International Journal of Agriculture, Environment and Food Sciences

JAEFS

e-ISSN : 2618-5946

DOI: 10.31015/jaefs.2019.4.7

www.jaefs.com

Research Article

Int J Agric Environ Food Sci 3(4):240-244 (2019)

Sequence data characterization and development of DNA markers for sesame (*Sesamum indicum* L.)

Ayse Gul Ince[1] **Mehmet Karaca**[2,*]

[1]Vocational School of Technical Sciences, Akdeniz University, 07070 Antalya, Turkey
[2]Department of Field Crops, Akdeniz University, 07070 Antalya, Turkey

[*]Corresponding Author: mkaraca@akdeniz.edu.tr

Abstract

New DNA sequencing techniques enable researchers obtain large quantity of sequence information, which are deposited in digital storage or could be further mined for other purposes. Sesame (*Sesamum indicum* L.) is one of very important oilseed crops, its seed oil contains many antioxidant properties making sesame the queen of oil crops. Today, molecular plant breeding technology is indispensable for plant breeders and seed producers. Compared to other seed oil crops the available number of microsatellite markers in sesame is still not sufficient enough for the development of polymorphic markers for breeding and genetic studies. Thus, new approaches or resources are needed for development of microsatellite markers for sesame. In the present study, we utilized a total of 45099 transcribed genomic DNA sequences/expressed sequence tags and mined these sequences for studying frequency of microsatellite motifs, ranging from di- to hexa-nucleotides with four to ten tandem repeats, and repeat numbers greater than 10. Using mined transcribed data, 42 putative microsatellite markers were developed and characterized at the sequence level. However, we did not confirm these markers and have no information about the level of their polymorphisms in sesame *in vitro*. We discussed the biological meaning of the motif lengths and repeat numbers in the sesame genome.

Keywords: Microsatellite, Motif length, Repeating motif, Primer pairs, Tandem repeats

Introduction

Sesame (*Sesamum indicum* L.), belonging to the family *Pedaliaceae*, is one of the most important oil seed crops in tropical and subtropical regions of Asia, Africa and South America. Today, approximately 2.74 million tons of sesame seeds are produced in approximately 6.1 million hectares worldwide. Major sesame seed producing countries are India, China, Ethiopia, Sudan, Myanmar and Uganda. Sesame is sometime referred to as the 'queen of oilseeds' due to its superior quality of oil among the major oilseed crops including peanut, soybean and rapeseed (Wei et al., 2014). Seeds of sesame contain about 50–60% edible oil and 25% protein, with antioxidants such as lignans, sesamolin and sesamin. Sesame has small diploid genome size of approximately 350 Mb and contains 2n = 26 chromosomes. Although sesame is among the first oilseed crop utilized by human and has many varieties and ecotypes adapted to various ecological conditions throughout the world, it is one of the neglected crops and not widely cultivated. One of the main reasons behind its limited cultivation is its mechanical harvesting difficulties such seed shedding, uninform maturation of the seeds and very low yield (Yen, 1990; Cheung et al., 2007; Ali-Al-Somain et al. 2017).

DNA markers have been proven useful in plant breeding and maintenance of germplasm collections. Although it is very behind from other important agricultural crops, the use of DNA marker technologies has been carried out to estimate the genetic variation in sesame germplasm using random amplified polymorphic DNA (RAPD, Pham et al., 2011), amplified fragment length polymorphism (AFLP, Ali et al., 2007), inter-simple sequence repeat (ISSR, Kumar and Sharma, 2011), sequence-related amplified polymorphisms (SRAP, Ali-Al-Somain et al., 2017) and simple sequence repeat or also known

