GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Short Communication

Development and characterization of novel transcriptome-derived microsatellites for genetic analysis of persimmon

C. Luo[1,2], Q.L. Zhang[1] and Z.R. Luo[1]

[1]Key Laboratory of Horticultural Plant Biology, Ministry of Education, Huazhong Agricultural University, Wuhan, China
[2]South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences, Key Laboratory of Tropical Fruit Biology, Ministry of Agriculture, Zhanjiang, China

Corresponding author: Z.R. Luo
E-mail: luozhr@mail.hzau.edu.cn

Genet. Mol. Res. 13 (2): 3013-3024 (2014)
Received September 24, 2013
Accepted March 28, 2014
Published April 16, 2014
DOI http://dx.doi.org/10.4238/2014.April.16.10

ABSTRACT. Oriental persimmon (*Diospyros kaki* Thunb.) (2n = 6x = 90) is a major commercial and deciduous fruit tree that is believed to have originated in China. However, rare transcriptomic and genomic information on persimmon is available. Using Roche 454 sequencing technology, the transcriptome from RNA of the flowers of *D. kaki* was analyzed. A total of 1,250,893 reads were generated and 83,898 unigenes were assembled. A total of 42,711 SSR loci were identified from 23,494 unigenes and 289 polymerase chain reaction primer pairs were designed. Of these 289 primers, 155 (53.6%) showed robust PCR amplification and 98 revealed polymorphism between 15 persimmon genotypes, indicating a polymorphic rate of 63.23% of the productive primers for characterization and genotyping of the genus *Diospyros*.

Transcriptome sequence data generated from next-generation sequencing technology to identify microsatellite loci appears to be rapid and cost-efficient, particularly for species with no genomic sequence information available.

Key words: *Diospyros kaki* Thunb.; Simple sequence repeats; Roche 454 sequencing

INTRODUCTION

Simple sequence repeats (SSRs), also called microsatellites, are considered the markers of choice for the study of plant populations because of their codominant nature, high level of polymorphism and reproducibility and because they are more informative than dominant marker data for the estimation of population structure and genetic diversity (Mariette et al., 2002). In general, SSRs are identified from genomic DNA, but their use is time-consuming and labor-intensive (Thiel et al., 2003). During the last few years, next generation sequencing (NGS) methods have become widely available and cost-effective. NGS data can be screened for the presence of molecular markers such as microsatellites and single nucleotide polymorphisms (SNPs).

Diospyros kaki Thunb. is native to China with abundant genetic resources. But genomic information and EST sequences are lacking, and molecular research is relatively poor for persimmon, compared with other fruit trees such as apple, pear, peach, citrus, and grape. High-throughput sequencing technologies developed in recent years provide a convenient way for the establishment of a rapid and efficient molecular research platform to handle huge datasets for molecular marker development. In an earlier study, we effectively obtained information on expressed sequence tags-simple sequence repeats (EST-SSRs) and targeted region amplified polymorphism (TRAP) using the 9467 *D. kaki* ESTs (Sablok et al., 2011; Luo et al., 2013), but the number of SSR markers was insufficient for genetic analysis of persimmon. In the present study, we focused on the identification and characterization of the microsatellite loci that will be able to delineate and infer genotyping in *D. kaki*, from transcriptome sequence data.

MATERIAL AND METHODS

Plant material and sample preparation

The flesh of 'Luotian-tianshi' (*D. kaki* Thunb.) treated with 5% ethanol for 3 days and control were used for transcriptome sequencing. Total RNA was extracted using TRIzol reagent (Invitrogen, USA) following manufacturer protocol. A total of 15 genotypes of *Diospyros* spp were used for SSR marker validation and included 5 Chinese, 9 Japanese (*D. kaki* Thunb.) and 1 related species (Table 1). Genomic DNA isolation was standardized using fresh leaves by the improved CTAB method. All samples were collected from the Persimmon Repository, Huazhong Agricultural University, Wuhan, China.

NGS 454 transcriptome sequencing and sequence assemblies

Approximately 1 μg RNA was used to generate double-stranded cDNA using the

SMART™ cDNA Library Construction kit (Clontech, USA). Finally, approximately 5 µg cDNA were used to construct a 454 library. Roche GS-FLX 454 pyrosequencing was conducted by the Oebiotech Company in Shanghai, China.

Table 1. Fifteen genotypes of *Diospyros* used in EST-SSR mining based on trancriptome data.

