

Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing

Ibrahim Celik · Visam Gultekin · Jens Allmer · Sami Doganlar · Anne Frary

Received: 10 July 2013 / Accepted: 23 January 2014 / Published online: 6 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Opium poppy (*Papaver somniferum* L.) is an important pharmaceutical crop with very few genetic marker resources. To expand these resources, we sequenced genomic DNA using pyrosequencing technology and examined the DNA sequences for simple sequence repeats (SSRs). A total of 1,244,412 sequence reads were obtained covering 474 Mb. Approximately half of the reads (52 %) were assembled into 166,724 contigs representing 105 Mb of the opium poppy genome. A total of 23,283 non-redundant SSRs were identified in 18,944 contigs (11.3 % of total contigs). Trinucleotide and tetranucleotide repeats were the most abundant SSR repeats, accounting for 49.0 and 27.9 % of all SSRs, respectively. The AAG/TTC repeat was the most abundant trinucleotide repeat, representing 19.7 % of trinucleotide repeats. Other SSR repeat types were AT-rich. A total of 23,126 primer pairs (98.7 % of total SSRs) were designed to amplify SSRs. Fifty-three genomic SSR markers were tested in 37 opium poppy accessions and seven *Papaver* species for determination of polymorphism and transferability. Intraspecific

polymorphism information content (PIC) values of the genomic SSR markers were intermediate, with an average 0.17, while the interspecific average PIC value was slightly higher, 0.19. All markers showed at least 88 % transferability among related species. This study increases sequence coverage of the opium poppy genome by sevenfold and the number of opium poppy-specific SSR markers by sixfold. This is the first report of the development of genomic SSR markers in opium poppy, and the genomic SSR markers developed in this study will be useful in diversity, identification, mapping and breeding studies in opium poppy.

Keywords GS-FLX pyrosequencing technology · Genetic diversity · Genomic SSRs · Microsatellites · *Papaver somniferum* L.

Introduction

Opium poppy (*Papaver somniferum* L.) produces several pharmaceutically important alkaloids and is an economically important member of the family Papaveraceae (Acharya and Sharma 2009). Benzylisoquinoline alkaloids are extracted from opium poppy and have extensive medicinal properties, including analgesic and narcotic (morphine), antitumor (noscapine), antitussive (codeine) and muscle relaxant (papaverine) effects (Facchini and De Luca 2008; Ziegler et al. 2009; Winzer et al. 2012). In addition, poppy seeds and their oil are edible (Schulz et al.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-014-0036-0) contains supplementary material, which is available to authorized users.

I. Celik · V. Gultekin · J. Allmer · S. Doganlar · A. Frary (✉)
Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Izmir 35430, Turkey
e-mail: annefrary@iyte.edu.tr

2004). Although Turkey currently ranks first in the world in terms of surface area of cultivated opium poppy fields with 49,000 hectares (48 % of the world total), the country ranks second in total world morphine production, with 150 tons (18 % of the world total), after Australia (23 %), due to the low morphine content of Turkish poppy (Turkish Soil Product Office 2009). An important contributing factor to the use of low-morphine cultivars is a lack of poppy-specific molecular tools for characterization of poppy germplasm and more efficient breeding (Gumuscu and Arslan 2008).

To date, molecular characterization of opium poppy has mainly involved the use of non-specific markers, such as amplified fragment length polymorphism (AFLP) (Saunders et al. 2001; Dittbrenner et al. 2008), random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (Acharya and Sharma 2009; Parmaksiz and Ozcan 2011) markers. The most comprehensive analysis of genetic diversity of opium poppy was done by Dittbrenner et al. (2008), who analyzed the genetic diversity of 300 accessions from the opium poppy world collection using AFLP markers. These authors also quantified major alkaloids (morphine, codeine, thebaine, papaverine and noscapine) in the accessions by high performance liquid chromatography. Straka and Nothnagel (2002) constructed the only genetic linkage map currently available for opium poppy using 77 AFLP and 48 RAPD markers. Thus, it is clear that opium poppy-specific markers would be a valuable contribution to diversity and mapping studies.

Simple sequence repeat (SSR) or microsatellite markers are iterations of one to six nucleotide motifs and are dispersed throughout the genome (Varshney et al. 2005; Jones et al. 2009). Although genic SSR markers are gene targeted and more conserved than genomic SSR markers, they are reported to have lower copy number and lower polymorphism information content (PIC) than genomic SSR markers (Li et al. 2002; Ellis and Burke 2007). Opium poppy-specific markers have recently been developed using the publicly available expressed sequence tag (EST) database [National Center for Biotechnology Information (NCBI), Bethesda, MD] to identify and design SSR primers (Lee et al. 2011; Selale et al. 2013). Lee et al. (2011) developed 22 opium poppy-specific genic SSR markers but tested only six of these for molecular

characterization of opium poppy. Selale et al. (2013) tested 67 EST–SSR primer pairs on Turkish opium poppy accessions, landraces and cultivars. In these two latter studies, the markers were validated for use in forensic identification of opium poppy (Lee et al. 2011) and genetic diversity analysis (Selale et al. 2013). In both cases, opium poppy-specific genic SSR markers revealed low genetic diversity. Given the low polymorphism and limited number of the poppy-specific genic SSR markers currently available, the aim of our study was to develop opium poppy-specific genomic SSR markers with pyrosequencing technology. Pyrosequencing is a low-cost, high-throughput sequencing method which can expedite SSR discovery (Abdelkrim et al. 2009). As part of this goal, we tested a subset of the SSR markers for polymorphism and transferability in opium poppy and related species. This validation of the markers was important to determine their usefulness for fingerprinting, diversity, mapping and breeding studies in opium poppy and related species.

