Mol Breeding (2016) 36:68
DOI 10.1007/s11032-016-0491-x

CrossMark

Genetic structure based on EST–SSR: a putative tool for fruit color selection in Japanese plum (*Prunus salicina* L.) breeding programs

M. González · E. Salazar · J. Castillo · P. Morales · I. Mura-Jornet · J. Maldonado · H. Silva · B. Carrasco

Received: 4 January 2016 / Accepted: 13 May 2016 / Published online: 23 May 2016
© Springer Science+Business Media Dordrecht 2016

Abstract *Prunus salicina* is one of the most economically important stone fruits. However, there is scarce genetic information available, which makes it difficult to implement marker-assisted selection (MAS) in genetic improvement programs. Recently, next-generation sequencing has greatly enhanced breeding program strategies, generating information associated with the identification of expressed sequence tag–simple sequence repeats (EST–SSRs) and single-nucleotide polymorphisms (SNPs), two of the most used molecular markers in MAS. Few studies have focused on developing EST–SSR markers considering both gene expression levels of contrasting phenotypes and specific transcription factors of metabolic pathways. This study investigated the transcriptome profile of *P. salicina* in fruits with contrasting skin colors, obtaining 54,224 unique contigs. From this dataset, 44 EST–SSRs have been generated, considering gene expression levels of contrasting phenotypes and specific transcription factor from three metabolic pathways: citric acid, carbohydrate metabolism and flavonoid pathways. Three EST–SSR markers developed from the putative flavonoid pathway transcription factors *PsMYB10*, *PsMYB1* and *PsbHLH35* were selected to determine genetic structure in 29 cultivars. This structure was contrasted with the genetic structure generated using genomic SNPs obtained by genotyping-by-sequencing (GBS). The analysis using SNPs identified two groups, while the use of selected EST–SSRs identified three. In contrast to the structure given by the SNPs, the EST–SSRs grouped all the yellow cultivars in one cluster, which was composed mainly of cultivars of this color. The EST–SSRs developed in this study may

Electronic supplementary material The online version of this article (doi:10.1007/s11032-016-0491-x) contains supplementary material, which is available to authorized users.

M. González (✉) · I. Mura-Jornet · B. Carrasco (✉)
Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Casilla 306, Macul, Santiago, Chile
e-mail: mpgonza2@uc.cl

B. Carrasco
e-mail: bcarrasco@uc.cl

E. Salazar · J. Castillo
Centro Regional de Investigación La Platina, Instituto de Investigaciones Agropecuarias, Santa Rosa 11610, La Pintana, Santiago, Chile

P. Morales
Institut of Plant Genetics, Gottfried Wilhelm Leibniz Universität, Herrenhäuser Straße 2, 30419 Hannover, Germany

J. Maldonado · H. Silva
Laboratorio de Genómica Funcional y Bioinformática, Departamento de Producción Agrícola, Facultad de Ciencias Agronómicas, Universidad de Chile, Santa Rosa 11315, La Pintana, Santiago, Chile

be considered as candidate markers to be evaluated in MAS strategies in genetic improvement programs.

Keywords *R2R3MYB* · *bHLH* · Genotyping-by-sequencing (GBS) · Marker-assisted selection (MAS)

Introduction

The Japanese plum (*Prunus salicina* L.) belongs to the Rosaceae and it is a diploid species ($2\times = 2n = 16$), with its center of origin in China, where it has been cultivated for several 1000 years and from there distributed around the world. During the last decades, the production of plums (both European and Japanese) increased ranking on the tenth place of the world fruit production (Hartmann and Neumüller 2009).

Molecular markers are extremely useful tools in plant genetics and breeding. The implementation of marker-assisted selection (MAS) in plant breeding programs helps to reduce costs and increase the efficiency of the use of space and time in the generation of new cultivars (He et al. 2014). Random amplified polymorphic DNA markers (RAPDs) (Ortiz et al. 1997; Shimada et al. 1999) and amplified fragment length polymorphism (AFLP) (Goulão et al. 2001) have been used for the identification of genotypes and the genetic diversity in Japanese plum. Mnejja et al. (2004) isolated 27 microsatellites or simple sequence repeats (SSRs) in *P. salicina,* and recently Carrasco et al. (2012) analyzed genetic diversity and correlation among 29 Japanese plum cultivars using a combination of inter simple sequence repeat (ISSR) and SSR markers, while Klabunde et al. (2014) genotyped forty-seven cultivars with eight microsatellite markers. A unique genetic map based on AFLP has been generated for this species (Vieira et al. 2005).

The development and implementation of high-throughput sequencing platforms (or next-generation sequencing platform, NGS) have facilitated the study of expression profiles of genes associated with traits with agronomic interest, also the use of strategies based on large numbers of SNPs for genetic analysis and genotyping. Genotyping-by-sequencing (GBS) has become an important cost-effective tool for genomics-assisted breeding in peach and other species of economic value (Bielenberg et al. 2015; Kim et al. 2016). Also RNA sequencing (RNAseq) is an accurate and sensitive tool for quantification, detecting both low and high levels of gene expression. Transcriptome studies have facilitated the understanding of metabolic pathways, as well as suggesting possible transcription factors associated with characters of interest such as fruit development, pigmentation and response to viral infections (Rubio et al. 2015; Wei et al. 2015).

NGS has strongly enhanced genetic improvement strategies based on quantitative trait loci (QTLs) and genome-wide association studies (GWASs) (Kang et al. 2015). NGS has also provided a large amount of information associated with the identification of expressed sequence tag–simple sequence repeats (EST–SSRs) and single-nucleotide polymorphisms (SNPs), two of the most used markers in MAS (Chen et al. 2015; Ceballos et al. 2015; Kaur et al. 2015; Trebbi et al. 2015).

EST–SSR markers are codominant, highly reproducible and polymorphic markers. With these properties, EST–SSRs have been used favorably for population genetic analysis, genetic mapping and molecular breeding in diverse species such as tea, lettuce, cabbage and mango (Yi et al. 2006; Simko 2009; Izzah et al. 2014; Dillon et al. 2014). Current research in MAS seeks to identify key genes and with these generate markers within or adjacent to the functional sequence. However, few studies have focused on the developing of EST–SSR markers, which consider gene expression levels of contrasting phenotypes and also key transcription factors of metabolic pathways that regulate the traits of interest. The use of both may be a powerful tool for the selection of target characters in breeding programs (Bosamia et al. 2015).

In this study, we report the first set of EST–SSR markers for *P. salicina* developed from specific genes associated with the tricarboxylic acid, glycolysis and flavonoid pathways, which determine different organoleptic properties of the fruit. Our objectives were: (1) to obtain the transcriptome from the skin of contrasting cultivars of Japanese plum, (2) develop and validate EST–SSRs and (3) evaluate the potentiality of using a set of the validated EST–SSRs as molecular markers related to fruit skin color.

Materials and methods

Plant material

All the plant material used in this investigation was collected from 29 stabilized cultivars in the Japanese plum orchard in the Estación Experimental Pirque (33°40′12″S–70°35′06″W) that belong to the Facultad de Agronomía e Ingenieria Forestal of the Ponticia Universidad Católica de Chile, Pirque, Región Metropolitana.

Fruits at harvest maturity of cultivars 'Angeleno' (red skin) and 'Lamoon' (yellow skin) were harvested in the 2011–2012 season. Skins were separated from the pulp and frozen in liquid nitrogen. Young leaves from the 29 cultivars were collected for DNA isolation. All of the plant material was stored at −80 °C until use.

RNA extraction, library construction, deep sequencing, de novo transcriptome assembly and annotation

Total RNA from the skin of 'Lamoon' (yellow) and 'Angeleno' (red) cultivars was isolated using the protocol of Meisel et al. (2005). Paired-end library preparation and high-throughput sequencing for each sample were performed at a contract sequencing facility (Macrogen, Inc. Seoul, South Korea). The cDNA library was sequenced using an Illumina HiSeq 2000 platform.

