



OPEN ACCESS

EDITED BY

Muhammad Ali Abid,
Chinese Academy of Agricultural
Sciences, China

REVIEWED BY

M. Z. Abdin,
Jamia Hamdard University, India
Ahsan Irshad,
Chinese Academy of Agricultural
Sciences, China
Alia Anwar,
University of the Punjab, Pakistan

*CORRESPONDENCE

Assadollah Ahmadikhah
a_ahmadikhah@sbu.ac.ir

SPECIALTY SECTION

This article was submitted to
Functional and Applied Plant
Genomics,
a section of the journal
Frontiers in Plant Science

RECEIVED 11 July 2022

ACCEPTED 17 October 2022

PUBLISHED 17 November 2022

CITATION

Ahmadi AJ and Ahmadikhah A (2022)
Occurrence of simple sequence
repeats in cDNA sequences of
safflower (*Carthamus tinctorius*)
reveals the importance of
SSR-containing genes for cell
biology and dynamic response
to environmental cues.
Front. Plant Sci. 13:991107.
doi: 10.3389/fpls.2022.991107

COPYRIGHT

© 2022 Ahmadi and Ahmadikhah. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the
copyright owner(s) are credited and
that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Occurrence of simple sequence repeats in cDNA sequences of safflower (*Carthamus tinctorius*) reveals the importance of SSR-containing genes for cell biology and dynamic response to environmental cues

Ahmad Jawid Ahmadi¹ and Assadollah Ahmadikhah^{2*}

¹Agronomy Department, Faculty of Agriculture, Higher Education Institute of Samangan, Samangan, Afghanistan, ²Department of Cell and Molecular Biology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran

Safflower (*Carthamus tinctorius*) is a diploid crop plant belonging to the family Asteraceae and is well known as one of important oilseed crops due to edible oil containing unsaturated fatty acids. In recent years it is gaining increased attention for food, pharmaceutical and industrial uses, and hence the updating its breeding methods is necessary. Genic simple sequence repeats (SSRs) in addition of being desire molecular markers, are supposed to influence gene function and the respective phenotype. This study aimed to identify SSRs in cDNA sequences and further analysis of the functional features of the SSR-containing genes to elucidate their role in biological and cellular processes. We identified 1,841 SSR regions in 1,667 cDNA sequences. Among all types of repeats, trinucleotide repeats were the most abundant (35.7%), followed by hexanucleotide (29.6%) and dinucleotide repeats (22.0%). Thirty five SSR primer pairs were validated by PCR reaction, detected a high rate of polymorphism (>57%) among safflower accessions, physically mapped on safflower genome and could clearly discriminate the cultivated accessions from wild relatives. The cDNA-derived SSR markers are suitable for evaluation of genetic diversity, linkage and association mapping studies and genome-based breeding programmes. Occurrence of SSR repeats in biologically-important classes of proteins such as kinases, transferases and transcription factors was inferred from functional analyses, which along with variability of their repeat copies, can endow the cell and whole organism the flexibility of facing with continuously changing environment, and indicate a structure-based evolution mechanism of the genome which acts as an up-to-dating tool for the cell and whole organism, which is realized in GO terms such as involvement of most SSR-containing genes in biological, cellular and

metabolic processes, especially in response to stimulus, response to stress, interaction to other organisms and defense responses.

KEYWORDS

blast, genic SSR, genome-based breeding, GO enrichment, *in silico*

Introduction

Safflower (*Carthamus tinctorius*, $2n=2x=24$) is ordinarily related to tribe (*Cynareae*), subfamily (*Tubulifloreae*), and a member of the *Asteraceae* family. The east of the Mediterranean region has known as an origin center for the genus (Ashri and Knowles, 1960). Vavilov (1951) proposed three centers of safflower's origin including India, Afghanistan, and Ethiopia; whereas Knowles (1969) defined seven origin center for safflower including Far East, India and Pakistan, the Middle East, Egypt, Sudan, Ethiopia and the last one in Europe (Knowles, 1969). The size of haploid genome of the genus '*Carthamus*' ranges from 1.36 GB (*C. gypsicola*) to 1.41 GB (*C. palaestinus*) (Garnatje et al., 2006). Cultivated safflower (*C. tinctorius*) is being widely grown across the world and globally is considered as an important crop because of its multi-purpose applications including edible oil products, red dye, and medicinal usages and the use in livestock feeding (Wu et al., 2021). In the mid last century, its recognition as a qualified oil source containing high-unsaturated fatty acid in comparison to other oil crops and industrial usages caused it to be widely cultivated in Asia, Europe, Australia and North America (Knowles, 1955; Weiss, 1971). It is especially popular for its high linoleic acid (LA) and flavonoid contents (Wu et al., 2021). Its two unsaturated fatty acids (linoleic acid (LA) and oleic acid (OA)) are essential for human health; for example it was reported that LA decreases serum cholesterol and risk of cardiovascular events (Ramsden et al., 2016; Marklund et al., 2019). On the other hand, the olive oil which is rich in OA, has a beneficial effect on cancer, autoimmune and inflammatory diseases, and also facilitates wound healing (Sales-Campos et al., 2013).

Different types of molecular markers (mostly arbitrary) have been used to reveal genetic diversity in safflower by several investigators; For example, random amplified polymorphic

DNA (RAPD) markers (Sehgal and Raina, 2005; Amini et al., 2008; Khan et al., 2009; Rehman et al., 2015; Talebi and Abhari, 2016), inter simple sequence repeat (ISSR) markers (Sabzalian et al., 2009; Bagmohammadi et al., 2012; Majidi and Zadhoush, 2014), amplified fragment length polymorphism (AFLP) markers (Sehgal and Raina, 2005; Johnson et al., 2007), retrotransposon markers (Ali et al., 2019) and single nucleotide polymorphism (SNP) markers (Zhao et al., 2021). Likewise, Mokhtari et al. (2013) and Golkar and Mokhtari (2018) used the sequence-related amplified polymorphism (SRAP) and start codon targeted (SCoT) markers to determine genetic variation and similarity among safflower species. Moreover, simple sequence repeat (SSR) markers have also been applied in safflower by other researchers for different purposes. For example, Kiran et al. (2017) used 44 SSR markers for genetic diversity evaluation, linkage disequilibrium and population structure analysis among 148 accessions of safflower. Moreover, assessment of genetic distinctiveness among Indian and Mexican cultivars of safflower (Kadirvel et al., 2017), and association mapping of agronomic traits (Ambreen et al., 2018) were reported.

Recently, genic-SSR or expressed sequence tag (EST)-SSR markers were used for investigating the genetic diversity of different crop plants. Developing EST-SSRs through data mining acquired from publicly available databases has made them fast, efficient and moderately inexpensive markers compared to the genomic-SSR markers, as well as their transferability across species (Gupta et al., 2003; Chapman et al., 2009; Singh et al., 2021). The microsatellites existing in the genic regions have a direct role in the genome organization, recombination, gene regulation, quantitative genetic variation, and evolution of genes (Hao, 2018). Gene-based SSRs are sometimes assigned putative function and hence also termed as functionally relevant SSRs (Mishra et al., 2021). Therefore, numerous researches attempted to develop and utilize genic-SSRs in multiple crops. For instance, the genic-SSRs were developed and reported for sugarcane (Cordeiro et al., 2001), barley, maize, rice, sorghum and wheat (Kantety et al., 2002; Gupta et al., 2003; Yu et al., 2004; Peng and Lapitan, 2005; Fu et al., 2006; Varshney et al., 2006), cotton (Park et al., 2005; Han et al., 2006), coffee (Aggarwal et al., 2007), citrus (Chen et al., 2006); barrel medic (Gupta and Prasad, 2009), sorghum (Srinivas et al., 2009), pigeonpea (Dutta et al., 2011), finger millet (Arya et al., 2013; Pandian et al., 2018), mungbean (Gupta et al., 2014), Japanese larch

Abbreviations: AFLP, amplified fragment length polymorphism; BP, biological process; BLAST, basic local alignment search tool; CC, cellular component; EST, expressed sequence tag; GO, gene ontology; ISSR, inter simple sequence repeat; KEGG, Kyoto encyclopedia of genes and genomes; LA, linoleic acid; MF, molecular function; OA, oleic acid; SSR, simple sequence repeat; QTL, quantitative trait loci; RAPD, random amplified polymorphic DNA; SCoT, start codon targeted; SRAP, sequence-related amplified polymorphism.

(Chen et al., 2015), mulberry (Thumilan et al., 2016), lettuce (Oliya et al., 2018). In a few investigations the genic-SSRs were reported for oil crops such as sunflower (Pashley et al., 2006), soybean (Hisano et al., 2007), cultivated groundnut (Mondal et al., 2012; Zhang et al., 2012) and sesame (Zhang et al., 2012). Nevertheless, a limited number of studies investigated EST-SSRs in safflower. For instance, more recently Singh et al. (2021) identified EST-based SSRs and used them for assessment of genetic diversity and cross-species transferability in safflower. The use of such markers may be due to their proper characters such as hyper-variability, multi-allelic nature, co-dominant inheritance, reproducibility, relative abundance, intensive genome coverage, chromosome-specific location, feasibility and high-throughput genotyping (Parida et al., 2009). Chapman et al. (2009) generated polymorphic EST-SSRs in safflower and assessed their cross-taxon utility across the genus and the family. EST-SSRs have also used for genetic purity analysis of safflower hybrids, as well (Naresh et al., 2009). Whereas, limited studies were available for the development of genic-SSR markers from safflower cDNA sequences in which the complete length of coding sequences of genes must be screened for presence of SSRs. The microsatellites located in coding regions of genes may impact the coded protein activity and thus can affect the expression of proteins, while their presence in non-coding regions (5'-UTR, 3'-UTR and introns) may be involved in gene regulation and transcription (Lawson and Zhang, 2006). According to the above, the objective of the present study was to develop novel cDNA-SSRs from safflower's cDNA sequences. Furthermore, we also aimed to characterize the genes containing these tandem repeats and their functions by molecular analyses.

Materials and methods

SSR detection

Initially, the full length cDNA sequences data set (39,718 entries) of safflower (*Carthamus tinctorius*) was downloaded from national center for biotechnology information (NCBI) website (www.ncbi.nlm.nih.gov). Subsequently, for detecting SSRs ≥ 18 nucleotides in length within the subject sequences, the R Biostrings package (matchPDict function) was used (the R scripts are presented in [Supplementary File S1](#)). Mononucleotide and all possible forms of dinucleotide (NN), trinucleotide (NNN), tetranucleotide (NNNN), pentanucleotide (NNNNN) and hexanucleotide (NNNNNN) repeats were extracted and stored in.txt format for further analyses.

