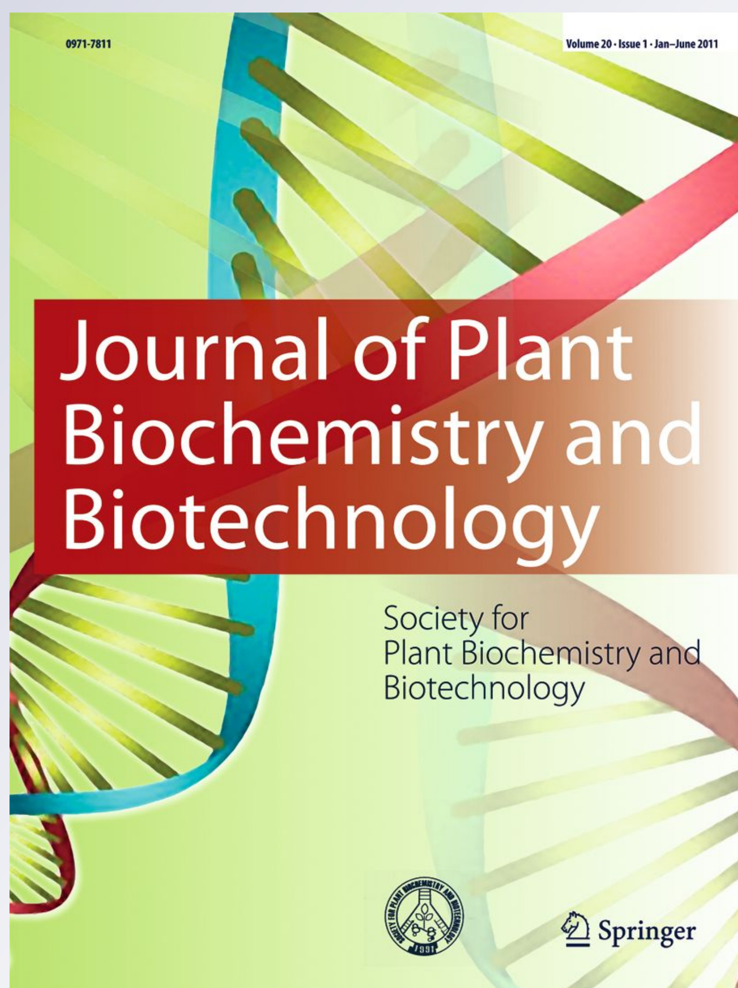
*Development of EST-SSR markers and their utility in revealing cryptic diversity in safflower (*Carthamus tinctorius *L.)*

Kalinati Narasimhan Yamini, Kamtam Ramesh, Vasupalli Naresh, Passoupathy Rajendrakumar, Kammili Anjani & Viswanathaswamy Dinesh Kumar

Journal of Plant Biochemistry and Biotechnology

ISSN 0971-7811

J. Plant Biochem. Biotechnol.
DOI 10.1007/s13562-012-0115-4



Springer

Your article is protected by copyright and all rights are held exclusively by Society for Plant Biochemistry and Biotechnology. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Springer

Author's personal copy

J. Plant Biochem. Biotechnol.
DOI 10.1007/s13562-012-0115-4

ORIGINAL ARTICLE

Development of EST-SSR markers and their utility in revealing cryptic diversity in safflower (*Carthamus tinctorius* L.)

Kalinati Narasimhan Yamini · Kamtam Ramesh · Vasupalli Naresh · Passoupathy Rajendrakumar · Kammili Anjani · Viswanathaswamy Dinesh Kumar

Received: 26 November 2011 / Accepted: 29 March 2012
© Society for Plant Biochemistry and Biotechnology 2012

Abstract With the aim of developing additional genomic resources in safflower, a set of 41,011 ESTs of safflower were mined for the presence of SSRs. 18,773 SSR containing ESTs (SSR-ESTs) were identified and were analyzed to remove redundant sequences leading to identification of 8,810 non-redundant SSR-ESTs (categorized into 6104 singletons and 2,706 contigs) having 13,085 non-redundant SSRs. The average number of non-redundant SSRs per EST was 0.32 and they predominantly consisted of dinucleotide (57.7 %), and trinucleotide (37.7 %) repeat motifs. 500 primer pairs were designed for the non-redundant EST-SSRs of which, 151 were tested. 60 markers which gave robust amplicons, were validated in a set of 19 *Carthamus* lines. A subset of EST-SSR markers, having average polymorphism information content (PIC) ≥0.4 could precisely elucidate the pedigree relatedness among these lines. Further, these markers exhibited high cross-species transferability to five other wild species of *Carthamus.* The markers reported here would be a valuable addition to existing safflower marker resources aiding in hastening its improvement.

Keywords EST-SSR · Safflower · Polymorphism · Transferability · Diversity

Abbreviations

DSC	Dice Standard Coefficiient
ISSR	Inter-simple Sequence Repeats
MAS	Marker-assisted Selection
NTSYS	Numerical Taxonomy System
PAGE	Polyacrylamide gel electrophoresis
UPGMA	Unweighted Pair-Group Method Arithmetic Average

K. N Yamini and K. Ramesh have contributed equally for the work.

Electronic supplementary material The online version of this article (doi:10.1007/s13562-012-0115-4) contains supplementary material, which is available to authorized users.

K. Ramesh · V. Naresh · K. Anjani · V. Dinesh Kumar (✉)
Directorate of Oilseeds Research,
Rajendranagar,
Hyderabad 500 030 Andhra Pradesh, India
e-mail: dineshkumarv@yahoo.com

K. N. Yamini
A.N.G.R. Agricultural University,
Regional Agricultural Research Station,
Warangal 506 007 Andhra Pradesh, India

P. Rajendrakumar
Directorate of Sorghum Research,
Rajendranagar,
Hyderabad 500 030 Andhra Pradesh, India

Introduction

Safflower (*Carthamus tinctorius* L., 2n=24) is one of the important oilseed crops of the world and has been grown commercially for its edible oil rich in polyunsaturated fatty acids. Increasing the productivity of this crop is of great importance also under the present scenario of huge demand for edible oil. Beneficial exploitation of safflower inbreds, to intensify and hasten breeding efforts in this crop, requires understanding of the genetic relatedness of inbreds and partitioning of genetic diversity among them. Recognition of genetically diverse inbreds is in turn essential to generate heterotic pools. Though morphological and biochemical markers have been used for the characterization of safflower genotypes (Carapetian and Estilai 1997), they are not

Published online: 15 May 2012

Author's personal copy

popular due to disadvantages like availability in limited numbers, environmental sensitivity and developmental stage specificity. Further, these markers are unable to elicit the hidden genetic diversity among phenotypically similar looking pedigree related inbreds.

But, DNA markers lack such limitations as well as provide precise and authentic markers for the assessment of genetic diversity. In safflower, molecular markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), inter-simple sequence repeats (ISSR) and amplified fragment length polymorphism (AFLP) have been used to assess the genetic diversity (Yazdi-Samadi et al. 2001; Sehgal and Raina 2005; Zhang et al. 2006a; Johnson et al. 2007; Yang et al. 2007).