The CLC Genome Workbench software (version 4.8) for de novo sequence assembly strategy was used to assemble reads into the contigs from a pool of all the paired-end short-read data (hybrid assembly), using the parameters: similarity = 0.95, length fraction = 0.7; insertion/deletion cost = 3 and mismatch cost = 3.

Functional annotation was performed through BLAST2GO software (Conesa et al. 2005) using the following databases: Refseq (Nr), GeneOntology (GO), InterProScan, and Kyoto Encyclopedia of Genes and Genomes (KEGG). All functional annotation was performed using alignment with *e* value $< 10E-5$.

RNA sequence analysis and differential expression in silico

Contigs over 1000 base pair (bp) arising from the hybrid de novo assembly were used as a reference set of transcripts for RNAseq analysis. Short-read sequence data from samples were separately mapped against the reference set of assembled transcripts using the CLC Genome Workbench RNAseq function (similarity = 0.95; length fraction = 0.7; maximum mismatches = 2; unspecific match limit = 10); paired reads counted as two and paired-end distances as described for the assembly.

For digital gene expression (DGE) analysis, we selected only contigs with at least five reads mapped to each hybrid contig, to increase the confidence of the results. The expression levels were compared using a *Z* test (Kal et al. 1999). For visual inspection, original expression values were transformed to Log_2 and then normalized by the quantile. Parameters of $p < 0.001$ and at least a fourfold difference were used for significant differential gene expression.

The main focus in the analysis was to obtain the patterns of gene expression belonging to primary metabolism, the citric acid cycles and the sugar metabolism pathways. From secondary metabolism, genes associated with the flavonoid pathway and related transcription factors were selected (Table 1).

SSR mining and primer design

Two strategies were used for microsatellite mining. The first consisted in the analysis and identification of simple tandem repeated motifs in the whole assembled transcriptome and selection of SSR sequences found in genes belonging to citric acid, sugar and flavonoid metabolic pathways. The second strategy was based on the recent study of Rahim et al. (2014), who described three paralog genes for *PpMYB10* in *P. persica*. Considering that these transcription factors have been directly associated with the pigmentation of skin and pulp, we explored their promoter sequences 3000 (bp) upstream from the +1 site using the genome sequence available in Phytozome (www.phytozome.net). The SSRLocator (Da Maia et al. 2008) software was used to search for repeat units of di-, tri-, tetra- to nonanucleotides in the transcriptome of *P.salicina* and in the *P. Persica* promotor sequences of the *PpMYB10.1*, *PpMYB10.2* and *PpMYB10.3* genes, using the software's suggested parameters. Primers were designed with the AmplifX software (version 1.7.0, by Nicolas Jullien; CNRS; Aix-Marseille University; http://crn2m.univ-mrs.fr/pub/amplifx-dist), using a minimum distance of 40-250 nucleotides from the tandem sequence identified.

Table 1 Digital gene expression profiles of flavonoid pathway genes and putative transcription factors regulators

Feature ID	Sequence description	Gene length	RPKM		Lemoon > angeleno	
			Angeleno	Lemoon	Fold change	FDR *p* value
Contig1142	Phenylalanine ammonia lyase (*PAL*)	2584	42.34	4.51	−9.39	0
Contig2053	Chalcone synthase (*CHS*)	1591	87.08	0.84	−102.72	0
Contig436	Chalcone isomerase (*CHI*)	792	23.51	42.64	1.81	0
Contig4405	Flavanone 3-hydroxylase (*F3H*)	2538	26.76	18.20	−1.48	0
Contig10314	Dihydroflavonol 4-reductase (*DFR*)	1493	28.93	1.24	−23.32	0
Contig10307	Anthocyanidin reductase (*ANR*)	3078	33.86	24.34	1.33	1.56E−06
Contig17478	Leucoanthocyanidin reductase (*LAR*)	313	0.60	0.12	−5.19	0.17
Contig481	Leucoanthocyanidin dioxygenase (*LDOX*)	1350	640.7	257.73	−2.49	0
Contig17535	Anthocyanidin-o-glucosyl transferase (*UFGT*)	245	26.77	2.37	−11.29	0
Contig522	Flavonoid o-methyl transferase (*OMT*)	1471	16.22	365.78	22.55	0
Contig12190	r2r3-Myb transcription factor myb10	963	10.61	0.11	−93.75	0
Contig15491	r2r3-Myb transcription factor myb123	1644	22.81	30.66	1.34	7.61E−12
Contig29731	r2r3-Myb transcription factor myb1	1622	1.58	1.88	1.19	0.43
Contig15252	Transcription factor bhlh35	2975	9.87	1.48	−6.74	0
Contig44812	Transcription factor hy5-like	967	0.27	1.61	5.92	3.84E−06

PCR amplification and validation of selected SSRs

DNA was isolated using the modified CTAB method as described by Barra et al. (2012) from frozen young leaf tissue of 29 cultivars. Considering the contribution of primary and secondary metabolism in different fruit quality traits, putative EST–SSRs identified in genes involved in the citric acid cycle, metabolism of sugars and the flavonoid pathway were selected. Twelve SSR sequences were identified from the promoters of the *PpMYB10.1*, *PpMYB10.2* and *PpMYB10.3* genes. A total of 72 primer pairs were screened for amplification and polymorphism using DNA from the 29 cultivars of *P. salicina* with different skin colors. The yellow skin cultivars analyzed were: 'Byron Gold', 'Shiro', 'Wickson', 'Lamoon', 'Norma' and 'Pink D' Light'; the cultivars with red skin were: 'Early Rosa', 'Catalina', 'Aurora', 'Angeleno', 'Royal Diamond', 'Larry Ann', 'Fortune', 'September King', 'Black Amber', 'Black Ruby', 'Elephant Heart', 'Rubusto', 'Ruby Queen', 'Laroda', 'Red Heart', 'Queen Rosa', 'Blue Gusto', 'Black Queen', 'Friar', 'MG-141', 'Sapphire', 'Santa Rosa' and 'Flavor Rich.'

Fragment amplification was performed following the method developed by Schuelke (2000), which uses three primers: a forward primer with an M13 (−21) tail at its 5′ end, a normal reverse primer and the universal M13 (−21) primer labeled with either 6-FAM, HEX or NED fluorochromes. PCR amplifications were performed in a 20 µl reaction mixture with 10 ng template DNA, 0.15 mM each dNTP; 1 × Taq polymerase reaction buffer; 1.0; 1.5 or 2.0 mM $MgCl_2$; 0.025 µM forward primer; 0.1 µM reverse primer; 0.1 µM M13 primer; and 0.35 U Taq DNA polymerase (5 U μl^{-1}, KapaBiosystems, Wilmington, Massachusetts). PCR cycles were as follows: initial denaturation at 94 °C for 5 min; 31 cycles of 30 s at 94 °C, annealing temperature specific to each primer pair for 45 s, extension at 72 °C for 45 s, followed by 16 cycles of 30 s at 94 °C, annealing at 53 °C for 45 s, extension at 72 °C for 45 s and a final extension at 72 °C for 30 min. Three PCR products labeled differentially were pooled and electrophoresed on the ABI 3130 × 1 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and sized precisely in base pair levels determined by an internal size standard (35–500 bp, GeneScan 500 LIZ; Applied Biosystems).

Data analysis

Allele binning and data statistics

Allele sizes were automatically calculated with GeneMapper v4.0 (Applied Biosystems) and reviewed

manually. Allele binning was done automatically with the Tandem program, which transforms allele sizes before rounding in order to compensate for allelic drift and compression at large fragment sizes (Matschiner and Salzburger 2009).