Sequence characterization

The FGENESH (<http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>) or GENESCAN (<http://hollywood.mit.edu/GENSCAN.html>)

software were used to predict existence of genes and putative coding open reading frames (ORFs) in cDNA sequences. The search options in both software were based on Arabidopsis code table, and the rest of the options were all set to default. The basic local alignment search tool for proteins (BLASTp) (protein BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used for identification of putative homologs of genes in protein databases. Consequently, the BLASTp hits with at least 65% coverage and 70% identity were chosen for further analysis.

gProfiler online tool (<http://biit.cs.ut.ee/gprofiler/gost/>) was used for gene ontology (GO) enrichment. For this purpose, *Arabidopsis thaliana* was selected as a model organism. The GO analysis for genes was done according to molecular function (MF), cellular component (CC), biological process (BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. An adjusted p value of <0.05 was considered for denoting significance. Analysis of promoter regions of the given genes was conducted using PlantPAN (<http://plantpan.itps.ncku.edu.tw/>). The STRING database (<https://string-db.org/>) was used for interaction analysis of proteins.

SSR primers, PCR assay and physical mapping

For validation of the identified SSR sequences and their use in subsequent analyses, 38 primer pairs were designed based on the flanking regions of SSR motifs ([Supplemental Table S1](#)). Where the SSR motif lied in a marginal boundary of cDNA sequence, the cDNA sequence was aligned to newly published genome sequence of safflower (Wu et al., 2021) (<https://safflower.scuec.edu.cn/>) to obtain the flanking SSR sequences to be used for primer design. Total DNA was extracted by cetyltrimethylammonium bromide (CTAB) method (Dellaporta et al., 1983) from seedlings of 11 safflower accessions including 9 cultivated accessions (*Carthamus tinctorius*) and 2 wild relatives (*C. lanatus* collected from Tehran Province, and *C. oxyacanthus* collected from Golestan Province). Polymerase chain reaction (PCR) reactions were done in 0.2 mL micro tubes. PCR mixture included 5 μ L of deionized water, 6 μ L PCR Master Mix (CinaClone Co.), 0.25 μ L of each primer (10 μ g) and 0.75 μ L of DNA template (15 ng). The PCR temperature profile was as following: amplification started at 94°C for 5 min and continued for 35 cycles (each cycle consisted of denaturation at 94°C for 35 s, annealing at 55°C for 30 s, extension at 72°C for 1 min) and the extension was done at 72°C for 7 min. The PCR products were separated by electrophoresis in 3% agarose gels and 1x TBE buffer containing 0.5 μ g mL⁻¹ of ethidium bromide. DNA bands became visible after staining by ethidium bromide and their photos were taken under ultraviolet light by a Gel-Doc system (BioRad, USA).

The physical location of SSR sequences was defined by searching the primer pairs and internal sequences against

newly published safflower genome (Wu et al., 2021) (<https://safflower.scu.ec.edu.cn/>) using R Biostrings package, matchProbePair and matchLRPatterns functions. The map was drawn using MapChart v.2 (Voorrips, 2002).

Results

Identified cDNA-SSRs

In 39,817 available safflower cDNA clones, 1,841 (4.62%) microsatellite repeats with SSR length of ≥ 18 n were detected (Supplemental Table S2). Frequency of the extracted repeats based on their types in the safflower cDNA sequences was different: 7 (0.38%) mononucleotides, 405 (22.0%) dinucleotides, 658 (35.74%) trinucleotides, 128 (6.95%) tetranucleotides, 98 (5.32%) pentanucleotides, and 545 (29.6%) hexanucleotides (Figure 1A). Among the cDNA sequences, the repeat numbers of SSRs varied from 3 to 28, but repeat numbers of 3-9 were predominant in the SSRs (72.6% of the total SSRs). Repeat numbers of 10-15 accounted for 22.9% of SSRs, repeat numbers of 15-20 accounted for 3.3% of SSRs and repeat numbers more than 20 just were accounted for 1.1% of the SSRs (Figure 1B).

Also we found 1, 4, 18, 20, 32 and 115 motif types of mono-, di-, tri-, tetra-, penta- and hexanucleotides in cDNAs of safflower (Supplemental Table S2). Longest SSR regions consisted of hexanucleotide (TCCATC)_n (with 16 and 18 repeats) which were found in cDNA clones EL382716.1 and EL381184.1, respectively. Next, AG and TC with 28 repeats were longest dinucleotide SSRs (in EL391333.1 and EL383067.1 clones, respectively). The more abundant motifs within each repeat type were also different. For example, as shown in Table 1, among all of 4 possible mononucleotide repeat types [(A)_n, (T)_n, (C)_n and (G)_n], only (T)_n was found in safflower cDNAs, while in the case of dinucleotide SSRs, (TC)_n and (AG)_n motifs were

prevalent. The motifs such as (TTC)_n, (ATC)_n and (AAG)_n in trinucleotide SSRs, (TTTG)_n and (ATTC)_n in tetranucleotide SSRs, (TTCTC)_n and (AATTC)_n in pentanucleotide SSRs, and lastly, the (TCCATC)_n and (GATGGA)_n motifs in hexanucleotide SSRs were more abundant repeats among all expected motifs for the repeat types (see Supplemental Table S2).

In total, 1,841 SSR regions were found in 1,667 cDNA clones (e.g. some cDNA sequences harbored more than one SSR region). The distribution of SSR regions is shown in Table 2. All mono-, tri-, tetra- and pentanucleotide SSRs (in total 1,434) were perfect (SSR repeat length ≥ 20 n), and only a small portion of dinucleotide SSRs (4 out of 405, <1%) were imperfect (SSR repeat length <20 n), in contrast most of hexanucleotide SSRs (403/545, ~74%) were imperfect (Table 3). All cDNAs which contained mono, tetra- and pentanucleotide motifs (in total 233) carried just one SSR region and in contrast, the cDNAs containing di-, tri- and hexanucleotide motifs carried one or more SSR regions (Table 3).

Results of BLAST searches

1,721 out of 1,841 SSR-containing cDNA sequences, could be identified as protein-coding sequences by ORF prediction tools. Peptide sequences obtained by the software were used as input for BLASTp searches. From these SSR-containing protein-coding sequences, 1,613 sequences showed significant hits with existing genes in database sequences, and after removing duplicates, 713 genes with unique gene IDs were selected for subsequent analyses (Supplemental Table S3).

652 out of 1,613 SSR-containing genes could be classified in 15 protein classes (Figure 2). Kinases, transferases and transporters were the most abundant protein classes in BLASTp hits. Transcription factors (TFs) also were more abundant protein classes in BLASTp hits, from which bHLH, bZIP, MYB and NAC domain containing proteins were most

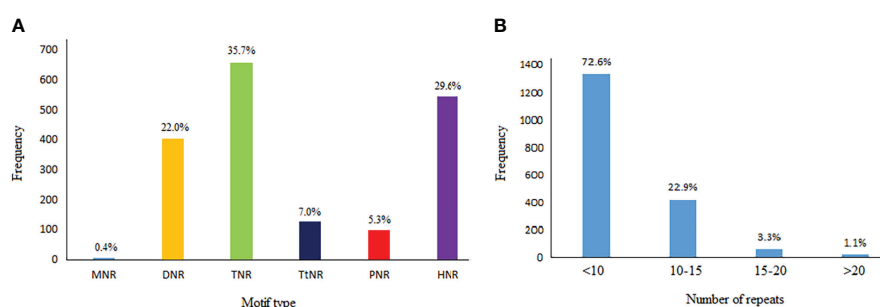


FIGURE 1

(A) Frequency distribution of 6 SSR motif types (viz. mono- to hexanucleotide repeats) in cDNA sequences of safflower. (B) The number of repeats vs. frequency of different motifs in identified SSR sequences. MNR, mononucleotide repeat; DNR, dinucleotide repeat; TNR, trinucleotide repeat; TtNR, tetranucleotide repeat; PNR, pentanucleotide repeat; HNR, hexanucleotide repeat.

TABLE 1 Number and frequency of different types of SSR motifs in safflower cDNA sequences (SSR length \geq 18 bp).

Repeat type (total number)	Motif	Frequency	%
Mononucleotide (7)	(T)n	7	100.0
	Other	0	0.0
Dinucleotide (405)	(TC)n	228	56.3
	(AG)n	83	20.5
	(AC)n	66	16.3
	Other	28	6.9
Trinucleotide (658)	(TTC)n,	212	32.2
	(ATC)n	127	19.3
	(AAG)n	108	16.4
	Other		42.1
Tetranucleotide (128)	(TTTG)n	56	43.8
	(ATTC)n	16	12.5
	Other	56	43.7
Pentanucleotide (98)	(TTCTC)n	18	18.4
	(AATTC)n	14	14.3
	Other	66	77.3
Hexanucleotide (545)	(TCCATC)n	32	5.9
	GATGGA	27	5.0
	(AACCT)n	25	4.6
	Other	461	84.5

Total number of each repeat type was written in parenthesis.

abundant TFs (Supplemental Table S3). However, the uncharacterized and hypothetical proteins were the other major portions of the analyzed SSR-containing genes that occupied 8.9% and 4.2%, respectively.

Gene ontology and KEGG enrichment of SSR-containing genes

Using gProfiler tool all genes with known IDs were analyzed and only 173 genes containing mono-, di- or tri-nucleotide SSRs had hits in GO analysis and hence, were enriched for GO terms. Two genes in mononucleotide type, 145 genes in dinucleotide type, and 26 genes in trinucleotide type were enriched for GO terms. The GO-enriched genes were classified into different

categories including molecular function (MF), biological process (BP) and cellular component (CC). 160, 157 and 157 genes were enriched for MF, BP and CC, respectively. In total, 322 GO terms were assigned to 173 genes: 55 GO terms for MF, 194 GO terms for BP and 73 GO terms for CC (Figure 3). The GO enrichment analysis and subclasses within each group with highest significant hits are presented in Supplemental Table S4. The results of gene ontology showed that the GO enriched genes mostly involve in biological (GO:0008150, 160 genes), cellular (GO:0009987, 132 genes) and metabolic processes (GO:0008152, 101 genes). Among molecular functions, binding (GO:0005488, 122 genes), catalytic activity (GO:0003824, 85 genes) and protein binding (GO:0005515, 78 genes), and among biological processes, organic substance metabolic process (GO:0071704, 96 genes), cellular metabolic process (GO:0044237, 94 genes),

TABLE 2 Distribution of SSR repeats in cDNA clones.

