

## Development of EST-SSR markers for diversity and breeding studies in opium poppy

HATICE ŞELALE, İBRAHİM ÇELİK, VISAM GÜLTEKİN, JENS ALLMER, SAMI DOĞANLAR and ANNE FRARY\*

Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir 35430, Turkey; \*Corresponding author, E-mail: annefrary@iyte.edu.tr

With 2 figures and 5 tables

Received October 12, 2012/Accepted January 29, 2013

Communicated by T. Debener

### Abstract

All publicly available opium poppy expressed sequence tag (EST) sequences, totalling 20 885, were assembled into unigenes and examined for simple sequence repeats (SSRs). Nearly 19% of the 14 957 unigenes contained SSRs with 4% harbouring more than one SSR. Average density of the SSRs was 1 SSR per 3.6 kb of non-redundant EST sequence. Trinucleotide SSRs were most frequently identified (39%), and many of the most prevalent motifs were AT-rich. Flanking primers were designed for 86% of the SSRs and 67 primer pairs were tested on 37 opium poppy accessions and seven related species. All markers were transferable to the related species. Polymorphism information content (PIC) values for the markers were intermediate for comparisons within opium poppy (average of 0.27) and slightly higher for comparisons across species (average of 0.29). The markers were found to be useful for diversity analysis as they successfully distinguished among Turkish opium poppy accessions and land races.

**Key words:** expressed sequence tag — genetic diversity — genic SSRs — microsatellites — *Papaver somniferum* L. — simple sequence repeat

Opium poppy (*Papaver somniferum*) is grown for its edible seeds, oil (Nergiz and Qiles 1994) and 80 different alkaloids (Schulz et al. 2004). The pharmaceutically valuable alkaloids obtained from opium poppy include morphine and codeine, which are narcotics and analgesics; thebaine, which is a mild analgesic and sedative; noscapine and narcotine, which are anti-tussives and apoptosis inducers; and papaverine, which is a vasodilator (Ye et al. 1998). Turkey currently accounts for 48% of the world's production, more than any other country. However, the country ranks second in world morphine production (18%) after Australia (23%) (Soil Product Office 2009). This discrepancy is due to low morphine yield of Turkish poppy. Because opium poppy serves as one of the most important renewable resources for pharmaceutical alkaloids, there is a growing demand for cultivars with high alkaloid content.

Despite the crop's significance, opium poppy has been the subject of few investigations with respect to its genetic variation. The molecular genetic diversity studies carried out to date have used non-specific markers such as AFLP (Saunders et al. 2001) and RAPD (Parmaksiz and Özcan 2011). RAPD and ISSR markers were also used by Hari et al. (2009) to analyse 24 accessions. In one of the most comprehensive studies, 300 accessions were analysed with AFLP markers and amounts of the major alkaloids were determined (Dittbrenner et al. 2008). For forensic identification of opium poppy, Lee et al. (2010) developed six

morphine-biosynthesis-specific PCR markers and a limited number (22) of EST-SSR markers, only six of which were tested for polymorphism in opium poppy (Lee et al. 2011). To date, only one poppy genetic linkage map was constructed with 77 AFLP and 48 RAPD markers (Straka and Nothnagel 2002). Clearly, there are very few poppy-specific molecular markers available for analysis of this important plant's genome. Development of such molecular markers is necessary for comprehensive analysis of the opium poppy genome for both breeding and molecular studies.

Simple sequence repeat (SSR) or microsatellite markers are sequence-specific markers and are widely used in the assessment of genetic diversity, construction of framework genetic maps, gene mapping, marker-assisted selection and comparative mapping studies. Simple sequence repeat markers are hypervariable, codominantly inherited, multiallelic, relatively abundant, reproducible, suitable for automation and high-throughput genotyping and provide extensive genome coverage (Gupta and Varshney 2000). Simple sequence repeats are tandem repeated motifs of 1–6 bp, which occur frequently in prokaryotic and eukaryotic genomes (Zane et al. 2002). Simple sequence repeats are classified depending upon the number of nucleotides per repeat unit (mono-, di-, tri-, tetra-, penta- or hexanucleotides), type of repeat unit (simple or compound, meaning that two basic repeat motifs are present together in various configurations) and location in the genome (nuclear, chloroplastic or mitochondrial) (Wang et al. 2009). Polymorphisms in SSRs are introduced as errors caused by single-stranded DNA slippage, unequal crossing over, gene conversion, mismatch/double-strand break repair and retrotransposition (Ellegren 2004). Simple sequence repeats are distributed throughout the nuclear, mitochondrial (Rajendrakumar et al. 2007) and chloroplast genomes (Chung et al. 2006) and are present in both coding and non-coding regions. These motifs constitute a significant proportion of the non-coding DNA in many organisms (Morgante et al. 2002) and are found in lower amounts and copy numbers in expressed sequences (Li et al. 2002). Despite this apparent limitation, approximately 14% of protein-coding sequences contain repetitive DNA regions with SSRs (Marcotte et al. 1999). According to EST (expressed sequence tag) database-mining studies in various organisms, untranslated regions (UTRs and introns) have more SSRs than exons. (Gao et al. 2003). It has been reported that SSRs in 3' UTRs are the most polymorphic at the cultivar level, while those in 5' UTRs are the most polymorphic between cultivars and species (Varshney et al. 2005a). Simple sequence repeats within exons are most polymorphic between species and genera (Scott

et al. 2000). Repeat motifs and number of repeats can also vary for SSRs located in different regions of the genome. For example, dinucleotide repeats are most common in the genomic DNA of many species, but are much less frequent in coding regions (Li et al. 2002). Moreover, exons have more trinucleotide SSRs than other repeats (Li et al. 2004). Simple sequence repeats located in coding regions can affect gene activity and even cause phenotypic changes among individuals (Li et al. 2004). Even if a SSR is located in a non-coding region of an EST (UTR or introns), it may still have a role in gene regulation or gene transcription (Lawson and Zhang 2006). Thus, EST-SSR markers are gene-targeted markers with known or putative functions, which can enable direct allele selection, if they are shown to be associated or responsible for a targeted trait (Sorrells and Wilson 1997). Moreover, the highly conserved nature of genes enables EST-SSR markers to be used in different species and allows genome alignment across distantly related species (Varshney et al. 2005b).

With advances in sequencing technology, a huge amount of EST data have been produced for various species including opium poppy. These data are available in databases and can be economically and easily mined for EST-SSRs. In this work, publicly available *P. somniferum* EST sequences (www.ncbi.nlm.nih.gov/) were assembled, annotated and analysed for the presence of SSRs. Identified SSRs were then used to design primers that were tested for their use and transferability in opium poppy and related species.

