DETERMINATION OF GENETIC DIVERSITY OF TURKISH SESAMUM (*Sesamum indicum* L.) BY USING AFLP MARKERS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Molecular Biology and Genetics

by
Pelin TEKİN

June 2011
İZMİR

We approve the thesis of **Pelin TEKİN**

_____
Assoc. Prof. Dr. Sami DOĞANLAR
Supervisor

_____
Prof. Dr. Anne FRARY
Co-supervisor

_____
Assist. Prof. Dr. Çağlar KARAKAYA
Committee Member

_____
Assist. Prof. Dr. Jens ALLMER
Committee Member

_____
Prof. Dr. Hülya İLBİ
Committee Member

29 June 2011

_____
Assoc. Prof. Dr. Ahmet KOÇ
Head of the Department of
Molecular Biology and Genetics

_____
Prof. Dr. Durmuş Ali DEMİR
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

It is an honor to express my thanks and gratitude to my advisor **Assoc. Prof. Dr. Sami DOĞANLAR** and my co-advisor **Prof. Dr. Anne FRARY** for their endless guidance, help, support, encouragement and patience during my Masters' work. I owe everything to them.

I would like to thank **Assis. Prof. Dr. Jens ALMER** and **Visam GÜLTEKİN** for their studies and help with my thesis.

I am very appreciative to my friend **İbrahim ÇELİK** for his great support, giving a hand in my studies and sharing his knowledge with me.

I want to thank my all friends in the **Plant Molecular Genetics Laboratory** for their being with me everytime, their kindness and help.

I would like to thank the Molecular Biology and Genetics Department for giving me an excellent education.

I want to thank to my friend **Fevzi Çoban** for his great encouragement and supportive helps during my thesis.

Finally, I am very grateful to **my family** with all my pleasure for their endless encouragement, understanding, patience and support.

ABSTRACT

DETERMINATION OF GENETIC DIVERSITY OF TURKISH SESAMUM (*Sesamum indicum* L.) BY USING AFLP MARKERS

Sesame (*Sesamum indicum* L.) belongs to the Pedaliaceae family. It is an important oil seed crop which is cultivated in tropical and subtropical areas of Asia and Africa. China is the largest producer of sesame seed in the world while Turkey ranks seventh and produces 21036 tonnes of sesame seed in a year. Although sesame's edible seed and high quality seed oil are important for both humans and the economy, there is not enough information about the sesame genome in the literature. Our aim was to determine the diversity of 158 Turkish sesame accessions by using the AFLP marker system and to design a new set of sesame-specific SSR markers from genomic sequence of *S. indicum*. The Turkish sesame accessions were tested with five AFLP primer combinations, as a result, 148 polymorphic fragments were obtained. The maximum similarity was 57% for the accessions and a good level of diversity was present in the sesame germplasm. Secondly, a genomic library of sesame was constructed. A total of 1.094.317 reads were obtained and 702.371 of them were clustered to 140.669 reads containing 93.365 nucleotides. A total of 3101 primer pairs were developed from flanking regions of SSRs with primers for dinucleotide (36,4%), tetranucleotide (29,3%), trinucleotide (23,1%), pentanucleotide (7,1%), and hexanucleotide (4,2%) repeats. These primers are the first genomic-SSR markers developed for sesame cultivars. SSRs have good reproducibility, high genome coverage, co-dominant inheritance, good transferability to close species and are multiallelic. The designed genomic-SSRs should be very useful for sesame mapping and diversity studies.

ÖZET

TÜRK SUSAMLARINDA (*Sesamum indicum* L.) AFLP MARKÖRLERİNİ KULLANARAK GENETİK ÇEŞİTLİLİĞİN BELİRLENMESİ

Susam (*Sesamum indicum* L.) Pedaliaceae ailesine aittir. Asya ve Afrika' nın tropikal ve tropikal iklime yakın iklimdeki bölgelerinde yetiştirilen önemli bir yağlı tohum bitkisidir. Türkiye yedinci olarak yılda 21036 ton susam tohumu üretirken Çin dünyadaki en büyük susam tohumu üreticisidir. Susamın yenilebilir ve yüksek kalitedeki tohumunun hem insanlar hem de ekonomi için önemli olmasına rağmen susam genomu ile ilgili literatürde yeterli bilgi bulunmamaktadır. Amacımız AFLP markör sistemini kullanarak 158 tane Türk susam hattındaki çeşitliliği belirlemek ve *S. indicum* genomundan susama özgü yeni bir SSR markör seti dizayn etmekti. Türk susam hatları beş AFLP primer kombinasyonu ile test edildi, sonuç olarak, 148 polimorfik allel elde edildi. Hatlar arasında en büyük benzerlik %57 olarak belirlendi ve susam hatlarında iyi bir seviyede benzerlik bulundu. İkinci olarak, susamdan bir genomik kütüphane oluşturuldu. 1.094.317 tane sekans parçası elde edildi ve 702.371 tanesi, 93.365 tane nükleotid içeren 140.669 sekans parçasına birleştirildi. Toplam olarak SSR' lara en yakın gen bölgelerinden, dinükleotid (36,4%), tetranükleotid (29,3%), trinnükleotid (23,1%), pentanükleotid (7,1%) ve heksanükleotid (4,2%) olmak üzere 3101 tane primer çifti elde edildi. Bu primerler susam çeşitlerinden elde edilen ilk genomik-SSR primerleri oldu. SSR'lar yüksek takrarlanabilir özellikte, genomu yüksek ölçüde kapsayan, eş baskın kalıtımı gösterebilen, yakın türlere yüksek ölçüde aktarılabilen çoklu allelik primerlerdir. Dizayn edilen genomik-SSR' lar susamın haritalanması ve çeşitliliğinin belirlenmesinde büyük ölçüde yararlı olacaktır.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Biology of Sesame (*Sesamum indicum* L.)

Sesame (*Sesamum indicum* L.) is an important oil seed crop which has been cultivated in tropical and subtropical areas of Asia and Africa since 3050-3500 B.C. (Mondal et al. 2010; Bedigian and Harlan 1986). Although it produces considerable seed and high quality, edible seed oil which are important for both humans and the economy, there is little information about the cultivated sesame genome in the literature (Li-Bin Wei et al. 2009). Seed production of sesame is lower than other oil seed crops because varieties are suitable for cultivation only in the environment where they were derived. Sesame is not an ideal crop. It is susceptible to drought, diseases, insect pests, and pathogens; it has non-synchronous maturity; indeterminate growth habit; lack of seed retention; profuse branching; and low harvest index (Mahajan 2007; Ashri 1994, 1998; Kumar et al. 2007). Nonetheless sesame is widely cultivated especially in east Asia and Africa.

S. indicum has a diploid chromosome number of 2n=26 and belongs to the Pedaliaceae family. Sesame is an autogamous species with populations comprised of homozygous individuals. Sesame seed is rich in polyunsaturated oils (Davidson 1999; Wood 1999); oleic and linoeic acids (Mondal et al. 2010); and antioxidants such as sesamin, sesamolin, sesamol, sesamolinol and squalene (Suja et al. 2004; Mohamed and Awatif 1998). These antioxidants make sesame seed oil stable (Brown 2001; Uzun et al. 2007). Sesame seed has one of the highest oil contents with 50-60% oil (Uzun et al. 2002; Arslan et al. 2007) which is nearly equal to the oil content of olive (Kapoor 1990). Sesame seed oil has a reducing effect on cholesterol in the blood and also lowers blood pressure (Sankat et al. 2005; Banerjee and Kole 2008).

1.2. Origin and Domestication of Sesame (*Sesamum indicum* L.)

Although there is still a debate about the origin of sesame, Bedigian and associates substantiated that the origin of *S. indicum* was in the Indian subcontinent and that sesame's wild populations belong to the Western Indian Peninsula and parts of Pakistan (Bedigian 2010). According to the results of reciprocal crosses of sesame and its wild progenitor, the progenitor taxon was discovered as *Sesamum orientale var. malabaricum* Nar. (Bedigian 2010). These crosses were also done by other researchers who obtained the same result as Bedigian (2010) (Bish et al. 2004; Kawase 2000; Hiremath and Patil 1999; Kumar 2003; Kumar and Hiremath 2008). A further study which used chloroplast DNA regions, verified that *S. indicum* and *S. orientale var. malabaricum* are very closely related to each other (Bedigian 2010; Olmstead pers. comm. 2010). In addition, Bhat (1999) proved this similarity by RAPD markers.

1.3. World Sesame Seed Production with Emphasis on Turkey

According to FAO (Food and Agricultural Organization of the United Nations) statistics from 2009, China is the largest producer of sesame in the world (Figure 1.1). Turkey produces aproximately 21000 tonnes of sesame seed in a year.

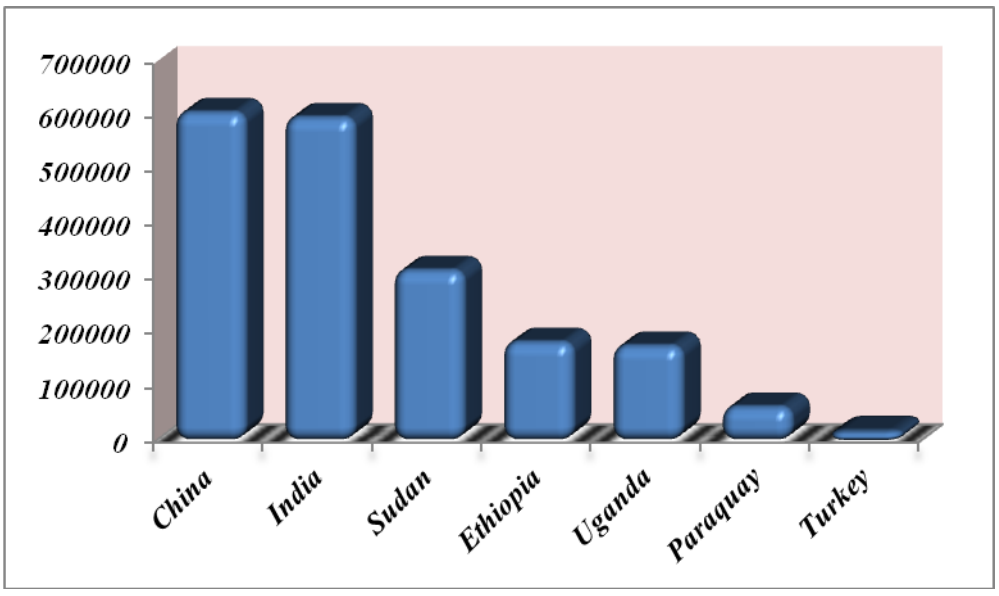

Figure 1.1. World production quantity of sesame seed.
(Source: FAOSTAT 2009)

Sesame is an important oilseed crop in Turkey. Anatolia played an important role in the migration of sesame between East and West, so it has variable sesame cultivars (Furat and Uzun 2010). There are also many sesame landraces in the country (Demir 1962). Sesame is used as a condiment in the making of cakes, cupcakes, cookies, biscuits and pastries.