SSR repeat	Total No.	No of cDNAs harboring SSRs	% cDNAs
Mononucleotide	7	7	100
Dinucleotide	405	396	97.8
Trinucleotide	658	635	96.5
Tetranucleotide	128	114	89.1
Pentanucleotide	98	93	94.9
Hexanucleotide	545	422	77.4
Sum	1,841	1,667	

TABLE 3 Distribution of perfect and imperfect SSRs, and cDNAs with 1 or more SSR regions in different types of repeated motifs.

SSR repeat	Imperfect	Perfect	cDNAs with 1 SSR region	cDNAs with >1 SSR region
Mononucleotide	0	7	7	0
Dinucleotide	4	401	395	10
Trinucleotide	0	658	648	10
Tetranucleotide	0	128	128	0
Pentanucleotide	0	98	98	0
Hexanucleotide	403	142	464	81
Sum	407	1434	1740	101

primary metabolic process (GO:0044238, 88 genes) and nitrogen compound metabolic process (GO:0006807, 80 genes) were more important subclasses; response to stimulus (GO:0050896) and biological regulation (GO:0065007) also are important biological processes. Among cellular components, cellular anatomical entity (GO:0110165, 152 genes), intracellular anatomical structure (GO:0005622, 136 genes) and organelle (GO:0043226, 124 genes) have highest importance. Note that the importance of given term(s) was deduced in view of the number of hits (genes listed for each class or subclass) in GO analysis. In KEGG enrichment analysis 67 genes were enriched in 8 different pathways, more important of which were metabolic pathways (KEGG:01100, including lipid metabolism and plant terpenoid biosynthesis), ubiquitin mediated proteolysis (KEGG:04120) and endocytosis (KEGG:04144) (Supplemental Table S4).

Since trinucleotide repeats are of functional importance (Krzyszosiak and Napierala, 2011; Hao, 2018) the downstream

analyses of the genes containing this type of repeats were conducted. Interestingly, all trinucleotide SSR regions which were found in 635 cDNAs, harbored perfect trinucleotide motifs (SSR length ≥ 20 n). 563 trinucleotide-containing genes showed significant hits in BLASTp, 258 (45.8%) of which could be grouped into 16 protein classes (Figure 4). Kinases, proteasome-related proteins and transferases were the most abundant protein classes which contained trinucleotide repeats in their respective coding sequences (CDS). ATPases and transcription factors (TFs) occupied next locations.

GO enrichment analysis indicated that most (77.8%) of annotated trinucleotide-containing genes have binding activity, followed by protein binding activity (55.6%) (Figure 5). 100% of the genes involve in biological processes, followed by cellular (74.7%) and metabolic (66.7%) processes. The considerable portion of genes also involve in organic substance metabolic process and response to stimulus (66.7%), followed by primary metabolic process and biological regulation (62.9%). Cellular

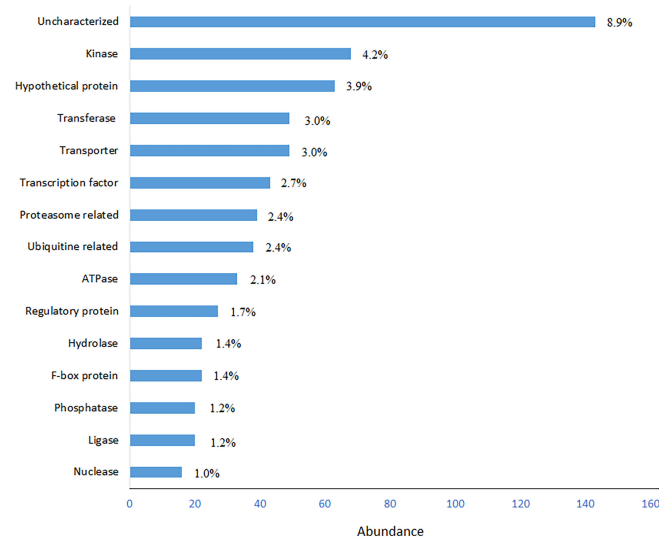
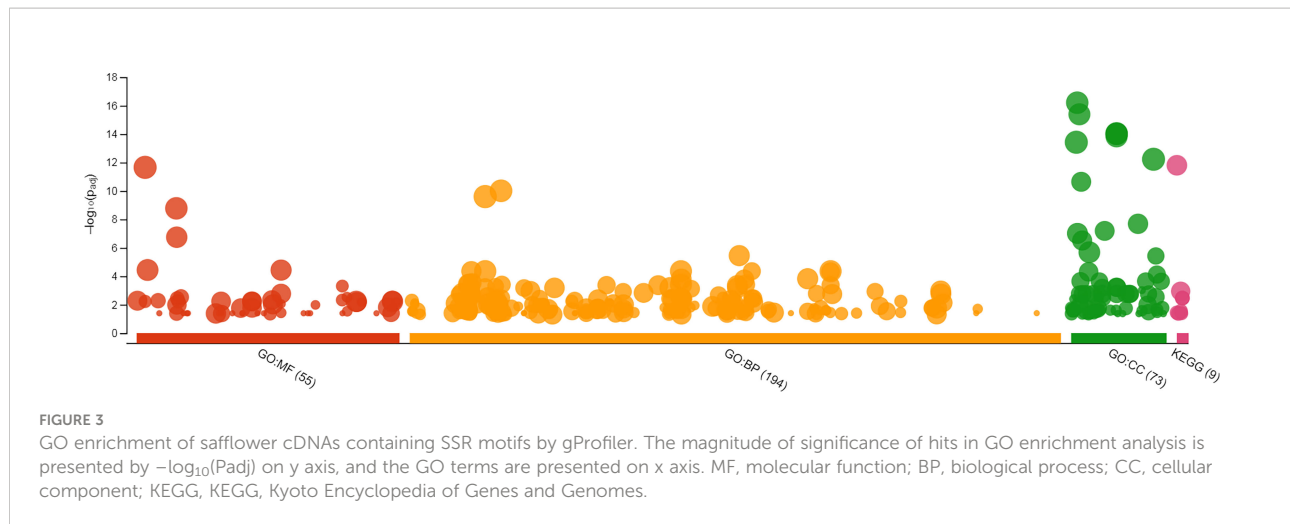


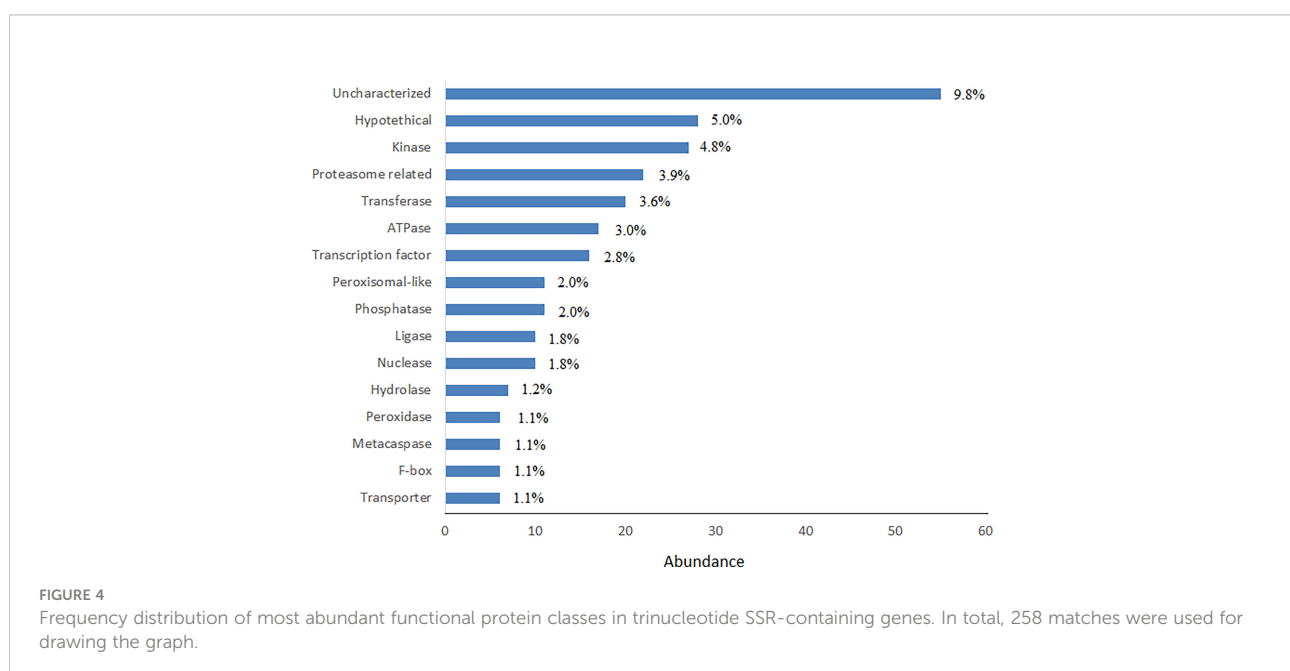
FIGURE 2

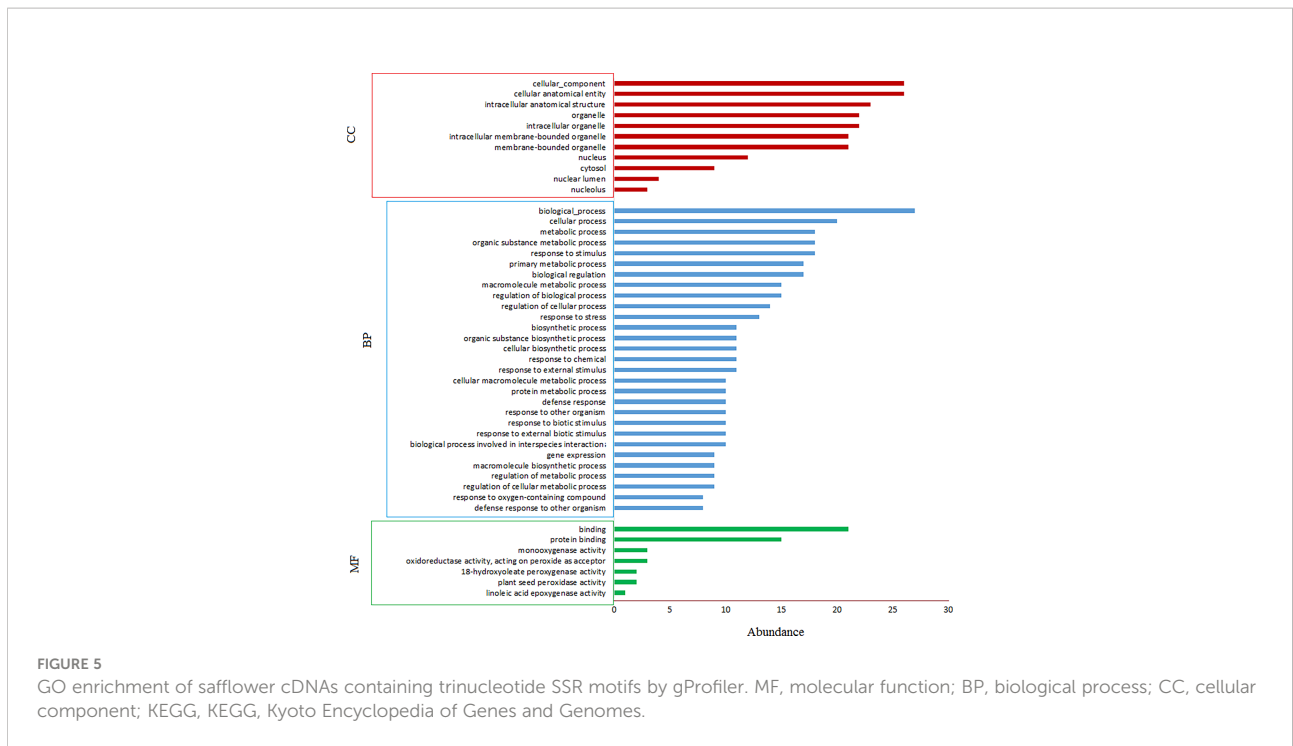
Frequency distribution of most abundant BLASTp hits of SSR-containing genes. Functional classes with ≥ 10 hits (in total, 652 hits) were used for drawing the graph.



