Transcriptome sequencing of Himalayan Raspberry (*Rubus ellipticus*) and development of simple sequence repeat markers

Samriti Sharma¹ · Rajinder Kaur¹ · Amol Kumar U. Solanke² · Himanshu Dubey² · Siddharth Tiwari³ · Krishan Kumar⁴

Received: 15 September 2018 / Accepted: 19 March 2019 / Published online: 30 March 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Key message *Rubus ellipticus* is a wild crop having less number of EST-SSR markers. First attempt was made towards the transcriptomics data analysis and generation of EST-SSR markers which were used in genetic diversity and transferability studies.

Abstract *Rubus ellipticus* is a raspberry with yellow fruits, native to tropical and subtropical India and Asia. Leaves of *Rubus ellipticus* 'Kumarhatti' collection were utilized for cDNA library construction. More than 15 million sequencing reads were generated using NextSeq 500 Illumina RNA-seq technology. The DNASTAR software was used for de novo assembly from which 7777 unigenes with an average length of 500 bp was obtained. These unigenes were annotated using public databases, including NCBI non-redundant and gene ontology. De novo assembly of *R. ellipticus* unigenes found the highest similarity to apple than to other members of Rosaceae. This is the first attempt to use the Illumina platform of RNA sequencing and de novo assembly for *R. ellipticus* without a reference genome. In this study, unigenes were used for SSR marker development. ESTs containing SSR motifs were extracted using an online Microsatellite Identification Tool (MISA). SSR primers were designed from the SSR containing 704 EST sequences using the Websat software. Total 304 EST-SSRs primers were successfully designed, out of which 68 randomly selected primer pairs were custom synthesized and used for validation. Real-time PCR was also performed to analyze the relationship of transcriptional factors with fruit ripening. Out of 68 primer pairs, 61 were found to be informative in *R. ellipticus*, whereas 65 primer pairs were informative in the five tested genera of Rosaceae, i.e., pear, peach, apple, rose, and strawberry with 95.3% and 93.5% polymorphism, leading to the conclusion that these marker systems are very efficient to carryout diversity and cross transferability study in Rosaceae genera.

Keywords *Rubus ellipticus* · Library preparation · Molecular markers · EST-SSR · Polymorphism and transferability study

✉ Samriti Sharma
1992samritisharma@gmail.com

Rajinder Kaur
rkaur_uhf@rediffmail.com

Amol Kumar U. Solanke
amolsgene@gmail.com

Himanshu Dubey
hemu.bt@gmail.com

Siddharth Tiwari
siddarhdna20@yahoo.com

Krishan Kumar
drkrishankumar@gmail.com

¹ Department of Biotechnology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Nauni, Solan, India

² National Research Center on Plant Biotechnology, Pusa Campus, New Delhi 110012, India

³ Department of Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, India

⁴ Department of Fruit Science, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Nauni, Solan, India

Introduction

Rubus ellipticus (subgenus: *Idaeobatus*) is a raspberry with yellow fruits, native to tropical and subtropical India and Asia. *Rubus* is one of the most diverse genera of the plant kingdom with over 600–800 species (Thompson, 1995). The main members of genus *Rubus* are raspberry, blackberry, and dewberries which together commonly known as Brambles. Chromosome number of raspberry is seven; however, polyploidy also exists. Ploidy level in this genus varies from diploid ($2n=2x=14$) to tetradecaploid ($2n=14x=98$), with odd-ploids and aneuploids (Moore 1984; Thompson 1995; Meng and Finn 2002).

Interest in the improvement of this crop is increasing in light of its nutraceutical value. The fruits are routinely used for the treatment of fever, cough, and sore throat (Saklani et al. 2012) and have also been reported as an effective treatment of diabetes (Verma et al. 2014). The transcriptional factors such as ethylene responsive factor (ERF), MYB, and WRKY regulate the development and ripening of fruits in highly coordinated and irreversible manner that involves various organoleptic, biochemical, and physiological modifications responsible for soft and edible fruit formation. The fruit contains anthocyanins, catechins, flavonols, flavones, ascorbic acid, and tannins which are protective against a variety of diseases, such as cardiovascular disease and epithelial cancer (Jennings 1988).

Development of new cultivars can be benefited from reliable markers linked to important traits, including fruit quality characteristics, flowering traits, disease resistance, and plant architecture (Van et al. 2012, 2014). Recent advances in molecular biology have provided us with the novel, more efficient tools to establish evolutionary and genetic relationships among plants by observing variations in DNA sequence using highly accurate and throughput molecular markers linking this information with phenotype.

Molecular markers designed from simple sequence repeats (SSR), tandem repeats of 1–6 nucleotides that frequently show co-dominant inheritance, are known to be highly variable even within species, and are transferable across taxa to a varying extent (Powell et al. 1996; Kaur et al. 2015; Samriti et al. 2017). Molecular markers have numerous applications in plant breeding, including the analysis of genetic diversity, cultivar identification, gene tagging, and Quantitative Trait Loci analysis. Very few simple sequence repeat (SSR) markers exist for *Rubus* in general and fewer are transferable between species (Castillo et al. 2010; Debnath et al. 2008; Bushakra et al. 2012).

Genetic linkage maps composed of various types of molecular markers are available for raspberry (Ward et al. 2013; Woodhead et al. 2008), and one is available for blackberry (Castro et al. 2013), however, not all marker types

used to construct these maps are transferable among taxa. Limited number of expressed sequence tag (EST-SSR) i.e. seven are present for *R. ellipticus* at the National Center for Biotechnology Information (NCBI) GenBank (Samriti et al. 2017). So, there is a necessity of more *R. ellipticus* molecular markers for identification and development of cultivar, Mapping of important traits, and to study basic genetic and genomic mechanism.