## Materials and Methods

**Plant materials:** The study was carried out with 37 poppy accessions grown in Turkey and seven closely related species of poppy. Nine poppy accessions were provided by Soil Product Office (TMO), while the remaining materials were provided by the Anatolian Agricultural Research Institute (AARI), Eskisehir, Turkey. The seven related species were as follows: *Papaver orientale* (Iran), *Papaver pseudoorientale* (Iran), *Papaver bracteatum* (Iran), *Papaver rhoeas* (Bulgaria), *Papaver umbonatum* (Turkey), *Papaver nudicaule* (Mongolia) and *Papaver armeniacum* (Armenia). The 44 accessions were grown in soil in a growth chamber at 25°C with a 16/8-h photoperiod until they produced enough leaf material for DNA extraction. Leaf tissue was combined from approximately 20 plants per accession, and genomic DNA was extracted using a CTAB method (Doyle and Doyle 1990).

**Data and preprocessing:** All available *P. somniferum* sequences (20,885 raw ESTs; Table 1) were downloaded from NCBI (www.ncbi.nlm.nih.gov/) on 10 October 2010. As EST sequences downloaded from NCBI may contain contamination from cloning vectors, adapters, linkers, or primer sequences, SeqClean (http://compbio.dfci.harvard.edu/tgi/software) was used to remove such potential impurities. The vector sequences that served as references for vector removal with SeqClean were downloaded from NCBI (http://ftp.ncbi.nih.gov/pub/UniVec/Univec, version from 11 August 2010). Furthermore, the default adapter and linker sequences for the Roche 454 pyro-sequencer were used as an additional database for SeqClean. The parameters of SeqClean were set

so that the program would not remove poly(A)/(T) tails as having poly (A)/(T) tails in the sequence can be beneficial (Morin et al. 2008). NCBI BLAST (Zhang et al. 2000) was used to analyse whether the filtering process successfully removed the vector sequences.

**Assembly and annotation:** Because the presence of poly(A)/(T) tails can complicate assembly, the tails of the preprocessed EST sequences were masked during assembly with MIRA software (Chevreux et al. 2004). The default settings for the parameters were used with a few changes as follows: uniform read distribution was off, masking poly(A)/(T) tails on, no quality file requirement on, quality clipping on, extra gap penalty on and non-IUPAC consensus pre seq type was forced. Blast2Go was used for ontological analysis of the EST assembly (Conesa et al. 2005).

**Primer design and PCR amplification:** BatchPrimer3 was used to identify SSRs and design primers flanking the SSRs (You et al. 2008). The default settings of the program were used with some adjustments: product size 100–300 bp, optimum 150 bp; primer size 18–27 nt, optimum 21 nt; primer T<sub>m</sub> 50–60°C, optimum 55°C; primer GC content 40–60%, optimum 50%. Single PCRs were carried out for the 44 accessions with 67 EST-SSR primers. The PCR mix contained 8 pmol of the reverse and M13-labelled primers and 2 pmol of the forward primer in a 25-μl reaction volume with 2.5 μl PCR buffer, 0.2 mM dNTPs, 2 mM MgCl<sub>2</sub>, 50–100 ng template DNA and 1 U *Taq* DNA polymerase. Conditions for all PCR amplifications were as follows: 94°C for 5 min, then 30 cycles of 94°C for 30 s, 58°C for 45 s and 72°C for 45 s, followed by eight cycles of 94°C for 30 s, 53°C for 45 s and 72°C for 45 s and a final extension at 72°C for 10 min. Afterwards, 3 μl of the PCR product was added to 27 μl of sample loading buffer (Beckman Coulter) and 0.5 μl 600-bp size standard (Beckman Coulter). Finally, the mixture for each accession was run on a Beckman CEQ8800 for capillary electrophoresis using frag3 method (capillary temperature, 50°C; denaturation at 90°C for 120 s; injection voltage, 2.0 kV for 3 s; separation voltage, 4.8 kV for 60 min).

**Data analysis:** For each accession, fragments amplified with the EST-SSR markers were scored as present (1) or absent (0). Fragments that occurred in fewer than four accessions were assumed to be rare and possibly unreliable and therefore eliminated from further analysis. The data were then used to compute the polymorphism information content (PIC) value for each polymorphic marker fragment according to the formula:  $PIC_i = 2f_i(1-f_i)$  where  $f_i$  is the frequency of band presence (Roldán-Ruiz et al. 2000). The marker data were also used to produce a dissimilarity matrix using the DICE coefficient as implemented in DARwin5 software (Perrier and Jacquemoud-Collet 2006). A dendrogram was then drawn using the unweighted neighbour-joining method. Bootstrap analysis was also performed with 1000 replications. A Mantel test statistic was used to determine the correlation of the dissimilarity matrix and the dendrogram.

## Results

### Assembly and annotation

A total of 20 885 raw EST sequence reads from a publicly available opium poppy EST database (NCBI) were cleaned using

Table 1: Sequence preprocessing and assembly statistics

Parameter	Raw ESTs	Cleaned ESTs	Contigs	Singlets
Minimum sequence length (nt)	50	101	135	101
Maximum sequence length (nt)	2 435	2 435	5 600	2 435
Average sequence length (nt)	821.5 ± 279.9	823.7 ± 276.5	1 213.5 ± 542.2	849.5 ± 283.6
Total no. sequences	20 885	20 795	2 297	12 660
Total no. bases	17 158 054	17 130 647	2 787 822	10 755 450

SeqClean software (Table 1). After cleaning, 20 795 ESTs were assembled with MIRA software (Chevreux *et al.* 2004). As a result, 8135 reads were assembled into 2297 contigs ranging from 135 to 5600 nucleotides in length with an average length of 1213.5 nt. A total of 12 660 reads were unassembled singletons ranging in size from 101 to 2435 nt with an average of 849.5 nt. Together, the contigs and singletons represented 14 957 unigenes covering 13.5 Mb of the opium poppy genome. Blast2Go was used for ontological analysis of the EST assembly. As the EST sequences were gathered from a public database and not generated purposefully, no direct conclusions can be drawn from such an analysis, other than that the biological processes enriched in Fig. 1 are of general importance. A total of 2456 unigenes (1906 singletons and 550 contigs) were assigned ontologies. The ontologies of singletons and contigs were very similar as reflected by the similar proportions for each category seen in Fig. 1. The majority of unigenes were assigned to cellular and metabolic processes (24 and 21% each, respectively) with a much lower proportion associated with stimulus response (approximately 12%). Biological regulation, and developmental and multicellular organismal processes were represented by approximately 6% of the unigenes, while the remaining categories each represented less than 5% of the unigenes.