anatomical entity, intracellular anatomical structure and organelle are more important cellular targets of the coded proteins of trinucleotide-containing genes. Interestingly, more than 44% of the proteins are localized to the nucleus (Figure 5). Based on KEGG analysis, two pathways were enriched for trinucleotide-containing genes including cutin, suberine and wax biosynthesis (KEGG:00073) and endocytosis (KEGG:04144) pathways. Because of more probable role of the first pathway in defense reactions, we would pay more attention to it. The core genes in the pathway include *Caleosin 3* (*RD20*) and *peroxygenase 2* (*PXG2*), both genes belong to Caleosin superfamily. A single cDNA clone in safflower (EL385976.1) which is mapped on chromosome 4 (at position 13.584-13.590 Mbp) carries *Caleosin 3*, and four cDNA clones (EL385118.1,

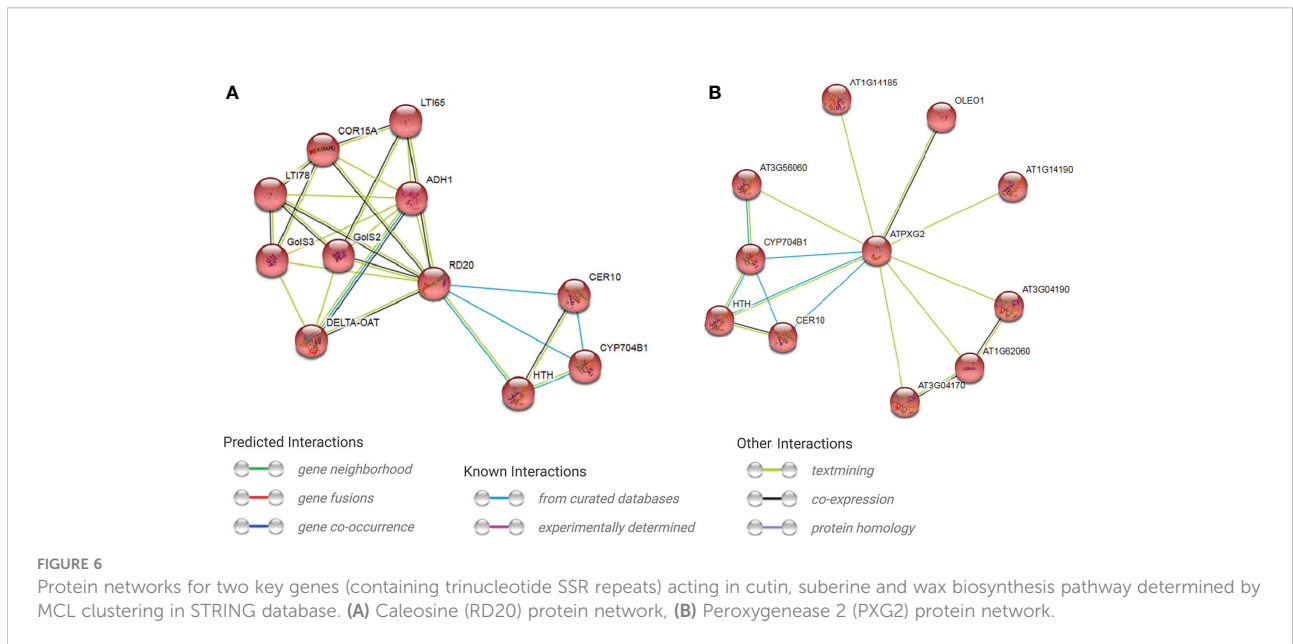
EL391927.1, EL400831.1 and EL409620.1) mapped on reverse strand of chromosome 6 (at position 17.892-17.931 Mbp) carry *peroxygenase 2* (Supplemental Table S5). Protein interaction analysis reveals the protein networks of these two genes using STRING database (Figures 6A, B). As seen in the figure, both interaction networks/modules share 3 genes including CER10 (3-oxo-5- α -steroid 4-dehydrogenase), HTH (Glucose-methanol-choline (GMC) oxidoreductase) and CYP704B1 (a Cytochrome P450). CER10 catalyzes the last of the four reactions of the long-chain fatty acids elongation cycle. It participates in the production of very long-chain fatty acids (VLCFAs) of different chain lengths that are involved in multiple biological processes as precursor. HTH has similarity to the mandelonitrile lyase family of FAD containing oxidoreductases





and is predicted to be secreted. It originally was identified as a mutation that causes floral organs to fuse together. CYP704B1 catalyzes omega-hydroxylation of long-chain fatty acids (LCFAs), implicating these molecules in sporopollenin synthesis. It is expressed in the developing anthers and is essential for pollen exine development. The complete annotations and descriptions of the proteins in these two modules presented in Supplemental Table S6. The analysis of

RD20 promoter region in PlantPAN (<http://plantpan.itps.ncku.edu.tw/>) showed that its promoter have cis-element binding sites for 3 transcription factors (viz. *ANAC019*, *ATHB-12* and *RD26*) with relatively high probability (>70%); the co-expression analysis also indicates all of the TFs are co-expressed with *RD20* (Supplemental Table S5). The *PXG2* promoter analysis showed that its promoter harbor many cis-element binding sites for 17 transcription factors (such as *PIL5*, *NF-AY8*, *WRKY 32*



and *OPB1*) with a high probability (>90%), which also showed co-expression with it (Supplemental Table S5).

Validation of SSR markers

Based on cDNA sequences which contained SSR motifs, thirty eight primer pairs were designed, synthesized and used in PCR reactions for amplification of DNA of 11 safflower accessions (9 cultivated accessions of *C. carthamus* species and 2 wild relatives including *C. lanatus* and *C. oxyacanthus* species). Three primer pairs were amplified only in 4 or 5 samples (which discarded in downstream analyses), and 35 primer pairs were amplified in majority of samples, from which 22 primer pairs produced polymorphic patterns in safflower accessions.

However, 12 out of these 22 SSRs were monomorphic among cultivated accessions (e.g. 26.3% were polymorphic among cultivated accessions). The polymorphic SSR markers had 2 to 3 alleles (Supplemental Table S7). All SSR markers were physically mapped on the newly published safflower genome by sequence alignment of their carrying cDNA sequences (Figure 7A). The safflower accessions were grouped into two distinct groups based on two-sided clustering by heatmap analysis (Figure 7B). All cultivated accessions placed in one group (g1) and 2 wild relatives (*C. lanatus* collected from Tehran Province and *C. oxyacanthus* collected from Golestan Province) obviously placed in a separate group (g2). The cultivated group g1 itself was clustered into two subgroups (subgroup g1.1: LRV-5151 and 34074, and subgroup g1.2: Arak 2811, mutant Arak 2811, Pi-50537, 3404, 541-5, Aceteria and Dinger).