Simple sequence repeats (SSRs) have become the marker of choice in plant genetics and breeding because of the codominant inheritance, relative abundance, multi-allelic nature, extensive genome coverage, high reproducibility, and simple detection (Powell et al. 1996). They are widely used for the assessment of genetic diversity, variety protection, molecular mapping, and marker-assisted selection (MAS), providing an efficient tool to link phenotypic and genotypic variation (Powell et al. 1996; Gupta and Varshney 2000; Varshney et al. 2005). Since development of genomic SSRs is time-consuming and expensive process involving generation of genomic libraries and sequencing of large numbers of clones (Eujayl et al. 2004), development of EST-SSR markers from the growing expressed sequence tags (ESTs) databases available in the public domain offers an easy and rapid alternative. However, the development of EST-SSR markers is limited to species for which this type of database exists. These markers have been successfully developed in several species such as triticeae (Zhang et al. 2006b), forage grasses (Mian et al. 2005), barley (Varshney et al. 2006), soybean (Hisano et al. 2007), cotton (Han et al. 2004; Park et al. 2005; Han et al. 2006), rice (La Rota et al. 2005), wheat (Yu et al. 2004; Peng and Lapitan 2005; Fu et al. 2006), sugarcane (Cordeiro et al. 2001), sorghum (Srinivas et al. 2009), pearl millet (Senthilvel et al. 2008), sunflower (Pashley et al. 2006) *etc*. About 1 to 5 % of ESTs from different plant species have been found to contain SSRs suitable for marker development (Kantety et al. 2002). EST-SSR markers are gene-tagged markers as they are directly associated with an expressed gene and, thus, completely linked with putative qualitative or quantitative trait locus alleles. EST-SSR markers are, therefore, superior and more informative compared to anonymous markers (Andersen and Lübberstedt 2003). Since they are present in or near coding DNA they would be more conserved, allowing cross-species transferability and a lower frequency of null alleles.

Very few SSR markers have been reported in safflower and more recently, EST-SSR markers have been reported (Chapman et al. 2009; Mayerhofer et al. 2010). However, in this study we report 452 novel safflower EST-SSRs which would be a valuable addition to the growing safflower genomic resources, a subset of which was able to reveal the cryptic diversity among 19 *Carthamus* lines.

Materials and methods

Plant material and DNA isolation

The plant material used in the present study consisted of 18 inbred lines of cultivated safflower (*C. tinctorius* L.; $2n=24$) and one wild species (*C. creticus*; $2n=44$). The description of plant material are listed in Table 1. These lines were grown under normal field conditions. In addition, five wild species of *Carthamus viz., C. turkestanicus, C. lanatus, C. creticus, C. oxyacanthus* and *C. glaucus* were used to assess the transferability of developed markers. Genomic DNA was isolated from 2 to 3 leaved seedlings as per the protocol of Doyle and Doyle (1987) with slight modification. The isolated DNA were used for PCR analysis.

Identification of non-redundant SSRs

A total of 41,011 EST sequences (accessed on June 6, 2008) of safflower were downloaded from NCBI EST database (http://www.ncbi.nlm.nih.gov/sites/entrez). Perfect dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats were identified and localized using the software, Simple Sequence Repeat Identification Tool (SSRIT) (Temnykh et al. 2001). The criterion for the identification of SSRs was fixed in such a way that only those repeats wherein the motif was repeated ≥4 times were identified resulting in a minimum repeat length of 8 (dinucleotide repeats) to 24 (hexanucleotide repeats) nucleotides. ESTs that contained SSRs were labeled as SSR-ESTs. In order to minimize redundancy, a cluster analysis was performed on the SSR-ESTs using the software, SEQUENCHER 4.7 (Gene Codes Corporation, Ann Arbor, MI). For assembling the SSR-ESTs, Assemble Automatically option was executed with default settings for minimum match % (85) and minimum overlap (20). This resulted in the identification of contigs and unassembled unique sequences (singletons). With respect to contigs, the longest sequence was selected from each cluster. These contig sequences along with the singletons formed the non-redundant SSR-ESTs. SSRs were classified considering the complementarities of the repeat motifs according to the classification given by Jurka and Pethiygoda (1995). This final set of SSRs in the non-redundant SSR-ESTs was used for marker development.

Author's personal copy

Table 1 Details of genotypes used in the present study

Code	Accession ID.	Specific traits
24B-90	MS 6(0)	Genetic male sterility, spiny
24B-18	[(MS 773xVB-19-3) VB −19-3]P_4	Genetic male sterility, non-spiny, very bold capitulum, resistance to *Fusarium* wilt
25B−7	CART10/79 (*C. creticus*)[a]	Wild species (*C. creticus*), high resistance to *Fusarium* wilt
15B−7	[(MS 773xVB−19-3) VB−19-3xCART10/79]P_1-P_{17}	Non-spiny, very bold capitulum
16B-11	[(MS 773xVB−19-3) VB−19-3xCART10/79]P_{32}-P_{14}	Spiny, very bold capitulum
16B-50	[(MS 773xVB-19-3) VB−19-3xCART10/79]P_{31}	Spiny, very bold capitulum
6B-33	[(MS 773xVB −19-3) VB−19-3xCART10/79]$P_6$35x2F	Spiny, very bold capitulum
9B-31	VB−19-3	Non-spiny, very bold capitulum, resistance to *Fusarium* wilt
14B-2	18–106	Non-spiny, high resistance to *Fusarium* wilt
9B-23	14–129	Non-spiny, moderate resistance to *Fusarium* wilt
18B-21	Ole-9-P_4	High oleic acid content, pinkish-white flower, resistance to *Fusarium* wilt, spiny
18B-22	Ole-9-P_1	High oleic acid content, pinkish-white flower, spiny
18B-24	Ole-9-Yellow	High oleic acid content, yellow flower, spiny
23B-121	W-05-2037-1-P_{18}	Spiny, high resistance to *Fusarium* wilt
23B-84	W-05-2050-9-P_{12}	Spiny, high resistance to *Fusarium* wilt
23B-69	W-05-912-19-P_{11}	Spiny, high resistance to *Fusarium* wilt
23B-126	W-05-2037-7-P_1	Spiny, high susceptibility to *Fusarium* wilt
16B-1	Derivative of (*C. tinctorius* x *C. lanatus*[a])x*C. palaestinius*[a]	Late maturity (140 days), thick wide and long basal leaves, spiny
24B-62	SFS-9998	Aphid tolerance, spiny, medium maturity

[a] wild species

Marker development

EST-SSRs with a repeat length of >20 nucleotides were used to design primers using the software Primer3 version 4.0 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) with the following parameters: primer length of 18–24 bases, an optimal T_m of 55°C with a minimum and maximum of 50°C and 60°C, respectively, and a 40–60 % GC content with a low chance of dimer or hair-loop formation. The range for PCR product length was set to be between 100 and 400 bp. A total of 500 primer pairs were designed (Supplementary Table 1). Of these 151 primer pairs were commercially synthesized (Bioserve, India) and tested in safflower. Sixty of these were validated in a set of 19 genotypes that included 18 safflower advanced breeding material and one *Carthamus creticus* accession (Table 1). The details of these validated markers *viz.,* locus designation, primer sequences, repeat motifs, allele attributes and PIC estimates, are summarized in Table 2.

Marker validation

PCR conditions for the selected 151 EST-SSR markers were standardized through gradient PCR with an annealing temperature range of 50–60°C, keeping other PCR components and parameters constant, using two safflower genotypes (MMS and C28-29-39-6-1). Based on this preliminary analysis, a set of 60 primer pairs that produced robust and sharp banding pattern were selected and were used for amplification of DNA isolated from 19 selected *Carthamus* genotypes. PCR was carried out in 10 μl reaction volume containing 25 ng of template DNA, 0.2 μM of each primer, 200 μM of each dNTP, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 1 unit of *Taq* polymerase in a 96-well thermal cycler (Eppendorf, Germany). The thermal profile was: 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 52–56°C (depending on the primers) for 30 s, 72°C for 1 min, and a final extension at 72°C for 7 min. Amplified products were resolved in 6 % non-denaturing polyacrylamide gel electrophoresis (PAGE) followed by silver staining (Panaud et al. 1996). The visualized SSR alleles were sized using the software utility Alphaease® (Alphainnotech, USA) with 50 bp ladder (MBI Fermentas, Lithuania) as the size standard.