#### Abundance and distribution of EST-SSRs

Simple sequence repeats were mined from the unigene sequences, and 3701 non-redundant SSRs were identified in 2818 unigenes (Table 2). Thus, 18.8% of the unigenes contained SSRs. Total SSR length varied up to 156 nucleotides. The longest SSR was a hexanucleotide repeat. Numerous unigenes contained multiple SSRs. For example, 446 unigenes (16% of the unigenes with SSRs) contained two SSRs, 125 unigenes (4%) contained three SSRs, while 32 (1%) and 14 (0.5%) unigenes

contained four or five SSRs, respectively. Only two unigenes (<0.1%) contained six or seven SSRs and only one unigene contained eight SSRs. More than one-third (39%) of the identified SSRs were trinucleotide repeats. Tetra- and mononucleotide repeats were also fairly common (22 and 19%, respectively). Dinucleotide repeats ranked fourth in abundance, accounting for 11% of the SSRs. Penta- and hexanucleotide repeats were rarer (6 and 3%, respectively), while hepta- and octanucleotides together accounted for less than 1% of the identified SSRs. Among mononucleotide repeats, (A/T)<sub>n</sub> was more common than (G/C)<sub>n</sub> and accounted for 57% of mononucleotide repeats (Table 3). Among dinucleotide repeats, (AT/TA)<sub>n</sub> and (TA/AT)<sub>n</sub> were the most abundant repeat motifs and represented 17 and 16% of dinucleotides, respectively. Among the trinucleotide repeats, (TTC/AAG)<sub>n</sub> and (AAG/TTC)<sub>n</sub> were most abundant with each accounting for approximately 6% of the trinucleotides. Only one motif accounted for more than 5% of tetra- and heptanucleotides. Both were AT-rich repeats: AAAT/TTTA and AA-ATA/TTTAT.

#### Marker evaluation and polymorphism detection

A total of 2147 primer pairs were designed for 86.4% of the identified SSRs. Of these, 93 EST-SSR primers were selected based on high G/C content, melting temperature of 55–60°C and so that no contig was represented by more than one primer. These primers were tested on a subset of opium poppy and related species accessions. Most of the primers (76, 82%) produced amplicons; however, 17 (18%) failed to amplify DNA. A total of 67 primers with clear amplification fragments on agarose gel were selected and tested in 37 *P. somniferum* accessions and seven related species to test their polymorphism in opium poppy and their transferability (Table 4). The *P. somniferum* accessions included 19 accessions representing current varieties and breed-

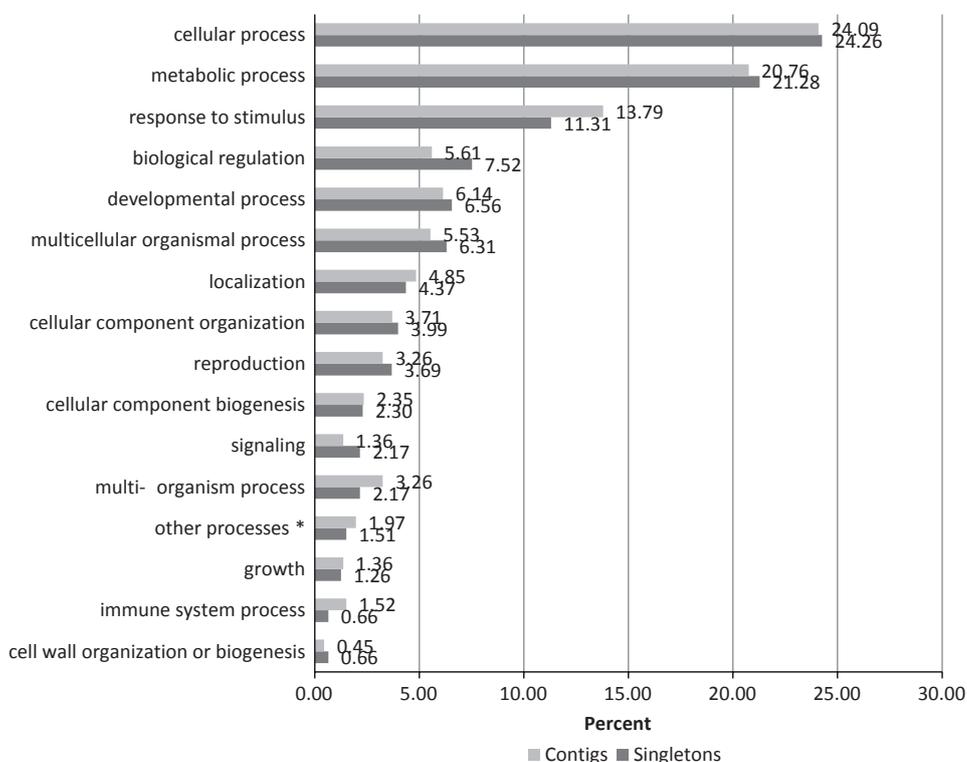


Fig. 1: Functional categories of annotated opium poppy unigenes

Table 2: Frequencies of microsatellite repeat types in the opium poppy EST database

Repeat types	Number of SSRs	Proportion of all SSRs (%)
Mononucleotide	697	18.8
Dinucleotide	407	11.0
Trinucleotide	1435	38.8
Tetranucleotide	819	22.1
Pentanucleotide	224	6.1
Hexanucleotide	93	2.5
Heptanucleotide	20	0.5
Octanucleotide	6	0.2
Total	3701	100

Table 3: The most abundant SSR motifs in the opium poppy EST database. Only motifs accounting for 5% or more of each repeat type (i.e. mono-, di-, trinucleotides, etc.) are included

SSR motif	Number of SSR motifs	Percentage of SSR motifs
A/T	400	57.4
AT/TA	71	17.4
TA/AT	66	16.2
AG/TC	54	13.3
GA/CT	51	12.5
TTC/AAG	89	6.2
AAG/TTC	84	5.9
GAA/CTT	82	5.7
AGA/TCT	78	5.4
AAAT/TTTA	47	5.7
AAATA/TTTAT	19	8.5

ing materials. In addition, 18 accessions represented older varieties/land races collected from different poppy-growing regions. The 67 primers yielded 562 polymorphic fragments in all accessions and 253 polymorphic fragments in *P. somniferum* accessions only (Table 5). Of the 67 markers, 65 (97%) were polymorphic in all accessions with an average fragment polymorphism of 55.5%. PIC values were calculated for each polymorphic marker using a method that gives a maximum value of 0.50 (Roldán-Ruiz et al. 2000). For all accessions, average (PIC) values ranged from 0 for psSSR036 and psSSR077 to 0.50 for psSSR003 with an average PIC of 0.29. When only *P. somniferum* accessions were considered, 14 of the 67 markers were non-polymorphic. Thus, 79% of the poppy EST-SSR markers were polymorphic within *P. somniferum* and average fragment polymorphism was 30%. For the intraspecific comparison, PIC values for the markers ranged from 0 to 0.50 with an average of 0.27. The most informative markers were psSSR006, psSSR010, psSSR023 and psSSR100, which each had a PIC value of 0.5 for *P. somniferum* accessions. PIC value was similar across all repeat types (dinucleotide, trinucleotide, etc.) ranging from 0.23 (heptanucleotide) to 0.32 (tetranucleotide) for all accessions with no significant differences based on motif length.