Code	Genotype	Scientific name	Ploidy	Astringent type*	Origin
1	Luotian-tianshi	*D. kaki* Thunb.	2n = 6x = 90	PCNA	China
2	Damopan	*D. kaki* Thunb.	2n = 6x = 90	PCA	China
3	Fuping-jianshi	*D. kaki* Thunb.	2n = 6x = 90	PCA	China
4	Gongcheng-shuishi	*D. kaki* Thunb.	2n = 6x = 90	PCA	China
5	Eshi 1	*D. kaki* Thunb.	2n = 6x = 90	PCNA	China
6	Jirou	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
7	Maekawa-Jirou	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
8	Fuyuu	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
9	Youhou	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
10	Matsumoto-wase	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
11	Xiangxi-tianshi	*D. kaki* Thunb.	2n = 6x = 90	PCNA	Japan
12	Nishimura-wase	*D. kaki* Thunb.	2n = 6x = 90	PVNA	Japan
13	Yama-fuji	*D. kaki* Thunb.	2n = 6x = 90	PVA	Japan
14	Hiratane-nashi	*D. kaki* Thunb.	2n = 9x = 135	PVA	Japan
15	Date plum	*D. lotus* L.	2n = 2x = 30	-	China

*Persimmon can be classified into four types depending on the nature of the fruit's astringency loss at maturity and the change in flesh color: PCNA = pollination constant non-astringent; PCA = pollination-constant astringent; PVA = pollination-variant astringent; PVNA = pollination-variant non-astringent.

A Perl program was written to remove vector sequences and polyA (T) tail from sequences, and the reads lower than 100 bp were removed before assembly. The high quality reads were then assembled with MIRA (Chevreux et al., 2004) to construct unique consensus sequences.

Identification of SSRs and design of SSR primers

All contigs or sequences longer than 100 bp were searched for microsatellites using the MISA software (MIcroSAtellite) (Thiel et al., 2003). The parameters defined for the identification of SSRs were ten minimal repeats for mono-, six minimal repeats for di-, and five minimal repeats for tri-, tetra-, penta-, and hexanucleotide. Primers were then designed using the PRIMER3 software (Rozen and Skaletsky, 2000).

Amplification of SSRs in persimmon

PCR amplification was conducted to validate the newly designed SSR primers. PCR amplifications were performed according to the protocol described in a previous report (Luo et al., 2013). PCR fragments were separated on 6% denaturing polyacrylamide gels followed by silver staining for visualization of the amplicons asper the protocol described by Bassam et al. (1991) with slight modification as per Charters et al. (1996).

Analysis of polymorphic loci

A total of 15 persimmon accessions were used to evaluate the degree of polymorphisms

in newly explored SSR markers. All the visible and clear polymorphic bands were scored in SSR profiles as 0 and 1 on the basis of the absence and presence of the band. Each polymorphic band at a particular position on the gel was considered to be a unique locus in the genome.

RESULTS

A total of 1,250,893 reads were generated from a run of Roche 454 GS-FLX sequencing at the Oebiotech Company of the persimmon transcriptome with a total length of 392,758,416 bp. After trimming the adaptor sequences and removing those shorter than 100 bp, the clean reads were assembled into 83,898 unigenes with an average size of 579 bp.

Of all unigenes (83,898) larger than 100 nucleotides, 8,743 unigenes contained one or more microsatellites, and a total of 42,711 SSR loci were identified (Table 2). The number of different repeat types is shown in Figure 1. Mononucleotide repeats (71.76%) and dinucleotide repeats (18.39%) comprised the two largest groups of repeat motifs. Less commonly found were trinucleotide repeats (8.78%), followed by tetranucleotide repeats (0.63%) and pentanucleotide repeats (0.23%), with hexanucleotide repeats being the least abundant at merely 0.20%. Among the dinucleotide tandem repeats AG/GA was by far the most common, while GAA/TTC was the most abundant in the trinucleotide repeats.

Table 2. SSR mining in transcriptome sequencing data.

	No.	Percentage (%)
Total number of unigenes examined	83,898	
Total size of unigenes examined (bp)	48,559,518	
Number of unigenes containing SSR	23,494	28.0
Total number of SSRs identified	42,711	
Average number of SSRs per 1 kb	0.9	
Number of unigenes containing 1 SSR	14,751	17.6
Number of unigenes containing more than 1 SSR	8,743	10.4

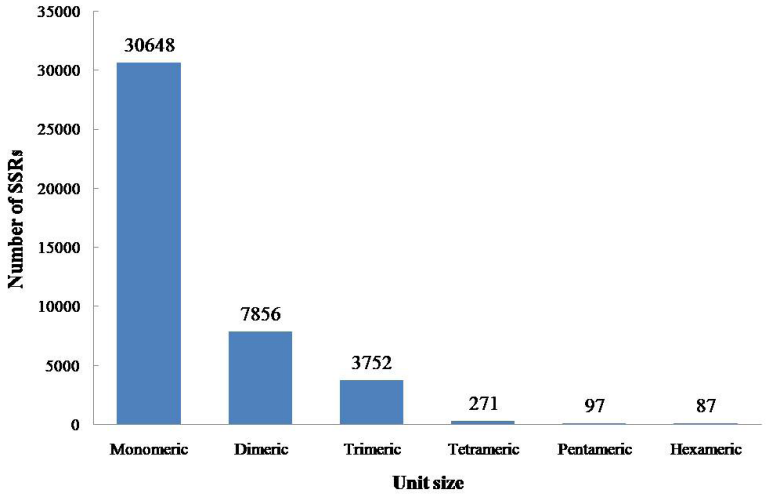

Figure 1. Distribution of SSR type and number from persimmon transcriptome sequencing data.