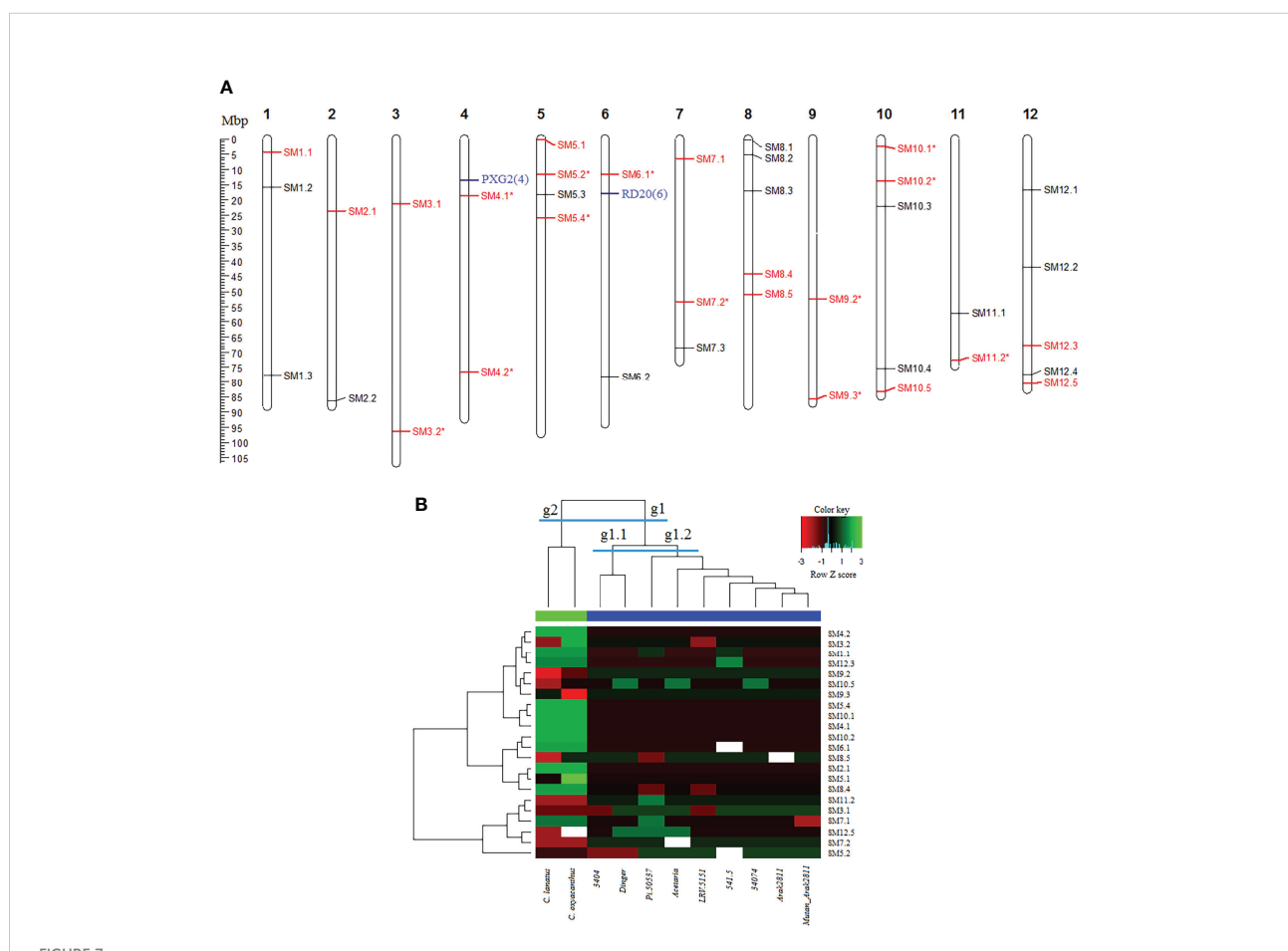


FIGURE 7

Analysis and validation of SSR markers in safflower. (A) Physical map positions of 38 cDNA-SSR markers in safflower genome. The map positions are in Mbp. The SSR marker names are presented on right-hand of each chromosome. Polymorphic markers are in red and the markers showing a monomorphic pattern among cultivated accessions are shown with an asterisk. Two key genes harboring trinucleotide SSRs are shown in blue. (B) Grouping of the cultivated and wild relative accessions of safflower by 22 polymorphic cDNA-SSR markers using two-sided clustering of heatmap method. Two distinct groups are defined: group g1 contains all cultivated accessions and group g2 contains 2 wild relatives (*C. lanatus* and *C. oxyacanthus*). The group g1 itself is clustered into two subgroups (subgroup g1.1 containing LRV-5151 and 34074, and subgroup g1.2 containing Arak 2811, mutant Arak 2811, Pi-50537, 3404, 541-5, Aceteria and Dinger). Green color: higher Z score values, red color: lower Z score values.

Discussion

In self-pollinated species like safflower there are few studies on genome structure and features, resulting to a limited number of SSR markers until now and hence, the exploitation of the marker-assisted breeding (MAB) potential is limited. For instance, 509 putative genomic SSR markers were acquired *via* next-generation sequencing (NGS), thirty of which were assessed in a diverse collection of safflower (Lee et al., 2014). In a more comprehensive NGS effort, Ambreen et al. (2015) identified 23,067 SSR regions in whole-genome sequences and validated 294 of them in a set of 23 safflower accessions. However, the identification of microsatellite sequences in genic regions of safflower, which can be done by *in silico* analysis, rarely was reported. These genic-based developed markers are functional molecular markers and cost-effective; because of their presence in conserved and semi conserved genic regions of the genome, they are expected to have high inter-specific transferability (Scott et al., 2000; Varshney et al., 2005). Genic SSR markers are the markers of choice and more robust in compare to non-genic SSRs for genetic investigation and QTL mapping (Agarwal et al., 2015). Genic SSRs, particularly cDNA-derived ones, are more reliable markers to be used in linkage and association mapping studies because they are not prone to recombination events between marker and associated trait, which decreases the occurrence of false positives in marker-assisted selection programs (Gupta et al., 2010).

Based on BLASTp results, >93% of SSR-containing cDNAs showed high similarity with the existing database sequences and only around 7% of them did not show any similarity with the existing database sequences. Similar results were reported in mulberry transcripts by Thumilan et al. (2016). 1,407 (81.5%) of cDNA-SSR sequences analyzed here had well-defined function including major protein classes such as kinase, transferase, transporter, transcription factor and regulatory protein (Figure 2). Thumilan et al. (2016) conducted a screening of mulberry EST sequences and identified SSR sequences in kinases, transcription factors, transporters and ribosomal proteins. Most frequent SSR-containing transcription factors in our study were bHLH, bZIP, NAC and MYB. Tranbarger et al. (2012) also identified SSRs in regulatory transcripts of oil plum (*Elaeis guineensis* Jacq.) including bZIP, zinc finger, MADS-box, and NAC-like transcription factors. Similar results reported by others in chickpea (Kujur et al., 2013), *Medicago truncatula* (Liu et al., 2015), mango (Mahato et al., 2016), jute (Satya et al., 2017) and tea (Parmar et al., 2022).

Repeat number of SSRs varied from 3 to 28, but repeat numbers of <10 were predominant in cDNA sequences. Similar results were reported for *Eucalyptus* genic SSRs by Liu et al. (2018). The effect of repeat number of SSRs in genic regions on the phenotypic performance of the genes was not studied

comprehensively in plants. A preliminary study on the effect of repeat number of SSRs showed that repeat number of CT motif in rice *Waxy* gene affected amylose content of seed endosperm (Prathepha, 2003), but more recently its effect on starch properties of endosperm was reported as a suspicious issue (Li et al., 2017).

In present research the TC, AG and AC motifs were the most abundant dinucleotide motifs, and TTC, ATC and AAG were the most abundant trinucleotide motifs (Table 1). Liu et al. (2018) also reported that among dinucleotide motifs, TC (21.3%) and AG (20.7%) were the most frequent SSR motifs in *Eucalyptus*, which is consistent with the results of present research. In tea transcripts, AG, AT and AC motifs were most prevalent among the dinucleotide repeats, and AAG, ACC and ATC motifs were the most abundant trinucleotide repeats (Parmar et al., 2022). As the results of the present study, the motifs with thymine (T) and cytosine (C) bases showed the highest contribution in all of the identified motifs (see Supplemental Table S2). This issue also reported by Zhou et al. (2016) in the *Dipteronia oliver*, which could be realized among the most abundant microsatellite repeats such as dinucleotide and trinucleotide motifs.

The trinucleotide motifs were most abundant SSR type (35.7%) followed by hexa- and dinucleotide motifs (29.6 and 22.0%, respectively) in safflower SSR-containing cDNAs, which these results are similar to those found in *Larix kaempferi* by Chen et al. (2015), in which the trinucleotide motifs were the most abundant (27.3%), followed by hexa- and dinucleotide motifs, respectively. Interestingly, it was reported that the trinucleotide repeats were more frequent in coding sequences of the human genes too (Krzyzosiak and Napierala, 2011). In contrast, mono- and dinucleotide motifs were predominant in EST-SSRs obtained by RNA-seq analysis of safflower under drought stress (92.30% and 5.12%, respectively) (Sudhakar et al., 2019). This difference may be due to the use of all cDNA sequences for deriving SSR motifs in our research and the use of only drought responding transcripts for deriving SSR motifs in the mentioned work. Parmar et al. (2022) analyzed tea transcripts and reported highest abundance of di- (61.5%) and trinucleotide SSR repeats (34.8%). In coding regions, but the small indels generated by SSRs can cause frameshifts that result in truncated or elongated proteins with disrupted protein function (Lin and Kussell, 2012). Although frameshift mutations are often regarded as deleterious, under fluctuating selection (a mode of natural selection characterized by the fluctuation of the direction of selection on a given phenotype over a relatively brief period of evolutionary time) but they can be advantageous (Kashi and King, 2006). However, trinucleotide repeats do not cause frameshifts when their repeat number is changed. Thus, they are better tolerated than other SSRs in translated sequences. When these repeats are long enough, they tend to form unusual DNA

structures that affect chromatin organization and DNA function (Krzyzosiak and Napierala, 2011).

Gene ontology analysis of SSR-containing genes indicated that 24.3% of genes with known gene IDs showed significant hits for GO terms. The GO enriched genes mostly involve in biological, cellular and metabolic processes; organic substance metabolic process (GO:0071704), cellular metabolic process (GO:0044237) and primary metabolic process (GO:0044238) were more important subclasses of cellular processes. These results partly were confirmed by KEGG enrichment analysis showing 41.8% of genes involve in metabolic pathways. Lulin et al. (2012) reported that most of safflower SSR-containing genes involve in various biological processes such as cellular and metabolic processes, which is consistent with the present study. In tea, involvement of most SSR-containing genes in metabolic processes and biological regulation was reported (Parmar et al., 2022). The most important molecular functions of the GO enriched genes in present study are the binding, catalytic activity and protein binding. Higher importance of the binding, enzymatic and transporter activities in EST-SSR sequences of safflower were reported earlier (Lulin et al., 2012). Similarly, the prevalence of catalytic, transcription regulation and transporter activities in SSR-containing transcripts of tea was reported (Parmar et al., 2022). GO enrichment of trinucleotide-containing genes almost followed the GO enrichment of total SSR-containing genes (see Supplemental Table S4 and Figure 5). The mild difference is in subclasses of cellular processes, where response to stimulus and biological regulation have more importance in the trinucleotide-containing genes compared to the total SSR-containing genes. Kinases, proteasome related proteins, transferases and transcription factors are among the more important classes of proteins that harbor trinucleotide SSRs in the encoding genes (Figure 4). Length variation of trinucleotide repeats within ORFs is thought to be the fast evolutionary track to generate novel protein sequences. Some functional classes of proteins, such as protein kinases and transcription factors, took advantage of this rapid means of change more often than other protein classes (Krzyzosiak and Napierala, 2011). However, it was suggested that DNA recombination, cell cycle, replication, DNA mismatch repair, and transcription are among the cellular processes impaired by long trinucleotide repeats (Krzyzosiak and Napierala, 2011).