Polymorphism information content

The polymorphism information content (PIC) is a measure of the effectiveness of a given DNA marker for detecting polymorphism. The PIC value for each EST-SSR marker

Author's personal copy

Table 2 Details of the Safflower EST-SSR markers developed and validated in the present study

Locus	Repeat Motif	Primer sequence (5′–3′)	N_a	PIC	H_e	GenBank Accession No.	Putative function
CtDES-2	$(ttc)_{14}$	**F**: GCTCTCTCTCTCTTCTCTCG	2	0.10	0.10	EL390273	Hypothetical protein
		R: GTGATTGCACATCAAGTAGC					Wound-responsive protein
CtDES-4	$(ttc)_{13}$	**F**: ATCGCCATTAAAGAAGAGG	2	0.10	0.10	EL394987	Hypothetical protein
		R: GTGATTGCACATCAAGTAGC					Wound-responsive protein
CtDES-5	$(ttc)_{12}$	**F**: GGGATACGCTTTCTTTCTTC	2	0.28	0.33	EL391215	Hypothetical protein
		R: AGCCATGGATTAGATGTTTG					Wound-responsive protein
CtDES-10	$(tga)_{12}$	**F**: ACGGGTAGATTTAAGGAAGG	3	0.54	0.62	EL411704	Hypothetical protein
		R: ACAATCCAACAGAGATTTGC					Fasciclin-like domain-containing protein
CtDES-12	$(tct)_{12}$	**F**: GGAACCCTAATCTTGTCTGC	3	0.50	0.57	EL373535	Hypothetical protein
		R: AGGAGTTTAGGAAGGAAACG					Aminocyclase
CtDES-13[a]	$(tct)_{11}$	**F**: AAGCAACCACTTACATTTCC	3	0.15	0.16	EL411999	Phytol kinase 1, chloroplast precursor
		R: CATGCTTCAATTTGTTCTCC					Hypothetical protein
CtDES-16	$(tct)_{10}$	**F**: AAGCAACCACTTACATTTCC	2	0.29	0.34	EL410757	Phytol kinase 1, chloroplast precursor
		R: CTTGCTTCAATTTGTTCTCC					Hypothetical protein
CtDES-19	$(tca)_{11}$	**F**: GAAGAAAAGCAGGAGTTGC **R**: GAGTCACCACACTGGAACC	3	0.56	0.63	EL409091	NAC domain protein
CtDES-24	$(gga)_{10}$	**F**: GTGGATACCTAGGGTTAGGG **R**: ACAAAAGAGGGGTTATCTGG	2	0.10	0.10	EL373321	Hypothetical protein
CtDES-25	$(gat)_{10}$	**F**: GAAATTTGGGTTGTACTTGG **R**: CAACCAAAACTAAGCATTCC	3	0.49	0.55	EL399841	Hypothetical protein
CtDES-26	$(gag)_{10}$	**F**: CTCAAAGTTCTAGGGTTAGGG	4	0.28	0.30	EL384966	Hypothetical protein
		R: ACCTCTAGGGTGTCTGTAAGG					FCA RNA binding protein
CtDES-27	$(gaa)_{13}$	**F**: CTCTTCTTGCAATTTCTTCG	3	0.40	0.45	EL373450	Hypothetical protein
		R: GAAGCTTCTCAAACTCATCC					ACR aminoacid binding protein
CtDES-28	$(gaa)_{12}$	**F**: GGACGAAAGAAAACAAAGG **R**: GATGTGGAGATAAAGGTTGG	2	0.10	0.10	EL510655	No significant similarity
CtDES-29	$(gaa)_{11}$	**F**: ACAAGAAAGTGAAGCAGTGG	2	0.10	0.10	EL404371	Hypothetical protein
		R: GCTGAATTGAAGAGGTTAGC					Chloroplast RNA binding protein
CtDES-30	$(gaa)_{10}$	**F**: AAGATGAGAGTGAAATTGAGC **R**: CTGGAGGGTAATTAGTCTGG	3	0.20	0.22	EL399842	Hypothetical protein
CtDES-33	$(cat)_{12}$	**F**: CGTTCTAGGACGACTACTCC **R**: ACTGCTTTTTGTCTCTTTCC	6	0.44	0.46	EL393817	Hypothetical protein
CtDES-34	$(cat)_{10}$	**F**: TATAAATCTCCCTCCATTCC	3	0.15	0.16	EL510679	Hypothetical protein
		R: ATCCAAACAATACGATCTGC					Isoamyl acetate-hydrolyzing esterase
CtDES-36	$(caa)_{12}$	**F**: CCATCGATCAGAACAAACC **R**: AAATAGACTCACGGTTGTGG	4	0.59	0.65	EL372999	Hypothetical protein
CtDES-38[a]	$(atg)_{10}$	**F**: GTTCAAGTGCAAAAGTCTCC **R**: GGTGTCTTGATAACAGAGTGC	2	0.10	0.10	EL385200	Hypothetical protein
CtDES-39	$(atc)_{10}$	**F**: TGACCTTCTGCTTCTTCTTC	4	0.47	0.51	EL380206	Hypothetical protein
		R: CTTCAGACGGAACAACTAGG					Xaa-Pro aminopeptidase
CtDES-43	$(aca)_{12}$	**F**: TCACTCTCTCTTGTTTCTTGC	3	0.48	0.55	EL385742	Hypothetical protein
		R: AAAGACCTGGGTGATAAAGG					Oligouridylate binding protein
CtDES-45	$(aca)_{10}$	**F**: ACAAACCCTAACCAACACC **R**: AAGGGAAGAAATAGACTCACG	8	0.74	0.77	EL374284	Hypothetical protein
CtDES-53	$(acaacg)_6$	**F**: CCGTACCTCCAAATCTCC **R**: TCTGATCATCGGAACTCG	4	0.65	0.70	EL375240	Hypothetical protein
CtDES-54	$(taatgc)_5$	**F**: TAGATGGAAACAGCTTCTCG **R**: ACGAGGAGAGAGAGAGACG	3	0.28	0.31	EL510201	Hypothetical protein

Author's personal copy

Table 2 (continued)