Materials and methods

Plant materials and DNA isolation

A total of 37 opium poppy accessions from Turkey and seven other *Papaver* species were used as plant material (Table 1). Eight opium poppy accessions were obtained from the Turkish Soil Product Office (TMO) and 29 accessions were obtained from the Anatolia Agricultural Research Institute (AARI), Eskisehir, Turkey. The related species were used to test transferability of the markers and included *Papaver orientale* (Iran), *P. pseudoorientale* (Iran), *P. bracteatum* (Iran), *P. rhoeas* (Bulgaria), *P. umbonatum* (Turkey), *P. nudicaule* (Mongolia) and *P. armeniacum* (Armenia). These accessions were obtained from the U.S. Department of Agriculture–Agricultural Research Service Plant Germplasm Inspection Station, Beltsville MD, USA. For genomic DNA isolation, each accession was planted in seedling plates. Plants were grown in the greenhouse (24–25 °C, approximately 33 % humidity). Total genomic DNA was isolated from leaf tissue bulked from ten plants per accession using a CTAB method (Doyle and Doyle 1990).

Table 1 Opium poppy accessions and *Papaver* species used in the study

Name	Source	Location	Landrace/cultivar/species
1290	AARI	–	Cultivar
1061	AARI	–	Cultivar
1259	AARI	–	Cultivar
1065	AARI	–	Cultivar
Kemerkaya	AARI	–	Cultivar
Tınaztepe	AARI	–	Cultivar
Zaferyolu	AARI	–	Cultivar
Anayurt	AARI	–	Cultivar
Afyon95	AARI	–	Cultivar
Ofis3	TMO	–	Cultivar
Ofis4	TMO	–	Cultivar
Ofis8	TMO	–	Cultivar
Ofis95	TMO	–	Cultivar
TM01	TMO	–	Cultivar
TM02	TMO	–	Cultivar
TM03	TMO	–	Cultivar
TM04	TMO	–	Cultivar
7	AARI	Sandıklı, Alagöz, Afyon	Landrace
10	AARI	Çeltik, Burdur	Landrace
14	AARI	Sivas	Landrace
15	AARI	Koçyatağı, Şuhut, Afyon	Landrace
19	AARI	Dişli, Afyon	Landrace
22	AARI	Anayurt, Şuhut, Afyon	Landrace
32	AARI	Afyon	Landrace
33	AARI	Ekinhisar, Sandıklı, Afyon	Landrace
37	AARI	Höyükli, Yalvaç, Isparta	Landrace
45	AARI	Simav, Kütahya	Landrace
59	AARI	Sülümenli, Afyon	Landrace
60	AARI	Koçyatağı, Afyon	Landrace
61	AARI	Şuhut, Afyon	Landrace
67	AARI	Alacami, Sandıklı, Afyon	Landrace
76	AARI	Göğen, Uşak	Landrace
89	AARI	Güre, Uşak	Landrace
92	AARI	Bolvadin, Afyon	Landrace
95	AARI	Acipayam, Denizli	Landrace
96	AARI	Kütahya	Landrace
103	AARI	Çay, Afyon	Landrace
PI 229617	AARI	Iran	<i>P. orientale</i>
PI 381612	AARI	Iran	<i>P. pseudoorientale</i>
PI 414784	AARI	Iran	<i>P. bracteatum</i>
W6 10919	AARI	Bulgaria	<i>P. rhoeas</i>
W6 11444	AARI	Turkey	<i>P. umbonatum</i>
W6 18131	AARI	Mongolia	<i>P. nudicaule</i>
W6 23866	AARI	Armenia	<i>P. armeniacum</i>

TMO Turkish Soil Product Office, AARI Anatolia Agricultural Research Institute

DNA sequencing

For sequencing, total genomic DNA of *P. somniferum* cv. 'Kemer kaya' was extracted using the Wizard Magnetic 96 Plant System (Promega Corp., Madison, WI) with the Beckman Coulter Biomek NX Workstation according to the manufacturer's instructions. Pyrosequencing was done with a Roche 454 GS-FLX sequencer and was performed by 454 Lifesciences Corp. (Branford, CT).

Data acquisition and pre-processing

The adapter and linker sequences were removed to facilitate genome assembly. Following this, the Standard Flowgram Format (SFF) data was converted to separate FASTA (Lipman and Pearson 1985) and quality files because most assembly tools cannot directly work on SFF files (Kumar and Blaxter 2010). The conversion was performed using an open source package of tools written in Python language, which is available at http://bioinf.comav.upv.es/seq_crumbs/download.html. We used the seq_crumbs tool from the package to perform the conversion with the default settings. The resulting FASTA and FASTQ format files were suitable for sequence assembly.

Sequence assembly

MIRA, a whole genome shotgun and EST sequence assembler (Chevreux et al. 2004), was used for sequence assembly because it allowed us to customize the assembly process in great detail and led to the best assembly among more than 100 trials with MIRA and four other assembly tools (Gultekin and Allmer, in preparation). Assembly quality was based on various parameters, such as the weighted median of contig lengths (N50), a commonly used measure. The most successful assembly by MIRA used non-default parameters which are described in Electronic Supplementary Material (ESM) 1. The sequences will be provided in a future publication as an annotated draft genome (Gultekin et al. in preparation).