Population genetic parameters for polymorphic loci including number of alleles (N_A), observed heterozygosity (H_o) and expected heterozygosity (H_e) were estimated in the R package Adegenet (Jombart 2008). Polymorphism information content (PIC) was calculated in Pic-heterozygosity_v1.1.pl—Perl script (Arias et al. 2009).

Population structure assessed by EST–SSRs and SNPs

Figure 1 shows the working procedure used in the study. To determine the potential use of EST–SSR-type markers designed from structural genes and transcription factors associated with specific metabolic pathways for phenotypes of interest, a discriminant analysis of principal components (DAPC) was used which is available in the R package Adegenet (Jombart 2008; Jombart et al. 2010). For DAPC analysis, polymorphic EST–SSRs for structural genes and transcription factors related to flavonoid pathways were used; being also incorporated those polymorphic EST–SSRs identified from promotor sequences of the *PpMYB10* paralog genes. A Bayesian information criterion (BIC) value returned by DAPC analysis was used to determine the number of groups (*K* value).

We considered that only some polymorphism found within in the studied genes could be use as molecular markers for fruit skin color selection because not all of them could explain the different patterns of anthocyanin accumulation. For this reason, the clustering ability of different combinations of EST–SSRs and promoter–SSR (hereafter SSRs) that better separate colored from uncolored fruit was tested.

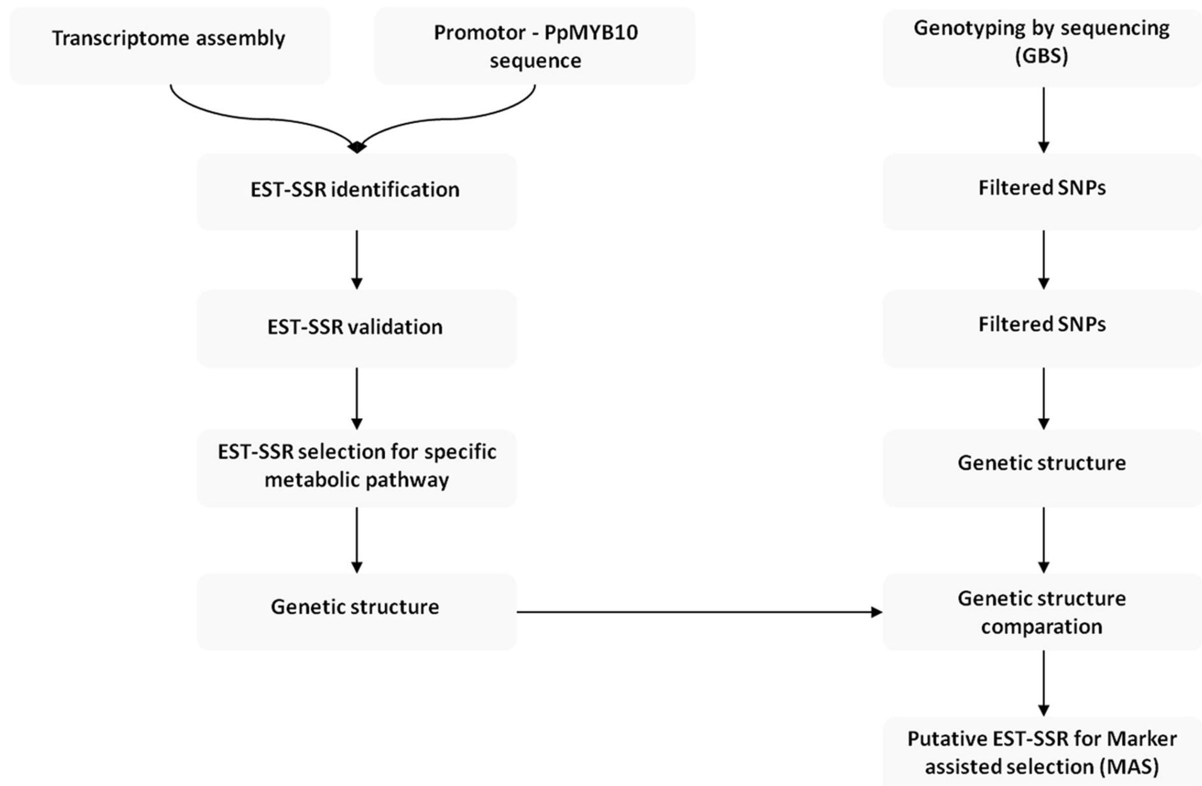

Fig. 1 Schematic representation of the procedure used to contrast genetic structure based on EST–SSR associated with specific metabolic pathways versus genetic structure given by high-density genomic markers generated by genotyping-by-sequencing (GBS)

Understanding the population genetic structure of the cultivars in which the EST–SSRs will be confirmed is fundamental to determine the possible relationships between these markers and the structure given by phenotypes associated with metabolic pathways that determine the organoleptic characteristics of interest (Fig. 1). SNPs generated by the GBS methodology were used to independently determine the population genetic structure of the 29 cultivars (Fig. 1). GBS was carried out at Cornell University Biotechnology Resource Center (BRC; Ithaca, NY, USA), according to the protocol described by Elshire et al. (2011). DNA of the 29 Japanese plum's cultivars was digested using ApeKI restriction enzyme. Each digested DNA sample was diluted to 100 ng/μl and submitted to BRC for analysis. Further details can be found at http://www.biotech.cornell.edu/brc/genomics-facility. In this regard, 251,463 SNPs were filtered using VCFtools (Danecek et al. 2011) and PLINK (Purcell et al. 2007) software packages. Using VCFtools, biallelic SNPs with quality score = 20 were obtained; meanwhile, with PLINK, the SNPs were filtered using the following criteria: minor allele frequency (maf) = 0.05; missingness per marker (geno) = 0.1; and missingness per individual (mind) = 0.1.

Genetic structure was determined using the Bayesian fastSTRUCTURE and DAPC (Jombart 2008; Jombart et al. 2010) software packages. Values of 2 to 12 were tested to determine the number of groups (*K*) using fastSTRUCTURE, and ten replicates were used for each value of *K*. With the DAPC, we determined the value of *K* using the BIC value given by the software. The critical membership value was set at 0.8.

Results

Library construction, deep sequencing and de novo transcriptome assembly

A total of 165,186,293 paired-end raw reads were used in hybrid assembly where 85,404,726 of them corresponded to the cultivar 'Lamoon' and 79,781,666 to 'Angeleno' (Supplementary Table 1). A total of 139,775,975 reads were filtered after removing low-quality reads and trimming the adapter sequences. The assembly generated 54,224 unique contigs, with an N50 of 1343 bp and mean length of 829 bp (Supplementary Table 1).

Metabolic pathway and transcription factor identification

Of the 54,224 contigs, 24,582 sequences (45.33 %) were annotated in at least one database. Of these, 19,873 and 19,738 were assigned to GO and InterPro databases, respectively. The data were classified into three categories: 17,536 (32.32 %) to biological processes; 16,025 (29.55 %) as cell components; and 16,908 (31.18 %) with molecular function (Supplementary Fig. 1). KEGG annotation was performed to determine the active metabolic pathways in plum skin. The results of the KEGG annotation determined that 8411 sequences were assigned to 142 metabolic pathways; 938 contigs were associated with antioxidant activities, especially vitamin metabolism (e.g., ascorbate, vitamin B6), carotenoids (β-carotene) and total phenols (phenylpropanoid pathway, phenylalanine, tyrosine and tryptophan biosynthesis, flavonoid biosynthesis; Supplementary Table 2). Our interest was to study the mechanisms associated with fruit skin color in the Japanese plum, with the focus on the identification of genes related to flavonoid biosynthesis. We were able to identify the following enzymes of this pathway: phenylalanine ammonia lyase (*PAL*); chalcone synthase (*CHS*); chalcone isomerase (*CHI*); flavanone 3-hydroxylase (*F3H*); dihydroflavonol 4-reductase (*DFR*); leucoanthocyanidin dioxygenase (*LDOX*); UDP-glucose/flavonoid 3-O-glucosyltransferase (*UFGT*); flavonoid O-methyltransferase (*OMT*); leucoanthocyanidin reductase (*LAR*); and anthocyanidin reductase (*ANR*). In addition, we identified a total of 355 putative transcription factors belonging to the families *R2R3MYB, MADS-box, bHLH, WD40, NAC, GATA, GRAS, TCP, bZIP* and *WRKY* (Supplementary Table 3).