1.4. Genetic Diversity

The study of plant genetic diversity is important for learning about the evolution of organisms, for assessing transferability of genes, and for improving desirable traits. The success of genetic improvement of *Sesamum indicum* L. by hybridization depends on the availabilty of diversity studies (Furat and Uzun 2010). One way to study sesame diversity is the measurement of polymorphism at the DNA level (Laurentin H. 2008).

There are only a few diversity studies about sesame in the literature. Dixit et al. (2005) studied diversity by developing 50 EST-SSR markers, ten of which were found to be polymorphic in 16 sesame accessions. Another study was done by Ercan et al. (2004) about diversity analysis of Turkish sesame populations using random amplified polymorphic DNA (RAPD) markers. Laurentin H.E. and Karlovsky P. (2006) showed the genetic relationships between 32 sesame accession representing different genotypes from Western Asia, India, China-Korea-Japan, Africa, and Central Asia. They obtained accessions from the Venezuelan Germplasm Collection. As a result of amplified length polymorphism (AFLP) experiments, 93% polymorphism was observed. Laurentin and collegues (2007) also investigated genomic and metabolic diversity in ten sesame accessions using AFLP technology. Ghulam and associates (2007) studied molecular diversity related with the geographical origin of sesame and morphological characteristics in 96 accesions of sesame which were collected throught the world. They used AFLP techniques.

1.5. Genetic Markers

Genetic markers are specific locations in the genome and correspond to polymorphisms in the DNA sequence. Genetic markers can be used to investigate how

organisms are genetically related. Genetic maps also can be developed using marker systems (Avise et al. 1995; Kumar et al. 1999). Development of high-density molecular mapping and sequencing have also demonstrated the possibility of identifying and tagging desired genes (Mohan et al. 1996; Kumar et al. 1999). Genetic markers can then be used to detect the gene and its function may be deduced by studying sequence diversity among allelles. There are basically two types of genetic markers: morpological markers and molecular markers (Kumar 1999).

1.5.1. Morphological Markers

Morphological markers are single genes with effects on the phenotype. Morphological markers are affected by environment, epistatic interactions and pleiotropic interactions. These single genes often interact in a dominant/recessive manner so heterozygosity or homozygosity cannot be detected. Thus, morphological markers often can not fully reveal genotype. They are inherited in a Mendelian manner but are limited in number (Kumar 1999; Jones et al. 2009; Mohan et al. 1996).

1.5.2. Molecular Markers

Molecular markers are divided into two groups. One group is biochemical markers like isozymes which are alternative forms of an enzyme. The other group is DNA markers (Kumar 1999). DNA markers are also called genic molecular markers (GMMs) (Varshney et al. 2007). DNA markers consist of two types: hybridization-based markers and PCR-based markers. Both types of markers can be either co-dominant or dominant (Varshney et al. 2007; Mohan et al. 1996; Kumar et al. 1999). Molecular markers provide the opportunity to detect polymorphic genes by not directly selecting the gene of interest. Instead, marker(s) linked to the gene of interest are selected. Therefore it is best to identify markers as close as possible to the target gene for efficient marker-assisted selection and transfer of the desired gene to other lines (Mohan et al. 1996). Polymorphism in the nucleotide sequence can be detected by many molecular marker techniques like restriction length polymorphism (RFLP), amplified length polymorphism (AFLP), simple sequence repeats (SSR), single nucleotide

polymorphism (SNP), cleaved amplified polymorphic sequences (CAPS), inter-simple sequence repeat amplification (ISSR), random amplified polymorphic DNA (RAPD), sequence characterized amplified region (SCAR), retrotransposon-microsatellite amplified polymorphism (REMAP), variable number tandem repeats (VNTR), sequence tagged sites (STS), and amplicon length polymorphisms (APLs) (Mohan et al. 1996; Kumar 1999; Varshney 2007; Somers et al. 2004).

Molecular marker technology also allows discovery of the genomic structure of various organisms and genotypic changes such as mutations, deletions, inversions and even single nucleotide differences (Jones et al. 2009). Markers allow prediction of genome organization and evolution and have practical application in plant breeding. They can also be used for phylogenetic analysis and positional cloning of genes (Kumar 1999). DNA marker technology allows plant breeders to overcome many difficulties of conventional breeding which involves several crosses and time consuming observation of the phenotype (Varshney et al. 2007; Somers et al. 2003; Kanazin et al. 2002).

1.5.2.1. Amplified Length Polymorphism (AFLP)

Amplified Length Polymorphism (AFLP) is a highly informative marker system, using restricted, amplified DNA fragments ranging from 60 to 600 base pairs. This technique was developed by Zabeau and Vos in 1993. The principle of this technique is specific PCR amplification of selected restriction fragments of DNA. Hundreds of amplified products are then simultaneously visualized by gel or capillary electrophoresis using fluorescent or radioactive labelling.

The steps for AFLP are initiated by digestion of the DNA sample with specific restriction enzymes (EcoRI/MseI). This is followed by ligation of primer specific adaptors to the restricted fragments. In the third step, the DNA fragments are amplified with twenty cycles of PCR. This is followed by a second selective PCR of fragments with specific adapters (Figure 1.2). One of the primers is labelled with fluorescent dye to allow visualization by laser reading during capillary electrophoresis. In the resulting output, the fragments are shown as peaks (Figure 1.3). The results are then scored for each peak of each sample, if the peak is present, the score is “1” and if the peak is absent the score is “0”.

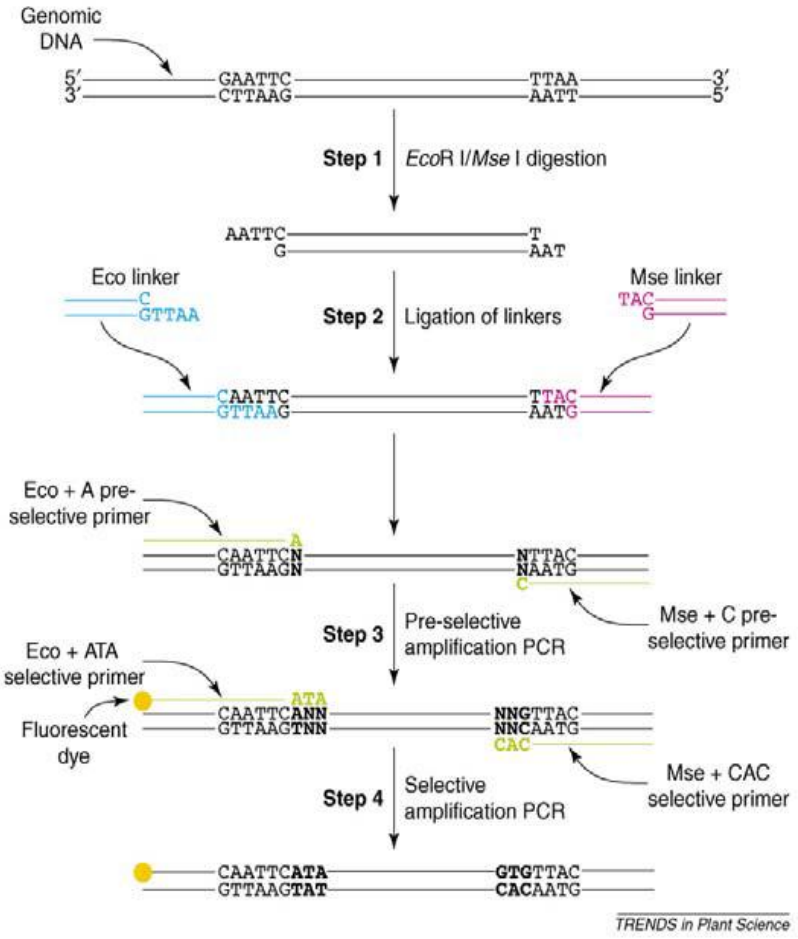

Figure 1.2. Basic steps of AFLP. Step I; Restriction by both rare and frequent cutter enzymes. Step II; Ligation by adaptors. Step III; Pre-selective PCR amplification with pre-selective primers having one more base. Step IV; Selective PCR amplification with selective primer combinations (Source: Meudt et al. 2007).

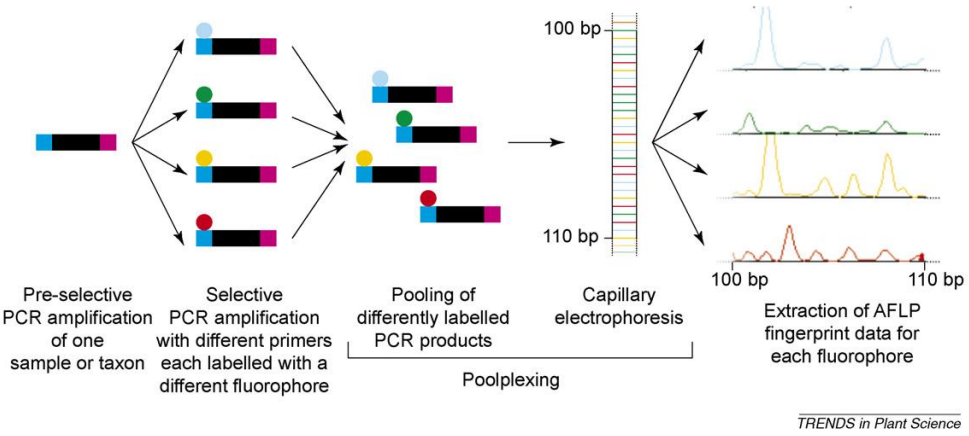

Figure 1.3. Capillary Electrophoresis. Labelling primers and visualizing on polyacrylamide gel (Source: Meudt et al. 2007).

The AFLP technique has some advantages as a molecular marker system including high reproducibility, high resolution and high multiplex ratio (Jones et al. 2009). Additionally, a small amount of DNA is sufficient to perform the analysis. Moreover, it is not necessary to know the DNA sequence of the sample. AFLP is a good technique to construct high density maps and study diversity (VanToai et al. 1997; Marsan et al. 2000; Sebastian et al. 2000) because it gives much polymorphic fragment information for each individual genotype (Mueller and Wolfenbarger 1999). On the other hand, AFLP is a dominant marker system so it does not provide information about allele frequencies.