The EST-SSR is a convenient source for gene-based SSR locus development which is more transferable across large taxonomic distances compared with genomic SSRs (Hendre and Aggarwal 2014; Sharma et al. 2016). EST-SSR markers are mainly developed from more conserved exonic regions, rather than from genomic sequence, and are, therefore, more likely to be transferable among taxa (Kaur et al. 2015). The goal of this research is to generate a useful set of EST-SSR markers to enable further genetic research into the raspberry genome and to increase the number of DNA sequences available for the Rosaceae research community and raspberry breeders who would like to use *R. ellipticus* as breeder's material.

Materials and methods

Plant material

Approximately 5 gm of young expanding leaves were harvested from a single healthy plant of *R. ellipticus* 'Kumarhatti'. This cultivar has small sized, yellow fruit, superior flavor, firmness, and freezing quality as compared to other maintained genotypes of *R. ellipticus* (Weber 2012). The collected leaves were frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

RNA isolation, cDNA library construction, and sequencing

Total RNA was extracted using Trizol method (Honaas and Kahn 2017). The poly-A mRNA was purified from total RNA using the oligo (dT) magnetic beads (Illumina, Inc), and library preparation was performed using TruSeq stranded mRNA library Prep kit. These short mRNA fragments were used as templates for first strand cDNA synthesis using random hexamer primers. The second strand cDNA was further synthesized using second strand mix and Act-D mix to facilitate RNA-dependent synthesis (Hyun et al. 2014a; Gutierrez et al. 2017). Then, these cDNA fragments were purified using Ampure XP beads followed by A-tailing and adapter ligation. The ligation reaction products were purified by agarose gel electrophoresis, and enriched by polymerase chain reaction (PCR) to create the final cDNA library (Hyun et al. 2014b). The PCR-enriched library was

analyzed in 4200 tape station system (Agilent Technologies) using high sensitivity D1000 screen tape for quantity and quality check. The library was sequenced from both ends using NextSeq 500 sequencing platform (Illumina, Inc).

The raw data were processed to obtain high-quality clean reads by removing adaptor sequences and the reads with more than 10% $Q < 20$ bases through trimmomatic v 0.35 (USADDELLAB.org). Transcriptome de-novo assembly was performed with DNASTAR with default parameters. Blast2GO was used to obtain gene ontology (GO) annotation of unigenes based on BLASTX hit against the non-redundant database with an E value cutoff of 10^{-5} (Ivamoto et al. 2017).

Unigene analysis and SSR development

SSR motifs were identified in assembled sequenced data using MISA (<http://pgrc.ipk-gatersleben.de/misa>) (Xu et al. 2013) and with FULL SSR (Metz et al. 2016). Only perfect SSRs of mono-, di-, tri-, tetra-, penta-, and hexa- nucleotide (nt) motifs with default parameters (number of uninterrupted repeat units) of more than 10, 7, 6, 5, 4, and 4, respectively, were targeted. Any SSR locus which was used to develop genetic markers should include a perfect repeat motif and two unique flanking sequences with 300 bp on each side of the repeats (Beier et al. 2017).

The forward and reverse primers were designed based on unique flanking sequences using websat (<http://www.wsmartsins.net/websat>) (You et al. 2010).

Experimental validation of polymorphic SSR markers

To assess the value of identified SSR markers, 68 primers were custom synthesized and chosen for polymorphism study. The samples used in this experiment included 21 *R. ellipticus* genotypes, seven strawberry genotypes, five pear genotypes, six peach genotypes, five rose samples, and eight apple genotypes. Genomic DNA was extracted from leaves using the CTAB methods (Kaur et al. 2015). Primers were synthesized from Eurofins genomics (GeNei™). The PCR was performed in a final reaction volume of 20 µl consisting of 1X Taq Buffer A (with 75 mM $MgCl_2$), 1 mM dNTP, 10 pico-moles of each forward and reverse primers, 1U Taq DNA polymerase, and 50 ng DNA in ProFlex™ Thermal Cycler (Applied Biosystem, Inc.). The PCR conditions were as follows: initial denaturation at 94 °C for 1 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing according to primer T_m for 1 min, extension at 72 °C for 2 min followed by final extension at 72 °C for 5 min (Samriti et al. 2017). The amplified DNA was mixed with 6X loading dye (Sisco Research Laboratories PVT. Ltd.) and electrophoresed in 2.0% agarose gel (Vaidya et al. 2015).

The gel was run for about two hours at a constant voltage of 5 V/cm under submerged conditions. Gel Documentation System (Syngene International Ltd.) was used to visualize PCR products and analyzed using the ABI Gene mapper software v4.0. The number of alleles was observed and the polymorphism information content (PIC) was calculated (Smith et al. 1997).

Quantitative real-time PCR

To validate sequencing results total RNA from six samples (Table 1) were isolated using Real genomics Hiyield™ total RNA mini kit. (Real Biotech Corporation Ltd.) Then, total RNA was reverse transcribed using Thermo Scientific Revert Aid first strand cDNA Synthesis Kit. Transcriptional factors gene-specific primers were constructed using websat (Schlautman et al. 2015) (Table 2). Real-time PCR reactions were carried out in triplicate using ABI PRISM 7700 Fast Real-time PCR system with ABI 7700 Sequence detector (Applied Biosystems, USA). The housekeeping gene 'Actin' was used to normalize the variant expression of selected genes. The conventional endpoint PCR was used to test primers for single band amplification. Melting curve study was carried out using qPCR consisted of 1X SYBR Green Master Mix (Applied Biosystems, USA), 10 pmol of each primer, 1 µl cDNA template and adjusted to 20 µl with nuclease-free water. Conditions followed during real-time PCR experiment were: initial denaturation at 95 °C for 30 s, followed by 32 cycles of denaturation at 95 °C for 30 s and extension at 55 °C for 30 s which is also by thermal dissociation curve. Primer details are mentioned in Table 2. The relative expression level was analyzed using $2^{-\Delta\Delta ct}$ method (Schmittgen et al. 2008).