To verify the effectiveness of the 289 newly designed pairs of SSR primers, they were used to amplify genomic DNA of 15 persimmon genotypes (Figure 2). Of these SSR primers, 155 showed robust PCR amplification and 98 of them revealed polymorphism between 15 persimmon genotypes, indicating a polymorphic rate of 63.23% of the primers for characterization and genotyping of the genus *Diospyros*. One hundred and fifty-five primer pairs (53.63%) yielded successful amplification with products of expected sizes. Of those amplifiable loci, 98 SSR markers (63.23%) were scorable and exhibited polymorphism between the persimmon genotypes examined (Table 3). For each of the polymorphic SSR markers, the sequences of the forward and reverse primers along with the expected PCR product sizes are listed in Table 3. The sequence data for the microsatellite markers were deposited in GenBank (accession No. KC425354-KC425451). The number of alleles detected by the 98 SSR loci ranged from 2 to 15 with an average of 6.11. The percentage of polymorphic bands ranged from 50.00 to 100.00% with an average of 91.38%. Furthermore, eleven SSR loci (DKES58, DKES71, DKES72, DKES76, DKES79, DKES80, DKES85, DKES86, DKES87, DKES89, and DKES96) could distinguish 'Fuyuu' from its bud sport, 'Matsumoto-wase', and 'Maekawa-Jirou' could be differentiated by the SSR loci DKES24 and DKES79 from 'Xiangxi-tianshi', which is believed to be the bud sport from 'Maekawa-Jirou'.

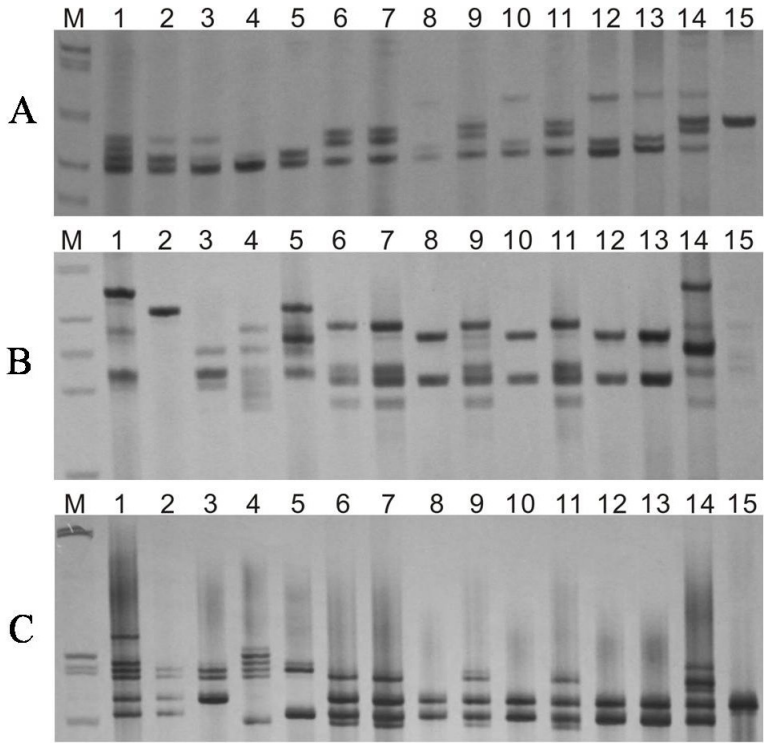

Figure 2. Representative examples of PCR products amplified using EST-SSR primer pairs and separated by electrophoresis on 6% denaturing polyacrylamide gels: *Lanes 1-15* = DNA samples from persimmon genotype Luotian-tianshi, Damopan, Fuping-jianshi, Gongcheng-shuishi, Eshi 1, Jirou, Maekawa-Jirou, Fuyuu, Youhou, Matsumoto-wase, Xiangxi-tianshi, Nishimura-wase, Yama-fuji, Hiratane-nashi, and Date plum, respectively; *lane M* = pBR322 DNA/*Msp*I marker. **A.** Primer DKES51; **B.** Primer DKES53; **C.** Primer DKES55.

Table 3. Characteristics of the 98 microsatellites in *Diospyros* L.