Cite this article as:
Ince, A.G., Karaca, M. (2019). Sequence data characterization and development of DNA markers for sesame (*Sesamum indicum* L.). Int. J. Agric. Environ. Food Sci., 3(4), 240-244
DOI: https://dx.doi.org/10.31015/jaefs.2019.4.7
Received: 28 September 2019 **Accepted:** 04 December 2019 **Published:** 24 December 2019
Year: 2019 **Volume:** 3 **Issue:** 4 (December) **Pages:** 240-244
Available online at : http://www.jaefs.com - http://dergipark.gov.tr/jaefs

as microsatellite (SSR, Yue et al. 2013). However, the use of DNA markers in sesame is limited at the genetic variation studies. Furthermore, most of DNA marker techniques used in sesame were carried out using random DNA markers along with a few co-dominant polymorphic DNA markers (Ali et al., 2007; Ince et al., 2010a; Kumar and Sharma, 2011; Wei et al., 2014).

Among DNA marker systems, microsatellites (SSRs) are considered the best DNA markers for understanding the genetic relationships, genetic mapping, hybrid detection and genetic purity testing of any given crop species, due to their abundance, random distribution within the genome, high polymorphism information content, high reproducibility and co-dominant nature (Ince et al., 2008; Ince et al., 2010a; Karaca and Ince, 2011). Microsatellites can be classified into genomic and expressed sequence tag (EST) based microsatellites, depending on their original sequences. Genomic microsatellites can be determined using costly, labor-intensive, and time-consuming traditional methods, but the inter-specific transferability of genomic microsatellites is limited. Whereas, EST-microsatellites obtained from expressed sequences tags have higher level of inter-specific transferability and polymorphism level in many plant species (Ince et al. 2010b). However, microsatellite markers have not been screened on whole-genome level in the *Sesamum* genus. Up to date, only 10 genomic microsatellite and 44 EST-microsatellite markers were developed in previous studies. Additionally, only one global transcriptomic analysis was performed (Wei et al., 2011). Development and utilization of microsatellite markers from ESTs deposited in public databases have several advantages such as identifiability of functional gene for suitable diagnostic markers, high transferability between species, suitable for medium-sized laboratory conditions, and their low development cost (Cloutier et. al., 2009; Karaca and Ince, 2011).

The main purpose of this study was to explore microsatellite sequences varying lengths and motifs in transcriptomic sequences of sesame. With this aim, a total of 45099 sequences in the transcribed region of sesame genome were analyzed. Additionally, new sets of microsatellite primer pairs were designed and these primer pairs could be very useful in sesame breeding and genetic studies.

Materials and Methods

Target sequences

A total of 45099 EST sequences of sesame (*Sesamum indicum* L.) stored in NCBI GenBank databases (ftp://ftp.ncbi.nih.gov/) were downloaded and stored in a personal computer. Data were analyzed to find ESTs with microsatellites or simple sequence repeats. In the present study, repeats in ESTs were identified using TRA 1.5 software (Bilgen et al., 2004). Tandem repeats with motif length di-, tri-, tetra-, penta and hexa-nucleotides were mined. The number of repeats in the motifs ranged from four to ten (Karaca et al., 2005).

Designing of microsatellite primer pairs

Microsatellite primer pairs were designed using Batch Primer 3 1.0 software (You et al., 2008) based on the following main parameters: GC content value was set between 40% and 80%, annealing temperature (Tm) was set between 59°C and 62°C, max self-complementary was set 4.00, max 3' self-complementary was set 3.00, and expected amplified product size was defined as 200–350 bp (Ince et al., 2010a). For primer pair selection, motif lengths were set di- to hexa-nucleotides. Other criteria including minimum number of microsatellite repeats for motifs di- to hexa-nucleotides were applied as previously used in Ince et al. (2010b).

Results and Discussion

In this study, we analyzed the pooled ESTs of sesame from NCBI EST databases and develop sets of EST-SSRs. DNA molecular markers play an important role in sesame breeding studies including identification of the genes responsible for desirable traits, determination of genetic variation, genetic relationships, genetic mapping, hybrid detection in germplasm (Karaca et al., 2017a).