Results and discussion

Rubus ellipticus cDNA library construction and functional characterization

The cDNA library of the yellow-fruited Himalayan raspberry 'Kumarhatti' was constructed from newly emerging leaves of a single, field-grown plant. Using Illumina sequencing

Table 1 List of collections/varieties used in the study

Sr. no.	Name of collection/variety	Name of Genus
1.	Kharkog-2	<i>Rubusellipticus</i>
2.	Kaithleeghat-3	<i>Rubusellipticus</i>
3.	Kumarhatti-1	<i>Rubusellipticus</i>
4.	Selva	Strawberry
5.	Sweet Charlie	Strawberry
6.	Torrey	Strawberry

Table 2 Details of primers of *Rubus ellipticus* used for the present study

Sr. no.	Primer name	Name of the gene	Primer sequence	Tm* (°C)	GC %*	Length in bp
1.	R1	AK317453.1(Actin gene)	F: CTACGAGGGTTTCTCTCTTCCT R: CTGCTCGTAGTCAACAGCAAC	54.8 54.4	50% 52%	22 21
2.	R2	Ethylene-responsive transcription factor 4	F: AACTTTACTTAGGTGCTCTCAACG R: TCATCAGCAGTACCATCAACAA	59.3 56.5	41.7% 40.9%	24 22
3.	R3	MYB-related 306-like	F: CTAGAACTCGAACCATTGTGGA R: GGGTTTTTCAGGTAATTCACAGC	58.4 58.4	45.5% 45.5%	22 22
4.	R4	WRKY transcription factor 35	F: TTGGATCAAGAAGGGACAAAAG R: CACATCCCCTGTGTACTCTCA	56.5 60.3	40.9% 50%	22 22
5.	R5	WRKY transcription factor 31	F: CCAAAACCCTGATTGACACTTT R: GCCATCTTCATTTGCTGATACA	56.5 56.5	40.9% 40.9%	22 22

technology, total 15,808,248 sequencing reads were generated. The high-quality raw reads were assembled into 7777 contigs (unigenes) (Sra accession: PRJNA510488) by DNASTAR (Eklblom and Wolf 2014). Contig lengths ranged from 122 bp to 3053 bp with N50 500 bp. The main purpose for creating a transcriptome of *R. ellipticus* was development of EST-based SSR markers; however, functional annotation

of the sequences is a useful supplement for selection of specific functional genes when testing with molecular markers.

Once the data were assembled, BLAST2GO was used to perform alignment and annotation. The Inter ProScan gave information regarding the functional analysis of proteins, classifying these proteins into families and predicting important sites and domains. BLAST was used for the comparison

Fig. 1 Basic alignment search tool data distribution pie chart

of the 7777 unigenes to the non-redundant protein database from the NCBI and yielded matches for 6814 (88%) unigenes. Annotation of gene function with BLAST (without hits) is 963, with BLAST hits was 1565, and 241 with gene ontology annotation (Fig. 1). The majority of best matches were found with fruit crops (more than 9000 BLAST hits similarity with *Malus Domestica* followed by *Prunus persica*). Inter ProScan analysis indicated that cytochrome P450 (IPR001128) (used heme to oxidize their substrate) and UDP-glucuronosyl/UDP-glucosyltransferase (IPR002213) (catalyzes addition of the glycosyl group from a UTP sugar to a small hydrophobic molecule) are very abundant in leaves of *R. ellipticus* followed by cytochrome P450, E-class group I (IPR002401) (Fig. 2).

The unigenes with significant BLAST hits were assigned GO terms and categorized as: biological processes (BP), molecular function (MF), and cellular components (CC) (Ashburner et al. 2000) (Fig. 3). The most abundant sub-level first gene ontology category was biological processes

which contained 2730 sequences, with most of the sequences involved in metabolic and cellular processes, followed by single organism process sequences. GO assignments for the category molecular function contained 2305 sequences, in which most of the sequences were responsible for the catalytic and binding activity, followed by transporter activity and structural molecule activity sequences. GO assignments for the category cellular component contained 2370 sequences assigned to cell and cell part, followed by organelles. A more detailed view of the BP, MF, and CC GO sub-levels revealed a significant fraction of genes related to biological processes such as macromolecules metabolism, organic substance metabolism, biosynthetic processes, and proteolysis. Within the category molecular function, binding-related subcategories such as ATP binding, ion binding, cation binding, nucleoside binding, and the molecules that contribute to the structural integrity of the ribosome were abundantly present. Finally, within the cellular component category, membrane and macromolecule complexes were

Fig. 2 Functional analysis of proteins using InterProScan. Results indicate that cytochrome P450 family proteins are abundantly present in *Rubus ellipticus*

Fig. 3 Unigene set was aligned to the gene ontology (GO) database and classified according to the three basic categories: biology process, molecular function, and cellular component. The most abun-

dant level was biological process followed by cellular component and molecular function

Table 3 Details of read mapping statistics with assembled unigenes

Global reference size	6,242,699
Number of reads	28,137,791
Mapped reads	28,137,791/100%
Unmapped reads	0/0%
Mapped paired reads	28,137,791/100%
GC percentage	44.67%
Mean coverage depth (standard deviation)	635.9831 (3186.32)
Mean mapping quality	238.55
Mean insert size (standard deviation)	208.6 (60.29)

enriched. To assess the transcriptome assembly, sequenced reads were aligned to the unigenes using bowtie2 program (Langmead et al. 2013). All the sequenced reads were aligned to the assembled unigene sets of *Rubus ellipticus* with mean coverage depth and mapping quality of 635.98 and 238.55, respectively (Table 3). More than 97% of the assembled unigenes showed at least 20× coverage with the sequenced reads (Fig. 4), this shows that the assembly produced by DNASTAR is of very good quality.