KEGG analysis of trinucleotide SSR-containing genes showed an important biosynthetic pathway [cutin, suberine and wax biosynthesis (KEGG:00073)] with 2 key genes including *caleosin 3 (RD20)* and *peroxygenase 2 (PXG2)*, which are isoform and belong to Caleosin superfamily. According to GO analysis the genes are enriched for molecular functions such as calcium ion binding, peroxygenase, lipase, monoxygenase and plant seed peroxidase activities (data not shown). Caleosin/peroxygenases (*CLO/PXG*) comprise small gene families found in a wide range of plant species; these gene families have seven

members in Arabidopsis genome and five members in rice genome (Khalil et al., 2014), and are present in the vast majority of viridiplantae taxa (Rahman et al., 2018). In plants, structural proteins, particularly caleosins and oleosins affect the biological functions of seed oil bodies (OBs) (Shao et al., 2019). Protein interaction analysis to find probable existing networks consisted of the two protein RD20 and PXG2 was resulted in identification of two protein interaction networks/modules (Figure 6A, B), which are interconnected by 3 common members including CER10, HTH and CYP704B1 (CHC submodule). It seems that each of RD20/or PXG2 plays a vital role in its module and acts as an intermediary factor between the two sides of same network; one side consisted of a three-protein submodule CHC, and the second side consisted of a larger set of proteins which presumably along with RD20/or PXG2 determines the specified action of each network. These family members have roles in in seed germination (Poxleitner et al., 2006), lipid droplet storage and abiotic stress tolerance (Khalil et al., 2014).

Interestingly, the proteins RD20 and PXG2 are targeted to lipid bodies. On the basis of annotations and descriptions of the proteins depicted in STRING database (Supplemental Table S6), the biological role of the two modules can be understood. The genes in both modules mainly cooperate to manipulate/elongate seed oil bodies, particularly under environmental stresses. Biological function of these two modules apparently is endowment of tolerance to abiotic stresses such as cold, drought and salt. Rahman et al. (2018) suggested that algal CLO/PXGs play a roles in lipid packaging and stress responses, and Khalil et al. (2014) reported the differential expression of the caleosin gene family in wheat during different developmental stages and also in response to environmental stresses. Three transcription factors (*ANAC019*, *ATHB-12* and *RD26*) are co-expressed with *RD20* and bind to its promoter *via* *sis*-element binding sites (Supplemental Table S5). As reported, these TFs play a vital role in tolerance to abiotic stresses such as drought and salt (Ré et al., 2014; Hur et al., 2015; Ye et al., 2017; Jiang et al., 2019; Duan et al., 2019; Sukiran et al., 2019). Seventeen transcription factors (such as *PIL5*, *NF-AY8*, *WRKY 32* and *OPB1*) are co-expressed with *PXG2* and bind to its promoter region (Supplemental Table S5). Some of these TFs have important roles in gibberellin response and germination (Oh et al., 2004; Oh et al., 2007) and floral transition and flowering (Wu et al., 2018; Chandler and Werr, 2020), but most of them improve tolerance to abiotic stresses such as drought (Singh and Laxmi, 2015; Li et al., 2016; Zhang et al., 2017; Huo et al., 2021), salt (Li et al., 2016) and photooxidative stress (Hackenberg et al., 2012) or combined biotic and abiotic stresses (Bai et al., 2018).

Most of the assayed SSR primer pairs designed based on flanking regions of SSR motifs within cDNA sequences, were robust (35 out of 38, >90%) and successfully amplified the DNA of the studied accessions; more than 57% of these genic

SSRs showed polymorphism among safflower accessions. In the case of SSR markers developed based on whole genome sequencing, lower polymorphism rate (31.6%) was reported (Ambreen et al., 2015). Lower polymorphism rates in EST-SSRs and transcript-SSRs also was reported by others (Kujur et al., 2013; Mahato et al., 2016; Satya et al., 2017; Parmar et al., 2022). However, if the two wild relatives are discarded and just the cultivated accessions considered, the polymorphism rate of the SSR marker set is ~26%, which is near the rates reported by others. The polymorphic SSRs could clearly differentiate cultivated accessions from wild relatives which is consistent with classic morphological taxonomy (Figure 7B). This shows the suitability of cDNA-SSR markers for purposes of variety identification and plant molecular systematics (Sharifi Tehrani et al., 2009; Kordrostami et al., 2016; Liu et al., 2018; Zhang et al., 2020).

Conclusion

Occurrence of SSR repeats in biologically-important classes of proteins such as kinases and transcription factors, and variability of SSR repeat copies, endow the cell and whole organism the flexibility of facing with continuously changing endogenous and exogenous environment. The fact indicates a structure-based evolution mechanism of the genome, which acts as an up-to-dating tool for the cell to face with surrounding environment. This conclusion is realized in GO terms such as involvement of most SSR-containing genes, particularly trinucleotide-containing ones, in biological, cellular and metabolic processes, and especially in response to stimulus, response to stress, interaction to other organisms and defense responses. Present study also shows that cDNA-derived SSRs have high robustness and detect a relatively high polymorphism rate among safflower accessions and hence, are suitable for evaluation of genetic diversity, linkage and association mapping studies and genome-based breeding programmes.

References

- Agarwal, G., Sabbavarapu, M. M., Singh, V. K., Thudi, M., Sheelamary, S., Gaur, P. M., et al. (2015). Identification of a non-redundant set of 202 in silico SSR markers and applicability of a select set in chickpea (*Cicer arietinum* L.). *Euphytica* 205 (2), 381–394. doi: 10.1007/s10681-015-1394-3
- Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., and 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 (2), 359. doi: 10.1007/s00122-006-0440-x
- Ali, F., Yilmaz, A., Nadeem, M. A., Habyarimana, E., Subaşı, I., Nawaz, M. A., et al. (2019). Mobile genomic element diversity in world collection of safflower (*Carthamus tinctorius* L.) panel using iPBS-retrotransposon markers. *PLoS One* 14 (2), e0211985. doi: 10.1371/journal.pone.0211985
- Ambreen, H., Kumar, S., Kumar, A., Agarwal, M., Jagannath, A., and Goel, S. (2018). Association mapping for important agronomic traits in safflower (*Carthamus tinctorius* L.) core collection using microsatellite markers. *Front. Plant Sci.* 9, 402. doi: 10.3389/fpls.2018.00402
- Ambreen, H., Kumar, S., Variath, M. T., Joshi, G., Bali, S., Agarwal, M., et al. (2015). Development of genomic microsatellite markers in *Carthamus tinctorius* L. (safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis. *PLoS One* 10 (8), e0135443. doi: 10.1371/journal.pone.0135443
- Amini, F., Saeidi, G., and Arzani, A. (2008). Study of genetic diversity in safflower genotypes using agro-morphological traits and RAPD markers. *Euphytica* 163 (1), 21–30. doi: 10.1007/s10681-007-9556-6
- Arya, L., Verma, M., Gupta, V. K., and Seetharam, A. (2013). Use of genomic and genic SSR markers for assessing genetic diversity and population structure in Indian and African finger millet (*Eleusine coracana* (L.) Gaertn.) germplasm. *Plant Syst. Evol.* 299 (7), 1395–1401. doi: 10.1007/s00606-013-0822-x

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

AhA conducted the research and prepared the draft of manuscript. AsA conceptualized and supervised the research and edited the final manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.991107/full#supplementary-material>