In the present study, tandem repeats varied in their repeats numbers. Analysis results indicated that the number of repeat containing sequences decreased as the repeat number in a motif increased as shown in Table 1. In total 6082 sequences containing 7376 tandem repeats were identified from 45099 sequences when lowest repeat number was set to 4. On the other hands, only 80 tandem repeats were determined when lowest repeat number was set to 10. As they can be seen in Table 1, about 13.5% of sequences contained four repeats while 0.17% sequences contained 10 or more repeat numbers. Based on sesame transcribed sequences we noted that majority of repeating numbers were between 4 and 6. This indicates that tandem repeats of di- to hexa-nucleotides containing four to seven repeats may have biological functions more than other repeats.

In the present study, tandem repeat types found in the sesame genome were represented in Table 2. Results indicated that repeat numbers within each motif ranged from di- to hexa-nucleotides. When repeat finding criteria of repeat numbers was set to a minimum number of 4 repeats, the most abundant type of repeat motifs was di-nucleotide (5278, 71.55%), followed by tri-nucleotide (1935, 26.23%), tetra-nucleotide (115, 1.56%), hexa-nucleotide (27, 0.37%), and penta-nucleotide (21, 0.28%) repeat units (Table 2). The number of microsatellites with greater than 10 tandem repeats was very low but di-nucleotides (67, 83.8%) were the most common, followed by tri-nucleotides (12, 15.0%), hexa-nucleotide (1, 1.25%) and tetra- and penta-nucleotides were not existed (Table 2).

Among the tandem repeats, di- and tri-nucleotides contained greater than 10 repeats while penta-nucleotides contained few number of sequences containing 5 and 6 repeats (Table 2). This indicated that di- and tri-nucleotides have higher frequency in part of transcribed portion of sesame genome. In general, the number of tandem repeat containing repeat types were di-nucleotides. This was the most interesting findings of the present study because other research conducted on *in silico* data mining reported that tri-nucleotide repeats were the most abundant type. However, there were inconsistent results between the higher frequencies between di- and tri-nucleotides among research. Inconsistency was probably due to the different number of repeat searching criteria used among the studies (Karaca et al., 2017a; Karaca et al., 2017b). Zhang

et al. (2012) reported that di-nucleotide motifs (48.01%) were the most abundant, followed by tri- (20.96%), hexa- (25.37%), penta- (2.97%), tetra- (2.12%), and mono-nucleotides (0.57%) in the 42566 unique-transcript sequences. Total length covered was 47987 kbp in the sesame genome. These authors identified a total of 7324 SSRs, with motifs greater than 15 bp. When the motif length was changed to greater than 18, the number of SSRs detected were 4440.

Table 1. Summary of tandem repeats in part of transcribed portion of sesame genome (4509 sequences)

# RN	# Sequences	# Repeat Strings	Repeat Index	Repeat Strings/ Sequences	Total TR	Repeat (%)
4	6082	7426	0.165	1.221	7376	13.486
5	1643	1866	0.041	1.136	1854	3.643
6	760	845	0.019	1.112	867	1.685
7	447	471	0.010	1.054	466	0.991
8	296	307	0.007	1.037	305	0.656
9	208	213	0.005	1.024	213	0.461
10	143	145	0.003	1.014	126	0.317
>10	79	80	0.002	1.013	80	0.175

RN: repeats number, TR: Tandem Repeats

Table 2**.** Frequency of tandem repeat motifs (RMs) in part of transcribed portion of sesame genome revealed by *in silico* research

	Number of Repeats in Motifs							
RMs	4	5	6	7	8	9	10	>10
Di	5278 (71.55%)	1330 (71.7%)	630 (75.2%)	377 (80.9%)	257 (84.3%)	184 (86.4%)	126 (86.9%)	67 (83.8%)
Tri	1935 (26.23%)	469 (25.3%)	174 (20.8%)	80 (17.2%)	45 (14.8%)	26 (12.2%)	16 (11.03%)	12 (15%)
Tetra	115 (1.56%)	36 (1.94%)	19 (2.27%)	4 (0.85%)	na	na	na	na
Penta	21 (0.28%)	8 (0.43%)	6 (0.72%)	na	na	na	na	na
Hexa	27 (0.37%)	8 (0.43%)	6 (0.72%)	3 (0.64%)	3 (0.98%)	3 (1.4%)	3 (2.07%)	1 (1.25%)
Total	7376	1854	867	466	305	213	126	80