SSR detection and primer design

The analysis of SSRs of the contig assemblies was performed with our in-house tool SiSeeR (<http://bioinformatics.iyte.edu.tr/index.php?n=Softwares.Si>

[SeeRHelp](#)) (Gultekin and Allmer, in preparation). The minimum number of repeats needed to identify perfect SSRs were: ten mononucleotide, six dinucleotide, four trinucleotide, four tetranucleotide, three pentanucleotide and three hexanucleotide repeats. Primer design was performed with the Primer3 (Rozen and Skaletsky 2000) (<http://frodo.wi.mit.edu/>) console application. A total of 19,046 contig sequences yielding 23,427 SSR sequences were converted from FASTA to the default Primer3 input format Boulder-IO. The Primer3 settings, which differed from the default settings, are described in ESM 1. In order to produce primers flanking the SSR sequences, Primer3's SEQUENCE_TEMPLATE switch was enabled, and values for the start and end positions of each SSR were generated. SSR and primer design data are available at <http://plantmolgen.iyte.edu.tr/research/>.

SSR amplification

Amplification of the opium poppy DNA with genomic SSR primers was carried out in 25- μ l reaction mixtures containing 1 \times PCR buffer, 3 mM MgCl₂, 0.125 mM deoxyribonucleotide triphosphates (dNTPs), 1 U Taq Polymerase, 2 pmol forward and reverse primers and 80 ng template DNA. The PCR cycling profile consisted of one cycle of 5 min at 94 °C, followed by 35 cycles of 45 s at 94 °C, 1 min at 55 °C (annealing) and 1 min at 72 °C, with a final extension step of 5 min at 72 °C. To prepare the PCR product for analysis by capillary electrophoresis, 3 μ l of the PCR product was added to 27 μ l of sample loading buffer (Beckman Coulter, Brea, CA). A size standard (0.5 μ l, 600 bp; (Beckman Coulter) was used per reaction in all runs. The mixture for each accession was then run on a Beckman CEQ8800 capillary electrophoresis device using the frag3 method (capillary temperature 50 °C, denaturation 90 °C for 120 s, injection voltage 2.0 kV for 3 s, separation voltage 4.8 kV for 60 min).

Data analysis

Amplified SSR loci were scored as present (1) or absent (0). Rare PCR fragments (<4 occurrences) were excluded from analysis because they might be unreliable. The binary data were used to calculate the PIC value for each marker fragment based on the formula: $PIC_i = 2f_i(1 - f_i)$ where f_i is the frequency of band presence (Roldan-Ruiz et al. 2000). A

dissimilarity matrix generated using the Dice coefficient was used to construct a dendrogram with the unweighted neighbor joining method using the Darwin computer program (Perrier and Jacquemoud-Collet 2006). The correlation of the dissimilarity matrix and the dendrogram was calculated using a Mantel test.

Results

Sequence assembly, SSR identification and primer design

Pyrosequencing resulted in 1,244,412 sequence reads covering more than 695 Mb (Table 2). After removal of the linkers and adapters, nearly 475 Mb remained, with an average sequence read length of 380.4 nucleotides (nt). A total of 649,267 reads representing 52 % of all sequences were assembled into 166,724 contigs encompassing 105 Mb of the opium poppy genome. The weighted median contig length (N50) of the assembly was 913 nt. Only contigs were used for SSR detection because singlet reads were shorter and, therefore, would not be as useful for designing effective primers.

A total of 23,283 non-redundant SSRs were identified in 18,944 contigs (11.3 % of total contigs). Approximately 16 % (3,135) of contigs contained more than one SSR (ESM 2). A maximum of ten and 11 different SSRs were identified in one contig each. SSR length ranged from 3 to 226 nt, with an average length of 13.4 ± 0.03 nt. The most abundant repeat types were trinucleotide SSRs, which accounted for 49.0 % of all SSRs (Table 3), followed by tetranucleotide repeats (27.9 %). The remaining repeat types each accounted for <10 % of the SSRs. Some motifs were more common than others (Table 4). The majority (82.2 %) of mononucleotide repeats consisted of A/T repeats. AT/TA was the most abundant

Table 3 Simple sequence repeat types in the opium poppy contig sequences

Motif length	Number of SSRs	Frequency (%)
Mononucleotide	842	3.6
Dinucleotide	2,114	9.1
Trinucleotide	11,406	49.0
Tetranucleotide	6,493	27.9
Pentanucleotide	1,600	6.9
Hexanucleotide	828	3.6
Total	23,283	100

SSR Simple sequence repeat

dinucleotide repeat motif and accounted for 50.4 % of these repeats. AAG/TTC was the most abundant (19.7 %) trinucleotide repeat. AT-rich repeats were also most common repeats in tetra-, penta- and hexanucleotide SSRs. Primers were designed for 23,126 SSRs; only 1.3 % of the SSR sequences were unsuitable for primer design.

Polymorphism, validation and transferability of genomic SSR markers

A total of 100 genomic SSR primer pairs were first tested on five opium poppy accessions (1290, 1061, 1259, 1065, cv. ‘Kemerkaya’). Of these primers, 96 (96 %) amplified products. A total of 53 genomic SSR markers which showed clear amplification following electrophoresis on agarose gel (ESM 3) were then tested in the 37 opium poppy accessions and seven *Papaver* species for determination of polymorphism and transferability (Table 1). Seventeen of the opium poppy accessions were named varieties while the others were landraces collected in Turkey. The 53 SSR primers generated 209 polymorphic fragments in all accessions and 90 polymorphic fragments in *P. somniferum* accessions (Table 5). The average