Reference genomes have been published for members of Rosaceae: apple (*Malus × Domestica* Borkh.) (Velasco et al. 2010), double haploid peach (*Prunus persica* L.)

Fig. 4 Fraction of unigenes covered (in Y-axis) with the X coverage in the X-axis

(Verde et al. 2013), European pear (*Pyrus communis* L.) (Chagne et al. 2014), Asian pear (*Pyrus bretschneideri* Rehd.) (Wu et al. 2013), and diploid strawberry (*Fragaria vesca* L.) (Shulaev et al. 2011), which is in the same subfamily (Rosoideae) as raspberry. Enough sequence conservation exists between raspberry and these genomes, so that *R. ellipticus* markers and primers designed from polymorphic regions can be transferable to these genera. The gene space, in particular, was conserved; therefore, the raspberry unigenes were aligned to the gene sets from apple, peach, and strawberry to evaluate the actual sequence conservation.

Each unigene was aligned to obtain the best possible alignment and found that *R. ellipticus* unigenes have a higher overall percent identity to apple than to the other gene sets (Fig. 5).

Data availability

The data are available under bioproject ID: SUB4927858 and SRA ID: SUB4927872.

Fig. 5 Basic local alignment search tool (BLAST) comparison of the 7777 *Rubus ellipticus* unigene set to the non-redundant (nr) protein database from the National Center for Biotechnology Information

(NCBI). Results indicate that the majority of the unigenes aligned to *Malus domestica*

The abundance of SSRs in *Rubus ellipticus* sequences

A large number of perfect SSR with mono-, di-, tri-, tetra-, penta-, and hexa- motifs were found, but their numbers varied with MISA and FULL SSR (Metz et al. 2016). The average numbers of SSRs were 2329 with MISA and 7870 with FULL SSR. Some of the reads could not be mapped on the non-redundant protein database from the NCBI resulting in lower genome coverage, and thus, fewer SSRs were identified as compared with other crops. A total of 2329 SSR loci were detected with MISA, including mono (1556)-, di (290)-, tri (336)-, tetra (10)-, penta (1)-, hexa (9)-, and complex nucleotide SSR motifs (117), respectively. The FULL SSR software detected 7870 SSR loci of which mono (4368)-, di (2322)-, tri (1122)-, tetra (47)-, penta (11)-, and hexa (0). The number of mononucleotide motifs identified is similar to MISA and FULL SSR software with 66.81% and 55.50%, respectively. The ratio of di-, tri-, tetra-, penta-,

and hexanucleotide motifs with MISA and FULL SSR was 12.45% and 29.51%; 14.43% and 14.26%; 0.43% and 0.60%; 0.043% and 0.14%; and 0.39% and 0, respectively. These results showed that there are many types of penta- and hexanucleotide SSRs but with low frequencies. We considered results of MISA for further analysis, because it produces a lesser number of mononucleotide SSR motifs. Perl script (Archive FAS System Biology) (Beier et al. 2017) was used to remove redundancy in the data, leaving 704 sequences out of 763 in FASTA format. A websat software (<http://www.wsmartins.net/websat>) has been chosen for designing of primers of di-, tri-, tetra-, and penta nucleotide motifs. A total of 304 EST-SSR primers were found and 68 were used in this polymorphism study.

Table 4 List of genotypes used in the study

Sr. no.	Accessions	Collection site
List of <i>Rubus ellipticus</i> collections		
1.	Badhu-2	Badhu (HP)
2.	Kumarhatti-1	Kumarhatti (HP)
3.	Kaithlighat-3	Kaithlighat (HP)
4.	Khadiyana	Khadiyana (HP)
5.	Kharkog-2	Kharkog (HP)
6.	Kukarigalu-2	Kukarigalu (HP)
7.	Kandaghat-8	Kandaghat (HP)
8.	Hiranagar-2	Hiranagar (HP)
9.	Kharkog-1	Kharkog (HP)
10.	Shoghi-5	Shoghi (HP)
11.	Kandaghat-1	Kandaghat (HP)
12.	Baurgaon-1	Baurgaon (UK)
13.	Hiranagar-1	Hiranagar (HP)
14.	Bhim-boot-1	Bhim-boot (HP)
15.	Nagangi	Nagangi (HP)
16.	Deothi-1	Deothi (HP)
17.	Badashahithul-1	Badashahithul (UK)
18.	Sanarali-2	Sanarali (HP)
19.	Guldi-1	Guldi (UK)
20.	Sarali-3	Sarali (HP)
21.	Majhgaon	Majhgaon (HP)
List of apple genotypes		
1.	Pinkplesant	Nauni, Solan H.P
2.	Tompkinstony	Nauni, Solan H.P
3.	Chaubattia ambrase	Nauni, Solan H.P
4.	Red June	Nauni, Solan H.P
5.	Royal Empire	Nauni, Solan H.P
6.	GingerGold	Nauni, Solan H.P
7.	Semi sweetRed	Nauni, Solan H.P
8.	Fanny	Nauni, Solan H.P
List of strawberry genotypes		
1.	Selva	Nauni, Solan H.P
2.	Douglas	Nauni, Solan H.P
3.	Torrey	Nauni, Solan H.P
4.	Confictura	Nauni, Solan H.P
5.	Shasta	Nauni, Solan H.P
6.	Sweet Charlie	Nauni, Solan H.P
7.	Brighton	Nauni, Solan H.P
List of pear genotypes		
1.	Jarainth	Nauni, Solan H.P
2.	Flemish Beauty	Nauni, Solan H.P
3.	Avast Sugar	Nauni, Solan H.P
4.	Baldwin	Nauni, Solan H.P
5.	Ayers Sugar	Nauni, Solan H.P
List of rose genotypes		
1.	High and magic (bicolor)	Nauni, Solan H.P
2.	High yellow magic (whitish pink)	Nauni, Solan H.P
3.	High and sparkling (Pink)	Nauni, Solan H.P
4.	High and peace (Pink)	Nauni, Solan H.P