There are some sesame AFLP studies in the literature. In Ghulam M. Ali, Sirato Yasumoto, and Masumi Seki-Katsuta's study (2007), 21 AFLP primer pairs were used to detect the geographical origin and morphological characteristics of *S. indicum* with 35% of the primer combinations found to be polymorphic. Using UPGMA they obtained two groups of cultivated sesame, one of them East Asian in origin while the other was of South Asian origin. In another study, AFLP was used for identifying a marker for the closed capsule mutant trait of sesame following bulked segregant analysis (Uzun et al. 2002). They screened 72 AFLP primer combinations for linkage and only one linked marker was identified for the mutant trait.

1.5.2.2. SSR As a Molecular Marker

Simple sequence repeats (SSRs) or microsatellites are very useful for applications in plant breeding because of their reproducibility, high genome coverage, co-dominant inheritance, transferability to close species, multiallelic nature and relative abundance. They are informative, polymorphic sequences that are short, species-specific and highly variable (hypervariable). SSRs are composed of 6 bp or shorter tandemly, middle repetitive sequences. SSRs occur in the genome with an estimated frequency ranging from one SSR every 29 to 50 kb (Lagercrantz et al. 1993; Morgante and Oliveriri 1993). These repeats may be dinucleotides [e.g. (AC)n]; trinucleotides [(TCT)n] or tetranucleotides [(TATG)n]. (Liu et al. 2005; Ma et al. 1966; Jones et al. 2009; Kahl 2001; Cardle et al. 2000; Vrashney et al. 2005). (AT_n) is the most common SSR in plant genomes (Ma et al. 1966). Polymorphism is detected according to the

number of repetitive sequences of the two allelles at one locus (Jones et al. 2007). The mechanisms of expansion and loss of simple sequence repeats in populations are replication slippage and unequal cross-over (Tautz 1984; Morgante et al. 2002). SSRs can be used as an anchor marker in comparative mapping and evolutionary studies. SSRs are also useful for marker-assisted-selection in many crop species and provide data to associate genotypic and phenotypic variation in mapping studies (Varshney 2000).

1.6. SSR Marker Development

1.6.1. EST-SSRs and Genomic-SSRs

SSR markers can be designed from EST sequences or genomic sequences. The design of SSR primers from ESTs is cheaper and easier than designing from genomic sequences because ESTs and other genic sequences are the byproducts of gene and EST sequencing projects and are publicly available. Because EST sequences are coding regions, their transferability in related species is higher than genomic sequences. On the other hand this feature can be a disadvantage of EST-SSRs because conserved sequences generally have less polymorphism than the whole genome. Because of their higher polymorphism, SSRs designed from genomic SSR motifs are especially useful for diversity and evolution studies, linkage mapping and comparative mapping. Also, assembling genomic sequences can be easier than for ESTs. For efficient SSR identification, ESTs must be reduced to unigenes. Because ESTs are transcribed regions and are obtained from cDNAs, high levels of expression, alternative splicing, and post-transcriptional regulation can result in more than one expressed mRNAs from one gene region. Thus, assembly of EST sequences before SSR identification prevents over-representation of certain genic regions.

1.6.2. Assembling

Assembly means bringing together of the unigenes which are initially restricted, cloned and sequenced (Figure 1.4). They should contain no contamination or non-native

sequences. Contamination may arise from adaptors which are left from the host genome or vector sequences. Contamination is reduced by vector trimming before assembly. One program that does assembly is CAP3 (Tang et al. 2008).

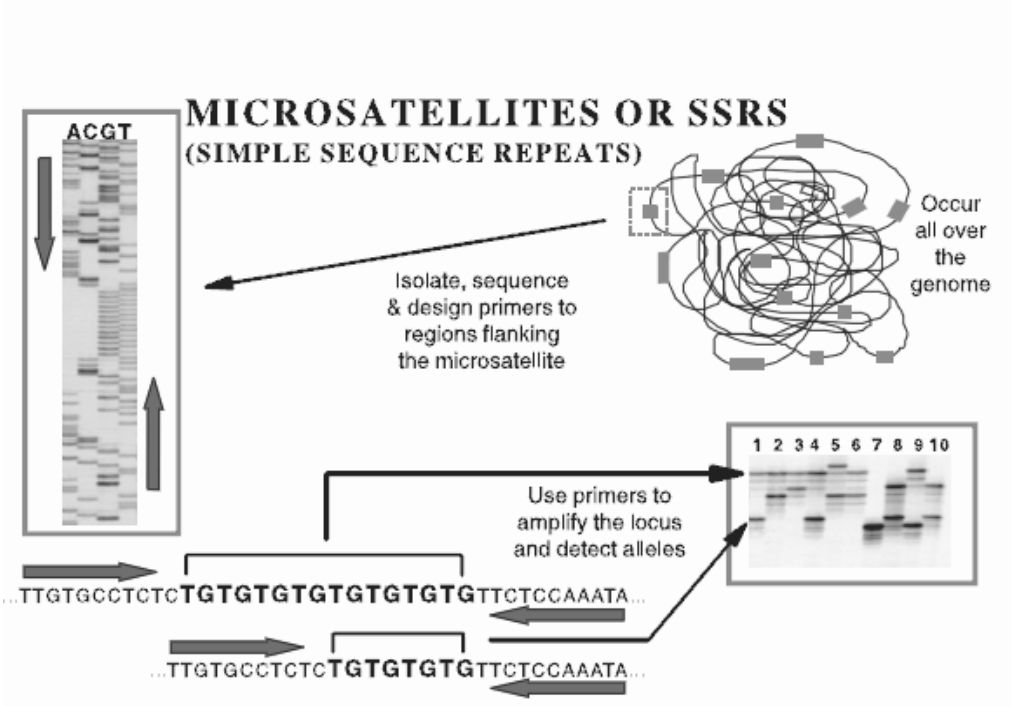

Figure 1.4. Schematic showing SSR location, isolation, sequencing, primer design,amplification and detection of the locus(Source: Jones et al 2007; Courtesy of Angela Karp).

1.6.3. Criteria of SSR Design

Assembled sequences of the host genome or ESTs are searched for SSRs and suitable primers flanking the SSR according to some criteria. These criteria are: length of the SSR motif (dinucleotide, trinucleotide, tetranucleotide or pentanucleotide repeats) for example $(AT)_n$; repeat number for example $(AT)_5$ (n=5); GC content and annealing temperature of the primers. Two factors determine total length of an SSR. One factor is the SSR repeat motif such as trinucleotide or tetra-, penta-, hexanucleotide. The other factor is the number of motif repeats in the sequence such as eight repeats of a dinucleotide or five repeats for a tetranucleotide. According to the literature, short total SSR length revealed a higher frequency of polymorphism than longer SSRs. (Tang et al. 2008; Varshney et al. 2005). The GC content, PCR product size and melting temperature (T_m) of each forward and reverse primer, primer size, and product size are

also very important features of reliable and correct primer design. Minimum 20% GC content is better for obtaining more stable primers which means that the primer can bind to the sequence more tightly. The difference between T_m value of forward and reverse primer should not be too great so that an appropriate annealing temperature for PCR conditions can be found (Li-Bin 2008; Tang et al. 2008). Primer size must be 18 nucleotides at least and the product size can range from 100 to 500 nucleotides (Tang et al. 2008).

1.6.4. Identification of EST-SSRs and Genomic-SSRs

The first identification of SSRs in plant gene sequences was done by Morgante and Oliveri in 1993 (Morgante and Oliveiri 1993). Since then SSR sequences are abundant in databases. During the initial years regular expression matches "BLASTN" tools were used in the form of FASTA or BLAST2. Currently, there are many databases which identify SSR motifs in sequence information, such as MIcroSAtellite "MISA", SSRFinder, SSRSEARCH, and BatchPrimer3 (Scott et al. 2000; Temnykh et al. 2000).

1.7. Frequency and Distribution of SSRs

SSR repeat motifs can be dinucleotide repeats (DNRs), trinucleotide repeats (TNRs), or tetranucleotide repeats (TTNRs) (Varshney et al. 2005). The density of SSRs in unigenes can depend on the structure and minimum length of the SSRs (Tang et al. 2008). Moreover the frequency of different length SSRs and repeat motifs depends on the criteria used to identify the SSRs and the database search program used (Varshney 2002). According to research, trinucleotide repeats were found to be most frequent, followed by DNRs or TTNRs in the EST sequences in many plants. As a result, the frequency of DNRs, TNRs and TTRNs are variable among plant species (Varshney et al. 2002; 2005). Moreover, the TNRs can have different lengths even in 3' UTR, 5'UTR and also in coding regions. Trinucleotide SSR motifs do not cause frameshift mutations and, therefore, may be found in such coding regions (Varshney et al. 2005; Tang et al. 2008). While polymorphism in 3'UTR-derived SSRs was greater

than in 5'UTR-derived SSRs (Scott et al. 2000), others have found more polymorphism in the 5'UTR than the 3'UTR (Tang et al. 2008).

1.8. Aim of the Study

Sesame (*S. indicum*) with its rich seed content is a very important crop in Turkey and all over the world. Despite its importance, studies about the diversity, domestication, evolution and molecular genetics of sesame are not sufficient.

Our aim was to test AFLP marker combinations on Turkish sesame accessions, examine the diversity of the accessions and see their distribution which is important for new breeding studies. For this goal, five AFLP combinations were used to determine the relationships among 158 sesame accessions. Furthermore, new genomic-SSR markers were developed from the sesame genome. These markers should be useful for future genetics and breeding studies in sesame.

CHAPTER 2

MATERIALS AND METHODS

2.1. Materials

2.1.1. Plant Materials

A total of 160 sesame accessions from the USDA-ARS Plant Germplasm Inspection Station, Bestville, Maryland, USA, were used for this study. An additional 27 sesame accession were included from Aegean Agricultural Research Institute, Menemen, İzmir. Pedigree numbers were given to each accession (Table 2.1). According to the purity of their DNA, 158 accessions were chosen for diversity analysis.

Ten seeds from each accession were planted and grown in soil containing peat moss, perlite and natural fertilizer. Seeds were germinated and grown in the growth chamber with 23-25° C, 24 h photoperiod and approximately 35% humidity at İzmir Institute of Technology, İzmir, Turkey.

Table 2.1. Turkish sesame accessions' pedigree numbers, names, sources and their locations.

Pedigree number	**Accession name**	**Source**	**Location**
170747 01 SD	*S. indicum* L.	USDA	Turkey
170745 01 SD	*S. indicum* L.	USDA	Turkey
170744 01 SD	*S. indicum* L.	USDA	Turkey
170743 01 SD	*S. indicum* L.	USDA	Turkey
170742 01 SD	*S. indicum* L.	USDA	Turkey
170739 01 SD	*S. indicum* L.	USDA	Turkey
170738 01 SD	*S. indicum* L.	USDA	Turkey
170737 01 SD	*S. indicum* L.	USDA	Turkey
170735 01 SD	*S. indicum* L.	USDA	Turkey
238487 01 SD	*S. indicum* L.	USDA	Turkey
238470 01 SD	*S. indicum* L.	USDA	Turkey

(cont. on next page)

Table 2.1. (cont.)