Primer code	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Size of expect PCR product (bp)	Total number of bands	Total number of polymorphic bands	Percent of polymorphic bands (%)
DKES1	KC425354	$(AG)_7$	F: GGCACAAGGGTTCATCTCAT R: GGGAAAAGGAGAGGGAACAG	60	270	2	1	50
DKES2	KC425355	$(CA)_9$	F: AGCTGATGCAAAAGAACGTC R: GTGCTCGTGGTGTGTATTGG	59	202	7	7	100
DKES3	KC425356	$(ATT)_5$	F: AGTTACATCAGTGCTGGGGG R: GTGATTAAGGACAGGGCCAA	59.9	208	4	4	100
DKES4	KC425357	$(TCT)_5$	F: ACCGATGCTCGATTGATTTC R: GATCACATACACCACTGCCG	59	182	2	1	50
DKES5	KC425358	$(TC)_8$	F: AAGAAACAGCAGCTACGGGA R: AGGGTGGGAGGAACTTCTGT	59	225	2	2	100
DKES6	KC425359	$(TC)_6$	F: TTGATCCTCTGAATCCGGTC R: AGAGGGAGGGAGCTGTGATT	60	100	5	5	100
DKES7	KC425360	$(AGA)_5$	F: TGTTGAAAGAGCGCAAAATG R: CAATCATGGCTCTCTCTCCC	59.8	157	7	7	100
DKES8	KC425361	$(CT)_7$	F: AAGCCAACTTCACTGTCGCT R: GGCGTTTGGGAGAAAACATA	60	232	2	1	50
DKES9	KC425362	$(TCA)_6$	F: CACTCTGGCTCTCATCACCA R: ACACGGAGAAGGTCCACAAC	60	163	6	4	66.7
DKES10	KC425363	$(TGT)_6$	F: CCCACATAGTGGGATAACGG R: GAAAGCTCCAATTTCCATGC	59.8	186	4	3	75
DKES11	KC425364	$(AGA)_{10}$	F: AATCCCATATCACCCAGTCG R: CGTGGAATATGATTGAGCCA	59.6	130	11	11	100
DKES12	KC425365	$(ATG)_8$	F: TCATTCCCACAAAAACCTCA R: CCTGTGCTTGGCCTAATGTT	59.5	180	9	9	100
DKES13	KC425366	$(TCT)_6$	F: AGCTTTGTCATCTGGGTGCT R: ACTCGAAGGGGCCAGTTATT	59.8	158	8	8	100
DKES14	KC425367	$(CTA)_6$	F: AGCATGGGTAATGTGGTGGT R: ACGGTATGCGATTGTTCTCC	60	266	6	6	100
DKES15	KC425368	$(CCG)_5$	F: GGTTCTTGAAGACTGCTGCC R: AGGTCAAAGGAGGAGGAGGA	60.5	192	4	3	75
DKES16	KC425369	$(AG)_6$	F: TGGCGTTGAGATTGTTGTGT R: GCAGTCTCTTTCCCAACTGC	60	201	5	5	100
DKES17	KC425370	$(AT)_9$	F: AAGGGCTTGATTGCAGACAG R: TGTTTCTGATGAAAAGCAAAAGA	60.2	129	10	10	100
DKES18	KC425371	$(TGC)_7$	F: AGCTTTCTGTGTCTCCCACA R: GGAAATCCATGCTCTCTCCA	59.2	165	6	6	100
DKES19	KC425372	$(TCA)_7$	F: TGCAGGGAACCTAATCTCTTC R: CGAGTTGGTGCAAGCATAGA	59	103	10	10	100
DKES20	KC425373	$(GA)_7$	F: GGAACTGCATTGATGCAAAA R: GTCAATTGGGCGTTTCTGTT	60	113	5	5	100

Continued on next page

Table 3. Continued.