Table 4 (continued)

Sr. no.	Accessions	Collection site
5.	Upper class Red	Nauni, Solan H.P
List of peach genotypes		
1.	Tropic Snow	Nauni, Solan H.P
2.	Prabhat	Nauni, Solan H.P
3.	EarliGrande	Nauni, Solan H.P
4.	FlordaPrince	Nauni, Solan H.P
5.	Tropic Sweet	Nauni, Solan H.P
6.	July Elberta	Nauni, Solan H.P

SSR marker validation for quality and polymorphism

A total of 68 primers of di- and tri nucleotide repeat motifs were randomly chosen for experimental validation in 52 genotypes (21 *R. ellipticus* samples, five pear genotypes, six peach genotypes, five rose genotypes, seven strawberry genotypes, and eight apple genotypes) (Table 4). Out of 68 primer pairs, 61 were found to be informative in *R. ellipticus*, whereas 65 primer pairs were informative in the five tested genera of Rosaceae (five pear genotypes, six peach genotypes, five rose genotypes, seven strawberry genotypes, and eight apple genotypes) with 95.3% and 93.5% polymorphism. Total amplified fragments were 647 in *R. ellipticus* and 1429 Rosaceae members. The average number of amplified fragments per species was 30.81 in *R. ellipticus* and 46.1 in Rosaceae members (Tables 5 and 6). In some earlier studies, a high level of polymorphism using SSRs has also been found, i.e., 100% in *R. ellipticus* (Samriti et al. 2017), 100% in blackberry (Dossett et al. 2012), 100% in peach (Li et al. 2013), and 98.36% in *Rubus* species (Thakur, 2013). The high level of polymorphism amplified primers designed from EST-SSRs demonstrated that this marker type was suitable for genetic diversity and cross transferability analysis.

NTSYSpc ver.2.02.h (Rohlf 2000) was used to obtain Jaccard's similarity coefficient matrix (Jaccard 1908) of 52 genotypes of Rosaceae family based on DNA amplification using EST-SSRs. Good range of genetic variability was observed in the species, because the value of coefficient ranged from 0.14 to 0.87. The highest value of similarity was 0.87, between 'High and magic' (bicolor) of Rose and 'Chaubattia ambrase' of Apple. The minimum similarity of 0.14 was obtained between 'Nagangi' of *Rubus ellipticus* and 'Red June' of Apple. Similarity coefficients obtained by Jaccard's coefficient were further used for UPGMA analysis and dendrogram using NTSYS.

52 genotypes of different genera of Rosaceae family were divided into two main clusters, 'E' and 'F' at 41% similarity. Cluster 'E' contained 50 genotypes viz 'Badhu-2', 'Khadiana', 'Heera nagar-2', 'Kumarhatti-1', 'Baurgaon-1',

Table 5 Summary of EST-SSR amplified products obtained in five members (31 genotypes) of Rosaceae family examined in study

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	G25	G26	G27	G28	G29	G30	G31
Total number of primers examined	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68	68
Number of informative primers	33	44	44	45	44	53	57	46	40	44	42	56	56	56	28	50	25	29	39	28	28	39	29	55	53	57	55	55	48	45	51
Number of polymorphic primers	33	44	44	45	44	53	57	46	40	44	42	56	56	56	28	50	25	29	39	28	28	39	29	55	53	57	55	55	48	45	51
Total number of bands	34	45	47	46	45	55	57	45	41	46	43	58	61	58	30	54	26	30	40	29	29	40	31	56	55	58	56	57	51	50	56
Average number of polymorphic bands per primer	0.50	0.66	0.69	0.68	0.66	0.81	0.84	0.66	0.60	0.68	0.63	0.85	0.90	0.85	0.44	0.79	0.38	0.44	0.59	0.43	0.43	0.59	0.46	0.82	0.81	0.85	0.82	0.84	0.75	0.74	0.82
Average number of amplified fragments per primer	1.03	1.02	1.07	1.02	1.02	1.03	1	0.98	1.03	1.05	1.02	1.04	1.09	1.04	1.07	1.08	1.04	1.04	1.03	1.04	1.04	1.03	1.07	1.02	1.04	1.02	1.02	1.04	1.06	1.11	1.10

Table 5 (continued)

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24	G25	G26	G27	G28	G29	G30	G31
Per cent transferability of <i>Rubus ellipticus</i> primers to other genotypes (%)	0.49	0.65	0.65	0.66	0.65	0.78	0.84	0.68	0.59	0.65	0.62	0.82	0.82	0.82	0.41	0.74	0.37	0.43	0.57	0.41	0.41	0.57	0.43	0.81	0.80	0.84	0.81	0.81	0.71	0.66	0.75

G₁ Jarenth pear, *G₂* Flemish beauty, *G₃* avast sugar, *G₄* baldwin pear, *G₅* ayers sugar, *G₆* tropic snow, *G₇* prabhat, *G₈* early grande, *G₉* florada prince, *G₁₀* tropic sweet, *G₁₁* july alberta, *G₁₂* high and magic (bicolor), *G₁₃* high yellow magic (whitish pink), *G₁₄* high and sparkling (pink), *G₁₅* high and peace (pink), *G₁₆* upper class red, *G₁₇* selva, *G₁₈* Douglas, *G₁₉* torrey, *G₂₀* confectura, *G₂₁* sasta, *G₂₂* sweet charlie, *G₂₃* brighten, *G₂₄* pinkplesant, *G₂₅* tompkinstony, *G₂₆* Chaubattia ambrase, *G₂₇* red june, *G₂₈* royal empire, *G₂₉* zinger gold, *G₃₀* semi sweet red, *G₃₁* fanny

Table 6 Summary of EST-SSR amplified products obtained from five genotypes of pear, six genotypes of peach, five genotypes of rose, seven genotypes strawberry, and eight genotypes of apple are examined in study