170722 01 SD	*S. indicum* L.	USDA	Turkey
170718 01 SD	*S. indicum* L.	USDA	Turkey
170717 01 SD	*S. indicum* L.	USDA	Turkey
170715 01 SD	*S. indicum* L.	USDA	Turkey
238469 01 SD	*S. indicum* L.	USDA	Turkey
238468 01 SD	*S. indicum* L.	USDA	Turkey
238466 01 SD	*S. indicum* L.	USDA	Turkey
238448 01 SD	*S. indicum* L.	USDA	Turkey
238447 01 SD	*S. indicum* L.	USDA	Turkey
238446 01 SD	*S. indicum* L.	USDA	Turkey
177071 01 SD	*S. indicum* L.	USDA	Turkey
177070 01 SD	*S. indicum* L.	USDA	Turkey
175908 01 SD	*S. indicum* L.	USDA	Turkey
175907 01 SD	*S. indicum* L.	USDA	Turkey
238449 01 SD	*S. indicum* L.	USDA	Turkey
238450 01 SD	*S. indicum* L.	USDA	Turkey
165021 01 SD	*S. indicum* L.	USDA	Turkey
167115 01 SD	*S. indicum* L.	USDA	Turkey
167224 01 SD	*S. indicum* L.	USDA	Turkey
167248 01 SD	*S. indicum* L.	USDA	Turkey
170733 01 SD	*S. indicum* L.	USDA	Turkey
238451 01 SD	*S. indicum* L.	USDA	Turkey
238453 01 SD	*S. indicum* L.	USDA	Turkey
238455 01 SD	*S. indicum* L.	USDA	Turkey
238456 01 SD	*S. indicum* L.	USDA	Turkey
238417 01 SD	*S. indicum* L.	USDA	Turkey
238416 01 SD	*S. indicum* L.	USDA	Turkey
179486 01 SD	*S. indicum* L.	USDA	Turkey
179484 01 SD	*S. indicum* L.	USDA	Turkey
179483 01 SD	*S. indicum* L.	USDA	Turkey
179482 01 SD	*S. indicum* L.	USDA	Turkey
238419 01 SD	*S. indicum* L.	USDA	Turkey
238420 01 SD	*S. indicum* L.	USDA	Turkey
238422 01 SD	*S. indicum* L.	USDA	Turkey
238435 01 SD	*S. indicum* L.	USDA	Turkey
238437 01 SD	*S. indicum* L.	USDA	Turkey

(cont. on next page)

Table 2.1. (cont.)

170711 01 SD	*S. indicum* L.	USDA	Turkey
170713 01 SD	*S. indicum* L.	USDA	Turkey
170714 01 SD	*S. indicum* L.	USDA	Turkey
238438 01 SD	*S. indicum* L.	USDA	Turkey
238439 01 SD	*S. indicum* L.	USDA	Turkey
238440 01 SD	*S. indicum* L.	USDA	Turkey
167343 01 SD	*S. indicum* L.	USDA	Turkey
238429 01 SD	*S. indicum* L.	USDA	Turkey
238428 01 SD	*S. indicum* L.	USDA	Turkey
238430 01 SD	*S. indicum* L.	USDA	Turkey
238431 01 SD	*S. indicum* L.	USDA	Turkey
238432 01 SD	*S. indicum* L.	USDA	Turkey
238433 01 SD	*S. indicum* L.	USDA	Turkey
238426 01 SD	*S. indicum* L.	USDA	Turkey
238423 01 SD	*S. indicum* L.	USDA	Turkey
238458 01 SD	*S. indicum* L.	USDA	Turkey
238434 01 SD	*S. indicum* L.	USDA	Turkey
170730 01 SD	*S. indicum* L.	USDA	Turkey
170729 01 SD	*S. indicum* L.	USDA	Turkey
205229 01 SD	*S. indicum* L.	USDA	Turkey
205225 01 SD	*S. indicum* L.	USDA	Turkey
205228 01 SD	*S. indicum* L.	USDA	Turkey
205227 01 SD	*S. indicum* L.	USDA	Turkey
238471 01 SD	*S. indicum* L.	USDA	Turkey
238473 01 SD	*S. indicum* L.	USDA	Turkey
238474 01 SD	*S. indicum* L.	USDA	Turkey
238475 01 SD	*S. indicum* L.	USDA	Turkey
238476 01 SD	*S. indicum* L.	USDA	Turkey
238477 01 SD	*S. indicum* L.	USDA	Turkey
238478 01 SD	*S. indicum* L.	USDA	Turkey
238479 01 SD	*S. indicum* L.	USDA	Turkey
238481 01 SD	*S. indicum* L.	USDA	Turkey
238482 01 SD	*S. indicum* L.	USDA	Turkey
238483 01 SD	*S. indicum* L.	USDA	Turkey
238485 01 SD	*S. indicum* L.	USDA	Turkey
238486 01 SD	*S. indicum* L.	USDA	Turkey

(cont. on next page)

Table 2.1. (cont.)

179481 01 SD	*S. indicum* L.	USDA	Turkey
240850 01 SD	*S. indicum* L.	USDA	Turkey
240848 01 SD	*S. indicum* L.	USDA	Turkey
240847 01 SD	*S. indicum* L.	USDA	Turkey
170726 01 SD	*S. indicum* L.	USDA	Turkey
170725 01 SD	*S. indicum* L.	USDA	Turkey
170724 01 SD	*S. indicum* L.	USDA	Turkey
170723 01 SD	*S. indicum* L.	USDA	Turkey
240846 01 SD	*S. indicum* L.	USDA	Turkey
240845 01 SD	*S. indicum* L.	USDA	Turkey
240844 01 SD	*S. indicum* L.	USDA	Turkey
238488 01 SD	*S. indicum* L.	USDA	Turkey
179035 01 SD	*S. indicum* L.	USDA	Turkey
179034 01 SD	*S. indicum* L.	USDA	Turkey
170710 01 SD	*S. indicum* L.	USDA	Turkey
170708 01 SD	*S. indicum* L.	USDA	Turkey
179033 01 SD	*S. indicum* L.	USDA	Turkey
179032 01 SD	*S. indicum* L.	USDA	Turkey
179031 01 SD	*S. indicum* L.	USDA	Turkey
177541 01 SD	*S. indicum* L.	USDA	Turkey
177540 01 SD	*S. indicum* L.	USDA	Turkey
170759 01 SD	*S. indicum* L.	USDA	Turkey
170758 01 SD	*S. indicum* L.	USDA	Turkey
170757 01 SD	*S. indicum* L.	USDA	Turkey
170755 01 SD	*S. indicum* L.	USDA	Turkey
170748 01 SD	*S. indicum* L.	USDA	Turkey
170749 01 SD	*S. indicum* L.	USDA	Turkey
170752 01 SD	*S. indicum* L.	USDA	Turkey
170753 01 SD	*S. indicum* L.	USDA	Turkey
170760 01 SD	*S. indicum* L.	USDA	Turkey
170761 01 SD	*S. indicum* L.	USDA	Turkey
170762 01 SD	*S. indicum* L.	USDA	Turkey
240852 01 SD	*S. indicum* L.	USDA	Turkey
240853 01 SD	*S. indicum* L.	USDA	Turkey
240854 01 SD	*S. indicum* L.	USDA	Turkey
240856 01 SD	*S. indicum* L.	USDA	Turkey

(cont. on next page)

Table 2.1. (cont.)

263373 01 SD	*S. indicum* L.	USDA	Turkey
263375 01 SD	*S. indicum* L.	USDA	Turkey
177072 01 SD	*S. indicum* L.	USDA	Turkey
204623 01 SD	*S. indicum* L.	USDA	Turkey
182295 01 SD	*S. indicum* L.	USDA	Turkey
182294 01 SD	*S. indicum* L.	USDA	Turkey
182293 01 SD	*S. indicum* L.	USDA	Turkey
179490 01 SD	*S. indicum* L.	USDA	Turkey
179489 01 SD	*S. indicum* L.	USDA	Turkey
179488 01 SD	*S. indicum* L.	USDA	Turkey
179488 01 SD	*S. indicum* L.	USDA	Turkey
179487 01 SD	*S. indicum* L.	USDA	Turkey
238465 01 SD	*S. indicum* L.	USDA	Turkey
238464 01 SD	*S. indicum* L.	USDA	Turkey
238463 01 SD	*S. indicum* L.	USDA	Turkey
170727 01 SD	*S. indicum* L.	USDA	Turkey
238462 01 SD	*S. indicum* L.	USDA	Turkey
238461 01 SD	*S. indicum* L.	USDA	Turkey
238460 01 SD	*S. indicum* L.	USDA	Turkey
170728 01 SD	*S. indicum* L.	USDA	Turkey
238459 01 SD	*S. indicum* L.	USDA	Turkey
170732 01 SD	*S. indicum* L.	USDA	Turkey
175906 01 SD	*S. indicum* L.	USDA	Turkey
174355 01 SD	*S. indicum* L.	USDA	Turkey
174354 01 SD	*S. indicum* L.	USDA	Turkey
174353 01 SD	*S. indicum* L.	USDA	Turkey
173101 01 SD	*S. indicum* L.	USDA	Turkey
173100 01 SD	*S. indicum* L.	USDA	Turkey
170769 01 SD	*S. indicum* L.	USDA	Turkey
170768 01 SD	*S. indicum* L.	USDA	Turkey
170767 01 SD	*S. indicum* L.	USDA	Turkey
170765 01 SD	*S. indicum* L.	USDA	Turkey
170764 01 SD	*S. indicum* L.	USDA	Turkey
170763 01 SD	*S. indicum* L.	USDA	Turkey
238445 01 SD	*S. indicum* L.	USDA	Turkey
238444 01 SD	*S. indicum* L.	USDA	Turkey

(cont. on next page)

Table 2.1. (cont.)