#### Assessment of genetic diversity in opium poppy with EST-SSR markers

A total of 562 polymorphic fragments from 67 EST-SSR markers were used to assess genetic diversity in the poppy germplasm. The data were used to compute a genetic similarity matrix using the Dice index and unweighted neighbour-joining algorithm. The resulting dendrogram (Fig. 2) was very highly correlated with the distance matrix as determined by a Mantel test ( $r = 0.9988$ ). The minimum dissimilarity between accessions

Table 4: Opium poppy accessions and related species used in study

Name	Source	Location
1290	AARI	–
1061	AARI	–
1259	AARI	–
1065	AARI	–
Kemerkaya	AARI	–
Aday1	AARI	–
Tinaztepe	AARI	–
Zaferyolu	AARI	–
Anayurt	AARI	–
Afyon95	AARI	–
Ofis3	TMO	–
Ofis4	TMO	–
Ofis8	TMO	–
Ofis95	TMO	–
Ofis96	TMO	–
TM01	TMO	–
TM02	TMO	–
TM03	TMO	–
TM04	TMO	–
5	AARI	Çeltik, Burdur
12	AARI	Yalvaçhöyüküklü, Isparta
16	AARI	Çamlıbel, Ulubey, Uşak
26	AARI	Dereşenek, Afyon
32	AARI	Afyon
33	AARI	Ekinhisar, Sandikli, Afyon
37	AARI	Yalvaçhöyüküklü, Isparta
40	AARI	Isparta
41	AARI	Koçyatağı, Şuhut, Afyon
45	AARI	Simav, Kütahya
48	AARI	Karayahşiler, Çivril, Denizli
62	AARI	Sülümenli, Afyon
72	AARI	Koçyatağı, Şuhut, Afyon
76	AARI	Gögen, Uşak
78	AARI	Alacami, Sandikli, Afyon
92	AARI	Bolvadin
96	AARI	Kütahya
111	AARI	Şuhut, Afyon
PI 229617 ( <i>P. orientale</i> )	AARI	Iran
PI 381612 ( <i>P. pseudoorientale</i> )	AARI	Iran
PI 414784 ( <i>P. bracteatum</i> )	AARI	Iran
W6 10919 ( <i>P. rhoeas</i> )	AARI	Bulgaria
W6 11444 ( <i>P. umbonatum</i> )	AARI	Turkey
W6 18131 ( <i>P. nudicaule</i> )	AARI	Mongolia
W6 23866 ( <i>P. armeniacum</i> )	AARI	Armenia

was 0.07 (that is 93% similarity), while the maximum dissimilarity was 0.80. Average dissimilarity was 0.31. The *Papaver* species formed a group that was separate from the *P. somniferum* accessions with very high bootstrap support (Fig. 2). Dissimilarity of opium poppy with the related species ranged from 0.47 to 0.79. The *P. somniferum* accessions fell into three main clusters. Group A contained 18 accessions all of which were land races collected from different poppy-growing locations in Turkey. This cluster had an average dissimilarity of 0.14 with minimum and maximum dissimilarities of 0.08 and 0.20, respectively. Group B contained two clusters encompassing 17 accessions, which consisted of named varieties and breeding materials. This group had an average dissimilarity of 0.15 with minimum and maximum dissimilarities of 0.07 and 0.21, respectively. The topology of group B had better bootstrap support than group A. Accessions 1061 and 1290 were the most genetically distinct opium poppy accessions and did not group with the rest of the accessions.

#### Discussion

EST-SSRs are abundant in the opium poppy genome. A total of 14 957 unigenes encompassing 13.5 Mb of the opium poppy

Table 5: EST-SSR markers and their polymorphism information content. Markers that amplified multiple loci in *P. somniferum* are indicated with an asterisk