Differential expression in silico

The DGE analysis performed between yellow-skinned and red-skinned cultivars found 4341 differentially expressed contigs: 2069 in 'Lamoon' and 2272 in 'Angeleno'. Neither the genes of the citric acid cycle nor those related to metabolism of sugars showed differences in expression in the analyzed tissues. However, the DGE showed that the flavonoid pathway

Table 2 Description of EST–SSR and promoter–SSR evaluated in this study

Locus	Description	Repeat motif	FW primer	RV primer	Allele range bp	Na	He	Ho	PIC
EST–*SSR from anthocyanin biosynthesis*									
EST-RG-Ps-25	Transcription factor hy5-like protein	(CT)16	TGGGGACTGACATGGCTTTA	TTGCTTTGCTGCCTCTACAC	158.00–172.00	7	0.70	0.62	0.92
EST-RG-Ps-26	Transcription factor bhlh35	(TC)18	CCCCACTCACTGTGACTAAA	ACTCATCGCCAATGTTGTCC	156.00–178.00	4	0.73	0.34	0.93
EST-RG-Ps-27	r2r3-myb Transcription factor myb1	(CAA)11	CTCCAATGTGAGCAGCTATG	GTTGCCATGAATTGGGTTTG	200.00–218.00	4	0.51	0.41	0.88
EST-RG-Ps-28	Chalconesynthase	(CT)10	TCTCCACTGCGTCTGCAACA	TGAGCCTTGCGAACTTCCT	130.00–160.00	9	0.83	0.59	0.96
EST-RG-Ps-29	r2r3-myb Transcription factor myb123	(TTTG)6	AGCTAGGAAGACCTCTTTGT	TGCCCATTTCACCGTGTT	149.00–169.00	4	0.63	0.72	0.91
EST-RG-Ps-30	Flavonoid O-Methyltransferase	(ATGCTG)2	GCGTACCATGGAGTGTCACATGTT	AGCGATCATGATCTTCCCACCTCT	276.00	1	0.00	0.00	0.75
EST-RG-Ps-31	Phenylalanine ammonia lyase	(TGAGAA)2	CCTGCAATGCCACTTACCCATTGA	AGGCAAAAGCACCTTCAGCTCT	164.00	1	0.00	0.00	0.75
EST-RG-Ps-32	r2r3-myb transcription factor myb10	(TCCTTT)2	ACCTATGTTGGGCTCCTTTTCC	AGCGTCATACCAGCTAAACCCA	208.00–212.00	3	0.54	0.00	0.89
SSR designed from promoter regions of paralogous PpMYB10									
p-RG-Ps-33	ppa016711m	(GAGTGT)2	GTTCGGCACATAGTTGCTTTG	GCGATCGGGTGTTCCAGTAAT	121.00–259.00	3	0.,26	0.29	0.85
p-RG-Ps-34	ppa016711m	(ATAAA)3	GTGTTTGGTCACTGGAGA	CTGGAAGTCTTCGAGCAATC	235.00–242.00	2	0.39	0.04	0.89
p-RG-Ps-35	ppa016711m	(TG)4	GGATCCGTGACATACAAACTTGTG	CCAATGGACCTGTTGAGGTACGAA	334.00–350.00	5	0.78	0.35	0.95
p-RG-Ps-36	ppa016711m	(TTTTCA)2	GGTACTTCAATTGTCCCGCGAATC	CACACATCACCTGGAATCCTGCTT	144.00–156.00	2	0.07	0.07	0.77
p-RG-Ps-37	ppa016711m	(T)12	TGACTTTTGAGATTGGATTGCAGA	TCACAAAATTCCCCTCGAGTCT	199.00–201.00	2	0.49	0.00	0.89
p-RG-Ps-38	ppa026640m	(ACTGTT)2	CCACACGTGCTGACACTAACAAAC	TTCAAAGCCTCCAGAAGCCACCTA	263.00–271.00	5	0.73	0.74	0.95
p-RG-Ps-39	ppa026640m	(GA)11	GTGTGGGTAAGGGAGGGAGTGT	GCGGATTCTCCATTACGGAGTACG	251.00–323.00	6	0.74	0.28	0.93
p-RG-Ps-40	ppa020385m	(ATTTTA)2	CGTCGGTCGATTCATAACTTCTGC	ACGGTCTAAGTGAAAACCGTCGTG	237.00	1	0.00	0.00	0.75
p-RG-Ps-41	ppa020385m	(AGTTTT)2	CGTCGGTCGATTCATAACTTCTGC	ACGGTCTAAGTGAAAACCGTCGTG	240.00	1	0.00	0.00	0.75
p-RG-Ps-42	ppa020385m	(TATATT)2	CGGTTTCCACACCAAAGTTTGTCG	TTTCAAGGGCAGTGACGTGCAT	258.00	1	0.00	0.00	0.77
p-RG-Ps-43	ppa020385m	(GTTA)3	CGACGGTAGTTAGTTAAGACGA	ACACTCCTGCCTGAGGATAA	222.00–290.00	4	0.30	0.00	0.88
p-RG-Ps-44	ppa020385m	(AGA)4	TTCGTAACCCTTGGGATCTCAG	CGACGTCATTCTCGTATAGTTC	276.00–308.00	5	0.58	0.97	0.89

is strongly expressed in ‘Angeleno’ and repressed in ‘Lamoon’. while genes *CHI*, *F3H* and *ANR* did not show differential expression; *PAL*, *CHS*, *DFR*, *LDOX* and *UFGT* genes were strongly overexpressed in this development stage (Table 1).

PCR amplification and validation of selected SSRs

A total of 2728 SSR sites were identified from the transcriptome (Supplementary Table 4), including four SSRs from the citric acid cycle, twenty from sugar metabolism (Supplementary Table 5) and eight from the synthesis and regulation of the flavonoid pathway (Table 2). Twelve SSRs were identified from the promotor sequences of the paralog *PpMYB10* genes described for *P. persica*; five, two and five SSRs were found in the genes ppa016711 m; ppa026640 m; and ppa020385 m, respectively (Table 2). The SSR identified from the transcriptome of *P. salicina* has a wide variety of repetitive motifs with dinucleotide sequences being the most frequent. SSR from the promotor sequences of the paralog *PpMYB10* genes also showed variability in the repeated motifs (Table 2).

From the selected sequences, 72 primer sets were designed; 44 of these produced consistent amplification. Of the 32 EST–SSRs designed, 12.5 % of the SSR were monomorphic in the 29 cultivars, while 25 % of the SSR-promoters were polymorphic in these cultivars. The polymorphic SSR showed a range of 2–10 alleles per locus. The number of alleles per locus, observed and expected heterozygosity and PIC values are given in both Table 2 and Supplementary Table 5.