Locus	Repeat Motif	Primer sequence (5′–3′)	N_a	PIC	H_e	GenBank Accession No.	Putative function
CtDES-57	$(ttggag)_5$	**F:** ATTCTTCGAGGGTTTTTCG **R:** GATGAAGATTTCACGAGAGG	5	0.69	0.73	EL405277	Protein phosphatase
CtDES-59	$(tcggat)_5$	**F:** TTGTCACTTCTATCCCTTCG **R:** GTGTTTATGGTGTTTTCTCG	4	0.28	0.29	EL398345	Hypothetical protein Fasciclin-like arabinogalactan protein
CtDES-60	$(tccatc)_5$	**F:** CTTATCTCCTTGTGTTCAACG **R:** AGAAACAGAAGAGCAAATGG	4	0.53	0.60	EL397025	Hypothetical protein
CtDES-63	$(cgccac)_5$	**F:** CTTCGGAGATTCGGAGAC **R:** ACCTCCATCACCATTGTAAC	6	0.58	0.62	EL395625	Mannose-6-phosphate isomerase
CtDES-64	$(ctttct)_5$	**F:** GAGTAGAAAACCCTGAATTGG **R:** AGGGTTAGTTGGTGAATGC	4	0.67	0.71	EL393948	Hypothetical protein
CtDES-65	$(atgaag)_5$	**F:** AGAAGAAGAAGAAGATCACG **R:** GAATCAACTCCTCCACTTCC	3	0.19	0.20	EL393291	Hypothetical protein
CtDES-66	$(aaaccc)_5$	**F:** ATTTCTTCTTCCTCCACTCC **R:** ACGAAGCCTTTGATTTTTCC	5	0.75	0.78	EL393206	Cytochrome P450 monooxygenase
CtDES-67	$(cggaga)_5$	**F:** TTAACGAAGAAGACGAGACC **R:** AAATTGAAGCAGGAGAAGC	2	0.10	0.10	EL392354	Hypothetical protein
CtDES-68	$(ctaacc)_5$	**F:** TGTTATTCTGACAGGGAAGC **R:** GTAGATTATGGTTGCACAAGG	6	0.70	0.73	EL391851	Hypothetical protein
CtDES-69	$(gcaata)_5$	**F:** GAAGAAGGAGATAGCAATGG **R:** AGAAGAAGAATCCGAAAAGG	3	0.41	0.46	EL390984	Transcription factor
CtDES-76	$(cagcca)_5$	**F:** GTGCAGCCAAATAACTTCC **R:** GGTCTCAAACTTTTCCTTAGC	3	0.26	0.28	EL374918	No good similarity
CtDES-77[b]	$(ag)_{24}$	**F:** ATCTGCCTCATCGCATGT **R:** GCATCACTGGTACCAAAACT	4	0.70	0.74	EL385094	No good similarity
CtDES-79	$(ag)_{23}$	**F:** ACTAATATGACCGTTACCTTCC **R:** GGAGATGAAGTGGGAATAGG	5	0.74	0.78	EL382280	Hypothetical protein
CtDES-83 [a]	$(ag)_{16}$	**F:** ATGAACTCTAGTCACCACTTGC **R:** GAAGAAGTGACCTCTATTTTGG	3	0.28	0.30	EL378019	Casein kinase II beta chain
CtDES-91 [a]	$(ct)_{16}$	**F:** CATTCCGTCATCTATTTTGC **R:** GAAGTAATCGACTAACCAACG	3	0.53	0.59	EL510507	Hypothetical protein
CtDES-94	$(ct)_{15}$	**F:** ACGGCGGTTTTCACTAGG **R:** ACACCAATAATCACGAATCC	5	0.53	0.56	EL406300	Hypothetical protein
CtDES-96	$(ct)_{15}$	**F:** CTTTTCGTGATCAACACTCC **R:** CAGGTAATACTGCCTCAGACC	4	0.28	0.29	EL381270	26 S proteasome regulatory particle triple-A ATPase subunit6
CtDES-98 [b]	$(ga)_{25}$	**F:** ACCTCACATGGCGAAGAG **R:** GATTTCCGGAATGAAACAG	3	0.35	0.38	EL404124	Pollen-specific C2 domain containing protein
CtDES-99	$(ga)_{21}$	**F:** TTCTCTACTCTTCACGATTTGG **R:** CCATCTGTCTTAAGCTGTTCC	3	0.52	0.59	EL399497	No significant similarity
CtDES-101	$(ga)_{17}$	**F:** CATCCAGGATTAAGAAGTGG **R:** CCTTCGATCCACATACTCC	5	0.76	0.80	EL395089	No significant similarity
CtDES-103	$(ga)_{15}$	**F:** GAATCCCCAACAACTAATCC **R:** GGTTTAGAGGACTTTGATTTCC	6	0.67	0.70	EL406961	Glycosyltransferase
CtDES-104	$(ga)_{15}$	**F:** TCCGTTCCTAACTGAATCC **R:** AGCTCAGATCAATCACTTTCC	5	0.66	0.71	EL390687	Glycosyltransferase
CtDES-106	$(tc)_{19}$	**F:** GGGGCTTTCTTTACTTCC **R:** TATTGCTGTTGTTGTCTAGGG	5	0.78	0.75	EL373260	Hypothetical protein MATH (meprin and TRAF-C homology) domain
CtDES-109	$(tc)_{18}$	**F:** CCTAAACACCCATTTGTGG **R:** CCTAAGCAGGATAGAATAAACG	4	0.58	0.63	EL376316	Lipid transfer protein Hypothetical protein
CtDES-113	$(tc)_{10}$	**F:** CTTCACACACTCACATTTCC **R:** GAGTTGTCCATGACTGTGC	5	0.32	0.35	EL374150	Triose phosphate/phosphate translocator
CtDES-114	$(tc)_{16}$	**F:** ACGATTCACAGCTCTCTCTC **R:** CTCTCCATAGTCGCCATAAC	4	0.24	0.25	EL510780	Hypothetical protein Dynamin-like protein

Author's personal copy

Table 2 (continued)

Locus	Repeat Motif	Primer sequence (5′–3′)	N_a	PIC	H_e	GenBank Accession No.	Putative function
CtDES-122	$(tttc)_7$	**F:** GGGATGAGACTGAGATCG **R:** GACAGTTTGGAAGGTGTAGC	3	0.37	0.41	EL411048	Eukaryotic translation initiation factor
CtDES-123	$(aaga)_7$	**F:** ATCTTTGGTCGAGCTTGG **R:** ATTCGATGATTCCATCTCC	2	0.10	0.10	EL407032	Hypothetical protein
CtDES-127	$(tctt)_7$	**F:** AAGACTTTTGACACCTTCTC **R:** CTTTTGGCCTCTGTCAATATAC	2	0.10	0.10	EL392963	Serine/threonine-protein kinase
CtDES-128	$(tctt)_7$	**F:** GAGATCTCTCTCCTTTTCTCC **R:** GAGTTAAAGCACGAACTATGC	4	0.69	0.74	EL409081	PHD-finger domain containing protein
CtDES-129	$(tttc)_7$	**F:** CTCTTTATTTCGACTGGAACTG **R:** ATGCTTGTTGTTGCCTTATC	2	0.27	0.33	EL388881	Oligouridylate binding protein
CtDES-137	$(tgttt)_5$	**F:** AGCTAGATTACTGATGCAGGAT **R:** GGCAGTTAGAACAACAATACAG	3	0.15	0.16	EL403686	Hypothetical protein
CtDES-139	$(tcttc)_5$	**F:** TTTGCGTGTCGATAATCC **R:** TATCCTCATCGTAACATCATCC	4	0.69	0.74	EL401029	Unknown protein
CtDES-141	$(attcg)_5$	**F:** AGTTGACCAAATTCAAGTCC **R:** CTAGATCGTTGTTGTTCTTCG	2	0.10	0.10	EL393162	Nitric oxide synthase interacting protein
CtDES-142	$(ttcat)_5$	**F:** AAGATCTCATCTGGGTTTCC **R:** AGAATGAATCAATGGGTAGG	4	0.45	0.49	EL392745	GTPase activator
CtDES-143	$(ctttt)_5$	**F:** ACCACCTCATGCTCTTACC **R:** AGCTATGAGTAGGAAGAATTGG	5	0.79	0.81	EL390871	Hypothetical protein

N_a Number of alleles; *PIC* Polymorphism Information Content; H_e Expected Heterozygosity

[a] These markers are also reported by Chapman et al. (2009)

[b] These markers are also reported by Mayerhofer et al. (2010)

was calculated by using the standard formula (Anderson et al. 1993):

$$\mathrm{PIC} = 1 - \Sigma \mathrm{P}_{ij}^2$$

where, P_{ij} is the frequency of the jth allele for the ith marker in a set of the investigated safflower genotypes and summation extends over k alleles detected for the ith marker.

Genetic diversity analysis

The EST-SSR alleles obtained with markers having more than or equal to the average PIC value (0.4) were converted into a binary matrix where each allele for the SSR loci was scored as present (1) or absent (0) as per NTSYS data format (Exeter Software, Setauket, NY). The resample function was used to create 9,999 randomization of the data set and used to obtain the bootstrap values. Genetic similarities were calculated among all possible pairs of accessions using the Dice Standard Coefficient (DC) and the corresponding UPGMA (Unweighted Pair-Group Method Arithmetic Average) phenograms were generated with the NTSYSpc (v2.2) program (Exeter Software, Setauket, NY).