Primer code	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Size of expect PCR product (bp)	Total number of bands	Total number of polymorphic bands	Percent of polymorphic bands (%)
DKES21	KC425374	$(CTC)_7$	F: CCCTTTGTACTTCGCCACTC R: TTCGATCAGATGGCCTCTTT	59.7	261	3	2	66.7
DKES22	KC425375	$(CT)_7$	F: CCAACCACCTTCCTCTTCCT R: ATGCCATCTGAACGTGATGA	60.2	209	6	6	100
DKES23	KC425376	$(GAA)_6$	F: ATGCTAGAGAAACGGGTGGA R: CAAACAAACTGGCAGCAAAA	59.7	204	5	5	100
DKES24	KC425377	$(CTT)_9$	F: CTTTATCGCTCCCATTTCCA R: GAATCCCCCGTAGATTCACA	59.8	213	7	7	100
DKES25	KC425378	$(TTC)_6$	F: TGGGGAGGAGTTGATGAAAG R: GAATCATGGGGGTAGCAAGA	59.9	233	10	10	100
DKES26	KC425379	$(ATA)_9$	F: CAACTATTGCACCGACAGGA R: AAAGGGAAAAGCATGCAGAA	59.7	195	11	11	100
DKES27	KC425380	$(TGA)_8$	F: ATAGAAAGCGCCTCCTCCTC R: CCGCTATGGTCACCAAGTTT	60	106	10	10	100
DKES28	KC425381	$(AG)_{15}$	F: GGTGAGCAGTCGCATCATAA R: TCTCCGTCTTCTGCTTCCAT	59.9	273	6	5	83.3
DKES29	KC425382	$(AC)_7$	F: AAGACCAGCAGCAGGAAAAA R: TGCATTGCATCATTTTCAGG	60.3	172	6	6	100
DKES30	KC425383	$(GA)_{13}$	F: TTCCGATTCCCAGAATCAAG R: AATGGGTTCAAGGGAGAAGC	60.2	269	4	4	100
DKES31	KC425384	$(AG)_6$	F: TCATTGATGCCTTCCTCCTC R: TGAGTCGTGTGGCTCTCTGT	59.8	268	4	3	75
DKES32	KC425385	$(GA)_9$	F: CTCTCTGGCGACTCTGACCT R: GTGCCAAATTAGAAAGCCCA	59.8	219	3	3	100
DKES33	KC425386	$(GAT)_5$	F: TTGTGGTTGGTGCTGTTGAT R: CCGGGAAACCAGTATTCTCA	60	207	3	2	66.7
DKES34	KC425387	$(CT)_{11}$	F: AAACGTATGCTCCCTCGCTA R: GGGGGAGAGAGAGAAGGAAA	59.8	156	9	9	100
DKES35	KC425388	$(GA)_{14}$	F: GCAACTGCAAGACGAAATGA R: TGTGCAGTTACAAAGCTGCC	60	266	3	3	100
DKES36	KC425389	$(CT)_8$	F: CGGGAGAGAGAGATTCAACG R: GAGAGAGGAGACTTCGGGCT	60	165	7	7	100
DKES37	KC425390	$(ATG)_9$	F: TCATTCCCACAAAAACCTCA R: CCTGTGCTTGGCCTAATGTT	59.5	192	10	10	100
DKES38	KC425391	$(AT)_{12}$	F: AAAAGAAGCTGATCTGCCCA R: TCCAACCAACCAAGTCTTCC	59.9	257	5	4	80
DKES39	KC425392	$(AGA)_8$	F: CGATGTACCCCGTTTTGTTT R: AGCAACCATGGCTCTCTCTC	59.6	185	5	5	100
DKES40	KC425393	$(AGT)_{12}$	F: GATGGATGGATGGAGATTGG R: GGCATTTTAATGCAATCAGGA	60	155	12	12	100

Continued on next page

Table 3. Continued.