Population structure based on SNPs

From GBS, a total of 251,463 SNPs were obtained; 19,153 biallelic SNPs were selected for posterior analysis, using maf = 0.05 and geno = 0.1. The Bayesian fastSTRUCTURE was performed with the 29 cultivars; the best separation of individuals was obtained with $K = 2$, suggesting that there is genetic differentiation among the evaluated accessions. DAPC was performed to confirm this classification,

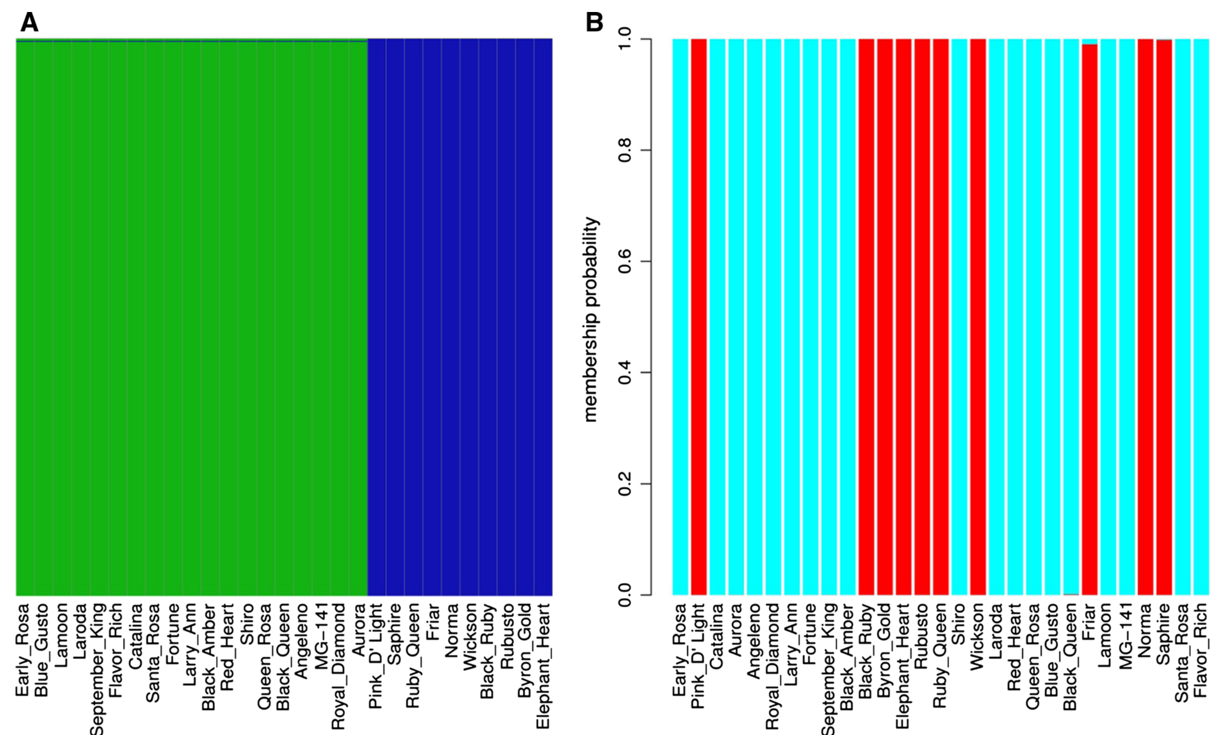

Fig. 2 Genetic structure using SNP data in twenty-nine *P. salicina* cultivars. **a** Genetic structure using the fastSTRUCTURE admixture model at $K = 2$. **b** Membership probability in DAPC analysis

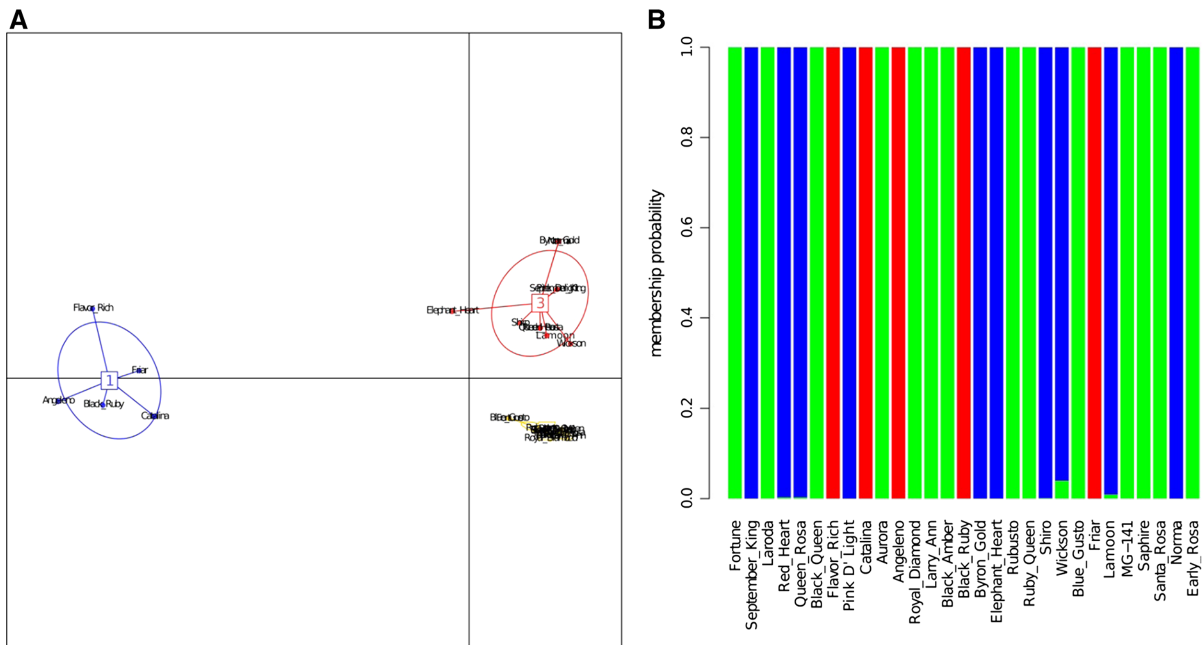

Fig. 3 Genetic structure using three selected EST–SSR data in twenty-nine *P. salicina* cultivars. **a** Genetic structure based on discriminant analysis of principal components (DAPC) of 29 *P. salicina* cultivars using flavonoid pathway EST–SSR: *PsMYB10*, *PsMYB1*, *PsbHLH35*. **b** Membership probability in genetic structure based on DAPC

again obtaining $K = 2$. The membership analysis performed with both tests showed that the cultivars in each cluster coincided with the individual analyses without exception (Fig. 2).

Clustering using genomic markers (SNPs) placed the six yellow-skinned cultivars (that do not accumulate anthocyanins) in both groups, with the 'Shiro' and 'Lamoon' cultivars separated from 'Pink D' Light', 'Wickson', 'Norma' and 'Byron Gold' (Fig. 2).

Population structure based on EST–SSRs

The genetic structure of the 29 cultivars using the seven EST–SSRs associated with the synthesis and regulation of the flavonoid pathway plus the nine promoter–SSRs found in the *PpMYB10* paralogs were analyzed. The structure analysis performed with DAPC using all the SSRs mentioned above differed from the population structure determined by the SNPs data; however, the grouping pattern was not associated with skin color (data not shown). For this reason, we considered that only some genes could be considered as molecular markers for fruit skin color selection. The best combination of EST–SSRs and promoter–SSR with the ability to separate colored from uncolored fruit skins was selected. EST–SSRs belonging to the transcription factors *PsMYB1*, *PsMYB10* and *PsbHLH35* generated three clusters of cultivars: two of red cultivars and a third that grouped all the yellow cultivars studied ('Lamoon', 'Norma', 'Byron Gold', 'Pink D' Light', 'Shiro' and 'Wickson'). The DAPC analysis and memberships are shown in Fig. 3.

Discussion

The NGS technologies of sequencing and transcriptome analysis have facilitated the understanding of the molecular mechanisms that are directly involved in traits of agricultural interest. GBS methodologies have strongly enhanced studies of genetic maps construction, QTLs and GWAS. Our study used GBS strategy to evaluate population genetic structure based on SNPs markers and then used RNAseq to identify candidate genes in order to select markers associated with fruit skin color in Japanese plum.