Results

Mining for ESTs containing SSR motifs

Among the 41,011 ESTs screened for presence of SSRs, a total of 18,773 SSR containing ESTs (SSR-ESTs), having 21,479 SSRs were identified. A redundancy analysis was performed on these SSR-ESTs in order to reduce overestimation of specific SSR types, since the above SSR estimates were based on a redundant EST dataset. This analysis categorized 18,773 SSR-ESTs into 6,104 singletons and 2,706 contigs (containing 12,669 SSR-ESTs). This resulted in the identification of a total of 8,810 non-redundant ESTs that contained 13,085 SSRs.

Frequency and distribution of SSRs in the ESTs of safflower

Considering the total ESTs analyzed in the present study, 21.48 % (8810) contained non-redundant SSRs (Supplementary Table 2). The majority (68.7 %) of the sequences had a single putative SSR, while the remaining (31.3 %) had 2 to 10 SSRs (Supplementary Table 3). The average number of SSRs per EST was 0.32 with one SSR containing EST for every 4.7 ESTs.

Author's personal copy

Dinucleotide (57.7 %) and trinucleotide (37.7 %) repeat motifs were more abundant compared to the other types of repeats which were detected in the frequency of 2.4 % (tetra-), 1.4 % (hexa-) and 0.8 % (penta-) (Supplementary Fig. 1). Among the dinucleotide repeats, AG/GA/CT/TC motif was the most common (59.2 %) followed by AC/CA/TG/GT (21.8 %) and AT/TA (18.4 %) motifs, whereas GC/CG motif was the least common (0.6 %) (Supplementary Fig. 2). Among the trinucleotide repeats, the motif AAG/AGA/GAA/CTT/TTC/TCT was the most common (28.4 %) followed by the motifs ATG/TGA/GAT/CAT/ATC/TCA (21.1 %) and ACC/CCA/CAC/GGT/GTG/TGG (14.2 %) whereas the motif ACG/CGA/GAC/CGT/GTC/TCG was the least common (2.3 %) (Supplementary Fig. 3; Supplementary table 4).

Polymorphism analysis of EST-SSR markers

Primers were designed targeting 500 SSRs with repeat length >20 nucleotides (Supplementary table 1). In a preliminary analysis, 151 of these primer pairs were amplified using two safflower genotypes (MMS and C28-29-39-6-1). Among these primer pairs, 99 gave expected sized amplicons, 40 gave nonspecific amplicons and 12 gave no amplification at all. Of the 99, about 60 markers giving robust and sharp amplicons were validated in a set of 19 safflower genotypes. A total of 211 alleles were generated by these EST-SSRs, of which 208 alleles were polymorphic. The number of alleles ranged from 2 to 8 with an average of 3.6 alleles per marker. In order to assess the informativeness of the EST-derived safflower SSR markers, the polymorphism information content (PIC value) was calculated for each of the 60 markers based on the 19 genotypes used in this study. PIC values of the 60 EST-SSRs ranged from 0.1 to 0.79 with an average of 0.41(Table 2).

Based on BLASTX analysis of the ESTs of the 60 markers validated in the present study, a putative function could be assigned to only 19 markers assuming a stringent threshold of <1.00E-20. About 60 % of the markers showed homology to hypothetical proteins with no assigned functions and some of the safflower SSR–ESTs showed significant homology to the annotated proteins of *Arabidopsis* (Table 2).

Genetic relationship among safflower genotypes

In order to assess the genetic relatedness among a set of 19 *Carthamus* lines (Table 1), 33 EST-SSR markers having more than or equal to the average PIC value (0.4) were used for cluster analysis based on Dice's similarity coefficient (DSC). These 33 loci covering 114 alleles could discriminate the 19 lines as evident by the dendrogram generated from DCS, which classified these genotypes into two major clusters (Fig. 1). The Dice's similarity coefficient ranged from a high of 0.76 to a low of 0.06 among these genotypes.

The wild species, *C. creticus* (25B-7) having on an average 13 % genetic similarity with the other 18 genotypes formed a separate major cluster-A. The 18 genotypes having 47 % average genetic similarity among them formed jointly another major cluster-B. Cluster-B was divided into two sub-clusters (B-I and B-II). These sub-clusters were further sub-divided based on genetic relatedness. In sub-cluster B-I, genotypes (16B-33, 15B-7 and 16B-11) grouped with their great grandparent (24B-90) and grandparent (9B-31). In sub-cluster B-II, the high oleic acid lines (18B-21, 18B-22 and 18B-24) which had common parents grouped together. Progenies of a cross involving wilt susceptible and resistant parents grouped together (23B-12, 23B-126, 23B-84 and 23B-69). All pedigree related inter-specific derivatives (15B-7, 16B-11, 16B-50, 6B-33 and 14B-2) were grouped relatively closer to each other than to other unrelated cultivated genotypes (Fig. 1).

Discussion

Safflower ESTs (41,011) were screened for the presence of SSRs and approximately 21.48 % of the ESTs were found to contain SSRs. This is higher than reports in other crops *i.e.,* coffee (18.5 %) (Aggarwal et al. 2007), grapes (2.5 %) (Scott et al. 2000), barley (2.8 %) (Varshney et al. 2006), wheat (7.41 %) (Peng and Lapitan 2005), sugarcane (2.88 %) (Cordeiro et al. 2001). However, these differences can be attributed to the different SSR search criteria, the size of the databases and the database mining tool(s) used. In general, when the minimum repeat length is 20 bp, microsatellites of various plant species are present in about 5 % of ESTs (Varshney et al. 2005).

The number of contigs and singletons identified in the present study was lower than that reported by Chapman et al. (2009) which might be due to the difference in the methods adopted for identification of SSRs. Though majority of the EST-SSR identification in most species involves assembly of ESTs followed by mining for SSRs in the non-redundant set of ESTs (as followed by Chapman et al. 2009), in the present study, mining for SSRs was done first, followed by assembling of the SSR containing EST and selecting only the non-redundant set of SSR-ESTs. Similar approach has been adopted in several studies (Ramu et al. 2009; Kumpatla and Mukhopadhyay 2005; Aggarwal et al. 2007).

Considering the repeats in different reading frames or on complimentary strand, we included all the four classes that are possible for dinucleotide, 10 for trinucleotide, and 33 for tetranucleotide repeats as reported by Jurka and Pethiygoda (1995). Among the SSR repeats, we observed preponderance of dinucleotide motifs (57.7 % of the total SSRs)