Table 2 Sequence preprocessing and assembly statistics

Parameter	Raw sequences	Cleaned sequences	Contigs
Total no. sequences	1,244,412	1,244,412	166,724
Minimum sequence length (nt)	52	40	40
Maximum sequence length (nt)	1,201	764	42,888
Average sequence length (nt)	558.7 ± 66	380.4 ± 142.5	731.0 ± 376.8
Total no. bases	695,339,138	474,398,321	105,303,268

Table 4 Most abundant SSR motif types in contigs

SSR motif	Number of SSRs	Percentage of SSR motifs
A/T	692	82.2
C/G	150	17.8
AT/TA	1,066	50.4
AG/TC	383	18.1
AC/TG	340	16.1
CT/GA	318	15.0
AAG/TTC	2,243	19.7
ACC/TGG	1,381	12.1
AAC/TTG	1,228	10.8
AAAT/TTTA	1,101	17.0
AAATA/TTTAT	300	18.8
AAAAAT/TTTTTA	88	10.6

Only motifs comprising more than 10 % of each repeat type are listed

number of amplified fragments per genomic SSR marker was 5 ± 0.01 , with a range of 1–13 fragments. A total of 48 SSR primers (95 %) were polymorphic in all accessions, with an average fragment polymorphism of 84 %. For all accessions, average PIC values ranged from 0.05 for psgSSR076 to 0.47 for psgSSR022, with an average PIC of 0.19. Fewer (60.4 %) SSR primers showed intraspecific polymorphism in *P. somniferum*, with an average fragment polymorphism of 63 %; the average intraspecific PIC value decreased to 0.17. Intraspecific PIC values ranged from 0.05 for five different markers to 0.49 for psgSSR22. In all analyses there was no significant correlation between PIC values and SSR motif types or lengths.

To ensure that the expected SSRs were amplified by the primers, we sequenced seven PCR products from cv. ‘Kemer kaya’. Only one of the products contained an unexpected SSR. Although these results indicate imperfect validation of the identified SSRs, the fact that the amplification product contained a polymorphic SSR is sufficient for use of this particular primer pair as a molecular marker.

Transferability of the 53 genomic SSR markers was tested in seven *Papaver* species: *P. bracteatum*, *P. pseudoorientale*, *P. orientale*, *P. nudicaule*, *P. armeniacum*, *P. rhoeas* and *P. umbonatum*. A high rate of transferability was observed in these species (ESM 4).

All 53 SSR markers yielded PCR products in *P. pseudoorientale*, 52 (98 %) yielded PCR products in *P. bracteatum* and *P. nudicaule*, 51 (96 %) markers yielded PCR products in *P. orientale* and *armeniaticum*, 49 (92 %) SSR markers yielded PCR products in *P. umbonatum*, and 47 (88.7 %) SSR markers yielded PCR products in *P. rhoeas*.

Genetic diversity analysis with genomic SSR markers in opium poppy

Low-frequency fragments (observed in <10 % of poppy accessions) were excluded from all analyses because these low-frequency fragments can be unreliable. A total of 209 high-quality, reproducible polymorphic fragments were used for our diversity analysis of opium poppy and related species. The binary presence/absence data were used to generate a distance matrix using the Dice coefficient to draw a dendrogram employing the unweighted neighbor-joining algorithm. As expected, a Mantel test showed a strong correlation ($r = 0.998$) between the distance matrix and dendrogram. Dissimilarity between accessions ranged from 0.008 to 0.48 (52 % similarity), with average dissimilarity of 0.14 (Fig. 1). Dissimilarity between opium poppy and related species ranged from 0.23 to 0.48. As expected, *Papaver* species clustered separately from *P. somniferum* accessions. The *P. somniferum* accessions fell into three clusters. Cluster 1 contained 13 landraces and 13 varieties/breeding lines, and the dissimilarity ranged from 0.01 to 0.08, with an average dissimilarity of 0.04. Cluster 1 had six subclusters (subclusters A–F), with cluster 1B containing only named varieties and cluster 1C containing only landraces of *P. somniferum*; varieties and landraces were intermixed in the other subclusters. Cluster 2 contained eight opium poppy accessions (four landraces and four varieties) and had an average dissimilarity of 0.06, with minimum and maximum dissimilarities of 0.03 and 0.16, respectively. Cluster 3 comprised three opium poppy landraces (59, 22 and 76) which were the most genetically distinct opium poppy accessions in the study. Clustering of the *Papaver* species did not match that obtained with the internal transcribed spacer nuclear ribosomal DNA and plastid trnL sequences (Carolan et al. 2006).