Here we report the first set of EST–SSRs and promoter–SSRs described for Japanese plum,

providing 37 new polymorphic markers and doubling the number of SSR reported to date for the species (Mnejja et al. 2004). These markers may potentially be transferred to other species of the genus *Prunus* and the Rosaceae. Gasic et al. (2009) evaluated 68 EST–SSRs in 14 species of the Rosaceae, finding a range of transferability between 25 and 59 %. Of the 44 markers evaluated, 12.5 % of the primers from EST–SSRs were monomorphic, which is less than the 26 % reported by Nishitani et al. (2009) for Japanese pear, while the promoter–SSR showed 25 % polymorphism.

Trinucleotide motifs are the most frequent in the EST–SSRs in purple yam (*Dioscorea alata*), cereal species and twenty-four tropical tree species (Varshney et al. 2002; Saski et al. 2015; Russell et al. 2014). In general, SSR occurrence in coding regions seems to be limited by non-perturbation of the reading frame, which is understandable in loci with repeat units that are a multiple of three. However, most of the polymorphic SSRs identified in the Japanese plum's transcriptome were dinucleotides, which finally were validated by allele binning using Tandem software. This coincides with the pattern observed in species such as *Aspidistra saxicola* (Huang et al. 2014) and common bean (*Phaseolus vulgaris*) (Blair and Hurtado, 2013). In addition, the same pattern was observed in rose, almond and peach (Jung et al. 2005), being our results consistent with the previous reported for Rosaceae. Although the dinucleotides motifs have not been linked to genetic functions, these SSR loci are those with the largest number of allelic variants (Vigouroux et al. 2002). In addition, a high level of SSR polymorphism increases the number of allelic variants, and depending on the position of the SSRs in the gene, these have a different effects. While, an SSR in the exon can affects the protein activity, SSR variations at the 5' UTR can affect the gene transcription and translation. When SSR variation occur in introns, different alternative splicing patterns can be obtained, which finally causes different phenotypic changes.

The mean number of alleles in polymorphic EST–SSRs and promoter–SSRs was 4.5. This is lower than the 5.7 alleles per locus reported by Mnejja et al. (2004) and the 12.1 reported by Carrasco et al. (2012) for genomic SSR, but similar to the 4.3 alleles per locus for genomic SSR reported for peach and cherry (Dirlewanger et al. 2002; Wünsch and Hormaza 2002), suggesting that in *Prunus*, the SSRs developed from functional sequences have a similar polymorphic levels that those developed from genomic regions.

The DGE analysis showed profiles of differential expression for structural genes of the flavonoid pathway (secondary metabolism) in the skin of Japanese plum. 'Lamoon' (yellow skin) showed a lower expression of the genes involved in the flavonoid pathway than did 'Angeleno' (red skin), which concurs with other species of the Rosaceae such as pear, strawberry, peach and cherry when a color difference is studied (Fischer et al. 2007; Salvatierra et al. 2010; Ravaglia et al. 2013; Wei et al. 2015). This metabolic pathway begins with the enzyme *CHS*, which condenses three units of malonyl-CoA with *p*-coumaroyl-CoA and is the first point of regulation. Then, the enzyme *CHI* generates a flavanone (naringenin) which can be transformed metabolically into flavones, isoflavones and dihydroflavonols. These molecules may be reduced to flavan 3,4-diols (leucoanthocyanidin), catalyzed by the enzyme *DFR*. Later, *LDOX* performs oxidation and dehydration steps to convert leucoanthocyanidin into anthocyanidin, which is modified by the enzyme *UFGT*, producing anthocyanins (Winkel-Shiley 2001).

Several processes must be carefully coordinated during fruit development. Anthocyanin accumulation is a key process in this stage. Previously, the transcription factor families *R2R3MYB*, *WRKY* (Ishida et al. 2007), *bZIP* (Shin et al. 2007; Stracke et al. 2010), *bHLH* (Baudry et al. 2004), *WD40* (Gonzalez et al. 2009) and NAC (Zhou et al. 2015) have been reported as having roles in the regulation of anthocyanin synthesis. The flavonoid pathway is highly coordinated, responding to a set of transcription factors with specific spatial and temporal expression (Xu et al. 2015). Current studies have focused on transcription factors of the *R2R3MYB*, *bHLH* and *WD40* families. Our DGE showed that *PsMYB10* and *PsbHLH35* are expressed differentially in favor of 'Angeleno', while *PsMYB1* did not show expression changes between the libraries. The analysis of genetic structure based on the EST–SSRs designed for these transcription factors grouped the yellow cultivars in only one cluster, in contrast to the result using SNPs markers, in which they were distributed in both of the clusters identified.

QTLs studies in peach and sweet cherry determined that the skin color would be explained by three genomics regions (Sooriyapathirana et al. 2010; Frett

et al. 2014). Specifically, Sooriyapathirana et al. (2010) identified a major QTL on linkage group 3 (LG3) which explained until to 74.0 % of fruit color phenotype, suggesting the presence of a major regulatory gene in this region. MYB10 has been mapped and colocated within the interval of this QTL (Sooriyapathirana et al. 2010; Frett et al. 2014) could be the major candidate gene involved in the skin and pulp color in fruits of *Prunus*. Our study found three allele variants in the EST–SSRs designed for the *PsMYB10* transcription factor. Recently, Tuan et al. (2015) described in peach different allelic variants of *PpMYB10.1*. Several insertion/deletion nucleotides (INDELs) from a sequences and SNPs were found in both 5′-upstream and 3′-downstream, identifying two allelic types *Pp*MYB10.1-1/*Pp*MYB10.1-2, which cause different *Pp*MYB10.1 proteins and determine the differences in the intensity of red skin color. At the same time, several studies have reported the importance of allele variants in this transcription factor in apple, indicating segregation of the color character in the progeny (Takos et al. 2006; Chagné et al. 2007; Espley et al. 2009; Lin-wang et al. 2010). Also, Ravaglia et al. (2013) proposed *PpMYB10* as a regulator of anthocyanin synthesis, controlling the expression of the enzymes *DFR* and *UFGT*, while in strawberry it has been proposed to have a regulatory role on *LDOX* and *UFGT* (Medina-Puche et al. 2014; Kadomura-Ishikawa et al. 2015). Espley et al. (2009) identified tandem repeat variations in the promotor zone of *MdMYB10* and determined the relation between these sequences and anthocyanin accumulation in the fruit pulp. We did not find a relation between the studied polymorphisms in promoter zones and the color of the skin; however, incorporating pulp color in future studies may provide a potential use of promoter–SSR in MAS studies.

The regulation of anthocyanin synthesis in strawberry has been associated with the negative regulation of the transcription factor *MYB1* on the enzymes *LDOX* and *UFGT* (Aharoni et al. 2001; Salvatierra et al. 2013). The participation of different transcription factors of the *bHLH* family has been validated in the genus *Prunus*; *bHLH3* was reported to be fundamental during the co-expression of *PpMYB10* in heterologous systems in peach (Ravaglia et al. 2013; Rahim et al. 2014). This information suggests possible participation of *PsMYB1*and *PsbHLH35* in anthocyanin regulation and synthesis in Japanese plum that would also explain the genetic structure generated with these markers.

Genetic structure based on EST–SSRs versus SNPs

The development of NGS technology has opened new opportunities to explore the relation between genotype and phenotype by raising the studies to the whole-genome level, improving the resolution of QTL studies and facilitating the development of markers inside of or adjacent to functional sequences (Varshney et al. 2014). In particular, the importance of EST–SSRs markers to determine genetic diversity, generate gene maps and perform studies of comparative genomics has been widely recognized (Yi et al. 2006; Simko 2009; Izzah et al. 2014; Dillon et al. 2014).