Primer code	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Size of expect PCR product (bp)	Total number of bands	Total number of polymorphic bands	Percent of polymorphic bands (%)
DKES41	KC425394	$(CT)_{11}$	F: TCTCCGTCTTCTGCTTCCAT R: GGTGAGCAGTCGCATCATAA	59.9	263	5	4	80
DKES42	KC425395	$(AG)_9$	F: AATTACACGCTCAGGGATG R: ATGCTTCTCAGGAGCTTCA	60	242	3	2	66.7
DKES43	KC425396	$(AAG)_6$	F: GAGGGGAGTGGAGGGTAGAG R: GAAGAATGAAACTGAGCGGC	60	132	13	13	100
DKES44	KC425397	$(AG)_9$	F: TGGTGGAGAGAAATGCAGAA R: AAACAATGGGAGGCTTTTAGC	59.4	154	6	6	100
DKES45	KC425398	$(CA)_{14}$	F: GAATTCGAATGGAACCAACC R: AAGCCCACACATGCCTAATC	59.5	239	12	11	91.7
DKES46	KC425399	$(TC)_{10}$	F: GGTGCAGAGTGGGTTAATGG R: ATCGATAAAGGGTCCTTGCC	60.3	211	9	8	88.9
DKES47	KC425400	$(GA)_{13}$	F: GAAGCCATCCAGTCCCAGTA R: ATACAGATCAATCGCCACCC	59.8	265	4	4	100
DKES48	KC425401	$(AG)_8$	F: TAACCCCTCCAAAACCTTCC R: ATGAAGGACGAGGTTGGTTG	60.5	241	8	8	100
DKES49	KC425402	$(TC)_{11}$	F: GGTGGTGTACCACAGGAAGG R: CACAGACAGACAGAGAGACGACA	60.4	215	5	5	100
DKES50	KC425403	$(TCT)_7$	F: ATCGGAGTTGCTTTGAGCAT R: AAAACACCACTTTTGCCGGTC	59.9	228	6	6	100
DKES51	KC425404	$(TCA)_8$	F: GAGGCAGAGGCATGAAAGTC R: GGTCTGAAGCTCGCACCTAC	60	208	5	5	100
DKES52	KC425405	$(TC)_8$	F: GCGAGATGAGAAGGGTTCAG R: GGAAGACGAACTTGCGAAAG	59.9	106	15	15	100
DKES53	KC425406	$(ATG)_8$	F: CATTCATTCCCACAAAACCC R: CCTGTGCTTGGCCTAATGTT	60	192	8	8	100
DKES54	KC425407	$(ACT)_7$	F: AAGTTCGTTCCGGGAATAC R: TGGGCATATTGGGTCTTGTT	60.2	182	10	10	100
DKES55	KC425408	$(TCT)_{12}$	F: CCCAGTTGTTCCAACGCTAT R: TGTGAAAATGGGGTGCCTAT	60.1	238	8	8	100
DKES56	KC425409	$(GA)_{13}$	F: ATAGGGACAACCCACACAGC R: CAGTGATGAGGCCAAGTGAA	59.8	151	6	6	100
DKES57	KC425410	$(ACA)_7$	F: AAGGCCAATAATGGATGCTG R: ATGATCATGCTTTGCTGCTG	59.9	267	4	4	100
DKES58	KC425411	$(GA)_{11}$	F: GGAGGGGGTTTCAAGTTTCT R: CTTGGACTTCAAGGTCTCCG	59.6	280	7	7	100
DKES59	KC425412	$(AG)_9$	F: ACGCAAGTTTGAAGTCGGAG R: TCTGCTTCTTGTTCCACACG	60.2	104	6	6	100
DKES60	KC425413	$(CT)_{12}$	F: GAAGCGCTGTGGGATAGAAG R: AGCGCTGTACTTTTGGTGGT	59.9	230	7	7	100

Continued on next page

Table 3. Continued.

Primer code	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Size of expect PCR product (bp)	Total number of bands	Total number of polymorphic bands	Percent of polymorphic bands (%)
DKES61	KC425414	$(TC)_9$	F: CCACAAGTTCCTCCTCTTGG R: CCCCAATCCATTCCTTCTTA	59.2	161	5	5	100
DKES62	KC425415	$(GAA)_8$	F: GAGGGATAAGAGGGAAGGTGA R: TTTGCTTTGATGTCATTGCC	59.6	274	3	2	66.7
DKES63	KC425416	$(AT)_{10}$	F: CTGGGCTCTTCACCGTAGAC R: CAAGTACGTTGCCTTGAGCA	59.9	248	4	4	100
DKES64	KC425417	$(TC)_{10}$	F: AACAGCTGATCATCGAGCAC R: ACCAAGCTCCTACCCTCCAT	59.5	211	4	4	100
DKES65	KC425418	$(TC)_9$	F: CGAGGTAGACGTAGGCGAAG R: AGAGTAGGTGGTGGTGGTGC	60	183	4	3	75
DKES66	KC425419	$(TC)_9$	F: GAAGTTGATTGCGCCTTCTC R: GAGAGGGGTGTGCTTTGTGT	60	135	9	9	100
DKES67	KC425420	$(CTT)_{15}$	F: CGTCTCCCCCTTTTCTTCT R: TCAGATAAGCGTGGACGATG	59.7	280	12	12	100
DKES68	KC425421	$(TCT)_{13}$	F: TCTTGGGCTTGGACTTGAAC R: TTGAACAACCGAACCACAAA	60.1	249	5	5	100
DKES69	KC425422	$(TCT)_8$	F: GGGACGATTTTTGTTTCCCT R: GTATTAACGGTGGAAGCCGA	60.1	204	8	8	100
DKES70	KC425423	$(GCT)_8$	F: CAGACTTGGAGAGAGGGACG R: GAGTGGGGCTGTCTGTCTGT	60.2	216	6	5	83.3
DKES71	KC425424	$(GGA)_7$	F: TTCGATCAGATGGCCTCTTT R: CCCTTTGTACTTCGCCACTC	59.7	261	4	4	100
DKES72	KC425425	$(TCG)_7$	F: GCTCGTTTTCTCAACTTGGC R: CCAGCAACAAACTCCTCGAT	60.1	218	5	5	100
DKES73	KC425426	$(GA)_{10}$	F: TCTGACCTTCCGATCAGACC R: AAGACCGTGACGACGAAGAC	60.2	221	6	5	83.3
DKES74	KC425427	$(GCT)_6$	F: GCCTTGATGGGAAGTTGGTA R: AACACCAGCAAGTGCAACAG	57.9	235	7	5	71.4
DKES75	KC425428	$(GCG)_6$	F: AAGGAAAGCGGTTTCGAGAG R: TCATCTCCACCAATCGATCA	60.4	250	3	3	100
DKES76	KC425429	$(AGG)_6$	F: ATCGATTCCATGTCCACCAT R: TGTCTTCTTCCCGCTTCAAT	59.9	213	10	9	90
DKES77	KC425430	$(CT)_{10}$	F: GCCTTCTCTTTCGCACCTTA R: ATCTTCGCTTTCTCTTCCCC	59.6	276	5	3	60
DKES78	KC425431	$(GA)_9$	F: TTTCGTCCTGTTTGTTTCCC R: AGAGCATGAACATGCAGCAG	60	227	10	10	100
DKES79	KC425432	$(CTT)_{14}$	F: CTTAATCGCTCCCATTTCCA R: GAATCCCCCGTAGATTCACA	59.9	229	10	10	100
DKES80	KC425433	$(GCT)_7$	F: GAGATTCTGTGTGCTGGCAA R: GAAGGAAATACCAGACGCCC	60.4	253	5	4	80