SSR marker	Repeat motif	All accessions		<i>P. somniferum</i>	
		No. Polymorphic/total no. fragments (%)	Average PIC $\pm$ SE	No. Polymorphic/total no. fragments (%)	Average PIC $\pm$ SE
psSSR002	(TTTG)3	14/41 (34)	0.30 $\pm$ 0.03	12/34 (35)	0.31 $\pm$ 0.04
psSSR003	(TTCT)3	2/2 (100)	0.50 $\pm$ 0.01	2/2 (100)	0.45 $\pm$ 0.01
psSSR004	(AAG)4	5/10 (50)	0.28 $\pm$ 0.05	2/8 (25)	0.23 $\pm$ 0
psSSR005	(TTAGG)5	2/25 (8)	0.16 $\pm$ 0.01	0/23 (0)	0
psSSR006	(TTA)5	5/5 (100)	0.22 $\pm$ 0.08	1/4 (25)	0.50
psSSR007	(AAAGTA)3	1/2 (50)	0.30	0/1 (0)	0
psSSR008*	(ATTA)3	17/25 (68)	0.33 $\pm$ 0.03	8/22 (36)	0.37 $\pm$ 0.04
psSSR009*	(TCAA)3	11/30 (37)	0.37 $\pm$ 0.02	6/30 (20)	0.38 $\pm$ 0.03
psSSR010	(AGAA)3	4/6 (67)	0.28 $\pm$ 0.08	1/2 (50)	0.50
psSSR012	(TAAT)3	7/7 (100)	0.29 $\pm$ 0.04	4/7 (57)	0.31 $\pm$ 0.07
psSSR013	(TTTTTC)3	6/12 (50)	0.41 $\pm$ 0.05	5/12 (42)	0.42 $\pm$ 0.06
psSSR014	(CT)8	5/12 (42)	0.36 $\pm$ 0.06	3/11 (27)	0.44 $\pm$ 0.01
psSSR016	(GTA)5	7/11 (64)	0.35 $\pm$ 0.04	4/11 (37)	0.39 $\pm$ 0.04
psSSR017*	(TTA)5	18/45 (40)	0.24 $\pm$ 0.03	4/45 (9)	0.36 $\pm$ 0.04
psSSR019*	(ATT)5	15/36 (42)	0.35 $\pm$ 0.03	9/36 (25)	0.39 $\pm$ 0.03
psSSR022	(CAAT)3	11/22 (65)	0.21 $\pm$ 0.02	1/16 (6)	0.24
psSSR023	(ATGAC)4	5/26 (19)	0.35 $\pm$ 0.05	4/12 (33)	0.50 $\pm$ 0.05
psSSR024	(TTTGT)3	5/14 (36)	0.22 $\pm$ 0.02	0/3 (0)	0
psSSR025	(TTAG)3	4/4 (100)	0.20 $\pm$ 0.07	0/4 (0)	0
psSSR032*	(TAC)4	24/24(100)	0.40 $\pm$ 0.01	15/24 (63)	0.39 $\pm$ 0.03
psSSR036	(CAATG)3	0/3 (0)	0	0/1 (0)	0
psSSR037	(GAA)4	5/10 (50)	0.34 $\pm$ 0.05	3/6 (50)	0.33 $\pm$ 0.05
psSSR039	(CAA)4	5/15 (33)	0.40 $\pm$ 0.03	5/14 (36)	0.40 $\pm$ 0.02
psSSR040*	(AAAG)4	10/12 (83)	0.45 $\pm$ 0.02	9/11 (82)	0.43 $\pm$ 0.03
psSSR041*	(ATAA)3	12/16 (75)	0.37 $\pm$ 0.03	7/15 (47)	0.36 $\pm$ 0.04
psSSR044*	(CTAG)3	10/11 (91)	0.38 $\pm$ 0.04	9/11 (82)	0.40 $\pm$ 0.03
psSSR046*	(TTTC)3	11/12 (92)	0.39 $\pm$ 0.03	9/18 (50)	0.44 $\pm$ 0.03
psSSR045	(TTC)5	11/13 (85)	0.24 $\pm$ 0.02	2/12 (17)	0.27 $\pm$ 0.00
psSSR047*	(ATGT)3	10/11 (91)	0.31 $\pm$ 0.04	3/9 (33)	0.31 $\pm$ 0.06
psSSR049	(AT)7	1/3 (33)	0.45	2/1 (50)	0.48 $\pm$ 0.00
psSSR051	(AGA)4	2/6 (33)	0.19 $\pm$ 0.03	0/2 (0)	0
psSSR052	(GATG)3	7/8(100)	0.27 $\pm$ 0.03	0/4 (0)	0
psSSR053*	(AAGA)3	12/16 (75)	0.38 $\pm$ 0.03	8/14 (57)	0.38 $\pm$ 0.02
psSSR054*	(TTTTG)3	18/23(78)	0.27 $\pm$ 0.02	2/9 (22)	0.27 $\pm$ 0.10
psSSR055	(TTATT)4	11/17 (65)	0.23 $\pm$ 0.03	1/9 (11)	0.49
psSSR057*	(AGA)4	10/11 (91)	0.32 $\pm$ 0.04	8/11 (73)	0.32 $\pm$ 0.03
psSSR058	(TGTACC)3	6/11 (55)	0.27 $\pm$ 0.04	1/9 (11)	0.31 $\pm$ 0.06
psSSR059*	(TTGG)3	9/9 (100)	0.41 $\pm$ 0.04	8/9 (89)	0.38 $\pm$ 0.03
psSSR060	(ATG)4	3/7 (43)	0.25 $\pm$ 0.04	0/1 (0)	0
psSSR061	(TGT)4	5/8 (62)	0.32 $\pm$ 0.07	1/4 (25)	0.47
psSSR063	(GTTG)3	3/6 (50)	0.39 $\pm$ 0.09	2/5 (40)	0.36 $\pm$ 0.05
psSSR064*	(TAAAAA)3	18/43 (42)	0.24 $\pm$ 0.02	7/30 (23)	0.22 $\pm$ 0.02
psSSR069*	(TAAA)3	8/13 (62)	0.41 $\pm$ 0.03	9/13 (69)	0.39 $\pm$ 0.05
psSSR071	(CT)6	8/10 (80)	0.30 $\pm$ 0.04	3/7 (43)	0.35 $\pm$ 0.15
psSSR072*	(AT)10	25/32 (78)	0.25 $\pm$ 0.02	4/23 (17)	0.31 $\pm$ 0.06
psSSR073	(TA)34	7/13(54)	0.33 $\pm$ 0.05	4/6 (67)	0.35 $\pm$ 0.07
psSSR074	(TTC)4	12/20 (60)	0.32 $\pm$ 0.04	12/29 (63)	0.35 $\pm$ 0.04
psSSR075*	(TA)6	20/27 (74)	0.23 $\pm$ 0.03	6/27 (22)	0.40 $\pm$ 0.05
psSSR076*	(AT)8	23/38 (61)	0.24 $\pm$ 0.02	2/17 (12)	0.35 $\pm$ 0.14
psSSR077	(ATGT)3	0/1 (0)	0	0/1 (0)	0
psSSR078	(GAA)4	8/16 (50)	0.23 $\pm$ 0.04	1/4 (25)	0.34
psSSR079*	(TG)6	7/9 (78)	0.30 $\pm$ 0.05	2/4 (50)	0.43 $\pm$ 0.00
psSSR080*	(AAT)4	9/22 (41)	0.29 $\pm$ 0.04	4/22 (18)	0.25 $\pm$ 0.07
psSSR081	(TAAA)3	5/18 (28)	0.34 $\pm$ 0.06	2/8 (25)	0.35 $\pm$ 0.11
psSSR082	(TAAA)3	4/15 (27)	0.42 $\pm$ 0.04	3/8 (38)	0.45 $\pm$ 0.05
psSSR084	(TAAA)3	10/27 (37)	0.30 $\pm$ 0.03	4/21 (19)	0.35 $\pm$ 0.06
psSSR087	(TTTC)3	5/7 (100)	0.35 $\pm$ 0.04	2/6 (33)	0.30 $\pm$ 0.07
psSSR089*	(CGTT)3	15/22 (68)	0.33 $\pm$ 0.03	9/23 (39)	0.30 $\pm$ 0.02
psSSR090	(CCTCT)3	2/4 (50)	0.39 $\pm$ 0.00	0/3 (67)	0
psSSR091	(TG)8	5/7 (71)	0.29 $\pm$ 0.04	3/6 (50)	0.28 $\pm$ 0.04
psSSR093	(ATCA)3	4/6 (67)	0.20 $\pm$ 0.10	1/4 (25)	0.49
psSSR094*	(ATC)4	9/17(53)	0.30 $\pm$ 0.02	0/15 (0)	0
psSSR095	(TAAA)4	3/10 (30)	0.20 $\pm$ 0.02	0/3 (0)	0
psSSR096	(GTCT)4	6/9 (67)	0.25 $\pm$ 0.03	0/5 (0)	0
psSSR097	(ATCC)4	1/4 (25)	0.42	1/2 (50)	0.47
psSSR099	(ATC)5	2/7 (29)	0.20 $\pm$ 0.04	0/5 (0)	0
psSSR100	(ACT)3	7/13 (54)	0.23 $\pm$ 0.04	1/11 (9)	0.50

