<td>103–123</td>
<td>10</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms2</td>
<td>Peach</td>
<td>162–190</td>
<td>5</td>
<td>0.5</td>
<td>0.2</td>
<td>0.4</td>
<td>UDAp-413</td>
<td>Apricot</td>
<td>135–196</td>
<td>12</td>
<td>0.8</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms3</td>
<td>Peach</td>
<td>207–217</td>
<td>7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
<td>UDAp-414</td>
<td>Apricot</td>
<td>168–188</td>
<td>8</td>
<td>0.8</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms4</td>
<td>Peach</td>
<td>168–196</td>
<td>6</td>
<td>0.8</td>
<td>0.9</td>
<td>0.7</td>
<td>UDAp-415</td>
<td>Apricot</td>
<td>164–186</td>
<td>10</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms5</td>
<td>Peach</td>
<td>178–192</td>
<td>5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.5</td>
<td>UDAp-416</td>
<td>Apricot</td>
<td>105–125</td>
<td>10</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms1</td>
<td>Peach</td>
<td>146–166</td>
<td>6</td>
<td>0.8</td>
<td>0.3</td>
<td>0.8</td>
<td>UDAp-418</td>
<td>Apricot</td>
<td>153–185</td>
<td>9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms2</td>
<td>Peach</td>
<td>187–189</td>
<td>2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>UDAp-419</td>
<td>Apricot</td>
<td>158–195</td>
<td>18</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>pchcms4</td>
<td>Peach</td>
<td>241–287</td>
<td>5</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>UDAp-420</td>
<td>Apricot</td>
<td>171–197</td>
<td>12</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>pchcms5</td>
<td>Peach</td>
<td>241–301</td>
<td>12</td>
<td>0.8</td>
<td>0.9</td>
<td>0.8</td>
<td>UDAp-421</td>
<td>Apricot</td>
<td>128–174</td>
<td>10</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
</tr>
</tbody>
</table>

n: number of alleles, He: expected heterozygosity, Ho: observed heterozygosity, PIC: polymorphism information content. *: Tetraploid cherry. Source articles for ssrPaCITA, PM and Pce, pchcms, UDA, aprigms and UDAp series are Lopes et al. (2002), Cantini et al. (2001), Sosinski et al. (2000), Testolin et al. (2004), Vilanova et al. (2006), and Messina et al. (2004), respectively.

Table 5. Characteristics of published gSSR loci in 24 apricot accessions.
identified in apricot (Decroocq et al., 2003; Hagen et al., 2004) and 17 identified in Japanese apricot (Li et al., 2010; Wang et al., 2014). Moreover, the new EST-SSRs contain tetra- and trinucleotide repeats presenting high potential for molecular characterization studies, since these kinds of SSR loci produce clear distinguishable peaks during allele sizing.

Comparison of the novel 72 polymorphic EST-SSRs to 54 polymorphic gSSRs shows that the average number of alleles and the PIC value produced by gSSRs (7.47, 0.65) are higher than those produced by EST-SSR (4.07, 0.43). This higher polymorphism level for gSSRs was expected, because it is known that genomic DNA has higher variability than the more conserved transcribed region (Nicot et al., 2004). Another reason for lower polymorphism levels of the new SSRs is more likely to be their repeat motifs. The new EST-SSRs are mostly tetra- and trinucleotide repeats, which were reported to be less polymorphic in other species, too, such as in pistachio (Topçu et al., 2016) and walnut (Ikhsan et al., 2016). Of the total of 88 published for Prunus, only 54 (62%) produced polymorphic bands in apricot, showing that not all publicly available primers are useable for breeding programs. Particularly for mapping purposes, there is a scarcity of reliable SSRs since most SSR loci do not show polymorphism for parents. For example, Ruiz et al. (2011) evaluated 187 SSRs and were able to map only 50 of them. Therefore, the number of primers is not enough since not all Prunus SSRs are always transferable in Prunus species, as our study showed, and more importantly are not polymorphic for parental accessions. Taken together, this shows that more reliable primers should be developed for apricot mapping studies.

In conclusion, a de novo transcriptome assembly for 7 apricot accessions, ('Stark Early Orange' (SEO), 'Hacıhaliloğlu' (HH), 'Perfection', 'İğdır', 'Roxana', 'Esen1', and 'Esen2') was performed. A high number of SSR loci in the transcriptomes were identified. Seventy-two polymorphic and reliable EST-SSR loci primers were developed and characterized, amplifying 24 diverse apricot accessions. The new EST-SSRs will be useful for molecular breeding researchers in apricot as well as in other Prunus fruits. Furthermore, the low transferability rate (62%) of published Prunus gSSRs to apricots demonstrated the need for more reliable SSR primers in molecular breeding studies of apricots, especially for reliable molecular map constructions.

Acknowledgment
This research was supported by the General Directorate of Agricultural Research and Policies (GDAR) of the Turkish Ministry of Food, Agriculture, and Livestock (Project Number TAGEM/12/AR-GE/10) and the Research Fund of Erciyes University (Project Number FCD-2015-6069).

References