In order to evaluate the capacity of a set of EST–SSRs or promoter–SSRs to separate plum cultivars according to fruit skin color, it was necessary to determine the genetic structure of the available germplasm bank. This was performed using GBS, which is a high-throughput, multiplex and short-read sequencing approach that reduces the complexity of a genome using restriction enzymes and generates high-density genome-wide markers (Elshire et al. 2011; He et al. 2014). In cultivated species, this strategy has improved the understanding of the genetic architecture of complex characters and its application in genetic improvement programs by means of association studies and genomic selection, especially in organisms that do not have a reference genome (Poland et al. 2012). This study utilized 19,153 SNP loci that were analyzed using fastSTRUCTURE and DAPC. fastSTRUCTURE is a Bayesian approximation to determine the structure of a population (Raj et al. 2014). DAPC is a multivariate method for the identification of genetic structure and clustering individuals that does not depend upon assumptions of Hardy–Weinberg equilibrium or linkage disequilibrium, maximizing the variance between groups and minimizing the variance within groups (Jombart et al. 2010). DAPC is a more accurate method to identify structure in populations that have undergone selection. Both tools produced $K = 2$, confirming the structure described by Carrasco et al. (2012) for the cultivars analyzed here, and improving the reliability of the results obtained in previous analyses.

Conclusion

The development of EST–SSR markers considering gene expression levels of contrasting phenotypes and specific transcription factors from metabolic pathways that regulate characters of interest could be a powerful tool for selection in breeding programs. The DGE analysis between contrasting cultivars for anthocyanin accumulation found expression profiles that will help in the understanding of the molecular mechanisms that regulate anthocyanin biosynthesis. There are few studies focused on developing EST–SSR markers considering expression profiles in structural genes of metabolic pathways or associated transcription factors. Our study evaluated the real potential of utilizing a set of EST–SSRs identified in key genes of the biosynthetic pathway of anthocyanins as molecular markers, using the presence or absence of these metabolites as test character.

Three EST–SSRs identified in the *PsMYB10*, *PsMYB1* and Ps*bHLH35* transcription factors identified a different pattern of grouping than that obtained by SNP loci alone. The three EST–SSRs showed three groups, one of which was composed mainly of yellow-skinned cultivars, and thus have the potential to be used as molecular markers to be evaluated in MAS strategies in genetic improvement programs for the selection of fruit skin color in Japanese plum and potentially other members of Rosaceae.

Acknowledgments This work was funded by the Comisión Nacional de Investigación Científica y Tecnológica, CONICYT. Scholarships Doctorado en Chile 2009 No. 21090118; Apoyo a la Tesis Doctoral 2011 No. 24110179; Tesis en la Industria 2013 No. 781211008 and FONDECYT/Regular No. 1120261. Consorcio Tecnológico de Industria Hortofruticula S.A. Fondo Instituto de Investigaciones Agropecuarias—Ministerio de Agricultura (INIA-MINAGRI) Project No. 501453-70.

References

Aharoni A, Ric de Vos CH, Wein M, Sun Z, Greco R, Kroon A, Mol JNM, O'Connell AP (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J 28:319–332. doi:10.1046/j.1365-313X.2001.01154.x

Arias RS, Ballard LL, Scheffler BE (2009) UPIC: perl scripts to determine the number of SSR markers to run. Bioinformation 3(8):352

Barra M, Salazar E, Sagredo B (2012) High-quality genome DNA extraction method for high-throughput genotyping analysis in populations of species with phenolic content such as *Solanum tuberosum* (L.) spp tuberosum or *Zea mays* (L.). Cienc Investig Agrar 39(3):593–601

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 39(3):366–380

Bielenberg DG, Rauh B, Fan S, Gasic K, Abbott AG, Reighard GL, Okie WR, Wells CE (2015) Genotyping by Sequencing for SNP-Based Linkage Map Construction and QTL Analysis of Chilling Requirement and Bloom Date in Peach [*Prunus persica* (L.) Batsch]. PLoS One 10(10):e0139406. doi:10.1371/journal.pone.0139406

Blair MW, Hurtado N (2013) EST–SSR markers from five sequenced cDNA libraries of common bean (*Phaseolus vulgaris* L.) comparing three bioinformatic algorithms. Mol Ecol Resour 13(4):688–695

Bosamia TC, Mishra GP, Thankappan R, Dobaria JR (2015) Novel and stress relevant EST derived SSR markers developed and validated in peanut. PLoS One 10(6):e0129127. doi:10.1371/journal.pone.0129127

Carrasco B, Díaz C, Moya M, Gebauer M, García-González R (2012) Genetic characterization of Japanese plum cultivars (*Prunus salicina*) using SSR and ISSR molecular markers. Cienc Investig Agrar 39:533–543

Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH (2015) Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theor Appl Genet 128(9):1647–1667

Chagné D, Carlisle CM, Blond C, Volz RK, Whitworth CJ, Oraguzie NC, Crowhurst RN, Allan AC, Espley RV, Hellens RP, Gardiner SE (2007) Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. BMC Genom 8(1):212. doi:10.1186/1471-2164-8-212

Chen H, Qiao L, Wang L, Wang S, Blair MW, Cheng X (2015) Assessment of genetic diversity and population structure of mung bean (*Vigna radiata)*germplasm using EST-based and genomic SSR markers. Gene 566(2):175–183

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676

Da Maia LC, Palmieri DA, De Souza VQ, Kopp MM, de Carvalho FIF, Costa de Oliveira A (2008) SSR Locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. Int J Plant Genom. doi:10.1155/2008/412696

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158. doi:10.1093/bioinformatics/btr330

Dillon NL, Innes DJ, Bally IS, Wright CL, Devitt LC, Dietzgen RG (2014) Expressed sequence tag-simple sequence repeat (EST–SSR) marker resources for diversity analysis of mango (*Mangifera indica* L). Diversity 6(1):72–87

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet 105:127–138

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5):e19379. doi:10.1371/journal.pone.0019379

Espley RV, Brendolise C, Chagné D, Kutty-Amma S, Green S, Volz R, Putterill J, Schouten HJ, Gardiner SE, Hellens RP, Allan AC (2009) Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell 21(1):168–183. doi:10.1105/tpc.108.059329

Fischer TC, Gosch C, Pfeiffer J, Halbwirth H, Halle C, Stich K, Forkmann G (2007) Flavonoid genes of pear (*Pyrus communis*). Trees 21(5):521–529

Frett TJ, Reighard GL, Okie WR, Gasic K (2014) Mapping quantitative trait loci associated with blush in peach [*Prunus persica* (L.) Batsch]. Tree Genet Genom 10(2): 367–381

Gasic K, Han Y, Kertbundit S, Shulaev V, Lezzoni AF, Stover EW, Bell RL, Wisniewski ME, Korban SS (2009) Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. Mol Breed 23(3):397–411

Gonzalez A, Mendenhall J, HuoY Lloyd A (2009) TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Dev Biol 325(2):412–421

Goulão L, Monte-Corvo L, Oliveira CM (2001) Phenetic characterization of plum cultivars by high multiplex ratio markers: amplified fragment length polymorphisms and inter-simple sequence repeats. J Am SocHortic sci 126(1):72–77

Hartmann W, Neumüller M (2009) Plum Breeding. In: Jain SM, Priyadarshan PM (eds) Breeding plantation tree crops: temperate species. Springer, New York, pp 161–231

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS): an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci. doi:10.3389/fpls.2014.00484

Huang D, Zhang Y, Jin M, Li H, Song Z, Wang Y, Chen J (2014) Characterization and high cross-species transferability of microsatellite markers from the floral transcriptome of Aspidistra saxicola (Asparagaceae). Mol Ecol Resour 14(3):569–577

Ishida T, Hattori S, Sano R, Inoue K, Shirano Y, Hayashi H, Shibata D, Sato S, Kato T, Tabata S, Okada K, Wada T (2007) Arabidopsis TRANSPARENT TESTA GLABRA2 is directly regulated by R2R3 MYB transcription factors and is involved in regulation of GLABRA2 transcription in epidermal differentiation. Plant Cell 19(8):2531–2543