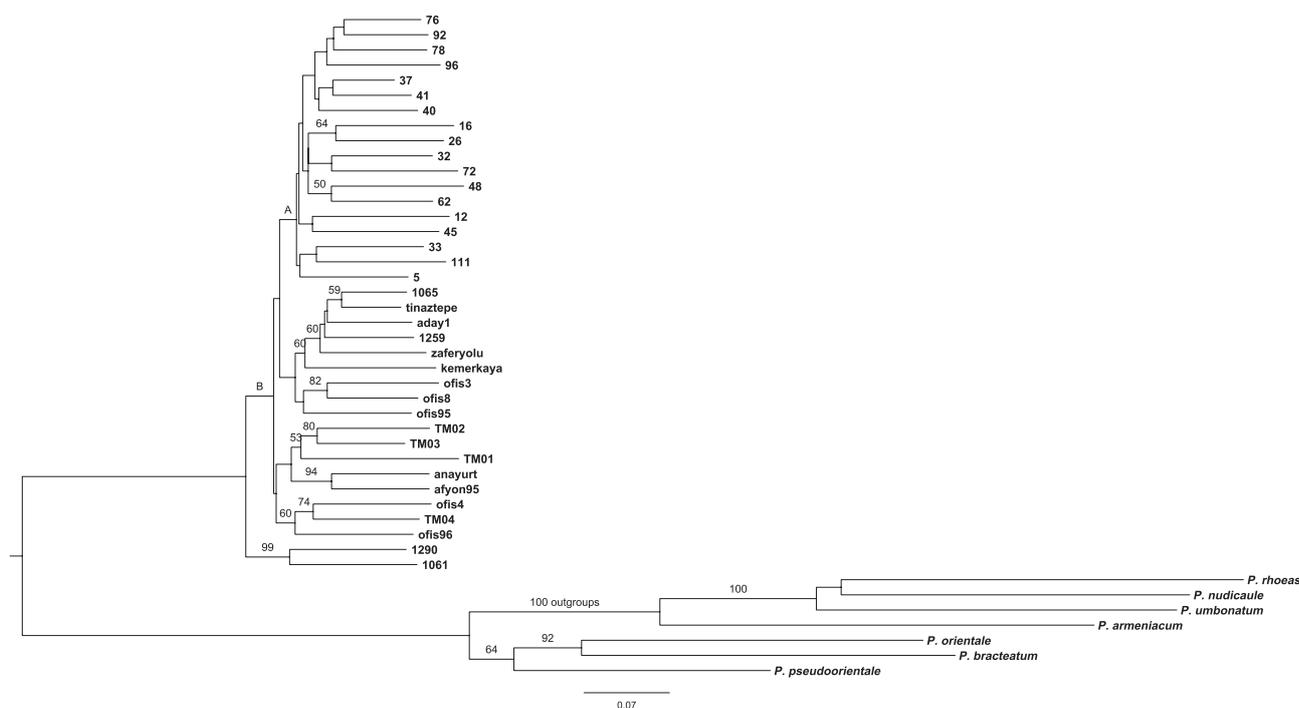


Fig. 2: Unweighted neighbour-joining dendrogram of the poppy accessions based on the Dice coefficient as calculated from the EST-SSR marker data. The accessions include 37 Turkish opium poppy accessions and seven related species used as outgroups. Bootstrap values (1000 replications) greater than 50% are indicated on branches

genome were detected in *P. somniferum* ESTs from the NCBI database. Thus, the analysed unigenes represented approximately 0.4% of the opium poppy genome (3724 Mb; Bennett and Smith 1976). Approximately 16% of the unigenes were assigned ontologies with more than half accounting for cellular and metabolic processes and response to stimuli. Desgagne-Penix et al. (2010) annotated 25% of opium poppy unigenes generated by pyrosequencing of a fungal-elicited opium poppy cell culture cDNA library. They also found that many annotated unigenes were involved in metabolism and cellular processes. The unigene sequences contained 3701 SSRs indicating an average density of one SSR every 3.6 kb of non-redundant EST sequences. In opium poppy, 18.8% of unigenes contained SSRs. This result is much higher than results obtained for other dicot species whose frequencies ranged from 2.6 to 10.6% (Kumpatla and Mukhopadhyay 2005, Varshney et al. 2005a). It is also higher than a previous value obtained for opium poppy, 13.7% (Lee et al. 2011). However, this previous value was based on unassembled ESTs and it is not clear whether the search tool allowed identification of multiple SSRs within one unigene. In the current work, 4% of the unigenes harboured more than one SSR. It is known that the frequency of EST-SSRs identified in genomes is dependent on detection criteria (Poncet et al. 2006); therefore, variability among studies is to be expected.

Trinucleotides were the most common SSRs in the opium poppy unigenes followed by tetra-, mono- and dinucleotides with very few SSRs larger than tetranucleotides. Trinucleotide repeats are the most common among monocot and dicot species (Varshney et al. 2005a). Non-trimeric repeats are less frequent because their occurrence in coding regions causes frameshift mutations, which can result in non-functional gene products (Kalia et al. 2011). As a result, many of these types of SSRs will be lost by negative selection. Lee et al. (2011) also found that trinucleotides were most abundant in opium poppy ESTs; however, in

their analysis dinucleotides were more abundant than tetranucleotides. In plants, the most frequent trinucleotide is AAG (Li et al. 2004) that, indeed, was the most common trinucleotide repeat in opium poppy ESTs (6.2% of trinucleotides). This motif was also most frequently identified in previous work in opium poppy (Lee et al. 2011). Many of the most prevalent motifs were AT-rich as was also observed by Lee et al. (2011) in opium poppy and by others in various plant species (Frery et al. 2005, Mun et al. 2006, Yi et al. 2006).