In the present study, tandem repeat contexts in the transcribed portion of sesame genome were also studied and findings were depicted in Table 3. We clearly noted that as the motif lengths increased from di- to hexa-nucleotides and the repeats numbers increased from four to ten, and greater than ten, motif contexts varied. Among the di-nucleotides AG/TC repeats contained 4, 8, 9, 10 and greater than 10 repeats were abundant while CT/AG repeats consisted of more repeats contained from 5 to 7 repeats in comparison to the other repeats. Tri-nucleotide motif context, CTT was the most abundant type in especially the repeats numbers from nine to ten. This finding revealed that tri-nucleotide motif context of CTT would be very important context and further transcriptomic studies should be focused on CTT context. CTT translates into leucine (L, Leu) amino acid that is an essential amino acid for human (Weber, 1990; Zhang et al., 2012; Karaca et al., 2017a; Karaca et al., 2017b). Tetra-nucleotide repeats ATAC/TATG consisted of more repeats comparison to other contexts (Table 3). Penta-nucleotides consisted of CTCCT/GAGGA while GCTCCC/CGAGGG is only one prominent sequence contexts in hexa-nucleotide repeats. However, there were no tetra- and penta-nucleotides with repeat number greater than seven in this study (Table 3).

In another study, Zhang et al. (2012) reported that the top four motif repeats for SSR marker were (AG/CT)n [1268 (34.51%)], (CA/TG)n [281 (7.65%)], (AT/AT)n [215 (5.85%)], and (GAA/TTC)n [131 (3.57%)] using RNA-Seq methods in sesame.

Finally, in the present study we also aimed to design microsatellite primer pairs for sesame using the transcribed sequences characterized in this study. A total of 42 microsatellite primer pairs from EST sequences were designed using criteria described in this study. These microsatellite primer pairs were called SUS primer pairs stand for Sesame Unique Simple Sequence Repeats (Table 4).

SUS primer pairs consisted of microsatellites ranging from di- to hexa-nucleotide repeats. Repeat numbers of microsatellite motifs varied from 6 to 17 for di-nucleotide motifs, 5 to 6 for tri-nucleotide motifs, 4 for tetra-nucleotide motifs, 3 for hexa-nucleotide motifs while only one penta-nucleotide motif consisted of 3 repeats. Melting temperature values of microsatellite primer pairs ranged from 59.32°C to 61.70°C. We strongly suggested that annealing temperature values of primer pairs used in polymerase chain reaction should be 3-4 degrees Celsius below the melting temperature given in Table 4 for efficient marker development.

Up to date, although several hundred EST-SSRs have previously been developed from EST sequences and utilized in sesame genetic diversity studies (Wei et al., 2011; Zhang et al., 2012; Wei et al., 2014). However, it is important to notice that polymorphic markers are not enough to use in marker-assisted plant breeding methods for important qualitative and quantitative traits in sesame. In this study, randomly selected a few SUS primer pairs were tested and results indicated that SUS markers developed in this study would be very useful in sesame breeding studies.