Author's personal copy

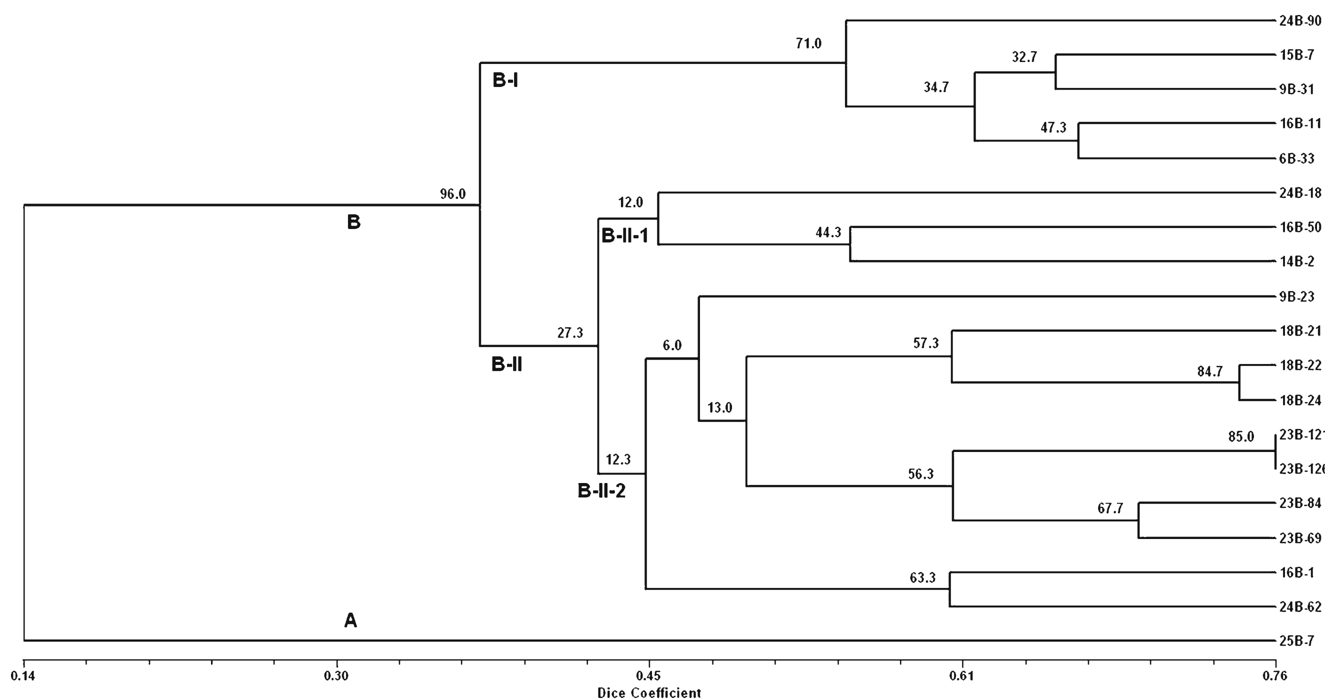


Fig. 1 Dendrogram of safflower genotypes based on EST-SSR polymorphism. 'A' and 'B' denote major clusters. Sub-clusters within 'B' are also indicated

followed by trinucleotide motifs (37.7 %), while in several crops trinucleotide motifs have been found to be more abundant (Scott et al. 2000; Thiel et al. 2003; Han et al. 2004; La Rota et al. 2005; Chen et al. 2006; Hisano et al. 2007; Chabane et al. 2007). Abundance of dinucleotide motifs has also been reported in peach, pumpkin, coffee, spruce and kiwi fruit (Fraser et al. 2004; Rungis et al. 2004; Xu et al. 2004; Aggarwal et al. 2007; Gong et al. 2008). As observed in other crops, a lower frequency of other types of repeats *i.e.,* 2.4 % (tetra-), 1.4 % (hexa-) and 0.8 % (penta-) were detected.

Among the dinucleotide repeats, the most frequent motif was AG/GA/CT/TC (59.2 %), as reported in other plant species (Gao et al. 2003; Kumpatla and Mukhopadhyay 2005; Temnykh et al. 2000; Kantety et al. 2002; Saha et al. 2003; Poncet et al. 2006). A di-nucleotide motif can represent multiple codons depending on the reading frame and translate into different amino acids. For example, the GA/CT motif can represent GAG, AGA, UCU and CUC codons in a mRNA population and translate into the amino acids Arg, Glu, Ala and Leu respectively. Ala and Leu are present in proteins at high frequencies of 8 % and 10 %, respectively (Lewin 1994). This could be one of the reasons why GA/CT motifs are present at such high frequencies in EST collections. Only 0.6 % of the dinucleotide repeats in this study had the CG/GC motif. Similar least frequency of this motif has been observed in coffee (Aggarwal et al. 2007; Poncet et al. 2006), sugarcane (Pinto et al. 2004) and many other monocot species (Kantety et al. 2002).

Among the trinucleotide repeats, AAG/AGA/GAA/CTT/TTC/TCT was the most abundant in safflower (28.4 %), and the abundance of this motif has been reported in majority of the dicotyledonous species including cotton, pine, soybean, *Arabidopsis* (Kumpatla and Mukhopadhyay 2005; Saha et al. 2003; Han et al. 2006; Liewlaksaneeyanawin et al. 2004; Cardle et al. 2000; Hisano et al. 2007), while GGC/CGG was most abundant in monocotyledonous species such as barley, maize, sorghum and rice (Kantety et al. 2002). However, it is difficult to compare the prevalence of a particular motif across different plant species due to the differences in the minimal motif-repeat criteria chosen in different studies. Abundance of trinucleotide motifs could be reflecting the fact that trinucleotide motifs in coding regions would not cause frame-shift mutation that could silence the gene but would result in a variation in amino acid residue number at the protein level (Cloutier et al. 2009)..

In our study, the AAG class (AAG/AGA/GAA/CTT/TTC/TCT) that can adopt triple-helical structures (Pearson and Sinden 1998) was predominant among tri repeats. It is indicated that while nucleotide composition might play an important role in the genesis of repeats, in the coding sequences, its effect on the structure and function of the encoded proteins would be a major selective force. The AAG class that can adopt triple-helical structures (Pearson

Author's personal copy

and Sinden 1998) are comparatively more numerous in *Arabidopsis*, *C. elegans*, and yeast and less numerous in human and *Drosophila* sequences.

Marker development and polymorphism of EST-SSR markers

An important feature of SSRs for genetic analysis is their high polymorphic nature leading to a large number of allelic variants across different genotypes. It is generally known that longer SSRs have a higher probability of being polymorphic (Cho et al. 2000; La Rota et al. 2005; Temnykh et al. 2001) and hence from marker utility perspective, it makes sense to start with SSRs with higher number of repeats first followed by the shorter stretches (La Rota et al. 2005). Although many previous reports have used threshold repeat lengths of 10–12 nucleotides, any preference(s) in genesis of repeats or variations in mutation rates are likely to be clearer at longer threshold lengths. Additionally, longer repeats being more unstable have implications in genome organization, genetic variation, protein evolution and disease on a relatively shorter evolutionary time scale (Katti et al. 2001). Based on this analogy, in the present study, though the criterion for identification of SSRs was fixed in such a way that all those repeats wherein the motif was repeated ≥4 times resulting in a minimum repeat length of 8 (di-repeats) to 24 (hexa-repeats) nucleotides were identified, primers were designed beginning with SSRs that had highest repeat length (108 bp in case of CtDES-48) till those SSRs which had at least 22 bp repeat length.

Similar to that reported for flax ESTs where 7.8 % of the designed primers failed to amplify (Cloutier et al. 2009), in our study too, 7.9 % of the safflower primers used for analysis did not amplify. This was in contrast to the reports in sugarcane and barley where approximately 40 % and 30–36 % of the designated EST-SSR primers failed, respectively (Cordeiro et al. 2001; Thiel et al. 2003; Varshney et al. 2006). These failures could be caused by primer mismatch, primers designed across splice sites or the presence of large introns within the target amplicons, the usage of questionable sequence information for primer development, or primers that were derived from chimeric cDNA clones. The lesser number of primers failing amplification could indicate a higher robustness of these markers. Several others have previously observed high quality markers from EST-SSRs and concluded that quality may be a byproduct of shorter repeat motif in genic regions as opposed to elsewhere in the genome (Eujayl et al. 2002; Woodhead et al. 2005; Pashley et al. 2006).