238442 01 SD	*S. indicum* L.	USDA	Turkey
238441 01 SD	*S. indicum* L.	USDA	Turkey
ORHANGAZİ-99	*S. indicum* L.	AARI	Turkey
TAN-99	*S. indicum* L.	AARI	Turkey
KEPSUT-99	*S. indicum* L.	AARI	Turkey
OSMANLI-99	*S. indicum* L.	AARI	Turkey
CUMHURİYET-99	*S. indicum* L.	AARI	Turkey
TR 45524	*S. indicum* L.	AARI	Adana
TR 45572	*S. indicum* L	AARI	Adıyaman
TR 39702	*S. indicum* L.	AARI	Antalya
TR 61609	*S. indicum* L.	AARI	Aydın
TR 38106	*S. indicum* L.	AARI	Balıkesir
TR 76589	*S. indicum* L.	AARI	Bilecik
TR 42870	*S. indicum* L.	AARI	Bursa
TR 68411	*S. indicum* L.	AARI	Çanakkale
TR 61927	*S. indicum* L.	AARI	Denizli
TR 45642	*S. indicum* L.	AARI	Diyarbakır
TR 38253	*S. indicum* L.	AARI	Edirne
TR 45599	*S. indicum* L.	AARI	Elazığ
TR 42145	*S. indicum* L.	AARI	Gaziantep
TR 39695	*S. indicum* L.	AARI	İçel
TR 52540	*S. indicum* L.	AARI	İzmir
TR 45543	*S. indicum* L.	AARI	Kahramanmaraş
TR 52533	*S. indicum* L.	AARI	Kars
TR 42635	*S. indicum* L.	AARI	Kırklareli
TR 50128	*S. indicum* L.	AARI	Kütahya
TR 45596	*S. indicum* L.	AARI	Malatya
TR 64094	*S. indicum* L.	AARI	Manisa
TR 45673	*S. indicum* L.	AARI	Mardin
TR 39716	*S. indicum* L.	AARI	Muğla
TR 37513	*S. indicum* L.	AARI	Siirt
TR 45707	*S. indicum* L.	AARI	Şanlıurfa
TR 38356	*S. indicum* L.	AARI	Tekirdağ
TR68905	*S. indicum* L.	AARI	Uşak
	S.alatum L.		

2.2. Methods

2.2.1. Genomic-SSR Design

To identify genomic SSR primers, Muganlı F4 sesame DNA was isolated with Promega Wizard Genomic DNA Purification kit. DNA amount was determined and quantified with 3% agarose gel electrophoresis and Nanodrop ND-1000 spectrophotometer. DNA was cloned by emulsion-based clonal amplification. After cloning, 454 pyrosequencing was used to sequence the genomic DNA. The output file was obtained as .ssf file format. For assembly, the Mimicking Intelligent Read Assembly (MIRA) (Chevreux et al. 1999) tool was chosen because MIRA has usage statistics and is distributed as free software under general public licence. The .ssf file format was not suitable for using MIRA software therefore, a ssf extract program was used to convert .ssf file to fasta format. To obtain the best assembly, a script, which determine the values of unassembled reads and maximum number of contigs, was improved. (bioinformatics.iyte.edu.tr/ACE/stats.zip). Afterwards, the contigs obtained with assembly by MIRA software were analyzed for primer design with BatchPrimer3 (Rozen and Skaletsky 2000; You et al. 2007) using the criteria shown in Table 2.2.

Afterwards, to filter the primers a script, which can be downloaded and distributed under the GNU GPL v.3.0 license from the website (bioinformatics.iyte.edu.tr/ACE/PConf.zip) was improved. This script provided analysis of primers individually by looking at each contig that had been used for each primer, GC content and the other properties because it read both the detailed contig information and the primer data file simultaneously.

Table 2.2. Primer criteria.

	Minimum	Optimum	Maximum
Product size	100	200	500
Primer size	18	21	25
Primer Tm	45	50	60
Max. Tm difference			2
Primer GC%	40	50	60

2.2.2. DNA Extraction

DNA extraction was applied to 2-4 leaf plantlets by using a modified CTAB-DNA extraction protocol from Laurentin et al. (2006) and Abdellatef et al. (2008). The protocol begins with the collection of the youngest leaves of the plant. The CTAB DNA extraction buffer was heated at 65° C, and β-mercaptoethanol (12 µl per 60 ml buffer) was added to prevent phenolic contamination. Buffer was added (250 µl) to the leaf samples and the plant tissue was homogenized. Ater homogenization, 500 µl buffer was added to the homogenate. The samples were incubated at 65° C for 40 minutes to degrade the cell wall, cell components and the nuclear envelope. After incubation, 750 µl chloroform-isoamylalcohol (24:1) mixture was added to the samples. They were shaken vigorously approximately 100 times and then centrifuged at 10000 rpm for 10 minutes. This step was repeated twice. Following the chloroform stage, the supernatants were taken to new tubes and approximately 500 µl cold isopropanol was added to each sample and inverted carefully to precipitate DNA. The samples were centrifuged at 10000 rpm for 5 minutes, then the isopropanol was removed and 500 µl 70% ethanol was added to the DNA. After cleaning with 70% ethanol, DNAs were dried then dissolved in 30 µl Tris-EDTA (TE) buffer. The amount and quality of DNAs were examined both by agarose gel electrophoresis and Nanodrop ND-1000 Spectrophotometer. The 230/280 and 260/280 ratios were screened for each sample of DNA by nanodrop. The amount of DNA was diluted to approximately 100 ng/µl for each sample. DNAs were stored at -20 °C in TE buffer.

2.2.3. AFLP Analysis

Amplified Fragment Length Polymorphism method was applied by using Invitrogen Core Reagent and Starter Kit and Promega Go-*Taq* DNA Polymerase enzyme. Analysis started with the restriction of genomic DNAs by a combination of a general cutter EcoRI; and a specific cutter, MseI. Initially, the DNA amount was adjusted to 300 ng/µl in 18µl dH_2O for each sample. For each sample, 2µl of enzyme mixture and 5µl 5x reaction buffer were added to DNA and the mixture was incubated at 37 °C for 2 hours then at 70 °C for 15 minutes to eliminate restriction enzyme activity. At the

second step, ligation of adapters was done by adding 24 µl adapter-ligation solution and 1µl T_4 DNA Ligase enzyme for each sample. Samples were incubated for 2 hours at 20 °C. The ligated DNA was diluted at a 1/10 ratio with TE buffer (10 µl ligated DNA + 90 µl TE buffer). At the third step, the diluted DNA was amplified in the Pre-amplification step of AFLP. Preamplification was done by adding 40 µl preamplification mixture, 5 µl 10X PCR buffer and 1 µl (500u *Go-Taq*) *Taq* polymerase enzyme to 5 µl diluted DNA of each sample. The PCR conditions were: 94 ° C for 30 seconds; 65 ° C for 30 seconds; 72 ° C for one minute of 20 cycles. The PCR product of preamplification PCR was diluted at a 1:50 ratio with TE buffer (3µl PCR product + 147 µl TE buffer). The last part of AFLP was selective PCR. Initially, the fluorescent labelled specific primer concentration (E-ACA primer concentration is 27.8 ng/µl) was adjusted. Two PCR mixes were prepared as mix1 and mix 2. Mix1 contained only the primers in a 1/9 ratio. Mix 2 consisted of 79 µl dH_2O, 20 µl 10X PCR buffer, 1µl Taq polymerase enzyme for 10 reactions. Then, 5µl of the diluted DNA from the preamplification step, 5µl mix 1, and 10 µl mix 2 were mixed. Selective PCR conditions were: one cycle of 30 seconds at 94 ° C, 30 seconds at 65 °C, one minute at 72 ° C, and 12 cycles of 30 seconds at 94 ° C, 30 seconds at 65 ° C (with a 0,7 ° C decrease), one minute at 72 ° C; the last 23 cycles of 30 seconds at 94 ° C, 30 seconds at 56 ° C, one minute at 72 ° C.

After selective PCR, 3 µl PCR product was taken and mixed with 27 µl sample loading solution (SLS), which contains formamide and prevents the DNA fragments from joining each other again. In addition, 0,5 µl 600 bp size standard (Beckman Coulter) was added. Then 50 µl mineral oil was added. Finally, the samples were run in CEQ 8800 Sequencer (Beckman Coulter) with 200 µl separation buffer. The capillary array with 8 columns was used. We obtained peaks for each combination when the process were done. Five combinations were used in this study (Table 2.3), from Laurentin and Karlovsky (2006). The electropherograms were scored visually.

Table 2.3. The AFLP combinations.

Combinations	M-CAT	M-CAG	M-CAC	M-CAA	M-CTC
E-ACA	√	√	√	√	√

2.2.4. Data Analysis

The result were scored as present (1) or absent (0). NTSYS-pc version 2-2 (Numerical Taxanomy Multivariate Analysis System, Exeter Software, Setauket N.Y.) software program was used to obtain a similarity matrix and a dendrogram. Construction of the similarity matrix was done by Jaccard similarity index. Jaccard similarity was calculated according to $J(i_1,i_2) = a/(a+b+c)$ formula (Kosman and Leonard 2004). The dendrogram was constructed with Unweighted Pair Group Method with Arithmetic Averages (UPGMA) and the SAHN clustering program. We applied Mantel test (1967) to compare the goodness of fit between similarity matrix and the dendrogram which was constructed by UPGMA. Polymorphism intention contents (PICs) were calculated for each allele of each combinations. PIC estimates the percent of polymorphism content of the combinations for each sample and compare them. PIC calculation was done according to $PIC_i = 2\ f_i\ (1 - f_i)$ formula: where PIC_i is the polymorphism information content of marker "i" , f_i is the frequency of band present and 1- f_i is the frequency of the band absence (Roldan-Ruiz I. et al. 2000. Furthermore 2D and 3D plot graphics were obtained by principal component analysis (PCA) in NTSYS).

CHAPTER 3

RESULT AND DISCUSSION

3.1 AFLP Analysis

AFLP analysis was applied to 158 Turkish sesame accessions which were chosen according to quality of their DNA (260/230 and 260/280 ratio must be between 1.85-2.25 at nanodrop results). Five specific AFLP primer combinations were tested on the accessions. Polymorphism was detected for all combinations (Table 3.1). Each combination commonly gave peaks between 70-400 nucleotides in size (Figure 3.1). The peaks were analysed visually to detect polymorphism between the accessions. A total of 148 polymorphic fragments were obtained. The most fragments (46) were obtained with combination E-ACA M-CAT followed by E-ACA M-CAC (34 fragments), E-ACA M-CAA (29 fragments), E-ACA M-CTC (25 fragments) and E-ACA M-CAG (14 fragments). PIC (Polymorphism Information Content) values were calculated for each allele. The highest average PIC value belonged to combination E-ACA M-CAA (0.35) and the lowest PIC value belonged to E-ACA M-CTC (0.27). Combination E-ACA M-CAT was also highly polymorphic with average PIC value of 0.32 (Table 3.2).