- Ashri, A., and Knowles, P. F. (1960). Cytogenetics of safflower (*Carthamus l.*) species and their hybrids. *Agron. J.* 52, 11–17. doi: 10.2134/agronj1960.00021962005200010004x
- Bagmohammadi, H., Pahlevani, M., AhmadiKahah, A., and Razavi, S. E. (2012). Genetic variation of safflower (*Carthamus tinctorius* L.) and related species revealed by ISSR analysis. *Plant Breed. Seed Sci.* 66, 139–150. doi: 10.2478/v10129-011-0064-4
- Bai, Y., Sunarti, S., Kissoudis, C., Visser, R. G., and van der Linden, C. (2018). The role of tomato WRKY genes in plant responses to combined abiotic and biotic stresses. *Front. Plant Sci.* 9, 801. doi: 10.3389/fpls.2018.00801
- Chandler, J. W., and Werr, W. (2020). A phylogenetically conserved APETALA2/ethylene response factor, ERF12, regulates arabidopsis floral development. *Plant Mol. Biol.* 102 (1), 39–54. doi: 10.1007/s11103-019-00936-5
- Chapman, M. A., Hvala, J., Strever, J., Matvienko, M., Kozik, A., Michelmore, R. W., et al. (2009). Development polymorphic and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theor. Appl. Genet.* 120 (1), 85–91. doi: 10.1007/s00122-009-1161-8
- Chen, X. B., Xie, Y. H., and Sun, X. M. (2015). Development and characterization of polymorphic genic-SSR markers in *Larix kaempferi*. *Molecules* 20 (4), 6060–6067. doi: 10.3390/molecules20046060
- Chen, C., Zhou, P., Choi, Y. A., Huang, S., and Gmitter, F. G. (2006). Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet.* 112 (7), 1248–1257. doi: 10.1007/s00122-006-0226-1
- Cordeiro, G. M., Casu, R., McIntyre, C. L., Manners, J. M., and Henry, R. J. (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160 (6), 1115–1123. doi: 10.1016/S0168-9452(01)00365-X
- Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* 1 (4), 19–21. doi: 10.1007/BF02712670
- Duan, P., Chen, J., Chen, Z., Zhang, J., Han, H., and Yao, L. (2019). Expression induction of a class of RD26 genes by drought and salinity stresses in maize. *Biologia* 74 (9), 1237–1246. doi: 10.2478/s11756-019-00286-z
- Dutta, S., Kumawat, G., Singh, B. P., Gupta, D. K., Singh, S., Dogra, V., et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) millspaugh]. *BMC Plant Biol.* 11 (1), 17. doi: 10.1186/1471-2229-11-17
- Fu, Y. B., Peterson, G. W., Yu, J. K., Gao, L., Jia, J., and Richards, K. W. (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 (7), 1239–1247. doi: 10.1007/s00122-006-0225-2
- Garnatje, T., Garcia, S., Vilatersana, R., and Vallès, J. (2006). Genome size variation in the genus *carthamus* (Asteraceae cardueae): systematic implications and additive changes during allopolyploidization. *Ann. Bot.* 97 (3), 461–467. doi: 10.1093/aob/mcj050
- Golkar, P., and Mokhtari, N. (2018). Molecular diversity assessment of a world collection of safflower genotypes by SRAP and SCoT molecular markers. *Physiol. Mol. Biol. Plants* 24 (6), 1261–1271. doi: 10.1007/s12298-018-0545-0
- Gupta, S. K., Bansal, R., and Gopalakrishna, T. (2014). Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) wilczek. *Euphytica* 195 (2), 245–258. doi: 10.1007/s10681-013-0993-0
- Gupta, P. K., Langridge, P., and Mir, R. R. (2010). Marker-assisted wheat breeding: present status and future possibilities. *Mol. Breed.* 26, 145–161. doi: 10.1007/s11032-009-9359-7
- Gupta, S., and Prasad, M. (2009). Development and characterization of genic SSR markers in medicago truncatula and their transferability in leguminous and non-leguminous species. *Genome* 52 (9), 761–771. doi: 10.1139/G09-05
- Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol. Genet. Genomics* 270 (4), 315–323. doi: 10.1007/s00438-003-0921-4
- Hackenberg, D., Keetman, U., and Grimm, B. (2012). Homologous NF-YC2 subunit from arabidopsis and tobacco is activated by photooxidative stress and induces flowering. *Int. J. Mol. Sci.* 13 (3), 3458–3477. doi: 10.3390/ijms13033458
- Han, Z., Wang, C., Song, X., Guo, W., Gou, J., Li, C., et al. (2006). Characteristics development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor. Appl. Genet.* 112 (3), 430–439. doi: 10.1007/s00122-005-0142-9
- Hao, D. C. (2018). *Ranunculales medicinal plants: biodiversity, chemodiversity and pharmacotherapy* (Cambridge, Massachusetts, United States: Academic Press).
- Hisano, H., Sato, S., Isobe, S., Sasamoto, S., Wada, T., Matsuno, A., et al. (2007). Characterization of the soybean genome using EST-derived microsatellite markers. *DNA Res.* 14 (6), 271–281. doi: 10.1093/dnares/dsm025
- Huo, T., Wang, C. T., Yu, T. F., Wang, D. M., Li, M., Zhao, D., et al. (2021). Overexpression of *ZmWRKY65* transcription factor from maize confers stress resistances in transgenic arabidopsis. *Sci. Rep.* 11 (1), 1–15. doi: 10.1038/s41598-021-83440-5
- Hur, Y. S., Um, J. H., Kim, S., Kim, K., Park, H. J., Lim, J. S., et al. (2015). Arabidopsis thaliana homeobox 12 (*ATHB 12*), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytol.* 205 (1), 316–328. doi: 10.1111/nph.12998
- Jiang, H., Tang, B., Xie, Z., Nolan, T., Ye, H., Song, G. Y., et al. (2019). GSK 3-like kinase BIN 2 phosphorylates RD 26 to potentiate drought signaling in arabidopsis. *Plant J.* 100 (5), 923–937. doi: 10.1111/tpj.14484
- Johnson, R. C., Kisha, T. J., and Evans, M. A. (2007). Characterizing safflower germplasm with AFLP molecular markers. *Crop Sci.* 47 (4), 1728–1736. doi: 10.2135/cropsci2006.12.0757
- Kadirvel, P., Ravi, D., Mukta, N., Montoya-Coronado, M. C. L., Ghuge, S. B., Singh, J., et al. (2017). Genetic distinctiveness of safflower cultivars of India and Mexico as revealed by SSR markers. *Plant Genet. Resour.* 15 (6), 474–487. doi: 10.1017/S1479262116000186
- Kantety, R. V., La Rota, M., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley maize rice sorghum and wheat. *Plant Mol. Biol.* 48 (5–6), 501–510. doi: 10.1023/A:1014875206165
- Kashi, Y., and King, D. G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22, 253–259. doi: 10.1016/j.tig.2006.03.005
- Khalil, H. B., Brunetti, S. C., Pham, U. M., Maret, D., Laroche, A., and Gulick, P. J. (2014). Characterization of the caleosin gene family in the triticeae. *BMC Genomics* 15 (1), 1–5. doi: 10.1186/1471-2164-15-239
- Khan, M. A., von Witzke-Ehbrecht, S., Maass, B. L., and Becker, H. C. (2009). Relationships among different geographical groups agro-morphology fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius*). *Genet. Resour. Crop Evol.* 56 (1), 19–30. doi: 10.1007/s10722-008-9338-6
- Kiran, B. U., Mukta, N., Kadirvel, P., Alivelu, K., Senthilvel, S., Kishore, P., et al. (2017). Genetic diversity of safflower (*Carthamus tinctorius* L.) germplasm as revealed by SSR markers. *Plant Genet. Resour.* 15 (1), 1–11. doi: 10.1017/S1479262115000295
- Knowles, P. F. (1955). Safflower-production processing and utilization. *Econ. Bot.* 9 (3), 273–299. doi: 10.1007/BF02859823
- Knowles, P. F. (1969). Centers of plant diversity and conservation of crop germ plasm: Safflower. *Econ. Bot.* 23 (4), 324–329. doi: 10.1007/BF02860678
- Kordrostami, M., Rabiei, B., and Hassani Kumbleh, H. (2016). Association analysis genetic diversity and haplotyping of rice plants under salt stress using SSR markers linked to SalTol and morpho-physiological characteristics. *Plant Syst. Evol.* 302 (7), 871–890. doi: 10.1007/s00606-016-1304-8
- Krzyzosiak, W. J., and Napierala, M. (2011). “Structural characteristics of trinucleotide repeats in transcripts,” in *Genetic instabilities and neurological diseases*, vol. 31. Eds. R. D. Wells and T. Ashizawa (Cambridge, Massachusetts, United States, Elsevier).
- Kujur, ALICE, Bajaj, DEEPAK, Saxena, M. S., Tripathi, SHAILESH, Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20 (4), 355–374. doi: 10.1093/dnares/dst015
- Lawson, M. J., and Zhang, L. (2006). Distinct patterns of SSR distribution in the arabidopsis thaliana and rice genomes. *Genome Biol.* 7 (2), R14. doi: 10.1186/gb-2006-7-2-r14
- Lee, G. A., Sung, J. S., Lee, S. Y., Chung, J. W., Yi, J. Y., Kim, Y. G., et al. (2014). Genetic assessment of safflower (*Carthamus tinctorius* L.) collection with microsatellite markers acquired via pyrosequencing method. *Mol. Ecol. Resour.* 14 (1), 69–78. doi: 10.1111/1755-0998.12146
- Li, K., Bao, J., Corke, H., and Sun, M. (2017). Association analysis of markers derived from starch biosynthesis related genes with starch physicochemical properties in the USDA rice mini-core collection. *Front. Plant Sci.* 8, 424. doi: 10.3389/fpls.2017.00424
- Li, W., Mallano, A. I., Bo, L., Wang, T., Nisa, Z., and Li, Y. (2016). Soybean transcription factor GmNFYB1 confers abiotic stress tolerance to transgenic arabidopsis plants. *Can. J. Plant Sci.* 97 (3), 501–515. doi: 10.1139/cjps-2016-0158
- Lin, W. H., and Kussell, E. (2012). Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Res.* 40 (6), 2399–2413. doi: 10.1093/nar/gkr1078
- Liu, W., Jia, X., Liu, Z., Zhang, Z., Wang, Y., Liu, Z., et al. (2015). Development and characterization of transcription factor gene-derived microsatellite (TFGM) markers in medicago truncatula and their transferability in leguminous and non-leguminous species. *Molecules* 20 (5), 8759–8771. doi: 10.3390/molecules20058759
- Liu, G., Xie, Y., Zhang, D., and Chen, H. (2018). Analysis of SSR loci and development of SSR primers in eucalyptus. *J. Forest. Res.* 29 (2), 273–282. doi: 10.1007/s11676-017-0434-3