In the present study, experiments carried out with 19 genotypes of *Carthamus*, to ascertain the number of allelic variants exhibited by the selected set of primers indicated that the number of alleles ranged from 2 to 8 with an average of 3.6 alleles per marker. When compared with other species this average was higher than soybean (2.8) and grass species (1.6–2.5) but not as high as for almond (5.45) (Mian et al. 2005; Hisano et al. 2007). Markers with higher PIC value could greatly aid the selection of loci likely to be informative in safflower and possibly also in other C*arthamus* species (Chapman et al. 2009). In our study, the markers analyzed had a wide range of PIC values with many of them having a PIC value of >0.5 indicating that highly polymorphic markers could be obtained from safflower. Dinucleotide motif EST-SSR markers showed higher average PIC value (0.53) compared to tri (0.32), tetra (0.31), penta (0.44) and hexanuceotide (0.47) motifs and similar observations have been reported in earlier studies (La Rota et al. 2005; Hisano et al. 2007; Cloutier et al. 2009).

In this study, the available safflower EST database (41,011 ESTs), at the time of initiation of this study (2008) were used for developing the EST-SSR markers. Independently Chapman et al. (2009) and Mayerhofer et al. (2010) have reported EST-SSR markers in safflower. Chapman et al. (2009) have developed 384 primer pairs of which 104 have been made available in the public domain (supplementary data doi:10.1007/s00122-009-1161-8) while Mayerhofer et al. (2010) have mapped 78 safflower EST-SSR markers (Supplementary data at http://genome.nrc.ca) on the *C. tinctorius* linkage map. However, a comparison of the EST-SSR markers developed by us with these reported markers (available on public domain) indicate that 452 markers (out of 500 developed) are new (Supplementary table 1) and not reported so far. Of the total 500 EST-SSR markers, 31 were common with those reported by Chapman et al. (2009) while 10 markers were common with those reported by Mayerhofer et al. (2010). Seven markers were found to be common among the three reports (Supplementary table 1). Also, Chapman et al. (2009) have reported only di, tri and tetra nucleotide EST-SSRs whereas our study includes penta- and hexa-nucleotide repeats. Minimum number of common SSRs among these studies could be due to differences in aspects such as number of ESTs used, class of repeat motifs identified, minimum number of repeat units identified and limited marker information available for comparison. Thus, 452 markers reported here are new and therefore could be a valuable genomic resource in developing saturated linkage map in safflower.

Genetic diversity among safflower genotypes

Potential of the selected EST-SSRs in revealing the genetic relatedness was ascertained using a set of 19 *Carthamus* genotypes consisting of 18 *C. tinctorius* inbred lines and one wild species (*Carthamus creticus* L.), as this species had been used in the development of some of the breeding lines chosen. These genotypes were chosen in such a way that they included inter-specific derivatives, breeding lines

Author's personal copy

selected for specific traits, lines obtained through different breeding methodologies and also lines related by descent (Table 1).

Clustering of genotypes tended to follow pedigree relatedness of genotypes. Ability of these markers to discriminate closely related lines is evident from the dendrogram (Fig. 1). The five genotypes in sub-cluster B-I were related by pedigree. The related genotypes *viz.,* 15B-7, 16B-11 and 6B-33 were grouped along with their great grand parent (24–90) and grand parent (9B-31). Their female parent (24B-18) was included in a nearby sub-cluster B-II-1 along with one of the related genotypes (16B-50). The genetic diversity revealed among these related genotypes might be due to use of different breeding and selection criteria in breeding of these individuals. Population approach was used for breeding16B-50 whereas pedigree method was followed for developing 16B-11. While 6B-33 was developed through selection of superior progenies of paired crosses between male sterile and fertile sister plants in single-plant-progeny rows, 15B-7 was developed through selection of superior single-plant progenies. The average genetic similarity of 16B-50 with the other related genotypes was very low (mean DSC= 0.38). This might be because of greater chance of accumulation of more fixed transgressive recombinants or allelic combinations when population approach was employed in development of 16B-50. Pedigree method (DCS=0.52), paired crossing between single-plant progenies (DCS=0.54) and single-plant progeny selection (DCS=0.54), maintained relatively high genetic similarity among the related genotypes. The selected traits and plant type might also have influenced the level of genetic similarity or diversity among related genotypes.

The three genotypes *viz.,* 6B-33, 15B-7 and 16B-11 were closer to their grandparent (9B-31) than 16B-50. This might be because of selection of grandparent's plant type during breeding of these genotypes.

The three high oleic lines *viz.,* 18B-21, 18B-22 and 18B-24 having common parents were grouped closely. The Dice's genetic similarity coefficient (DSC) among them ranged from 0.55 to 0.74. Pedigree based clustering was also observed in grouping of 23B-21, 23B-126, 23B-84 and 23B-69. These were progenies of a cross between *Fusarium* wilt susceptible and resistant parents. 23B-121 and 23B-126 are near isogenic lines whereas 23B-84 is a sister line and 23B-69 is a backcross progeny. These results clearly showed that EST-SSR markers could precisely assess genetic similarity or distance among closely related genotypes.

The genotype 9B-23 is an inter-specific derivative of a cross between cultivated species (*C. tinctorius L.*) and wild species (*C. turkestanicus*); it predominantly resembles the cultivated species with slight resemblance to *C. turkestanicus* in basal leaf morphology. Its inclusion in an individual group close to cultivated species genotypes was in conjunction with its morphology. The genotype 14B-2 was a derivative of *C. tinctorius* x *C. creticus* (25B-7) cross. It was grouped independently near to its paternally related 16B-50. All pedigree related inter-specific derivatives (15B-7, 16B-11, 16B-50, 6B-33 and 14B-2) were grouped relatively closer to each other than to other unrelated cultivated species genotypes. The precise classification of inter-specific derivatives indicates high robustness of the EST-SSR markers in identifying the cryptic polymorphism. Two genotypes *viz.,* 16B-1 and 24B-62, which do not share lineage with other genotypes, were grouped separately. The genotype 16B-1 is a derivative of *C. tinctorius* x *C. lanatus* x *C. palaestinus*. It was the most distant one in cluster-B (mean DSC=0.37), its clustering into cluster-B might be because of its predominant resemblance to *C. tinctorius.*

Hence, the EST-SSR markers could effectively show the degree of genetic relatedness among these 19 genotypes. This information could aid in designing crosses for specific breeding programs. Similar utility of EST-SSRs for assessment of the genetic relatedness among the breeding material of flax has been reported (Cloutier et al. 2009). The EST SSRs could also be applied in assessing the genetic diversity, population structure (Eujayl et al. 2002; Simko 2009; Xinquan et al. 2005), in varietal identification (Wiesner et al. 2001; Roose-Amsaleg et al. 2006), and in the construction of linkage maps (Mayerhofer et al. 2010). Recently, we had demonstrated the utility of EST-SSR markers in assessing the genetic purity of safflower hybrids (Naresh et al. 2009).

Cross species transferability

Transferability of EST-derived markers at different taxonomic levels has been demonstrated earlier (Scott et al. 2000; Cordeiro et al. 2001; Feingold et al. 2005). Recently, studies on the feasibility of transferring markers such as non-genic SSR markers and genic markers *i.e.,* intron fragment length polymorphism (IFLP) markers, resistance gene candidates (RGC)-based markers and EST-SSR markers from sunflower to safflower have indicated a low rate of transferability for non-genic SSRs and a better rate of transferability

Table 3 Transferability of safflower EST-SSR markers among the wild species of *Carthamus*

Species	Total markers	No. of markers amplified	Transferability (%)
C. turkestanicus	60	57	95.0
C. lanatus	60	57	95.0
C. creticus	60	59	98.3
C. oxyacanthus	60	56	93.3
C. glaucus	60	53	88.3

Author's personal copy

for the genic markers (Garcia-Moreno et al. 2010). It is reasonable to expect higher rate of transferability across *Carthamus* species if genic markers developed in safflower are used. In the present study, the EST-SSRs were found readily transferable to other species of *Carthamus* (*C. turkestanicus, C. lanatus, C. creticus, C. oxyacanthus* and *C. glaucus*) (88.3 % to 98.3 %) (Table 3). This was in corroboration with the fact that EST-SSR markers, being derived from expressed genes are more conserved and have a higher level of transferability to related species than anonymous markers (Asp et al. 2007). They are, therefore, useful as anchor markers for comparative mapping across species, comparative genomics, and evolutionary studies (Kashi and King 2006; Kantety et al. 2002; Thiel et al. 2003; Scott et al. 2000; Saha et al. 2004). The developed safflower markers might display polymorphism in a broader *Carthamus* species collection if analyzed further and could facilitate search for novel alleles in a broader gene pools of related *Carthamus* species. These markers would be useful to understand the introgression of genes from wild species into cultivated safflower as well serve as tools to construct genetic maps of wild species.