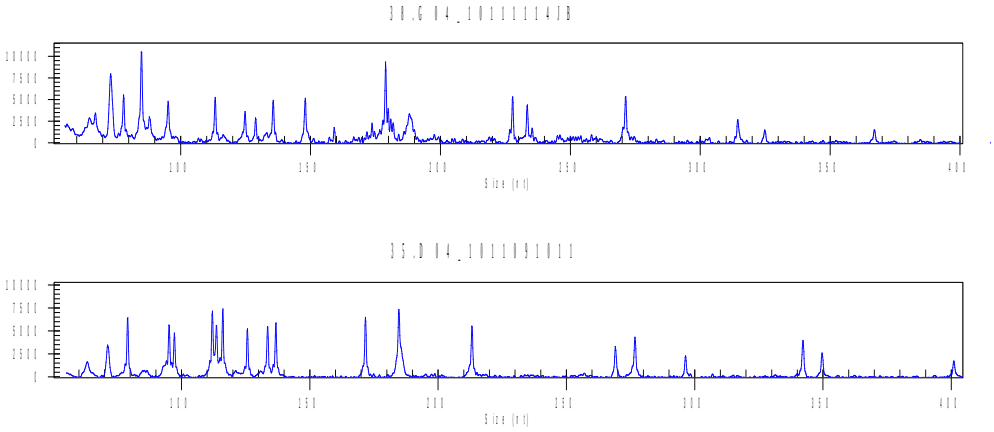

Figure 3.1. AFLP results of two samples from E-ACA M-CAC combination.

Table 3.1. The characteristics of AFLP combinations used to detect diversity between sesame accessions.

Total Number of Combinations	Number of Polymorphic Combinations	Total Number of Polymorphic Bands	Percent Polymorphism
5	5	148	100%

Table 3.2. Polymorphism information content of primer combinations.

Combinations	Number of Fragments	Average PIC ± SE	Range of PIC
E-ACA M-CAG	14	0.29 ± 0.64	0.08-0.48
E-ACA M-CAC	34	0.29 ± 0.03	0.01-0.50
E-ACA M-CAA	29	0.35 ± 0.02	0.06-0.50
E-ACA M-CTC	25	0.27 ± 0.03	0.01-0.44
E-ACA M-CAT	46	0.32 ± 0.02	0.08-0.50

The AFLP scores were used to construct a similarity matrix by applying NTSYSpc version 2.2 using Jaccard similarity matrix. The UPGMA method was used to convert the matrix to a dendrogram with NTSYS program. The Mantel goodness of fit test made comparision between the matrix and the dendrogram. The R value of the test was 0.79 indicating a good fit between the matrix and the dendrogram. The dendrogram contained three main clusters A, B, C, respectively (Figure 3.2). Similarity among the sesame accessions varied between 57% and 95%. The dendrogram was difficult to read because of the number of accessions, therefore the clusters were re-drawn and are shown separately. Cluster A contained 27 sesame accessions and their similarity index ranged from 61% to 89%. Two different main groups were found within cluster A with approximately 62% similiarity (Figure 3.3). The most crowded cluster was cluster B which contained 120 accessions with similarity ranging from 62% to 95% (Figure 3.4). Cluster C contained 11 accessions and genetic similarity ranged from 45% to 76% (Figure 3.5). Cluster A and B contained three important sesame accessions called African, Korea and *Sesamum alatum* L.. The African accession was detected in cluster B and had approximately 68% similarity to Turkish lines. This result indicated that the African accession was not very different from the Turkish accessions. The Korean accession and *Sesamum alatum* L. which was our outgroup in this study were found in Cluster C. *S. alatum* had approximately 59% to similarity Turkish lines

but 76% similarity to the Korean accession. The dendrogram confirmed *S. alatum* as an outgroup and also indicated that the Korean accession is genetically dissimilar to Turkish accessions and, therefore, may be a good parent for mapping population development.

Principle component analysis (PCA) was performed using NTSYS. The first three eigen values explained 21.9%, 7.5% and 6.3% of the variaton. The PCA did not show any distinct clustering of the accessions (Figure 3.6).

Previous diversity studies about sesame used AFLP and RAPD marker techniques. Laurentin and his colleagues studied the relationship between metabolic and genomic diversity in sesame (*S. indicum*) using eight AFLP marker combinations in ten sesame accessions showed 95% polymorphism with the 308 AFLP markers and the metabolic diversity tested by HPLC (2008). According to their results the genomic and the metabolic patterns differed indicating that selection plays an important role in evolution of metabolic events of sesame. So, when the selectible markers are used on any crop, they should be complemented with their metabolic profiles. In another study which was done by Laurentin and Karlovsky genetic relationships and genomic diversity in 32 sesame (*S. indicum*) accessions was studied using eight AFLP marker combinations (2006). They obtained 95% polymorphism with 457 fragments and a dendrogram whose similarity ranged from 39% to 58% and 20% total variation in allele frequencies corresponds to differences among groups. Furthermore, they indicated that there were no similarities between morphological features and their genomic origins. In another study, 38 Turkish sesame accessions were tested for genetic diversity with 39 RAPD markers by Ercan and his colleagues (2003). They found 78% polymorphism. Ghulam and his team detected 35% polymorphism among 96 sesame accessions from all over the world using 21 AFLP marker combinations (2007). We detected 148 polymorphic fragments with 5 AFLP combinations with 98% polymorphism on 158 Turkish sesame accessions. AFLP is more advantagous than the RAPD marker system to study polymorphism because of its high genome coverage and resolution. We obtained more polymorphism with fewer primers than Ercan's study. On the other hand, the diversity of Turkish sesames was proven both in our study and Ercan's. In general worldwide variation is not very high according to Ghulam' study. However, we found that there is a good variation between Turkish sesame (*Sesamum indicum* L.) lines.

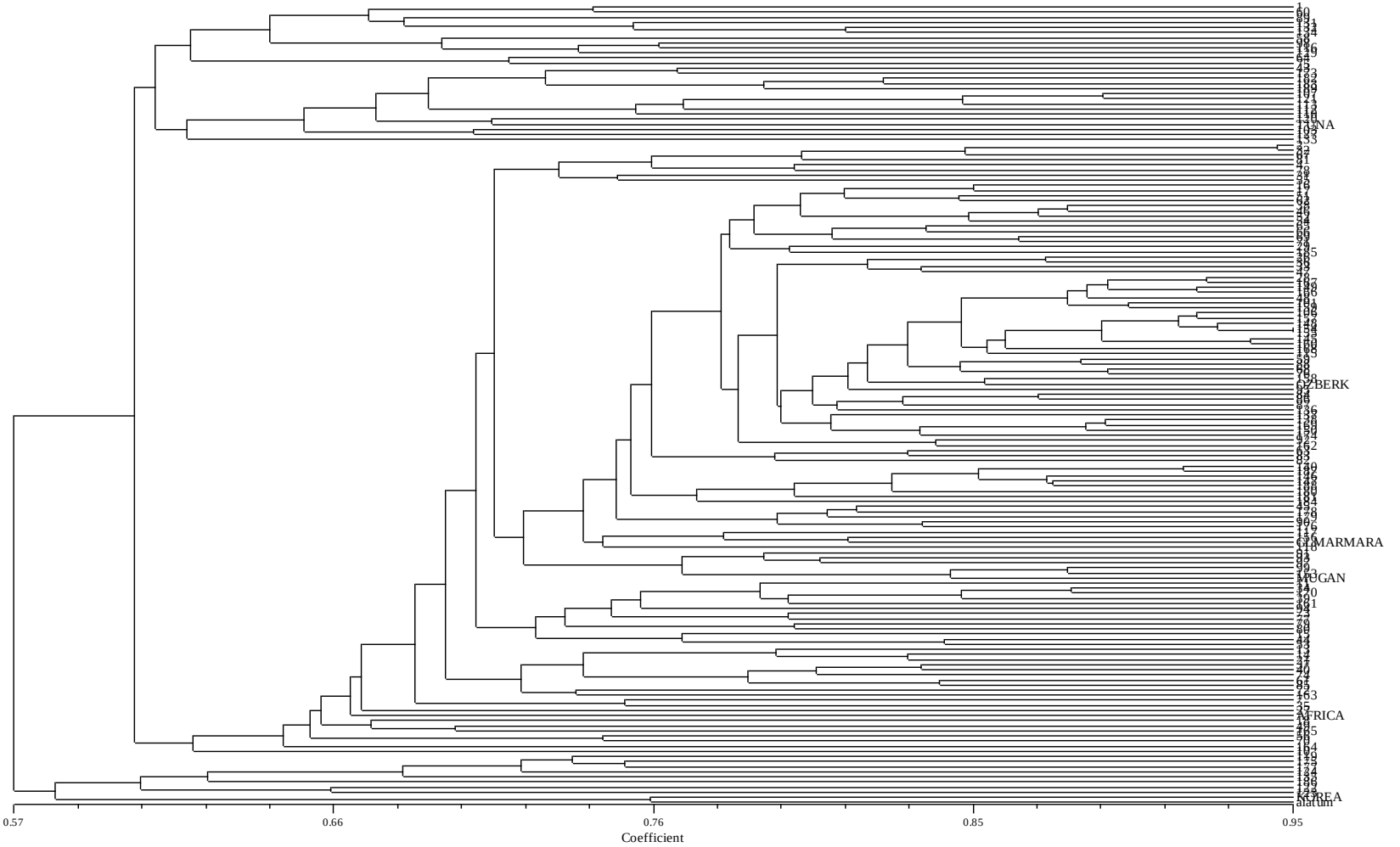

Figure 3.2. The phylogenetic tree containing 158 sesame acccessions construted by UPGMA method and Jaccard coefficient.