Izzah NK, Lee J, Jayakodi M, Perumal S, Jin M, Park BS, AhnK Yang TJ (2014) Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. capitata L.) and construction of an EST-based genetic map. BMC Genom 15(1):149. doi:10.1186/1471-2164-15-149

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11): 1403–1405

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94. doi:10.1186/1471-2156-11-94

Jung S, Abbott A, Jesudurai C, Tomkins J, Main D (2005) Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. Funct Integr Genom 5(3):136–143

Kadomura-Ishikawa Y, Miyawaki K, Takahashi A, Masuda T, Noji S (2015) Light and abscisic acid independently regulated FaMYB10 in Fragaria × ananassa fruit. Planta 241:953–965

Kal AJ, van Zonneveld AJ, Benes V, van den Berg M, Koerkamp MG, Albermann K, Strack N, Ruijter JM, Richter A, Dujon B, AnsorgeW Tabak HF (1999) Dynamics of Gene Expression Revealed by Comparison of Serial Analysis of Gene Expression Transcript Profiles from Yeast Grown on Two Different Carbon Sources. Mol Biol Cell 10(6):1859–1872

Kang YJ, Lee T, Lee J, Shim S, Jeong H, Satyawan D, Kim MY, Lee SH (2015) Translational genomics for plant breeding with the genome sequence explosion. Plant Biotechnol J. doi:10.1111/pbi.12449

Kaur S, Panesar PS, Bera MB, Kaur V (2015) Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. CRC CR Rev Food Sci 55(1):41–49

Kim C, Guo H, Kong W, Chandnani R, Shuang LS, Paterson AH (2016) Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci 242:14–22. doi:10.1016/j.plantsci.2015.04.016

Klabunde GHF, Dalbó MA, Nodari RO (2014) DNA fingerprinting of Japanese plum (*Prunus salicina*) cultivars based on microsatellite markers. Crop Breed Appl Biotechnol 14(3):139–145

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21(6):991–1007

Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol 10:50. doi:10.1186/1471-2229-10-50

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982–1983

Medina-Puche L, Cumplido-Laso G, Amil-Ruiz F, Hoffmann T, Ring L, Rodríguez-Franco A, Caballero JL, Schwab W, Muñoz-Blanco J, Blanco-Portales R (2014) MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria × ananassa fruits. J Exp Bot 65(2):401–417

Meisel L, Fonseca B, González S, Baeza-Yates R, Cambiazo V, Campos R, Gonzalez M, Orellana A, Retamales J, Silva H (2005) A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. Biol Res 38:83–88

Mnejja M, Garcia-Mas J, Howad W, Badenes ML, Arús P (2004) Simple sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and

transferable to peach and almond. Mol Ecol Notes 4(2):163–166

Nishitani C, Terakami S, Sawamura Y, Takada N, Yamamoto T (2009) Development of novel EST–SSR markers derived from Japanese pear (*Pyrus pyrifolia*). Breed Sci 59(4):391–400

Ortiz A, Renaud R, Calzada I, Ritter E (1997) Analysis of plum cultivars with RAPD markers. J HorticSci 72(1):1–10

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2):e32253. doi:10.1371/journal.pone.0032253

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81(3):559–575. doi:10.1086/519795

Rahim MA, Busatto N, Trainotti L (2014) Regulation of anthocyanin biosynthesis in peach fruits. Planta 240(5):913–929

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197(2):573–589

Ravaglia D, Espley RV, Henry-Kirk R, Andreotti C, Ziosi V, Hellens RP, Costa G, Allan AC (2013) Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biol 13:68. doi:10.1186/1471-2229-13-68

Rubio M, Rodríguez-Moreno L, Ballester AR, Moura MC, Bonghi C, Candresse T, Martínez-Gómez P (2015) Analysis of gene expression changes in peach leaves in response to Plum pox virus infection using RNA-Seq. Mol Plant Pathol 16(2):164–176

Russell JR, Hedley PE, Cardle L, Dancey S, Morris J, Booth A, Odee D, Mwaura L, Omondi W, Angaine P, Machua J, Muchugi A, Milne I, Dawson IK (2014) Tropitree: an NGS-based EST–SSR resource for 24 tropical tree species. PLoS One 9(7):e102502. doi:10.1371/journal.pone.0102502

Salvatierra A, Pimentel P, Moya-Leon MA, Caligari PDS, Herrera R (2010) Comparison of transcriptional profiles of flavonoid genes and anthocyanin contents during fruit development of two botanical forms of *Fragaria chiloensis* ssp. chiloensis. Phytochemistry 71(16):1839–1847

Salvatierra A, Pimentel P, Moya-León MA, Herrera R (2013) Increased accumulation of anthocyanins in *Fragaria chiloensis* fruits by transient suppression of FcMYB1 gene. Phytochemistry. 90:25–36

Saski CA, Bhattacharjee R, Scheffler BE, Asiedu R (2015) Genomic resources for water yam (*Dioscorea alata* L.): analyses of EST-sequences, de novo sequencing and GBS libraries. PLoS One 10(7):e0134031. doi:10.1371/journal.pone.0134031

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234

Shimada T, Hayama H, Haji T, Yamaguchi M, Yoshida M (1999) Genetic diversity of plums characterized by random amplified polymorphic DNA (RAPD) analysis. Euphytica 109(3):143–147

Shin J, Park E, Choi G (2007) PIF3 regulates anthocyanin biosynthesis in an HY5 dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. Plant J 49(6):981–994

Simko I (2009) Development of EST–SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). J Hered 100(2):256–262. doi:10.1093/jhered/esn072

Sooriyapathirana SS, Khan A, Sebolt AM, Wang D, Bushakra JM, Lin-Wang K, Allan AC, Gardiner SE, Chagné D, Iezzoni AF (2010) QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). Tree Genet Genom 6(6):821–832

Stracke R, Favory JJ, Gruber H, Bartelniewoehner L, Bartels S, Binkert M, Funk M, Weisshaar B, Ulm R (2010) The Arabidopsis bZIP transcription factor HY5 regulates expression of the PFG1/MYB12 gene in response to light and ultraviolet-B radiation. Plant Cell Environ 33(1): 88–103

Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR (2006) Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol 142(3):1216–1232

Trebbi D, Papazoglou EG, Saadaoui E, Vischi M, Baldini M, Stevanato P, Cettul E, Sanzone AP, Gualdi L, Fabbri A (2015) Assessment of genetic diversity in different accessions of Jatrophacurcas. Ind Crop Prod 75:35–39. doi:10.1016/j.indcrop.2015.06.051

Tuan PA, Bai S, Yaegaki H, Tamura T, Hihara S, Moriguchi T, Oda K (2015) The crucial role of PpMYB10. 1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biol 15:280. doi:10.1186/s12870-015-0664-5

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoSBiol 12(6):e1001883. doi:10.1371/journal.pbio.1001883

Vieira EA, Nodari RO, Dantas ACDM, Ducroquet JPHJ, Dalbó M, Borges CV (2005) Genetic mapping of Japanese plum. Crop Breed Appl Biot 5(1):29–37

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19(8):1251–1260

Wei H, Chen X, Zong X, Shu H, Gao D, Liu Q (2015) Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). PLoS One 10(3): e0121164. doi:10.1371/journal.pone.0121164

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126(2):485–493

Wünsch A, Hormaza, JI (2002) Molecular characterisation of sweet cherry (*Prunus avium* L.) genotypes using peach

[*Prunus persica* (L.) Batsch] SSR sequences. Heredity 89(1):56–63

Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci 20(3):176–185

Yi G, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST–SSRs and an SSR-based linkage map. Theor Appl Genet 114(1):113–130

Zhou H, Lin-Wang K, Wang H, Gu C, Dare AP, Espley RV, He H, Allan AC, Han Y (2015) Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. Plant J 82:105–121