Table 4. List of SUS primer pairs developed from transcribed part of sesame genome

ID	Accession #	Sequences (5'Þ3')	RT	T	GC	S
SUS01F	JZ971789.1	TCGTCGTCTCTCAGCTCTCTC	$[AG]_{17}$	60.03	57.14	244
SUS01R	JZ971789.1	CGTGTATTGCTTTCCCTACCTC		60.02	50.00	
SUS02F	JZ971779.1	AGACGGTTGGGTCCTCTCAT	$[GT]_{9}$	60.91	55.00	243
SUS02R	JZ971779.1	TTTATCCAGACAAGCCAGCAG		60.39	47.62	
SUS03F	JZ971778.1	CAAAGGTGTCAATCTTAGCAAGG	$[GCA]_{5}$	60.17	43.48	228
SUS03R	JZ971778.1	CCACCCTCCCAAAACTCTTT		60.33	50.00	
SUS04F	JZ971778.1	CAAAGGTGTCAATCTTAGCAAGG	$[ATA]_{5}$	60.17	43.48	228
SUS04R	JZ971778.1	CCACCCTCCCAAAACTCTTT		60.33	50.00	
SUS05F	JZ971764.1	TATCAGCTTGCCACTTCCTTC	$[AT]_{6}$	59.47	47.62	260
SUS05R	JZ971764.1	CAACAATAGCAGCAGCATCAA		60.03	42.86	
SUS06F	JZ971762.1	GAGATCAAGAACGGCGGACT	$[GAAT]_{4}$	61.70	55.00	249
SUS06R	JZ971762.1	CATTTACATCAGAGACACCCACA		59.91	43.48	
SUS07F	JZ971752.1	AGCTTCCACTAGCAACAGCAA	$[CAA]_{5}$	60.20	47.62	219
SUS07R	JZ971752.1	TCAGTAGCTTGACCCCTTCTG		59.48	52.38	
SUS08F	JZ971751.1	AAGCGGTCATGTTTCTGCTAA	$[TCT]_{5}$	59.90	42.86	275
SUS08R	JZ971751.1	GAAGGGGTATTGGAAAGCAAC		59.83	47.62	
SUS09F	JZ971716.1	CAACAATCCAAACACAGTAGAAGC	$[GCT]_{5}$	60.10	41.67	254
SUS09R	JZ971716.1	CCAAGGACGAGAAGAAGAAGAA		59.99	45.45	
SUS10F	JZ971716.1	GCAAGCAACCAACGGTAGAGT	$[CTT]_{5}$	61.59	52.38	204
SUS10R	JZ971716.1	GATCGAGAAGATCAAGGACAAGA		59.84	43.48	
SUS11F	JZ971716.1	CAACAATCCAAACACAGTAGAAGC	$[TCGCTG]_{3}$	60.10	41.67	254
SUS11R	JZ971716.1	CCAAGGACGAGAAGAAGAAGAA		59.99	45.45	
SUS12F	JZ971714.1	TTCCGGCACTGACTTTAACA	$[TCT]_{5}$	59.32	45.00	229
SUS12R	JZ971714.1	AGGCGAGAAGACTGATTGGAT		60.23	47.62	
SUS13F	JK755189.1	GACCGCACAAAGCATTACAAG	$[GACGAG]_{3}$	60.68	47.62	247
SUS13R	JK755189.1	CCATGTTAAGCCAATCTTCCA		59.95	42.86	
SUS14F	JK755186.1	CTCTTGACATGCCCGAACTAC	$[GACGAG]_{3}$	59.75	52.38	245
SUS14R	JK755186.1	GCGTGTGATGCACTCTTTCTT		60.46	47.62	
SUS15F	JK755185.1	CTTGAAAGCAAACTCGACCAG	$[CT]_{7}$	60.04	47.62	227
SUS15R	JK755185.1	GCAGCTCATCTTGCACTTGA		60.30	50.00	
SUS16F	JK755185.1	CTTGAAAGCAAACTCGACCAG	$[TC]_{9}$	60.04	47.62	227
SUS16R	JK755185.1	GCAGCTCATCTTGCACTTGA		60.30	50.00	
SUS17F	JK755181.1	GCATCTATCTCTCCCGGTTCT	$[TC]_{8}$	59.69	52.38	272
SUS17R	JK755181.1	TCCTTCGATTGGCTTACAAGA		59.83	42.86	
SUS18F	JK755181.1	TAGATGGCTCGATTACCCTCA	$[CTTG]_{4}$	59.68	47.62	269
SUS18R	JK755181.1	CTCTCAAGTGGACGCAAAGAC		60.04	52.38	
SUS19F	JK755164.1	CTGCTGTTGCTGCTGTAAATG	$[AGA]_{5}$	59.69	47.62	236
SUS19R	JK755164.1	CCGCGACTTCTTTTTCTTCTT		60.01	42.86	
SUS20F	JK755222.1	GGGAGTGATTAGGGTTTGCTC	$[AG]_{6}$	59.95	52.38	233
SUS20R	JK755222.1	TTGAAGCGGAGGAGTTAAGGT		60.25	47.62	
SUS21F	JZ191039.1	AGATTTCTCGGCTGAAACAGG	$[TGGA]_{4}$	60.75	47.62	242