#### Polymorphism and transferability of EST-SSRs in opium poppy and related Papaver

Most (82%) of the synthesized primers amplified products in opium poppy. Possible explanations for EST-SSR primers that failed to produce PCR products include the presence of a large intron in the genomic sequence, primers that extend across a splice site or primers that were derived from chimeric cDNA clones (Varshney et al. 2005a). EST-SSRs are reported to have more limited polymorphism than genomic SSRs (Varshney et al. 2005a). In our study, we obtained an intermediate level of polymorphism of the EST-SSR markers among wild species and opium poppy (PIC = 0.29) and only a slightly lower level of polymorphism within opium poppy (PIC = 0.27). A previous study with the same type of markers demonstrated similar levels of polymorphism among opium poppy cultivars (Lee et al. 2011). It is reported that longer SSRs and those with longer repeat motifs are more polymorphic than other repeat motifs in many organisms including rice and tomato (Cho et al. 2000, Temnykh et al. 2000, Frery et al. 2005). However, in our study, no significant correlations of PIC values with motif or full SSR length were observed.

In this study, we detected more DNA fragments per EST-SSR primer than the previously published study for opium poppy

(Lee et al. 2011). However, Lee et al. (2011) did not use markers that produced more than three fragments, thereby limiting fragment number. Multilocus amplification can be due to polyploidy and heterozygosity in opium poppy, which was reported previously in wheat (Yu et al. 2004). *P. somniferum* ( $2n = 22$ ) is described as an aneupolyploid derived from species with  $x = 7$  (Lavania and Srivastava 1999). In this case, individual accessions may have as many as six unique fragments for each single-locus marker. A total of 23 markers (34%) amplified more than six fragments in opium poppy accessions indicating their multilocus nature (marked with asterisks in Table 5). Although multilocus markers provide more fragments per amplification reaction, such markers must be used with caution when allelism is important (e.g. in mapping). It is also possible that non-specific fragments could be amplified by some of the EST-SSR primers. To ensure that the expected SSRs were amplified by the primers, eight PCR fragments were sequenced. Of these, 75% contained the expected SSRs, while the remaining 25% contained SSRs but with a slightly different motif than expected (e.g. CCT instead of CT; data not shown).

As has been reported for other species, the EST-SSR markers developed in this study were highly transferable within the *Papaver* genus (Kalia et al. 2011). EST-SSRs are expected to be more transferable than genomic SSRs because they originate from more conserved portions of the genome. Nearly all (97%) of the 67 markers developed in our study were useful for genetic analyses in *Papaver somniferum*, *P. bracteatum*, *P. pseudoorientale*, *P. orientale*, *P. nudicaule*, *P. armeniacum*, *P. rhoeas* and *P. umbonatum*. A previously published study for opium poppy demonstrated low transferability among related species as only two of six primers were transferable among the *Papaver* and *Eschscholzia* genera (Lee et al. 2011). Transferability of markers among species suggests the presence of highly conserved genes with important biological, cellular and/or molecular functions. Transferability of EST-SSRs is especially important in a genus like *Papaver*, which is under-utilized and has limited sequence data available.

#### Use of EST-SSRs for diversity analysis in *Papaver*

Despite their intermediate level of polymorphism, the EST-SSRs were suitable for distinguishing among opium poppy accessions. This confirms the work of Lee et al. (2011) who found that EST-SSRs were suitable for forensic applications to differentiate opium poppy genotypes. However, this previous work was limited in scope as only six loci were tested and the authors indicated a need for the development of additional markers. In the current work, 79% (53 of 67) of EST-SSRs were useful in distinguishing Turkish poppy accessions. This value is most likely an underestimate because infrequent alleles were excluded from the analysis in case they were artefacts. Retesting of the markers with these fragments should indicate whether or not they are valid alleles. Identification of such rare alleles is important for distinguishing among germplasm accessions. According to dendrogram analysis, the markers were able to differentiate opium poppy from other poppy species. Thus, the opium poppy EST-SSR markers can be used to determine both intra- and interspecific differences among poppy accessions. Overall, the markers showed limited genetic diversity within opium poppy accessions and indicated that opium poppy breeding materials and current accessions were distinct from land races collected from different locations in Turkey. However, further discussion

of the importance of this separation requires additional marker analysis and better bootstrap support for dendrogram topology.

#### Conclusions

Opium poppy is an important source of many pharmaceutical alkaloids; however, it is a little-studied crop. Poppy-specific molecular markers are needed to learn more about the genetic control of opium poppy's useful traits and to improve them by targeted breeding efforts. In this study, EST-SSRs were mined from an opium poppy EST database. As a result, 65 of 67 markers were verified for their polymorphism and their usefulness for diversity analysis. Thus, the work increased the number of EST-SSR markers for opium poppy and *Papaver* genus analyses by approximately 10-fold. The markers will be useful for diversity analysis of additional opium poppy accessions to select appropriate materials for future breeding efforts to increase genetic variability in poppy. They can also be used in forensic analyses to differentiate opium poppy from other types of poppy. In addition, the markers will be useful for genetic mapping and eventual marker-assisted selection in the crop.

#### Acknowledgements

We thank Dr. Hüseyin Camci, Dr. Arzu Köse and Dr. Ferda Koşar of the Anatolia Agricultural Research Institute for plant material. This work was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) project no. 109O797.

#### References

- Bennett, M. D., and J. B. Smith, 1976: Nuclear DNA amounts in angiosperms. *Phil. Trans. Royal. Soc. London B.* **274**, 227–274.
- Chevreux, B., T. Pfisterer, and B. Drescher, 2004: Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–59.
- Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCouch, W. D. Park, N. Ayres, and S. Cartinhour, 2000: Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 713–722.
- Chung, A. M., J. E. Staub, and J. F. Chen, 2006: Molecular phylogeny of Cucumis species as revealed by consensus chloroplast SSR marker length and sequence variation. *Genome* **49**, 219–229.
- Conesa A., S. Götz, J. M. Garc a-G omez, J. Terol, M. Tal n, and M. Robles, 2005: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.
- Desgagne-Penix, I., M. F. Khan, D. C. Schriemer, D. Cram, J. Nowak, and P. J. Facchini, 2010: Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC Plant Biol.* **10**, 252–269.
- Dittbrenner, A., U. Lohwasser, H. P. Mock, and A. B rner, 2008: Molecular and phytochemical studies of *Papaver somniferum* in the context of infraspecific classification. *Acta. Hort.* **799**, 81–88.
- Doyle, J. J., and J. E. Doyle, 1990: Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15.
- Ellegren, H., 2004: Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**, 435–445.
- Frary, A., Y. Xu, J. Liu, S. Mitchell, E. Tedeschi, and S. D. Tanksley, 2005: Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. *Theor. Appl. Genet.* **111**, 291–312.
- Gao, L. F., J. F. Tang, H. W. Li, and J. Z. Jia, 2003: Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* **12**, 245–261.