Continued on next page

Table 3. Continued.

Primer code	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Size of expect PCR product (bp)	Total number of bands	Total number of polymorphic bands	Percent of polymorphic bands (%)
DKES81	KC425434	$(AAG)_7$	F: ACAGGCTGGACCCTTATCCT R: ATGCCACTGGTTTTTGCTA	60.3	245	3	2	66.7
DKES82	KC425435	$(CGC)_7$	F: TCATCTCCACCAATCGATCA R: AAGGAAAGCGGTTTCGAGAG	60.4	256	3	2	66.7
DKES83	KC425436	$(GAG)_6$	F: GAGAGTGAGAAAAGCGGTGG R: AAATCTGGCAAACATAGCCG	60	202	3	1	33.3
DKES84	KC425437	$(TTC)_7$	F: TATCAATGGACTGGCAAGCA R: CGGCCAACAAAATTGAACTAA	60.1	265	5	5	100
DKES85	KC425438	$(GA)_9$	F: CGCAGAAGAGGATAGTTGGC R: GACTTTCCCACTACCCGACA	59.9	250	5	5	100
DKES86	KC425439	$(AG)_{10}$	F: ACCCGTATCTGCAACTCGAC R: CCAGTGGCGTATCTAAGGGA	60.1	158	4	4	100
DKES87	KC425440	$(AG)_{14}$	F: CCGACATCAACTACAGCCAA R: TCGTTTCATTTTTCAATTTCCC	59.9	266	3	3	100
DKES88	KC425441	$(TC)_{10}$	F: AGGACTGCTCAGCCACTGTT R: CATCCACACACAATGATGCC	60.4	203	2	2	100
DKES89	KC425442	$(GA)_{13}$	F: GGAGGGTGTTTGAGAAGCTG R: GAAATTACACGCACGCACAT	59.7	245	9	9	100
DKES90	KC425443	$(GAT)_8$	F: CCATAGAGGGTTTTGCTCCA R: GTTCTTCAGCGGCCTTCTTT	60.4	150	3	2	66.7
DKES91	KC425444	$(TCT)_7$	F: GGCCAGACAAGGTTGTGATT R: CCGAAACAAGACAGAGAGCC	59.9	197	5	4	80
DKES92	KC425445	$(CT)_{13}$	F: AATCCAAGCTCACTGAACGG R: CCCTATGGAATGCAGGAAGA	60.1	162	6	6	100
DKES93	KC425446	$(TTG)_7$	F: TTCATCTAGCCAACCCCATC R: AAAGTAACCCAAGGATGCCC	59.9	208	5	5	100
DKES94	KC425447	$(TC)_{10}$	F: AGCTACCAGTTCCAACCACG R: AAAAGGTACCCCGAAACACC	60.1	213	6	4	66.7
DKES95	KC425448	$(GA)_{16}$	F: GCACAATGCTTGTGAAAGGA R: ACTGATAACCCCAGCACCAC	59.8	159	5	5	100
DKES96	KC425449	$(TC)_{11}$	F: CCTGATACACCCGATGCTCT R: CAGTGAATTCCAAGCGTCAA	59.9	181	8	8	100
DKES97	KC425450	$(TTC)_8$	F: GGACATGCGTGTGCTCTTTA R: ACGGTTGATAGGCAAAGCAC	59.9	211	4	4	100
DKES98	KC425451	$(AAG)_8$	F: TCACGTTTTTGTCGTTACGC R: ACCAATCAGACCTCAGCCAC	59.9	227	5	5	100