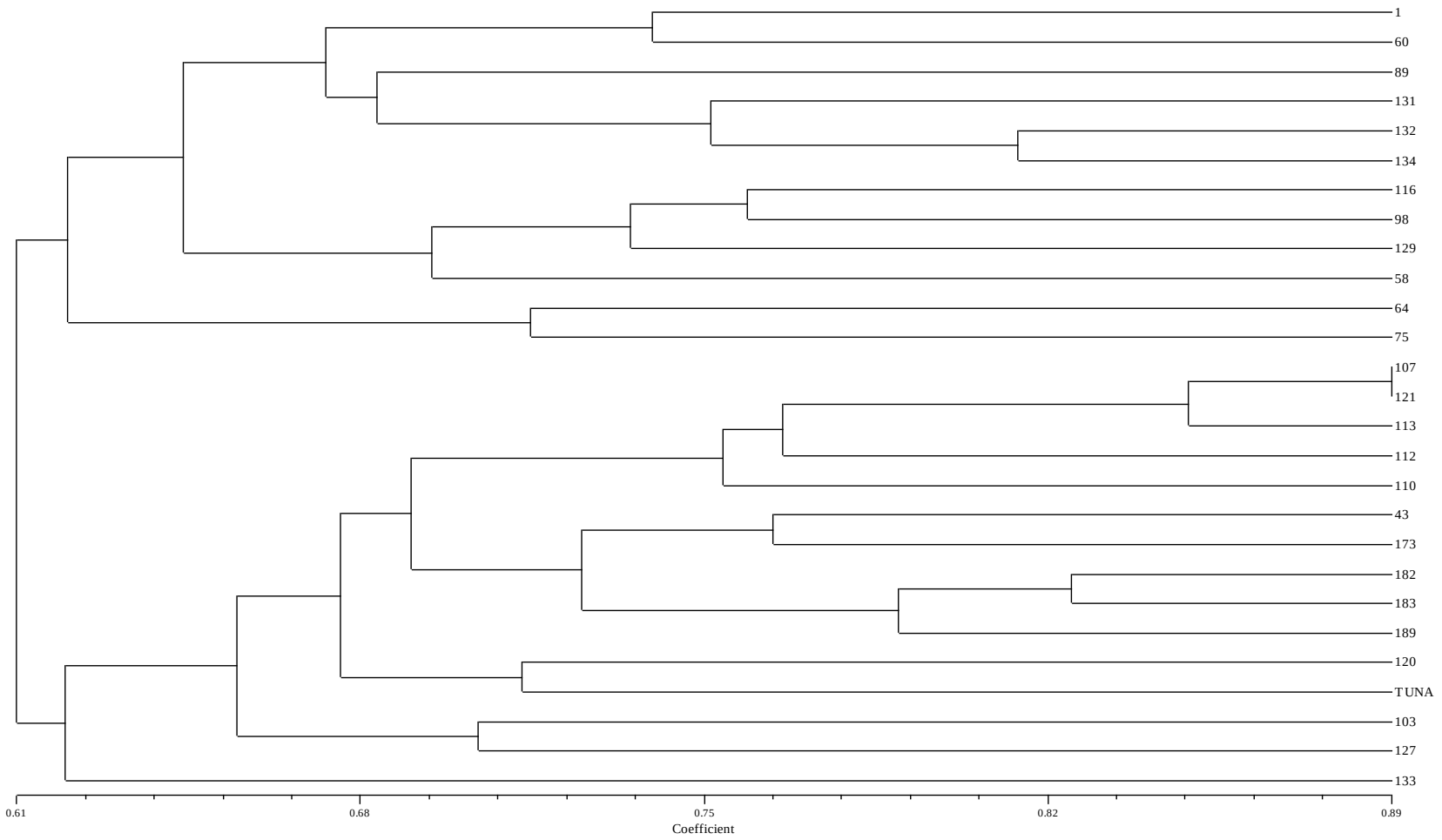

Figure 3.3. Cluster A from figure 3.2 re-drawn for legibility.

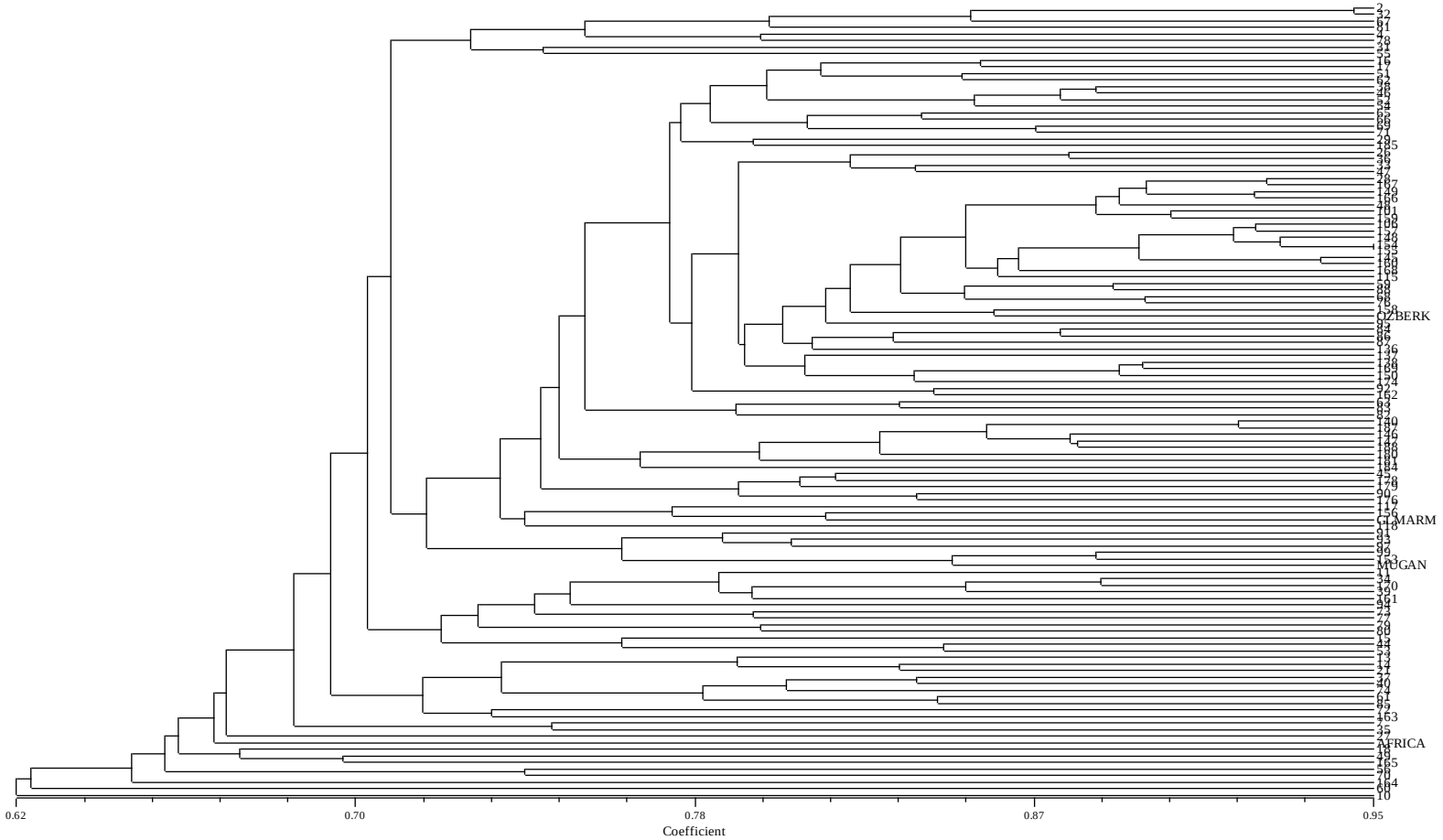

Figure 3.4. Cluster B from figure 3.2 re-drawn for legibility.

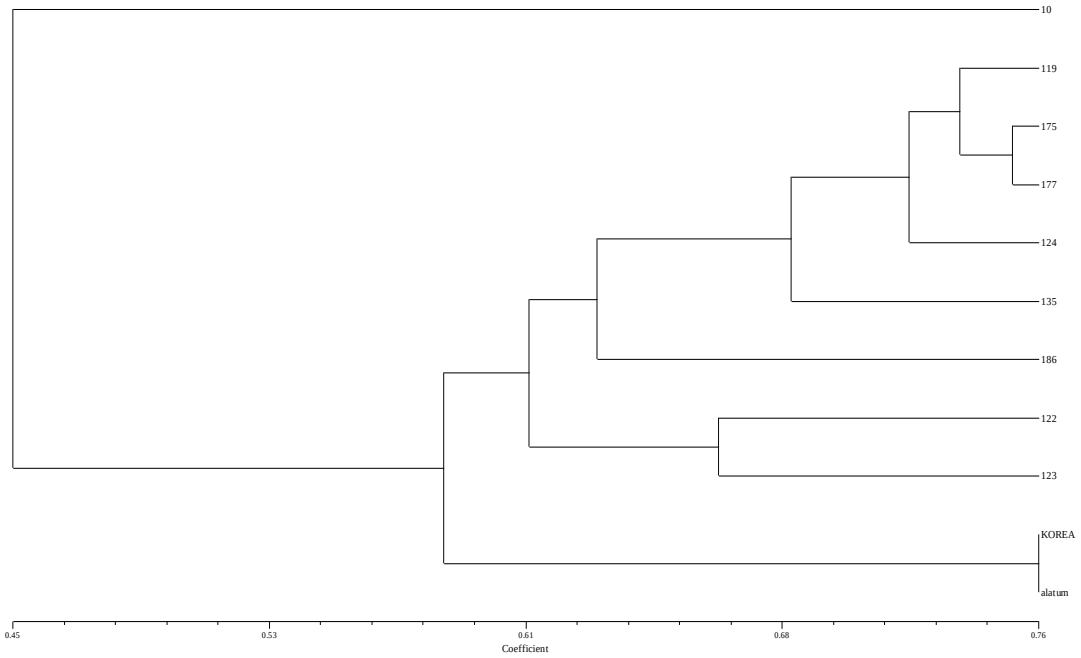

Figure 3.5. Cluster C from figure 3.2 re-drawn for legibility.

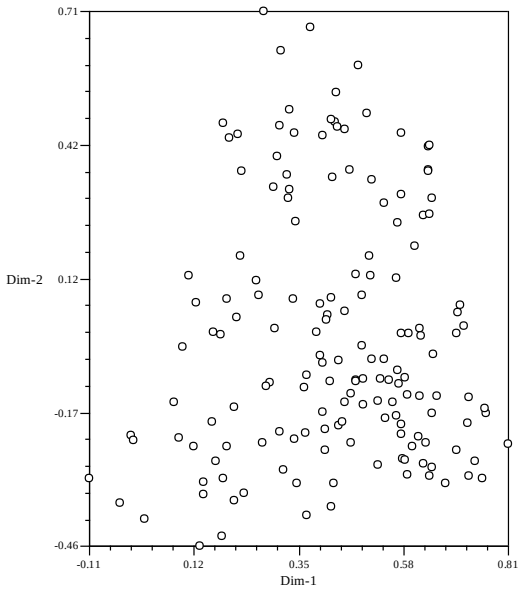

Figure 3.6. 2D PCA of data matrix of sesame accession discrimination.

3.3. Genomic-SSR Marker Design

Pyrosequencing of genomic sesame DNA resulted in 1.094.317 reads. The reads contained 623.365.931 nucleotides (Table 3.3). In all, 64% (702.371) of the reads could be assembled by MIRA software. The remaining 36% (391.946) of total reads were unassembled reads. The reads were used to construct 140.669 contigs with minimum contig length of 40 reads and maximum length of 1200 reads. Cumulative contig length was 56.288.476 nucleotides (Table 3.3). A total of 11.608 SSR motifs were found in the contigs and 3.101 primer pairs were developed from flanking regions of SSRs by using BatchPrimer3. The most common repeats were dinucleotide (36,4%) and then tetranucleotide (29,3%), trinucleotide (21,3%), pentanucleotide (7,1%) and finally hexanucleotide (4,2%) (Figure 3.7). Total number of different motifs were 12 for dinucleotide, 59 for trinucleotide, 204 for tetranucleotide, 229 for pentanucleotide and 351 for hexanucleotide repeats. Among the dinucleotide repeats AT/TA (28%) motif had the highest frequency indicating a high frequency of this motif in the genome.