SUS21R	JZ191039.1	ATACATCGCTCGCATCAAAAC		60.12	42.86	
SUS22F	HO710201.1	GGGGCTGAGAATTTGAGAGAG	$[AG]_{6}$	60.33	52.38	279
SUS22R	HO710201.1	GGCCTCTTAGTTGACCAGACA		59.34	52.38	
SUS23F	HO710187.1	GCAAAAGGTGAGGGATGAACT	$[GAT]_{5}$	60.49	47.62	253
SUS23R	HO710187.1	CTGCTGGATGTCAGTTTCCTG		60.84	52.38	
SUS24F	HO710180.1	AATCCACAACCCTCTTCTTCG	$[CT]_{11}$	60.48	47.62	247
SUS24R	HO710180.1	CTATCCTGCGGACCCATTATC		60.67	52.38	
SUS25F	HO710174.1	CGGTCCTGCAAGTGAAGATAA	$[GA]_{6}$	60.26	47.62	223
SUS25R	HO710174.1	TCATAAAGACACCACACCACACT		59.43	43.48	
SUS26F	HO710172.1	TCTGTTGTAGGCGGAAAGTGT	$[GA]_{7}$	59.79	47.62	274
SUS26R	HO710172.1	ACGAGCAGTTTGTTTGGTACG		60.22	47.62	
SUS27F	HO710172.1	TCTGTTGTAGGCGGAAAGTGT	$[GA]_{7}$	59.79	47.62	274
SUS27R	HO710172.1	ACGAGCAGTTTGTTTGGTACG		60.22	47.62	
SUS28F	HO710168.1	GGAGACATTCTTGGCATGGT	$[TGA]_{5}$	59.93	50.00	248
SUS28R	HO710168.1	CCCTTCAGACCTCAACTCCTC		60.24	57.14	
SUS29F	HO710168.1	GGAGACATTCTTGGCATGGT	$[GAT]_{5}$	59.93	50.00	248
SUS29R	HO710168.1	CCCTTCAGACCTCAACTCCTC		60.24	57.14	
SUS30F	HO710168.1	GGAGACATTCTTGGCATGGT	$[CTG]_{6}$	59.93	50.00	248
SUS30R	HO710168.1	CCCTTCAGACCTCAACTCCTC		60.24	57.14	
SUS31F	HO710163.1	GAGGTACGGAGAGCACAAGC	$[AGA]_{5}$	60.02	60.00	243
SUS31R	HO710163.1	GCAAGCAACCAACGGTAGAGT		61.59	52.38	
SUS32F	HO710163.1	GAGGTACGGAGAGCACAAGC	$[CAG]_{5}$	60.02	60.00	243
SUS32R	HO710163.1	GCAAGCAACCAACGGTAGAGT		61.59	52.38	
SUS33F	HO710163.1	GAGGTACGGAGAGCACAAGC	$[CAGCGA]_{3}$	60.02	60.00	243
SUS33R	HO710163.1	GCAAGCAACCAACGGTAGAGT		61.59	52.38	
SUS34F	HO710139.1	GCTTGTGCTATGCTGCCTTAT	$[CAC]_{5}$	59.54	47.62	245
SUS34R	HO710139.1	ACTTCTTTTAACCACCGACAGC		59.70	45.45	
SUS35F	HO710139.1	AAATGGGCATGACAGGAGAC	$[CCACGG]_{3}$	59.93	50.00	249
SUS35R	HO710139.1	GTGTAATTGTGAACGGGCCTA		59.87	47.62	
SUS36F	JK085550.1	AATAGTAAAACGGATCGCTTGG	$[AGAC]_{4}$	59.54	40.91	257
SUS36R	JK085550.1	ATCATGTATCCCCGCTGCTT		61.76	50.00	
SUS37F	JK085549.1	TCTTCTCCGTGTCCGAAAGT	$[CACT]_{4}$	59.84	50.00	253
SUS37R	JK085549.1	CATCTTCACATGCTTTAGCCTCA		61.64	43.48	
SUS38F	JK085527.1	CTGGGCGCTCTATCGTTTAG	$[AAG]_{5}$	60.00	55.00	240
SUS38R	JK085527.1	TGAACCTTGACGCTTTGATCT		59.87	42.86	
SUS39F	JK085514.1	GCCAGAAGTGTGAGTGATGC	$[GAAGC]_{3}$	59.43	55.00	230
SUS39R	JK085514.1	TGATTTACCGAGTTTCCGATG		59.95	42.86	
SUS40F	JK085504.1	GGGGACTCACTCACTCACTCA	$[CT]_{8}$	60.30	57.14	256
SUS40R	JK085504.1	AGCCATCTTTATTGGGAGGTG		60.33	47.62	
SUS41F	JK085497.1	TTCATCACCCTCACCTGTTTC	$[TTC]_{5}$	59.96	47.62	242
SUS41R	JK085497.1	GTATTTGATCCCACACGCATT		59.70	42.86	
SUS42F	JK085451.1	GGGCTGGATTTTTATTTGTTCC	$[CATC]_{4}$	60.85	40.91	265
SUS42R	JK085451.1	CTTAGAGGGAAAGGGGGAGTT		59.95	52.38	