Table 5 Polymorphism information content of genomic markers

SSR marker	Repeat motif	All accessions		<i>P. somniferum</i>	
		Number of polymorphic fragments/total no. fragments (%)	PIC ^a	Number of polymorphic fragments/total no. fragments (%)	PIC ^a
psgSSR002 ^b	(ATG/TAC) ₄	5/7 (71)	0.12 ± 0.02	1/7 (14)	0.06
psgSSR005	(CATCTG/GTAGC) ₃	2/2 (100)	0.21 ± 0.01	2/2 (100)	0.21 ± 0.02
psgSSR006	(AACA/TTGT) ₃	4/5 (80)	0.31 ± 0.06	4/5 (80)	0.25 ± 0.09
psgSSR008 ^b	(AAG/TTC) ₈	8/8 (100)	0.28 ± 0.05	6/8 (75)	0.20 ± 0.08
psgSSR013 ^b	(ATA/TAT) ₄	13/13 (100)	0.18 ± 0.01	1/8 (13)	0.05
psgSSR015	(CAG/GTC) ₄	5/6 (83)	0.17 ± 0.03	0/5 (0)	0
psgSSR021	(TA/AT) ₆	3/3 (100)	0.15 ± 0.09	3/3 (100)	0.15 ± 0.10
psgSSR022	(TGG/ACC) ₄	1/3 (33)	0.47	1/2 (50)	0.49
psgSSR023 ^b	(TGTC/ACAGT) ₃	7/7 (100)	0.29 ± 0.05	5/7 (71)	0.22 ± 0.09
psgSSR024	(TTC/AAG) ₆	9/9 (100)	0.16 ± 0.03	3/5 (60)	0.1 ± 0.03
psgSSR027	(TA/AT) ₆	1/3 (33)	0.20	0/2 (0)	0
psgSSR029	(TCAT/AGTA) ₃	0/2 (0)	0	0/2 (0)	0
psgSSR030	(AACA/TTGT) ₃	1/1 (100)	0.13	1/1 (100)	0.10
psgSSR033	(CAAA/GTTT) ₃	3/4 (75)	0.22 ± 0.02	1/3 (33)	0.10
psgSSR034	(TGG/ACC) ₄	4/5 (80)	0.21 ± 0.09	2/3 (67)	0.27 ± 0.22
psgSSR036	(CCAA/GGTT) ₃	1/3 (33)	0.20	0/6 (0)	0
psgSSR037 ^b	(GAA/CTT) ₁₀	9/9 (100)	0.20 ± 0.06	5/9 (56)	0.25 ± 0.08
psgSSR038	(TGAT/ACTA) ₃	6/7 (86)	0.24 ± 0.04	2/4 (50)	0.23 ± 0
psgSSR039 ^b	(ACAAC/TGTTG) ₄	11/11 (100)	0.21 ± 0.02	5/10 (50)	0.18 ± 0.01
psgSSR040	(TGT/ACA) ₄	3/4 (75)	0.18 ± 0.03	0/3 (0)	0
psgSSR041 ^b	(TCTTA/AGAAT) ₃	10/10 (100)	0.29 ± 0.03	10/10 (100)	0.13 ± 0.04
psgSSR042	(TTCA/AAGT) ₄	5/5 (100)	0.16 ± 0.03	0/4 (0)	0
psgSSR046	(TGAT/ACTA) ₃	3/4 (75)	0.17 ± 0.07	2/4 (50)	0.19 ± 0
psgSSR047	(AGA/TCT) ₄	5/5 (100)	0.19 ± 0.08	2/5 (40)	0.30 ± 0.20
psgSSR050	(GAA/CTT) ₄	0/3 (0)	0	0/3 (0)	0
psgSSR053	(TA/AT) ₇	2/5 (40)	0.12 ± 0.08	1/5 (20)	0.05
psgSSR054	(TCGT/AGCA) ₃	4/4 (100)	0.13 ± 0.02	0/2 (0)	0
psgSSR055	(TC/AG) ₇	3/5 (60)	0.13 ± 0.05	0/4 (0)	0
psgSSR058	(AATA/TTAT) ₃	1/2 (50)	0.13	0/2 (0)	0
psgSSR059	AAAT/TTTA) ₃	5/6 (83)	0.20 ± 0.02	0/1 (0)	0
psgSSR060	(TCT/AGA) ₄	2/3 (67)	0.30 ± 0.17	1/2 (50)	0.44
psgSSR061	(AAAT/TTTA) ₃	3/3 (100)	0.27 ± 0.02	3/3 (100)	0.1 ± 0.05
psgSSR062	(AGAC/TCTG) ₃	2/3 (67)	0.07 ± 0.02	0/3 (0)	0
psgSSR064	(ATA/TAT) ₄	1/1 (100)	0.20	1/1 (100)	0.05
psgSSR067 ^b	(TTCT/AAGA) ₃	9/9 (100)	0.23 ± 0.02	3/7 (43)	0.19 ± 0.07
psgSSR068	(TTCT/AAGA) ₃	0/1 (0)	0	0/1 (0)	0
psgSSR069	(CAAT/GTTA) ₄	5/6 (83)	0.20 ± 0.02	0/5 (0)	0
psgSSR070	(ATT/TAA) ₅	2/2 (100)	0.22 ± 0.05	0/1 (0)	0
psgSSR071	(TATTC/ATAAG)	2/2 (100)	0.13 ± 0.00	0/2 (0)	0
psgSSR074 ^b	(GTTT/CAAA) ₃	12/12 (100)	0.22 ± 0.03	11/11 (100)	0.19 ± 0.03
psgSSR076	(AATG/TTAC) ₃	1/1 (100)	0.05	1/1 (100)	0.05

Table 5 continued

SSR marker	Repeat motif	All accessions		<i>P. somniferum</i>	
		Number of polymorphic fragments/total no. fragments (%)	PIC ^a	Number of polymorphic fragments/total no. fragments (%)	PIC ^a
psgSSR077	(ATC/TAG) ₄	5/5 (100)	0.14 ± 0.04	1/4 (25)	0.05
psgSSR078	(GTATT/CATAA) ₃	3/4 (75)	0.15 ± 0.02	1/3 (33)	0.15
psgSSR079	(GGAA/CCTT) ₃	0/2 (0)	0	0/2 (0)	0
psgSSR080	(GGAA/CCTT) ₃	6/6 (100)	0.29 ± 0.05	4/4 (100)	0.29 ± 0.08
psgSSR082	(AAT/TTA) ₄	1/2 (50)	0.20	0/2 (0)	0
psgSSR085	(ATTT/TAAA) ₃	5/5 (100)	0.18 ± 0.04	1/4 (25)	0.1
psgSSR90	(TGT/ACA) ₄	5/6 (83)	0.16 ± 0.04	0/2	0
psgSSR093	(AAT/TTA) ₄	2/2 (100)	0.07 ± 0.02	2/2 (100)	0.05 ± 0
psgSSR094	(GCA/CGT) ₄	3/3 (100)	0.28 ± 0.03	3/3 (100)	0.15 ± 0.04
psgSSR098	(TGTTG/ACAAC) ₃	0/1 (0)	0	0/1 (0)	0
psgSSR099	(ATC/TAG) ₄	2/4 (50)	0.22 ± 0.05	0/2 (0)	0
psgSSR100	(GTG/CAC) ₄	4/5 (80)	0.14 ± 0.04	1/4 (25)	0.15