	<i>Rubus ellipticus</i>					Pear	Peach	Rose	Strawberry	Apple
Total number of primers examined	68	68	68	68	68	68	68	68	68	68
Number of informative primers	61	58	40	58	61	62	58	65	65	65
Number of polymorphic primers	59	40	58	67	39	45	51	36	36	36
Total number of scorable bands	64	58	40	67	67	71	59	71	71	71
Number of polymorphic bands	61	40	40	42	42	50	49	41	41	41
Number of monomorphic bands	3	18	18	25	25	21	10	30	30	30
Average number of polymorphic bands per primer	0.90	0.59	0.59	0.62	0.62	0.74	0.72	0.60	0.60	0.60
Total number of amplified fragments	647	217	43.4	47.8	287	261	225	436	436	436
Average number of amplified fragments per accession	30.81	43.4	3.74	4.71	4.71	52.2	32.1	54.5	54.5	54.5
Average number of amplified fragments per informative primer	10.61	3.74	68.9%	62.7%	62.7%	4.20	3.9	6.7	6.7	6.7
Per cent of total polymorphic bands	95.3%	68.9%	68.9%	62.7%	62.7%	70.4%	83.05%	57.8%	57.8%	57.8%
Per cent transferability of <i>Rubus ellipticus</i> primers to other genotypes	–	67.8%	67.8%	66.1%	66.1%	76.3%	86.4%	61.0%	61.0%	61.0%

P₁: Badhu-2 P₂: Kumarhatti-1, P₃: Kaithlighat-3; P₄: Khadiyana; P₅: Kharkog-2; P₆: Kukarigalu-2; P₇: Kandaghat-8; P₈: Hiranagar-2; P₉: Kharkog-1; P₁₀: Shoghi - 5; P₁₁: Kandaghat-1; P₁₂: Baurgaon-1; P₁₃: Hiranagar-1; P₁₄: Bhim-boot-1; P₁₅: Nagangi; P₁₆: Deothi -1; P₁₇: Badashahithul-1; P₁₈: Sanarali-2 P₁₉: Guldi-1; P₂₀: Sarali-3; P₂₁: Majhgaon, P₂₂: Jarenth pear; P₂₃: Flemish beauty, P₂₄: Avast sugar; P₂₅: Baldwin pear; P₂₆: Ayers Sugar; P₂₇: Tropic snow; P₂₈: Prabhat; P₂₉: Early grande; P₃₀: Florada prince; P₃₁: Tropic sweet; P₃₂: July alberta; P₃₃: High and magic (bicolor); P₃₄: High yellow magic (whitish pink); P₃₅: High and sparkling (Pink); P₃₆: High and peace (Pink); P₃₇: Upper class Red; P₃₈: Selva; P₃₉: Douglas; P₄₀: Torrey; P₄₁: Confectura; P₄₂: Sasta; P₄₃: Sweet Charlie; P₄₄: Brighten; P₄₅: Pinkpleasant; P₄₆: Tompkinstony; P₄₇: Chaubattia ambrose; P₄₈: Red june; P₄₉: Royal empire; P₅₀: Zinger gold; P₅₁: Semi sweet red; P₅₂: Fanny.

Fig. 6 Dendrogram of 52 genotypes of different genera of Rosaceae family based on EST-SSR analysis

‘Deothi -1’, ‘Kharkog-2’, ‘HeeraNagar-1’, ‘Kandaghat-8’, ‘Kharkog-1’, ‘Saroli-3’, ‘Badashahithul-1’, ‘Majhgaon’, ‘Shoghi-5’, ‘Guldi-1’, ‘Sanarali-2’, ‘Bhim-boot-1’ ‘Kaithlee ghat-3’, ‘Kukorigolu-2’ ‘Jarenth Pear’, ‘Flemish beauty’, ‘Avast sugar’, ‘Baldwin’, ‘Ayers sugar’, ‘tropic snow’, ‘Prabhat’, ‘Early grande’, ‘Florida PRINCE’, ‘Tropic SWEET’, ‘July Alberta’, ‘High and magic’, ‘High yellow magic’, ‘High and sparkling’, ‘Upper class red’, ‘Douglas’, ‘Torrey’, ‘Confectura’, ‘Sasta’, ‘Sweet Charlie’, ‘Brighten’, ‘Pinkpleasant’,

‘Tompkinstony’, ‘Chaubattia ambrose’, ‘Red june’, ‘Royal empire’, ‘Zinger gold’, ‘Semi-sweet red’, ‘Fanny’ ‘High and peace’, and ‘Selva’, while cluster ‘F’ contained only two genotypes viz ‘Kandaghat-1’, ‘Nagangi’. Maximum similarity, i.e., 87%, was found between ‘High and magic’ of Rose and ‘Chaubattia ambrose’ of Apple showing that these two are relatively closely related (Fig. 6). This indicates that the degree of transferability is inversely proportional to the genetic distance between the species for generating the

Fig. 7 Spatiotemporal real-time PCR expression analysis of role of transcriptional factors in fruit ripening. The gene expression was normalized with actin gene of *Arabidopsis thaliana*. Results indicate maximum expression of ERF in ‘Kumarhatti-1’ collection of *Rubus*

ellipticus and minimum in Selva cultivar of strawberry. The maximum transcript expression was observed in ‘sweet charlie’ cultivar of strawberry with WRKY and MYB

EST-SSRs and the species to which these are to be transferred (Cipriani et al. 1999; Decrooq et al. 2003 and Mnejja et al. 2010).

The present study reports the development of *Rubus ellipticus* EST-SSR markers, which proved polymorphic and transferable in Rosaceae genera, i.e., pear, peach, apple, rose, and strawberry. Transferability from a species will help in comparative genetic analysis such as comparative mapping and evolutionary studies. In our study, EST extraction from *Rubus ellipticus* cDNA library has been found to give high polymorphism in different rosaceous species and can be used without wasting resources and time for developing the same from these species.