- Lulin, H., Xiao, Y., Pei, S., Wen, T., and Shangqin, H. (2012). The first illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7 (6), e38653. doi: 10.1371/journal.pone.0038653
- Mahato, A. K., Sharma, N., Singh, A., Srivastav, M., Singh, S. K., Singh, A. K., et al. (2016). Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety 'Amrapali' (*Mangifera indica* L. *PLoS One* 11 (10), e0164325. doi: 10.1371/journal.pone.0164325
- Majidi, M. M., and Zadhoush, S. (2014). Molecular and morphological variation in a world-wide collection of safflower. *Crop Sci.* 54 (5), 2109–2119. doi: 10.2135/cropsci2013.12.0850
- Marklund, M., Wu, J. H., Imamura, F., Del Gobbo, L. C., Fretts, A., De Goede, J., et al. (2019). Biomarkers of dietary omega-6 fatty acids and incident cardiovascular disease and mortality: an individual-level pooled analysis of 30 cohort studies. *Circulation* 139 (21), 2422–2436. doi: 10.1161/CIRCULATIONAHA.118.038908
- Mishra, N., Tripathi, M. K., Tiwari, S., Gupta, N., and Sharma, A. (2021). Validation of drought tolerance gene-linked microsatellite markers and their efficiency for diversity assessment in a set of soybean genotypes. *Current Journal of Applied Science & Technology* 40(25):48–57.
- Mokhtari, N., Rahimmalek, M., Talebi, M., and Khorrami, M. (2013). Assessment of genetic diversity among and within carthamus species using sequence-related amplified polymorphism (SRAP) markers. *Plant Syst. Evol.* 299 (7), 1285–1294. doi: 10.1007/s00606-013-0796-8
- Mondal, S., Badigannavar, A. M., and D'Souza, S. F. (2012). Development of genic molecular markers linked to a rust resistance gene in cultivated groundnut (*Arachis hypogaea* L. *Euphytica* 188 (2), 163–173. doi: 10.1007/s10681-011-0619-3
- Naresh, V., Yamini, K. N., Rajendrakumar, P., and Kumar, V. D. (2009). EST-SSR marker-based assay for the genetic purity assessment of safflower hybrids. *Euphytica* 170 (3), 347–353. doi: 10.1007/s10681-009-9995-3
- Oh, E., Kim, J., Park, E., Kim, J. I., Kang, C., and Choi, G. (2004). PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in *Arabidopsis thaliana*. *Plant Cell* 16 (11), 3045–3058. doi: 10.1105/tpc.104.025163
- Oh, E., Yamaguchi, S., Hu, J., Yusuke, J., Jung, B., Paik, I., et al. (2007). PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in *Arabidopsis* seeds. *Plant Cell* 19 (4), 1192–1208. doi: 10.1105/tpc.107.050153
- Oliya, B. K., Kim, M. Y., and Lee, S. H. (2018). Development of genic-SSR markers and genetic diversity of Indian lettuce (*Lactuca indica* L.) in south Korea. *Genes Genomics* 40 (6), 615–623. doi: 10.1007/s13258-018-0660-x
- Pandian, S., Satish, L., Rameshkumar, R., Muthuramalingam, P., Rency, A. S., Rathinapriya, P., et al. (2018). Analysis of population structure and genetic diversity in an exotic germplasm collection of *Eleusine coracana* (L.) gaertn. using genic-SSR markers. *Gene* 653, 80–90. doi: 10.1016/j.gene.2018.02.018
- Parida, S. K., Dalal, V., Singh, A. K., Singh, N. K., and Mohapatra, T. (2009). Genic non-coding microsatellites in the rice genome: characterization marker design and use in assessing genetic and evolutionary relationships among domesticated groups. *BMC Genomics* 10 (1), 1–14. doi: 10.1186/1471-2164-10-140
- Park, Y. H., Alabady, M. S., Ulloa, M., Sickler, B., Wilkins, T. A., Yu, J., et al. (2005). Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol. Genet. Genomics* 274 (4), 428–441. doi: 10.1007/s00438-005-0037-0
- Parmar, R., Seth, R., and Sharma, R. K. (2022). Genome-wide identification and characterization of functionally relevant microsatellite markers from transcription factor genes of tea (*Camellia sinensis* (L.) o. kuntze. *Sci. Rep.* 12 (1), 1–14. doi: 10.1038/s41598-021-03848-x
- Pashley, C. H., Ellis, J. R., McCauley, D. E., and Burke, J. M. (2006). EST databases as a source for molecular markers: lessons from helianthus. *J. Heredity* 97 (4), 381–388. doi: 10.1093/jhered/esl013
- Peng, J. H., and Lapitan, N. L. (2005). Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Funct. Integr. Genomics* 5 (2), 80–96. doi: 10.1007/s10142-004-0128-8
- Poxleitner, M., Rogers, S. W., Lacey Samuels, A., Browse, J., and Rogers, J. C. (2006). A role for caleosin in degradation of oil-body storage lipid during seed germination. *Plant J.* 47 (6), 917–933. doi: 10.1111/j.1365-313X.2006.02845.x
- Prathepha, P. (2003). Characterization of Waxy microsatellite classes that are closely linked to the rice Waxy gene and amylose content in Thai rice germplasm. *Songklanakarinn J. Sci. Technol.* 25, 1–8.
- Rahman, F., Hassan, M., Rosli, R., Almousally, I., Hanano, A., and Murphy, D. J. (2018). Evolutionary and genomic analysis of the caleosin/peroxygenase (CLO/PXG) gene/protein families in the viridiplantae. *PLoS One* 13 (5), e0196669. doi: 10.1371/journal.pone.0196669
- Ramsden, C. E., Zamora, D., Majchrzak-Hong, S., Faurot, K. R., Broste, S. K., Frantz, R. P., et al. (2016). Re-evaluation of the traditional diet-heart hypothesis: Analysis of recovered data from Minnesota coronary experiment, (1968–1973). *Br. Med. J.* 353, 1246. doi: 10.1136/bmj.i1246
- Ré, D. A., Capella, M., Bonaventure, G., and Chan, R. L. (2014). Arabidopsis ATHB7 and ATHB12 evolved divergently to fine tune processes associated with growth and responses to water stress. *BMC Plant Biol.* 14 (1), 1–4. doi: 10.1186/1471-2229-14-150
- Rehman, H., Rabbani, M. A., Shinwari, Z. K., and Akbar, F. (2015). RAPD markers based genetic diversity of safflower (*Carthamus tinctorius* L.) germplasm. *Pakistan J. Bot.* 47, 199–204. doi: 10.2174/1389557511313020003
- Sabzalain, M. R., Mirlohi, A., Saeidi, G., and Rabbani, M. T. (2009). Genetic variation among populations of wild safflower *Carthamus oxyacanthus* analyzed by agro-morphological traits and ISSR markers. *Genet. Resour. Crop Evol.* 56 (8), 1057–1064. doi: 10.1007/s10722-009-9426-2
- Sales-Campos, H., Reis de Souza, P., Crema Peghini, B., Santana da Silva, J., and Ribeiro Cardoso, C. (2013). An overview of the modulatory effects of oleic acid in health and disease. *Mini Rev. Med. Chem.* 13 (2), 201–210.
- Satya, P., Chakraborty, A., Jana, S., Majumdar, S., Karan, M., Sarkar, D., et al. (2017). Identification of genic SSRs in jute (*Corchorus capsularis* malvaceae) and development of markers for phenylpropanoid biosynthesis genes and regulatory genes. *Plant Breed.* 136 (5), 784–797. doi: 10.1111/pbr.12514
- Scott, K. D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E. M., Lee, L. S., et al. (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100 (5), 723–726. doi: 10.1007/s001220051344
- Sehgal, D., and Raina, S. N. (2005). Genotyping safflower (*Carthamus tinctorius*) cultivars by DNA fingerprints. *Euphytica* 146 (1-2), 67–76. doi: 10.1007/s10681-005-8496-2
- Sharifi Tehrani, M., Mardi, M., Sahebi, J., Catalán, P., and Díaz-Pérez, A. (2009). Genetic diversity and structure among Iranian tall fescue populations based on genic-SSR and EST-SSR marker analysis. *Plant Syst. Evol.* 282 (1), 57–70. doi: 10.1007/s00606-009-0207-3
- Shao, Q., Liu, X., Su, T., Ma, C., and Wang, P. (2019). New insights into the role of seed oil body proteins in metabolism and plant development. *Front. Plant Sci.* 10, 1568.
- Singh, D., and Laxmi, A. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front. Plant Sci.* 6, 895. doi: 10.3389/fpls.2015.00895
- Singh, K. N., Parveen, S., Kaushik, P., Goel, S., Jagannath, A., Kumar, K., et al. (2021). Identification and validation of in silico mined polymorphic EST-SSR for genetic diversity and cross-species transferability studies in safflower. *J. Plant Biochem. Biotechnol.* 31 (1), 168–177. doi: 10.1007/s13562-021-00673-1
- Srinivas, G., Satish, K., Madhusudhana, R., and Seetharama, N. (2009). Exploration and mapping of microsatellite markers from subtracted drought stress ESTs in *Sorghum bicolor* (L.) moench. *Theor. Appl. Genet.* 118 (4), 703–717. doi: 10.1007/s00122-008-0931-z
- Sudhakar, C., Thippeswamy, M., Sivakumar, M., Sudhakarbabu, O., and Dudhe, M. Y. (2019). In silico mining of EST-SSRs from the drought tolerant ESTs in safflower. *J. Proteomics Bioinf.* 12, 8. doi: 10.35248/0974-276X.19.12.506
- Sukiran, N. L., Ma, J. C., Ma, H., and Su, Z. (2019). ANAC019 is required for reproductive development under drought stress in *Arabidopsis*. *Plant Mol. Biol.* 99 (1), 161–174. doi: 10.1007/s11103-018-0810-1
- Talebi, R., and Abhari, S. A. (2016). Evaluation of genetic diversity in safflower (*Carthamus tinctorius* L.) using agro-morphological fatty acid composition and ISSR molecular markers. *Res. J. Biotechnol.* 11, 7.
- Thumilan, B. M., Sajeevan, R. S., Biradar, J., Madhuri, T., Nataraja, K. N., and Sreeman, S. M. (2016). Development and characterization of genic SSR markers from Indian mulberry transcriptome and their transferability to related species of moraceae. *PLoS One* 11 (9), e0162909. doi: 10.1371/journal.pone.0162909
- Tranbarger, T. J., Sangsrakru, D., Morcillo, F., James Tregear, W., Tragoonrun, S., and Billotte, N. (2012). SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *elaeis guineensis*. *BMC Plant Biol.* 12 (1), 1–12. doi: 10.1186/1471-2229-12-1
- Varshney, R. K., Graner, A., and Sorrells, M. E. (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005
- Varshney, R. K., Grosse, I., Hähnel, U., Siefken, R., Prasad, M., Stein, N., et al. (2006). Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theor. Appl. Genet.* 113 (2), 239–250. doi: 10.1007/s00122-006-0289-z

- Vavilov, N. I. (1951). The Origin, variation, immunity and breeding of cultivated plants (translated by S.K. Chestitee). *Chronica Botanica* 13, 1–366.
- Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Heredity* 93 (1), 77–78. doi: 10.1093/jhered/93.1.77
- Weiss, E. A. (1971). Castor sesame and safflower. (London, UK: Leonard hill books) 901.
- Wu, M., Liu, D., Abdul, W., Upreti, S., Liu, Y., Song, G., et al. (2018). PIL5 represses floral transition in arabidopsis under long day conditions. *Biochem. Biophys. Res. Commun.* 499 (3), 513–518. doi: 10.1016/j.bbrc.2018.03.179
- Wu, Z., Liu, H., Zhan, W., Yu, Z., Qin, E., Liu, S., et al. (2021). The chromosome-scale reference genome of safflower (*Carthamus tinctorius*) provides insights into linoleic acid and flavonoid biosynthesis. *Plant Biotechnol. J.* 19 (9)1725–1742. doi: 10.1111/pbi.13586
- Ye, H., Liu, S., Tang, B., Chen, J., Xie, Z., Nolan, T. M., et al. (2017). RD26 mediates crosstalk between drought and brassinosteroid signalling pathways. *Nat. Commun.* 8 (1), 1–3. doi: 10.1038/ncomms14573
- Yu, J. K., La Rota, M., Kantety, R. V., and Sorrells, M. E. (2004). EST derived SSR markers for comparative mapping in wheat and rice. *Mol. Genet. Genomics* 271 (6), 742–751. doi: 10.1007/s00438-004-1027-3
- Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., et al. (2012). *De novo* assembly and characterisation of the transcriptome during seed development and generation of genic-SSR markers in peanut (*Arachis hypogaea* L. *BMC Genomics* 13 (1), 90. doi: 10.1186/1471-2164-13-90
- Zhang, X., Qin, H., Xie, W., Ma, Y., and Sun, W. (2020). Comparative population genetic analyses suggest hybrid origin of *Rhododendron pubicostatum* an endangered plant species with extremely small populations endemic to yunnan China. *Plant Diversity* 42 (4), 312–318. doi: 10.1016/j.pld.2020.06.012
- Zhang, T., Tan, D. F., Zhang, L., Zhang, X. Y., and Han, Z. X. (2017). Phylogenetic analysis and drought-responsive expression profiles of the WRKY transcription factor family in maize. *Agric. Genet.* 3, 99–108. doi: 10.1016/j.aggene.2017.01.001
- Zhang, H., Wei, L., Miao, H., Zhang, T., and Wang, C. (2012). Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13 (1), 316. doi: 10.1186/1471-2164-13-316
- Zhao, H., Li, Y., Petkowski, J., Kant, S., Hayden, M. J., and Daetwyler, H. D. (2021). Genomic prediction and genomic heritability of grain yield and its related traits in a safflower genebank collection. *Plant Genome* 14 (1) e20064. doi: 10.1002/tpg2.20064
- Zhou, T., Li, Z. H., Bai, G. Q., Feng, L., Chen, C., Wei, Y., et al. (2016). Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia oliver* (Aceraceae). *Molecules* 21 (3), 166. doi: 10.3390/molecules21030166