^a Polymorphic information content (PIC) is presented as the average ± standard error

^b Markers that amplified more than one locus in *P. somniferum*

Discussion

Simple sequence repeat markers are frequently used in plant genetic studies because they are easy to amplify, highly reproducible, polymorphic and often multiallelic (Varshney et al. 2005). In the past, detection of genomic SSRs and subsequent conversion to markers was expensive and time-consuming, involving the construction and screening of genomic DNA libraries followed by the sequencing of candidate clones (Zalapa et al. 2012). The advent of next-generation sequencing technologies, such as pyrosequencing, has redefined this process. Much of the work is now performed in silico with wet laboratory experiments confined to sequencing and SSR validation. As a result, many more SSR markers can be identified quickly and at a lower cost. This approach is especially promising for minor crop, tree and weed species that have been often ignored in the area of molecular marker development [e.g. cranberry (Zhu et al. 2012), black alder (Lepais and Bacles 2011) and waterhemp (Lee et al. 2009)].

Distribution of SSRs in the opium poppy genome

We used high-throughput pyrosequencing to develop genomic SSR markers in opium poppy. The assembly

of the relatively long sequences (average length 731 nt in assembly) resulted in more than 160,000 contigs covering 105 Mb of the opium poppy genome. The contigs provided 2.83 % coverage of the opium poppy genome, which has been reported to contain 3,724 Mb of DNA (Bennett and Smith 1976). As expected, this is much higher than the coverage obtained from EST unigene assembly (0.4 %; Selale et al. 2013) and provided a sevenfold increased coverage of the opium poppy genome. Average density of genomic SSRs was one SSR every 4.5 kb of genomic DNA, which is in the range expected for plant species (Cavagnaro et al. 2010). Genomic SSRs occurred less frequently than non-redundant genic SSRs (every 3.6 kb in EST sequences; Selale et al. 2013) in the opium poppy genome. This difference is in concurrence with the results of Morgante et al. (2002) and of the same magnitude (1.25-fold) as the difference in frequency observed in rice, soybean and sorghum (1.2- to 2-fold, Cavagnaro et al. 2010).

In our study, trinucleotides were the most prevalent SSR type in opium poppy genomic DNA, accounting for nearly half (48.7 %) of all SSRs identified. Trinucleotide repeats are also the most frequently identified SSR type in *Arabidopsis*, rice, soybean and sorghum genomic DNA (Cavagnaro et al. 2010). The



Fig. 1 Unweighted neighbor-joining dendrogram of the poppy accessions constructed by genomic simple sequence repeat markers

frequency distribution of the other SSR types in opium poppy also matched that of other species, with the frequency being tetranucleotide > dinucleotide > pentanucleotide (Cavagnaro et al. 2010). The frequencies of SSRs in opium poppy genomic DNA were similar to those observed in EST sequences, with the exception of mononucleotide repeats which ranked fifth in genomic DNA and third in genic DNA (Selale et al. 2013). The abundance of trinucleotides in genic DNA is hypothesized to be the result of purifying selection which eliminates any SSRs causing frameshift mutations. However, it is unknown if selection is involved in the distribution of SSR types in genomic DNA.

Among the different motif types, AT-rich motifs were often the most common. This has also been observed in genomic DNA of other dicot plant species (Cavagnaro et al. 2010) as well as in opium poppy genic DNA (Selale et al. 2013). AAG/TTC was the most common trinucleotide motif, which is in agreement with results reported for other dicots, including cucumber, soybean, *Arabidopsis* and grape (Cavagnaro et al. 2010). AAG/TTC was also the most frequent trinucleotide

detected by Selale et al. (2013) in their study of opium poppy genic sequences and was described as the most frequent genic trinucleotide in plants by Li et al. (2004).

The similarities between genomic and genic SSR types and motifs in opium poppy may, in part, be due to the presence of coding sequences in the genomic DNA contig assembly. In future studies, the contig sequences could be annotated to determine the percentage of protein-coding DNA and to determine any redundancy in the SSRs identified in our two studies (i.e. present study and Selale et al. 2013).

Polymorphism and transferability of genomic SSR markers

Most of the SSR primers (96 %) amplified PCR products. The amplification success of the genomic SSR markers was slightly higher than that for the genic SSR markers (82 %) developed by Selale et al. (2013). A high rate of successful amplification can be due to high-quality sequence data and the appropriate primer parameters, such as high GC content. In our study, the

genomic SSR markers detected an intermediate level of polymorphism, with an average PIC value of 0.19 among the *Papaver* species and opium poppy accessions, and a slightly lower level of intraspecific polymorphism (average PIC 0.17). Polymorphism of the genomic SSR markers was lower than that of previously developed genic SSR markers (Lee et al. 2011; Selale et al. 2013). Although genomic SSRs are often reported to have higher levels of polymorphism than genic SSRs (Varshney et al. 2005), Tian et al. (2012) recently showed that genic SSR markers were more polymorphic than genomic SSR markers in *Coreoperca whiteheadi*.