Expression pattern of transcript related to the fruit ripening

Transcriptional factors play a substantial role in controlling cellular processes and ethylene responsive factor (ERF). MYB and WRKY transcription factors are involved in controlling various processes such as response to abiotic and biotic stresses, development, differentiation, metabolism, and defense (Ambawat et al. 2013). Expression analysis of three transcriptional factors (ERF, MYB, and WRKY) in ripened fruits of *R. ellipticus* genotypes/cultivars and strawberry was performed (Tables 1 and 2). The transcript normalization was carried out against the housekeeping actin

gene. Higher expression of ERF genes (contig 5053) was observed in 'Kumarhatti' (Fig. 7). The lower expression of ERF was observed in strawberry 'Selva'. A high transcript level of WRKY (contig 5509 and 6191) and MYB (contig 5166) was noted in strawberry 'Sweet Charlie', while lower expression was observed in 'Kaithleeghat-3' (Light yellow colored fruit) of *R. ellipticus* collection. These results showed differential expression of transcription factor genes in fruit ripening among different genotypes/cultivars of *R. ellipticus* and strawberry.

Conclusion

The development of *R. ellipticus* specific EST-SSR markers is the first step towards gene-derived markers which will be used in linkage map construction. In the present study, 68 primer pairs were randomly selected from 304 unigene-SSRs, derived from leaf tissue of 'Kumarhatti'. These EST-SSRs gave high polymorphism in *R. ellipticus* thus can be used to other members of Rosaceae family for comparative mapping and evolutionary studies. Considering the various advantages of SSRs, we selected SSRs as a best and effective genetic marker for the study of *Rubus* species.

Author contribution statement Dr. SS designed, performed the experiments, and wrote the manuscript. Dr. RK was involved in planning and supervised the work. Dr. AKUS and Mr. HD contributed to the interpretation of the results, Dr. ST helped to carry out the RT-PCR analysis and interpretation of the RT-PCR results, and Dr. KK provided plant material for research

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants* 19(3):307–321
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* 25(1):25–29
- Beier S, Thiel T, Munch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33(16):2583–2585
- Bushakra JM, Stephens MJ, Atmadajaja AN, Lewers KS, Symonds VV, Udall JA (2012) Construction of black (*Rubus occidentalis*) and red (*Rubus idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theor Appl Genet* 125:311–327
- Castillo NRF, Reed BM, Graham J, Fernandez F, Bassil NV (2010) Microsatellite markers for raspberry and blackberry. *J Am Soc Hortic Sci* 13:271–282
- Castro P, Stafne ET, Clark JR, Lewers KS (2013) Genetic map of the primocane-fruiting and thornless traits of tetraploid blackberry. *Theor Appl Genet* 126:2521–2532
- Chagne D, Crowhurst RN, Pindo M, Thrimawithana A, Deng C, Ireland H (2014) The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS One* 9(4):e92644
- Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlanzer E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach (*Prunus persica* (L.) Batsch): isolation, characterization and cross-species amplification in *Prunus*. *Theor Appl Genet* 99:65–72
- Dale A, Moore PP, McNicol RJ, Sjulín TM, Burmistrov LA (1993) Genetic diversity of red raspberry varieties throughout the world. *J Am Soc Hortic Sci* 18:119–129
- Debnath SC (2008) Inter simple sequence repeat (ISSR) markers and pedigree information to assess genetic diversity and relatedness within raspberry genotypes. *J Fruit Ornament Plant Res* 7:1–17
- Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor Appl Genet* 106:912–922
- Dossett M, Bassil NV, Lewer KS, Finn CE (2012) Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. *Genet Resour Crop Evol* 10:40–57
- Eklom R, Wolf JBW (2014) A field guide to whole genome sequencing, assembly and annotation. *Evol Appl* 7:1026–1042
- Gutierrez E, Villaraco AG, Lucas JA, Gradillas A, Manero FJG, Solano BR (2017) Transcriptomics, targeted metabolomics and gene expression of blackberry leaves and fruits indicate flavonoid metabolic flux from leaf to red fruit. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2017.00472>
- Hendri PS, Aggarwal RK (2014) Development of genic and genomic SSR markers of robusta coffee (*Coffea canephora* Pierre Ex A. Froehner). *PLoS ONE* 9(12):e113661. <https://doi.org/10.1371/journal.pone.0113661>
- Honaas L, Kahn E (2017) A practical examination of RNA isolation methods for European pear (*Pyrus communis*). *BMC Res Notes* 10:1–8
- Hyun TK, Lee S, Kumar D, Rim Y, Kumar R, Lee SY, Lee CH, Kim JY (2014a) RNA-seq analysis of *Rubus idaeus* cv. Nova: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. *Plant Cell Rep* 33:1617–1628
- Hyun TK, Lee S, Rim Y, Kumar R, Han X, Lee SY, Lee CH, Kim JY (2014b) De novo RNA sequencing and metabolite profiling to identify genes involved in anthocyanin biosynthesis in Korean black raspberry (*Rubus coreanus* Miquel). *Plos One* 9:e88292
- Ivamoto ST, Reis O, Domingues DS, Santos TBD, Oliveira FFD, Pot D, Leroy T, Vieira LGE, Carazzolle MF, Pereira GAG, Pereira LFP (2017) Transcriptome analysis of leaves, flowers and fruits perisperm of *Coffea arabica* L. Reveals the differential expression of genes involved in raffinose biosynthesis. *PLoS One* 12:e0169595. <https://doi.org/10.1371/journal.pone.0169595>
- Jaccard P (1908) Nouvelles Recherches sur la distribution florale. *Bulletin de la Société Vaudoise des Sciences Naturelles* 44:223–270
- Jenning DL (1988) Raspberry and blackberries: their breeding, disease and growth. Academic press, London, p 230
- Kaur R, Shilpa Vaidya E, Kumar K (2015) Development, characterization and transferability of peach genic SSRs to some Rosaceae species. *Adv Res* 3:165–180
- Langmead B, Salzberg SL (2013) Fast gapped-read alignment with Bowtie 2. *Nat Method* 9(4):357–359
- Li XW, Meng XQ, Jia HJ, Yu ML, Ma RJ, Wang LR, Cao K, Shen ZJ, Niu L, Tian JB, Chen MJ, Xie M, Arus P, Gao ZS, Aranzana MJ (2013) Peach genetic resources: diversity, population structure and linkage disequilibrium. *BMC Genet* 14:1–16