ID: Primer Identification, SUS stands for Sesamum indicum L., RT: Repeat Type, T: Temperature of Melting (°C), GC: Guanine-Cytosine content (%) and S: Expected amplified product size (bp).

Conclusions

In the present study, results revealed that transcribed portion of sesame genome contained significant amount of tandem repeats. Frequency of tandem repeat motifs and the number of repeats in motif varied. Data mining analysis revealed that di-nucleotides were dominant repeat types used in all searching criteria in sesame transcribed genome. Among the di-nucleotides AG/TC repeat motifs were dominant in comparison to the other repeat motifs. However, especially in the tri-nucleotide repeats CTT motif contexts were frequently existed in the repeats numbers from nine to greater than ten. Microsatellite primer pairs developed in the present study could be used in sesame breeding programs. Main breeding goals of sesame are higher yield and better quality oil seed. Genetic relationships and diversity among sesame germplasm have been searched mostly using RAPD, ISSR, AFLP, etc. Marker-assisted selection and molecular breeding studies in sesame have lagged behind in comparison with other plants. Thus, a rapid and cost-effective approach to develop molecular markers for sesame is required. Polymorphic primer pairs (SUS markers) could be used as tools in marker-assisted selection and sesame germplasm identification.

References

Ali, G.M., Yasumoto, S. and Katsuta, M.S. (2007). Assessment of genetic diversity in sesame (*Sesamum indicum* L.) detected by amplified fragment length polymorphism. Electronic Journal of Biotechnology, 10, 12-23 [CrossRef]

Ali-Al-Somain, B.H., Migdadi, H.M., Al-Faifi, S.A., Muharram, A.A., Mohammed, N.A. and Refay, Y.A. (2017). Assessment of genetic diversity of sesame accessions collected from different ecological regions using sequence-related amplified polymorphism markers. 3 Biotech, 7, 82 [CrossRef]