Many of the genomic SSR markers in our study amplified multiple fragments (average of 5 fragments per marker). This is most likely the result of polyploidy in opium poppy. *Papaver somniferum* ($2n = 22$) is an aneuallopolyploid and is hypothesized to have originated from species with $x = 7$ (Lavania and Srivastava 1999). Therefore, a single copy SSR locus may amplify up to six fragments. In our study, nine SSR markers (17 %) had more than six fragments, indicating that these SSR markers originate from multiple loci. (These markers are labeled with an asterisk in Table 5). In our study, the number of fragments amplified by genomic SSR markers was slightly higher than that amplified by six genic SSR markers developed by Lee et al. (2011), who obtained an average of 2.8 ± 0.5 fragments (mean \pm standard error; range 2–5). However, in their study, Lee et al. (2011) did not use markers that produced more than three fragments, thereby limiting fragment number. These authors also reported that the average fragment number for the genomic SSRs was lower than that for the genic SSRs which amplified an average of 8.4 fragments per SSR (Selale et al. 2013). This result can be explained by the fact that many of the genic SSR markers (23 SSR markers) were multiallelic while only nine genomic SSRs were multiallelic in this study. Additional genic and genomic markers should be tested to determine if this difference is real or a sampling artifact.

High transferability of the genomic SSR markers to related *Papaver* species was observed, varying from 88.7 to 100 % depending on the species tested. A high transferability rate has also been reported for genic SSR markers in opium poppy (Selale et al. 2013). This high rate of transferability indicates that there are conserved regions among *Papaver* species. High transferability of both types of SSR markers is valuable for *Papaver*

species given the limited molecular tools available for this genus.

Genetic diversity analysis with genomic SSR markers

The genomic SSR markers developed in this study can be used in opium poppy identification. Although rare alleles were excluded, a total of 32 SSR primers (60.4 %) identified in our study were useful for Turkish opium poppy identification. Retesting of rare alleles is needed to confirm whether or not these are reproducible and suitable for opium poppy identification. Dendrogram analysis also showed that genomic SSR markers were suitable for differentiating opium poppy from *Papaver* species. Thus, these markers can be used to analyze intra- and interspecific genetic diversity of opium poppy. Landraces and named varieties were intermixed in the dendrogram of genomic SSR markers. This differs from the results reported by Selale et al. (2013) obtained with nearly the same accessions using genic SSR markers in which named varieties clustered separately from landraces. This difference may be the result of artificial selection pressure on genic SSRs. Chabane et al. (2005) reported that genic SSR markers provided clearer separation between wild and cultivated barley than genomic SSR markers, as was observed in our studies. Although the topology of the dendrogram for genomic SSR markers was different from the dendrogram based on genic SSR markers, Mantel test results showed that there was a very high correlation ($r = 0.98$) between the distance matrices for the two marker types. Therefore, genomic and genic SSR markers give consistent results in opium poppy.

In conclusion, pyrosequencing and mining of the opium poppy genome allowed identification of numerous SSR repeats which can be exploited for marker development. These markers are highly transferable within the genus *Papaver* and are an important addition to the repertoire of molecular tools available for genetic studies and breeding in opium poppy.

Acknowledgments The authors thank Dr. Hüseyin Camci, Dr. Arzu Köse and Dr. Ferda Kosar from the Anatolia Agricultural Research Institute for providing plant material. We are also grateful to Ali Üncü and Ayşe Dülger Üncü for SSR validation by sequencing. This study was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) project no. 1090797.