DISCUSSION

One major application of NGS in population genetics is the development of a large numbers of gene-based genetic markers such as SSRs and SNPs, of which a few are currently available in the public database for persimmon. SSR markers have been explored from NGS data for many plant species (Wang et al., 2010; Dutta et al., 2011; Zhu et al., 2012; Kaur et al., 2012). In the present study, Roche 454 sequencing technology was applied to sequence the persimmon transcriptome. With a sequencing run, a total of 1,250,893 reads with a length of 392,758,416 nucleotides were sequenced. To identify potential SSR loci, the unigenes were searched *in silico* using MISA. The average frequency of the transcriptomic-derived SSRs was one in 0.9 kb, which was clearly higher than that reported for the EST-SSRs in barley (one per 6.3 kb) (Thiel et al., 2003), sweet potato (one per 7.99 kb) (Wang et al., 2010) and sesame (one per 6.55 kb) (Zhang et al., 2012a). Furthermore, it has been emphasized that the frequency of SSRs is correlated with many factors, such as SSR mining software, SSR finding criteria, data-set size, and different materials (Varshney et al., 2005). Except for the mononucleotide repeat unit, dinucleotide repeat units were predominant, followed by tri-, tetra-, penta-, and hexa-nucleotide repeat units, which was consistent with the theory reported by Kaur et al. (2012). However, trinucleotide repeat units were predominant in peanut (Zhang et al., 2012b), red pepper (Lu et al., 2012), field pea and faba bean (Kaur et al., 2012). AG/CT motifs were abundant among the dinucleotide EST-SSRs, while AAG/CTT motifs were predominant in trinucleotide EST-SSRs. The relative proportion of different SSR motif types in persimmon observed was similar to that for other plant species (Gao et al., 2003; Chen et al., 2006; Feng et al., 2009).

A subset of the putative SSRs was evaluated against 15 persimmon genotypes, and 98 SSR loci exhibited repeat length polymorphism. The polymorphic rate of EST-SSR in the present study for persimmon was 63.23%, which was higher than that in sesame (11.59%) (Zhang et al., 2012a), pigeonpea (12.91%) (Dutta et al., 2011), faba bean (29.6%) (Kaur et al. 2012), peanut (40.63%) (Zhang et al., 2012b), field pea (46.5%) (Kaur et al., 2012), and sweet potato (51.09%) (Wang et al., 2010).

In summary, Roche 454 sequencing was utilized to expedite the process of microsatellite discovery in the present study. Ninety-eight polymorphic SSRs were added to the existing repertoire of molecular markers in persimmon. These markers characterized here will provide important resources for the identification, parentage analysis, and analysis of genetic diversity in persimmon.

ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of China (#31171929, #31071771) and the Special Fund for Agro-scientific Research in the Public Interest (#201203047).

REFERENCES

Bassam BJ, Caetano-Anolles G and Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80-83.

Charters YM, Robertson A, Wilkinson MJ and Ramsay G (1996). PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5'-anchored simple sequence repeat (SSR) primers. *Theor. Appl. Genet.* 92: 442-447.

Chen C, Zhou P, Choi YA, Huang S, et al. (2006). Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet.* 112: 1248-1257.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, et al. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* 14: 1147-1159.

Dutta S, Kumawat G, Singh BP, Gupta DK, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 11: 17.

Feng SP, Li WG, Huang HS and Wang JY (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23: 85-97.

Gao LF, Tang JF, Li HW and Jia JZ (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* 12: 245-261.

Kaur S, Pembleton LW, Cogan NO, Savin KW, et al. (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13: 104.

Lu FH, Cho MC and Park YJ (2012). Transcriptome profiling and molecular marker discovery in red pepper, *Capsicum annuum* L. TF68. *Mol. Biol. Rep.* 39: 3327-3335.

Luo C, Zhang F, Zhang QL, Guo DY, et al. (2013). Characterization and comparison of EST-SSR and TRAP markers for genetic analysis of the Japanese persimmon *Diospyros kaki*. *Genet. Mol. Res.* 12: 2841-2851.

Mariette S, Le C, V, Austerlitz F and Kremer A (2002). Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Mol. Ecol.* 11: 1145-1156.

Rozen S and Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.

Sablok G, Luo C, Lee WS, Rahman F, et al. (2011). Bioinformatic analysis of fruit-specific expressed sequence tag libraries of *Diospyros kaki* Thunb.: view at the transcriptome at different developmental stages. *3 Biotechnol.* 1: 35-45.

Thiel T, Michalek W, Varshney RK and Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411-422.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Wang Z, Fang B, Chen J, Zhang X, et al. (2010). *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11: 726.

Zhang H, Wei L, Miao H, Zhang T, et al. (2012a). Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13: 316.

Zhang J, Liang S, Duan J, Wang J, et al. (2012b). *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* 13: 90.

Zhu H, Senalik D, McCown BH, Zeldin EL, et al. (2012). Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor. Appl. Genet.* 124: 87-96.