- Meng R, Finn C (2002) Determining ploidy level and nuclear DNA content in *Rubus* by flow cytometry. *J Am Soc Hortic Sci* 127:767–775
- Metz S, Cabrera JM, Rueda E, Giri F, Amavet P (2016) FullSSR: Microsatellite finder and primer designer. *Adv Bioinform*. <https://doi.org/10.1155/2016/6040124>
- Mnejja M, Garcia MM, Audergon MJ, Arus P (2010) *Prunus* microsatellite marker transferability across Rosaceous crops. *Nat Genet* 30:194–200
- Moore JN (1984) Blackberry breeding. *Hortic Sci* 19:183–185
- Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- Rohlf FJ (2000) NTSYSpc numerical taxonomy and multivariate analysis version 2.0.h. Applied Biostatistics Inc., New York, p 25
- Saklani S, Chandra S, Badoni PP, Dogra S (2012) Antimicrobial activity, nutritional profile and Phytochemical screening of wild edible fruit of *Rubus ellipticus*. *Int J Med Aromat Plants* 2(2):269–274
- Samriti Kaur R, Shilpa Malhotra EV, Poonam Thakur D, Kumar K (2017) Assessment of genetic diversity in *Rubus ellipticus* (Smith) using molecular markers. *Proc Indian Nat Sci Acad* 83(3):669–679
- Schlautman B, Fajardo D, Bougie T, Wiesman E, Polashock J, Vorsa N, Steffan S, Zalapa J (2015) Development and validation of 697 novel polymorphic genomic and EST-SSR markers in the American cranberry (*Vaccinium macrocarpon* Ait.). *Molecules* 20:2001–2013
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C_T method. *Nat Protoc* 3:1101–1108
- Sharma N, Kaur R, Vaidya E (2016) Potential of RAPD and ISSR markers for assessing genetic diversity among *Stevia rebaudiana* Bertoni accessions. *Indian J Biotechnol* 15:95–100
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 43(2):109–116
- Smith JS, Chin EC, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Zeigler J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theor Appl Genet* 95:163–173
- Thakur R (2013) Studies on development of genic-SSRs in raspberry (*Rubus ellipticus* Smith.) and their transferability across related species. MSc. Thesis. Dr Y S Parmar University of Horticulture and Forestry, Nauni, Solan, India, p 58
- Thompson MM (1995) Chromosome numbers of *Rubus* species at the National clonal germplasm repository. *HortScience* 30:1447–1452
- Vaidya E, Kaur R, Kumar K, Sharma N (2015) Exploitation of *Malus* ESTs for development of SSR markers after insilico analysis. *J Appl Bot Food Qual* 88:164–169
- Van DT, Noordijk Y, Dubos T, Bink M, Meulenbroek B, Visser R (2012) Microsatellite allele dose and configuration establishment (MADCE): an integrated approach for genetic studies in allopolyploids. *BMC Plant Biol* 12:25–34
- Van DT, Pagliarini G, Pikunova A, Noordijk Y, Yilmaz-Temel H, Meulenbroek B (2014) Genomic rearrangements and signatures of breeding in the allo-octoploid strawberry as revealed through an allele dose based SSR linkage map. *BMC Plant Biol* 14:55–64
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet* 42(10):833–839
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F, Zuccolo A, Rossini L, Jenkins J, Vendramin E, Meisel LA, Decroocq V, Sosinski B, Prochnik S, Mitros D, Policriti A, Cipriani G, Donadini L, Ficklin S, Goodstein DM, Xuan P, DelFabbro C, Aramini V, Copetti D, Gonzalez S, Horner DS, Falchi R, Lucas S, Mica E, Maldonado J, Lazzari B, Bielenberg D, Pirone R, Miculan M, Barakat A, Testolin R, Stella A, Tartarini S, Tonutti P, Arus P, Orellana A, Wells C, Main D, Vizzotto G, Silva H, Salamini F, Schmutz J, Morgante M, Rokhsar DS (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45(5):487–494
- Verma R, Gangarde T, Punasiya R, Ghulaxe C (2014) *Rubus fruticosus* (blackberry) use as an herbal medicine. *Pharmacogn Rev* 8:101–104
- Ward J, Bhargoo J, Fernandez F, Moore P, Swanson J, Viola R (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genom* 14:22–35
- Weber CA (2012) Raspberry variety review. Cornell Cooperative Extension, Cornell University, Ithaca
- Woodhead M, McCallum S, Smith K, Cardle L, Mazzitelli L, Graham J (2008) Identification, characterisation and mapping of simple sequence repeat (SSR) markers from raspberry root and bud ESTs. *Mol Breed* 22:555–563
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 23(2):396–408
- Xu J, Liu L, Xu Y, Chen C, Rong T, Ali F, Zhou S, Wu F, Liu Y, Wang J, Cao M, Lu Y (2013) Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Res* 20:497–509
- You FM, Wanjugi H, Huo N, Lazo GR, Luo MC, Anderson OD, Dvorak J, Gu YQ (2010) RJ Primers: unique transposable element insertion junction discovery and PCR primer design for marker development. *Nucleic Acids Res* 38:313–320