References

- Abdelkrim J, Robertson BC, Stanton JA, Gemmill NJ (2009) Fast cost-effective development of species-specific microsatellite markers by genomic sequencing. *Biotechniques* 46:185–192. doi:[10.2144/000113084](https://doi.org/10.2144/000113084)
- Acharya HS, Sharma V (2009) Molecular characterization of opium poppy (*Papaver somniferum*) germplasm. *Am J Infect Dis* 5:148–153
- Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B Biol Sci* 274:227–274
- Carolan JC, Hook ILI, Chase MW, Kadereit JW, Hodkinson TR (2006) Phylogenetics of *Papaver* and related genera based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. *Ann Bot* 98:141–155. doi:[10.1093/aob/mcl079](https://doi.org/10.1093/aob/mcl079)
- Cavagnaro PF, Douglas AS, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genom* 11:569. doi:[10.1186/1471-2164-11-569](https://doi.org/10.1186/1471-2164-11-569)
- Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Gen Resour Crop Evol* 52:903–909. doi:[10.1007/s10722-003-6112-7](https://doi.org/10.1007/s10722-003-6112-7)
- Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14:1147–1159. doi:[10.1101/gr.1917404](https://doi.org/10.1101/gr.1917404)
- Dittbrener A, Lohwasser U, Mock HP, Börner A (2008) V International symposium on the taxonomy of cultivated plants. In: Groendijk-Wilders N, Alexander C, Berg RGVD, Hetterscheid WLA (eds) Molecular and phytochemical studies of *Papaver somniferum* in the context of infraspecific classification. *ISHS Acta Horticulturae*, Wageningen, pp 81–88
- Doyle JJ, Doyle JE (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* 99:125–132. doi:[10.1038/sj.hdy.6801001](https://doi.org/10.1038/sj.hdy.6801001)
- Facchini PJ, De Luca V (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J Cell Mol Biol* 54:763–784. doi:[10.1111/j.1365-313X.2008.03438.x](https://doi.org/10.1111/j.1365-313X.2008.03438.x)
- Gumuscu A, Arslan N (2008) Researches on heterosis on yield and yield components of some poppy. *Tarım Bilimleri Dergisi* 14:365–373
- Jones N, Ougham H, Thomas H, Pasakinskiene I (2009) Markers and mapping revisited: finding your gene. *New Phytol* 183:935–966. doi:[10.1111/j.1469-8137.2009.02933.x](https://doi.org/10.1111/j.1469-8137.2009.02933.x)
- Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. *BMC Genom* 11:571. doi:[10.1186/1471-2164-11-571](https://doi.org/10.1186/1471-2164-11-571)
- Lavania UC, Srivastava S (1999) Quantitative delineation of karyotype variation in *Papaver* as an measure of phylogenetic differentiation origin. *Curr Sci* 77:429–435
- Lee RM, Thimmapuram J, Thinglum KA, Gong G, Hernandez AG, Wright CL, Kim RW, Mikel MA, Tranel PJ (2009) Sampling the waterhemp (*Amaranthus tuberculatus*) genome using pyrosequencing technology. *Weed Sci* 57:463–469. doi:[http://dx.doi.org/10.1614/WS-09-021.1](https://doi.org/http://dx.doi.org/10.1614/WS-09-021.1)
- Lee EJ, Jin GN, Lee KL, Han MS, Lee YH, Yang MS (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. doi:[10.1111/j.1556-4029.2011.01810.x](https://doi.org/10.1111/j.1556-4029.2011.01810.x)
- Lepais O, Bacles CFE (2011) De novo discovery and multiplexed amplification of microsatellite markers for black alder (*Alnus glutinosa*) and related species using SSR-enriched shotgun pyrosequencing. *J Hered* 102:627–632. doi:[10.1093/jhered/esr062](https://doi.org/10.1093/jhered/esr062)
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 11:2453–2465. doi:[10.1046/j.1365-294X.2002.01643.x](https://doi.org/10.1046/j.1365-294X.2002.01643.x)
- Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* 21:991–1007
- Lipman DJ, Pearson WR (1985) Rapid and sensitive protein similarity searches. *Science* 227:1435–1441. doi:[10.1126/science.2983426](https://doi.org/10.1126/science.2983426)
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30:194–200. doi:[10.1038/ng822](https://doi.org/10.1038/ng822)
- Parmaksiz I, Ozcan S (2011) Morphological, chemical and molecular analyses of Turkish *Papaver* accessions (Sect. *Oxytona*). *Turk J Bot* 35:1–16. doi:[10.3906/bot-1003-39](https://doi.org/10.3906/bot-1003-39)
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/darwin>
- Roldan-Ruiz I, Dendauw J, Bockstaele EV, Depicker A, Loose MD (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Mol Breed* 6:125–134. doi:[10.1023/A:1009680614564](https://doi.org/10.1023/A:1009680614564)
- Rozen S, Skaletsky HJ (2000) Bioinformatics methods and protocols: methods in molecular biology. In: Krawetz S, Misener S (eds) Primer3 on the WWW for general users and for biologist programmers. Humana Press, Totowa, pp 365–386
- Saunders JA, Pedroni MJ, Penrose LDJ, Fist AJ (2001) AFLP analysis of opium poppy. *Crop Sci* 41:1596–1601. doi:[10.2135/cropsci2001.4151596x](https://doi.org/10.2135/cropsci2001.4151596x)
- Schulz H, Baranska M, Quilitzsch R, Schütze W (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. doi:[10.1039/B408930H](https://doi.org/10.1039/B408930H)
- Selale H, Celik I, Gultekin V, Allmer J, Doganlar S, Frary A (2013) Development of EST-SSR markers for diversity and breeding studies in opium poppy (*Papaver somniferum* L.). *Plant Breed* 132:344–351. doi:[10.1111/pbr.12059](https://doi.org/10.1111/pbr.12059)
- Straka P, Nothnagel T (2002) A genetic map of *Papaver somniferum* L. based on molecular and morphological markers. *J Herbs Spices Med Plants* 9:235–241. doi:[10.1300/J044v09n02_a](https://doi.org/10.1300/J044v09n02_a)
- Tian CX, Liang XF, Yang M, Zheng HZ, Dou YQ, Cao L (2012) Isolation and characterization of novel genomic and EST-SSR markers in *Coreoperca whiteheadi* Boulenger and cross-species amplification. *Int J Mol Sci* 13:13203–13211. doi:[10.3390/ijms131013203](https://doi.org/10.3390/ijms131013203)

- Turkish Soil Product Office (2009) Opium poppy report. Turkish Soil Production Office, Ankara
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotech* 23:48–55. doi:[10.1016/j.tibtech.2004.11.005](https://doi.org/10.1016/j.tibtech.2004.11.005)
- Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA (2012) A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Sci Exp* 336:1704–1708. doi:[10.1126/science.1220757](https://doi.org/10.1126/science.1220757)
- Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot* 99:193–208. doi:[10.3732/ajb.1100394](https://doi.org/10.3732/ajb.1100394)
- Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman J, Bassil N, Hummer K, Simon PW, Zalapa JE (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor Appl Genet* 124:87–96. doi:[10.1007/s00122-011-1689-2](https://doi.org/10.1007/s00122-011-1689-2)
- Ziegler J, Facchini PJ, Geissler R, Schmidt J, Ammer C, Kramell R, Voigtländer S, Gesell A, Pienkny S, Brandt W (2009) Evolution of morphine biosynthesis in opium poppy. *Phytochemistry* 70:1696–1707. doi:[10.1016/j.phytochem.2009.07.006](https://doi.org/10.1016/j.phytochem.2009.07.006)