Bilgen, M., Karaca, M., Onus, A.N. and Ince, A.G. (2004). A software program combining sequence motif searches with keywords for finding repeats containing DNA Sequences. Bioinformatics, 20, 3379-3386 [CrossRef]

Cheung, S.C., Szeto, Y.T. and Benzie, I.F. (2007). Antioxidant protection of edible oils. Plant Foods Human Nutrition, 62, 39-42 [CrossRef]

Cloutier, S., Niu, Z., Datla, R. and Duguid, S. (2009). Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theoretical Applied Genetic, 119, 53-63 [CrossRef]

Ince, A.G., Karaca, M., Bilgen, M. and Onus, A.N. (2008). Digital differential display tools for mining microsatellite containing organism, organ and tissue. Plant Cell Tissue and Organ Culture, 94, 281-290 [CrossRef]

Ince, A.G., Onus, A.N. and Karaca, M. (2010a). Polymorphic microsatellite markers transferable across *Capsicum* species. Plant Molecular Biology Reporter, 28, 285-291 [CrossRef]

Ince, A.G., Karaca, M. and Onus, A.N. (2010b). CAPS-microsatellites: use of CAPS method to convert non-polymorphic microsatellites into useful markers. Molecular Breeding, 25, 491-499 [CrossRef]

Karaca, M, Bilgen, M, Onus, A.N., Ince, A.G. and Elmasulu, S.Y. (2005). Exact tandem repeats analyzer (ETRA): a new program for DNA sequence mining. Journal of Genetics, 84, 49-54 [CrossRef]

Karaca, M. and Ince, A.G. (2011). New non-redundant microsatellite and CAPS-Microsatellite markers for cotton (*Gossypium* L.). Turkish Journal of Field Crops, 16, 172-178 [URL]

Karaca, M., Ince, A.G., Uygur-Göçer, E. and Aydin, A. (2017a). Microsatellite distribution and densities in promoter and gene body entities of some plant genes. The Journal of Scientific and Engineering Research, 4, 154-163 [URL]

Karaca, M., Ince, A.G., Uygur-Göçer, E. and Aydin, A. (2017b). Descriptive statistics and PIC values of genomic and transcriptomic-microsatellites in several plant species. The Journal of Scientific and Engineering Research, 4, 236-246 [URL]

Kumar, V. and Sharma, S.N. (2011). Comparative potential of phenotypic, ISSR and SSR markers for characterization of sesame (*Sesamum indicum* L.) varieties from India. Journal of Crop Science and Biotechnology, 14, 163-171 [CrossRef]

Pham, T.D., Geleta, M., Bui, T. M., Bui, T.C., Merker, A. and Carlsson, A.S. (2011). Comparative analysis of genetic diversity of sesame (*Sesamum indicum* L.) from Vietnam and Cambodia using agro-morphological and molecular markers. Hereditas, 148, 28-35 [CrossRef]

Weber, J.L. (1990). Informativeness of human (dC-dA)n·(dG-dT)n polymorphisms. Genomics, 7, 524-530 [CrossRef]

Wei, W., Qi, X, Wang, L., Zhang, Y., Hua, W., Li, D., Lv H. and Zhang W. (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics, 12, 451 [CrossRef]

Wei, X., Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X. and Zhang, J. (2014). Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. Molecules, 19, 5150-5162. doi: 10.3390/molecules19045150

Yen, G.C. (1990). Influence of seed roasting process on the changes in composition and quality of sesame (*Sesamum indicum* L.) oil. Journal of the Science of Food and Agriculture, 50, 563-570 [CrossRef]

You, F.M., Huo, N., Gu, Y.Q., Luo, M., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J. and Anderson, O.D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics, 9, 253 [CrossRef]

Yue, W., Wei L., Zhang T., Li C., Miao H. and Zhang H. (2013). Genetic diversity and population structure of germplasm resources in sesame (*Sesamum indicum* L.) by SSR markers. Acta Agronomica Sinica, 38, 2286-2296 [CrossRef]

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, 316 [CrossRef]