- Gupta, P. K., and R. K. Varshney, 2000: The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163–185.
- Hari, S., A. Sharma, and V. Sharma, 2009: Molecular characterization of opium poppy (*Papaver somiferum*) germplasm. *American J. Inf. Dis.* **2**, 155–160.
- Kalia, P. J., M. K. Rai, S. Kalia, R. Singh, and A. D. Dhawan, 2011: Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177**, 309–334.
- Kumpatla, S. P., and S. Mukhopadhyay, 2005: Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **48**, 985–998.
- Lavania, U. C., and S. Srivastava, 1999: Quantitative delineation of karyotype variation in *Papaver* as a measure of phylogenetic differentiation origin. *Curr. Sci.* **77**, 429–435.
- Lawson, M. J., and L. Zhang, 2006: Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* **7**, R14.
- Lee, E. J., I. K. Hwang, N. Y. Kim, K. L. Lee, M. S. Han, Y. H. Lee, M. Y. Kim, and M. S. Yang, 2010: An assessment of the utility of universal and specific genetic markers for opium poppy identification. *J. Forensic Sci.* **55**, 1002–1008.
- Lee, E. J., G. N. Jin, K. L. Lee, S. M. Han, Y. H. Lee, and M. S. Yang, 2011: Exploiting expressed sequence tag databases for the development and characterization of gene-derived simple sequence repeat markers in the opium poppy (*Papaver somniferum* L.) for forensic applications. *J. Forensic Sci.* **56**, 1131–1135.
- Li, Y. C., A. B. Korol, T. Fahima, A. Beiles, and E. Nevo, 2002: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* **11**, 2453–2465.
- Li, Y. C., A. B. Korol, T. Fahima, and E. Nevo, 2004: Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* **21**, 991–1007.
- Marcotte, E. M., M. Pellegrini, T. O. Yeates, and D. Eisenberg, 1999: A census of protein repeats. *J. Mol. Biol.* **293**, 151.
- Morgante, M., M. Hanafey, and W. Powell, 2002: Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* **3**, 194–200.
- Morin, R. D., M. Bainbridge, A. Fejes, A. Hirst, M. Krzywinski, T. J. Pugh, H. McDonald, R. Varhol, S. J. M. Jones, and M. A. Marra, 2008: Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**, 81–94.
- Mun, J. H., D. J. Kim, H. K. Choi, J. Gish, F. DeBelle, J. Mudge, R. Denny, G. Endre, O. Saurat, A. M. Dudez, G. B. Kiss, B. Roe, N. D. Young, and D. R. Cook, 2006: Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics* **172**, 2541–2555.
- Nergiz, C., and S. Qtl, 1994: The proximate composition and some minor constituents of poppy seeds. *J. Sci. Food Agric.* **66**, 17–120.
- Parmaksiz, I., and S. Özcan, 2011: Morphological, chemical, and molecular analyses of Turkish *Papaver* accessions (Sect. *Oxytona*). *Turk. J. Bot.* **35**, 1–16.
- Perrier, X., and J. P. Jacquemoud-Collet, 2006: DARwin software. Available at: <http://darwin.cirad.fr/darwin>.
- Poncet, V., M. Rondeau, C. Tranchant, A. Cayrel, S. Hamon, A. Kochko, and P. Hamon, 2006: SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol. Gen. Genomics* **276**, 436–449.
- Rajendrakumar, P., A. K. Biswal, S. M. Balachandran, K. Srinivasarao, and R. M. Sundaram, 2007: Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. *Bioinformatics* **23**, 1–4.
- Roldán-Ruiz, I., J. Dendauw, E. Van Bockstaele, A. Depicker, and M. De Loose, 2000: AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol. Breed.* **6**, 125–134.
- Saunders, J. A., M. J. Pedroni, L. D. J. Penrose, and A. J. Fist, 2001: AFLP Analysis of opium poppy. *Crop Sci.* **41**, 1596–160.
- Schulz, H., M. Baranska, R. Quilitzsch, and W. Schütze, 2004: Determination of alkaloids in capsules, milk and ethanolic extracts of poppy (*Papaver somniferum* L.) by ATR-FT-IR and FT-Raman spectroscopy. *Analyst* **129**, 917–920.
- Scott, K. D., P. Eggler, G. Seaton, M. Rosetto, E. M. Ablett, L. S. Lee, and J. R. Henry, 2000: Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* **100**, 723–726.
- Soil Product Office, 2009: Opium poppy report. Turkish Soil Production Office, Ankara.
- Sorrells, M. E., and W. A. Wilson, 1997: Direct classification and selection of superior alleles for crop improvement. *Crop Sci.* **37**, 691–697.
- Straka, P., and T. Nothnagel, 2002: A genetic map of *Papaver somniferum* L. based on molecular and morphological markers. *J. Herbs Spices Med. Plants* **9**, 235–241.
- Temnykh, S., W. D. Park, N. Ayres, S. Cartinhour, S. N. Hauck, L. Lipovich, Y. G. Cho, T. Ishii, and S. R. McCouch, 2000: Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**, 697–712.
- Varshney, R. K., A. Graner, and M. E. Sorrells, 2005a: Genic microsatellite markers in plants: features and applications. *Trends Biotech.* **23**, 48–55.
- Varshney, R. K., R. Sigmund, A. Borner, V. Korzun, N. Stein, M. E. Sorrells, P. Langridge, and A. Graner, 2005b: Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat rye and rice. *Plant Sci.* **168**, 195–202.
- Wang, M. L., N. A. Barkley, and T. M. Jenkins, 2009: Microsatellite markers in plants and insects. Part I. Applications of biotechnology. *Genes Genomes Genomics* **3**, 54–67.
- Ye, K., Y. Ken, N. Keshava, J. Shanks, J. A. Kapp, R. R. Tekmal, J. Petros, and H. C. Joshi, 1998: Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induces apoptosis in dividing cells. *Proc. Natl Acad. Sci.* **95**, 1601–1606.
- Yi, G., J. M. Lee, S. Lee, D. Choi, and B. D. Kim, 2006: Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theor. Appl. Genet.* **114**, 113–130.
- You, F. M., N. Huo, Y. Q. Gu, M. C. Luo, Y. Ma, D. Hane, G. R. Lazo, J. Dvorak, and O. D. Anderson, 2008: BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9**, 253.
- Yu, J. K., T. M. Dake, S. Singh, D. Bensch, W. Li, B. Gill, and M. E. Sorrells, 2004: Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome* **47**, 805–818.
- Zane, L., L. Bargelloni, and T. Patarnello, 2002: Strategies for microsatellite isolation: a review. *Mol. Ecol.* **11**, 1–16.
- Zhang, Z., S. Schwartz, L. Wagner, and W. Miller, 2000: A greedy algorithm for aligning DNA sequences. *J. Comp. Biol.* **7**, 203–214.