In conclusion, development of SSR markers from EST databases saves both cost and time, once a sufficient amount of EST sequences are available. The present study is an effort towards the adding more number of EST-SSR markers to the to the safflower marker database available on the public domain as well as to demonstrate their utility in revealing the cryptic genetic relationships among advanced breeding lines which could greatly aid in selecting parents for crop improvement. In the near future, these markers could be applied in aiding and hastening safflower crop improvement. The new markers reported in this study along with those reported earlier for safflower (Chapman et al. 2007; Chapman et al. 2009; Mayerhofer et al. 2010) have the potential of facilitating tagging of important traits as well as mapping studies for safflower and its close relatives.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Asp T, Frei UK, Didion T, Nielsen KK, Lubberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol 7:36

Carapetian J, Estilai A (1997) Genetics of isozyme coding genes in safflower. In: Corleto A, Mundel HH (eds). Proceedings of the Fourth International Safflower Conference. Bari, Italy, pp 235–237

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chabane K, Varshney RK, Graner A, Valkoun J (2007) Generation and exploitation of EST-derived SSR markers for assaying molecular diversity in durum wheat populations. Genet Resour Crop Evol 55:869–881

Chapman MA, Chang J, Weisman D, Kesseli RV, Burke JM (2007) Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theor Appl Genet 115:747–755

Chapman MA, Hwala J, Strever J, Matvienko M, Kozik A, Michelmore RW, Tang S, Knapp SJ, Burke JM (2009) Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). Theor Appl Genet 120:85–91

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet 119:53–63

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross-transferable to *Erianthus* and sorghum. Plant Sci 160:1115–1123

Doyle JJ, Doyle JN (1987) A rapid DNA isolation protocol for small quantity of fresh leaf tissue. Phytochem Bull 19:11–15

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor Appl Genet 111:456–466

Fraser LG, Harvey CF, Crowhurst RN, De Silva HN (2004) EST-derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108:1010–1016

Fu YB, Peterson GW, Yu JK, Gao L, Jia J, Richards KW (2006) Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theor Appl Genet 112:1239–1247

Gao LF, Tang J, Li H, Jia J (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed 12:245–261

Garcia-Moreno MJ, Velasco L, Perez-Vich B (2010) Transferability of non-genic microsatellite and gene-based sunflower markers to safflower. Euphytica 175:145–150

Gong L, Stift G, Kofler R, Pachner M, Lelley T (2008) Microsatellites for the genus *Cucurbita* and an SSR-based genetic linkage map of *Cucurbita pepo* L. Theor Appl Genet 117:37–48

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 272:308–327

Author's personal copy

Han ZG, Wang CB, Song XL, Guo WZ, Guo JY et al (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSR in allotetraploid cotton. Theor Appl Genet 112:430–439

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271–281

Johnson RC, Kisha TJ, Evans MA (2007) Characterizing safflower germplasm with AFLP molecular markers. Crop Sci 47:1728–1736

Jurka J, Pethiygoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. J Mol Evol 40:120–126

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol Biol 48:501–510

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22:253–259

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18:1161–1167

Kumpatla SP, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998

La Rota M, Kantety RV, Yu JK, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23

Lewin B (1994) Genes V. Oxford University Press, New York

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Mayerhofer R, Archibald C, Bowles V, Good AG (2010) Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. Genome 53:266–276

Mian MA, Saha MC, Hopkins AA, Wang ZY (2005) Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. Genome 48(4):637–47

Naresh V, Yamini KN, Rajendrakumar P, Dinesh Kumar V (2009) EST-SSR marker based assay for the genetic purity assessment of safflower hybrids. Euphytica 170:347–353

Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, el-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton *Wber* loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428–441

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: lessons from *Helianthus*. J Hered 97:381–388

Pearson CE, Sinden RR (1998) Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. Curr Opin Struct Biol 8:321–330

Peng JH, Lapitan NL (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics 5:80–96

Pinto LR, Oliveira KM, Ulian EC, Garcia AA, de Souza AP (2004) Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. Genome 47:795–804

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko A, Perla H (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

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Ramu P, Kassahun B, Senthilvel S, Ashok Kumar C, Jayashree B, Folkertsma RT, Reddy LA, Kuruvinashetti MS, Haussmann BI, Hash CT (2009) Exploiting rice-sorghum synteny for targeted development of EST-SSRs to enrich the sorghum genetic linkage map. Theor Appl Genet 119:1193–204

Roose-Amsaleg C, Cariou-Pham E, Vautrin D, Tavernier R, Solignac M (2006) Polymorphic microsatellite loci in *Linum usitatissimum*. Mol Ecol Notes 6:796–799

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet 109:1283–1294

Saha S, Karaca M, Jenkins JN, Zipf AE, Ramesh OU, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang LJ, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Sehgal D, Raina SN (2005) Genotyping safflower (*Carthamus tinctorius*) cultivars by DNA fingerprinting. Euphytica 146:67–76

Senthilvel S, Jayashree B, Mahalakshmi V, Sathish Kumar P, Nakka S, Nepolean T, Hash CT (2008) Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biol 8:119

Simko I (2009) Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). J Hered 100:256–62

Srinivas G, Satish K, Madhusudhana R, Seetharama N (2009) Exploration and mapping of microsatellite markers from subtracted drought stress ESTs in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 118:703–717

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

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

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

Varshney RK, Hoisington DA, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. Trends Biotechnol 24:490–499

Wiesner I, Wiesnerova D, Tejklova E (2001) Effect of anchor and core sequence in microsatellite primers on flax fingerprinting patterns. J Agric Sci 137:37–44

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Mol Ecol 14:1681–1695

Xinquan Y, Peng L, Zongfu H, Zhongfu N, Qixin S (2005) Genetic diversity revealed by genomic-SSR and EST-SSR markers among

Author's personal copy

common wheat, spelt and compactum. Progress Nat Scien 15:1745–5391

Xu Y, Ma R, Xie H, Liu J, Cao M (2004) Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome 47:1091–1104

Yang YX, Wu W, Zheng YL, Chen L, Liu RJ, Huang CY (2007) Genetic diversity and relationships among safflower (*Carthamus tinctorius* L.) analyzed by inter-simple sequence repeats (ISSRs). Genet Resour Crop Evol 54:1043–1051

Yu JK, La Rota M, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Genet Genomics 271:742–751

Yazdi-Samadi B, Maali Amiri R, Ghannadha MR, Abd-Mishani C (2001) Detection of DNA polymorphism in landrace populations of safflower in Iran using RAPD-PCR technique. In Bergman J, Mundel HH (eds) Proceedings of the Fifth International Safflower Conference. Williston, ND, Sidney MT, p. 163

Zhang L, Huang BB, Kai GY, Guo ML (2006a) Analysis of inter-specific variation of Chinese *Carthamus tinctorius* L. using AFLP markers. Yao Xue Xue Bao 41:91–96

Zhang LY, Ravel C, Bernard M, Balfourier F, Leroy P, Feuillet C, Sourdille P (2006b) Transferable bread wheat EST-SSRs can be useful for phylogenetic studies among the Triticeae species. Theor Appl Genet 113:407–418