To date there is only one study about EST-SSR design for sesame accessions (Wei et al. 2008). Wei and colleagues designed 50 EST-SSR markers and applied them to 36 sesame accessions. They found that the most common motif repeat was dinucleotide repeats (AG/TC). Our result was also in good agreement with their results in terms of the motif repeat length.

Table 3.3. The read, contig lengths and the numbers of reads per contig.

Total read length	623.365.931
Minimum read length	47
Median read length	560
Maximum read lenth	1200
Total contig length	56.288.476
Minimum length of contig	40
Median length of contig	327
Max length of contig	93365
Minimum number of reads per contig	1
Median number of reads per contig	3
Maximum number of reads per contig	39907

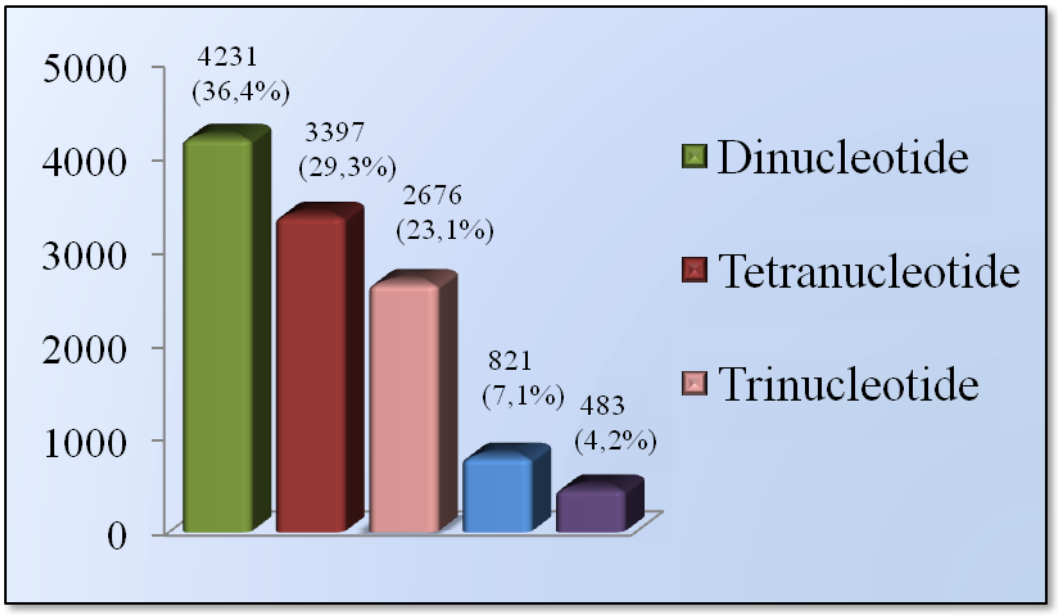

Figure 3.7. Types of SSRs identified.

The first 100 designed SSR markers were synthesized and tested on eleven different Turkish sesame lines and one outgroup (*S. alatum* L.). Two of the samples gave polymorphic bands for one primer (Figure 3.8). The product sizes and melting temperatures (55 °C) of the designed genomic-SSR markers were as expected.

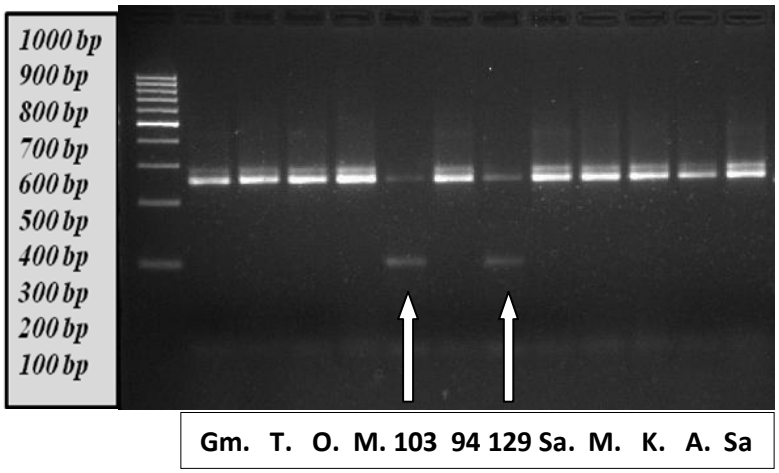

Figure 3.8. Agarose gel electrophoresis one of the designed genomic-SSR primers (Gm: Gölmarmara, T: Tuna, O: Ozberk, M: Muganli, S.a: *Sesamum alatum*, K: Kore, A: Africa).

CHAPTER 4

CONCLUSION

Sesame is an important crop in the world in terms of its rich oil seed content and its significance to people's diet and economy in the developing world. In spite of its importance, there have not been many studies about the sesame genome. Thus the literature is limited for sesame diversity and mapping studies. Therefore in this study we aimed to contribute to the information about sesame by designing new genomic-SSR markers and applying new AFLP combinations to test the dversity of Turkish sesame accessions.

The AFLP results that we obtained indicated that there is a maximum of 57% similarity among Turkish lines. Overall the results of this study may be used in germplasm conservation efforts as well as in selecting lines for breeding and mapping populations.

In this study, sesame genomic DNA sequence were used for SSR primer design and 3101 genomic-SSR primers were successfully obtained. Among them dinucleotides and tetranucleotides were the most commonly identified SSRs, 36.4% and 29.3%, respectively. In nucleotide repeats, AT/TA dinucleotide frequency was the highest. The genomic-SSR markers have the chance to show high polymorphism, because they cover both coding and noncoding regions of the genome.

This is the first study to develop genic-SSRs from sesame therefore these markers will be very helpful for sesame diversity and mapping studies. In future research the genic-SSR primers will be tested for polymorphism in sesame accessions.

,

REFERENCES

Abdellatef, E, R Sirelkhatem, M M Mohamed Ahmed, K H Radwan, and M M Khalafalla. 2008. Study of genetic diversity in Sudanese sesame (Sesamum indicum L .) germplasm using random amplified polymorphic DNA (RAPD) markers. *Journal of Biotechnology* 7, no. 24: 4423-4427.

Alkes, L., Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, David Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38, no. 8 (August): 904-0-909. doi:10.1038/ng1847 http://www.nature.com/naturegenetic

Anon 1984. simple sequences are ubiquitous repetitive components of eukaryotic genomes,1984 Nucleic acids research ,volume 12 number 10. *Nucleic Acids Research* 12, no. 10: 4127-4138.

Banerjee, P. P., and P. C. Kole. 2009. Analysis of genetic architecture for some physiological characters in sesame (Sesamum indicum L.). *Euphytica* 168, no. 1 (January): 11-22. doi:10.1007/s10681-008-9871-6. http://www.springerlink.com/index/10.1007/s10681-008-9871-6.

Bedigian, Dorothea. 2003. Evolution of sesame revisited : domestication , diversity and prospects. *Genetic Resources and Crop Evolution*: 779 - 787.

Bedigian, Dorothea. 2010. Characterization of sesame (Sesamum indicum L.) germplasm: a critique. *Genetic Resources and Crop Evolution* 57, no. 5 (April): 641-647. doi:10.1007/s10722-010-9552-x. http://www.springerlink.com/index/10.1007/s10722-010-9552-x.

Cardle, L, L Ramsay, D Milbourne, M Macaulay, D Marshall, and R Waugh. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156, no. 2 (October): 847-54. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461288&tool=pmcentrez&rendertype=abstract.

Dixit, Anupam, Ming-Hua Jin, Jong-Wook Chung, Jae-Woong Yu, Hun-Ki Chung, Kyung-Ho Ma, Yong-Jin Park, and Eun-Gi Cho. 2005. Development of polymorphic microsatellite markers in sesame (Sesamum indicum L.). *Molecular Ecology Notes* 5, no. 4 (December): 736-738. doi:10.1111/j.1471-8286.2005.01048.x. http://doi.wiley.com/10.1111/j.1471-8286.2005.01048.x.

El-bramawy, Mohamed A S. 2004. Inheritance of Resistance to Fusarium Wilt in Some Sesame Crosses under Field Conditions 42, no. 3: 99-105.

Ercan A., Gulhan, Melih Taskin, and Kenan Turgut. 2004. Analysis of genetic diversity in Turkish sesame (Sesamum indicum L.) populations using RAPD markers*. *Genetic Resources and Crop Evolution* 51, no. 6 (September): 599-607. doi:10.1023/B:GRES.0000024651.45623.f2. http://www.springerlink.com/openurl.asp?id=doi:10.1023/B:GRES.0000024651.45623.f2.

Ghulam M.Ali, 2007. Assessment of genetic diversity in sesame (Sesamum indicum L .) detected by Amplified Fragment Length Polymorphism markers. *Electronic Journal of Biotechnology* 10, no. 1. doi:10.2225/vol10-issue1-fulltext-16.

Seymus, F., and Uzun B. 2010. The use of agro-morphological characters for the assessment of genetic diversity in sesame (Sesamum indicum L .) 3, no. 3: 85-91.

Kumar, P. I. S, Mohammad A Kram, Madhu V Ajpeyi, and Rajendra Lal S Rivastava. 2007. Screening and development of resistant sesame varieties against phytoplasma. *Agriculture* 60, no. 2: 303-304.

Jones, N., Helen Ougham, Howard Thomas, and Izolda Pasakinskiene. 2009. Markers and mapping revisited: finding your gene. *The New phytologist* 183, no. 4 (January): 935-66. doi:10.1111/j.1469-8137.2009.02933.x. http://www.ncbi.nlm.nih.gov/pubmed/19594696.

Sabrao. 2007. Establishment of a Core Collection of World Sesame (Sesamum. *Source* 39, no. 1: 53-64.

Kosman, E, and K J Leonard. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular ecology* 14, no. 2 (February): 415-24. doi:10.1111/j.1365-294X.2005.02416.x. http://www.ncbi.nlm.nih.gov/pubmed/15660934.

Laurentin, Hernán E, and Petr Karlovsky. 2006. Genetic relationship and diversity in a sesame (Sesamum indicum L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC genetics* 7 (January): 10. doi:10.1186/1471-2156-7-10. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1434769&tool=pmcentrez&rendertype=abstract.

Laurentin, Hernán, Astrid Ratzinger, and Petr Karlovsky. 2008. Relationship between metabolic and genomic diversity in sesame (Sesamum indicum L.). *BMC genomics* 9 (January): 250. doi:10.1186/1471-2164-9-250. http://www.ncbi.nlm.nih.gov/pubmed/18510719.

Hiremath S. C., C. G. Patil, K. B. Patil and M. H. Nagasampige. 2007. Genetic diversity of seed lipid content and fatty acid composition in some species of Sesamum L . (Pedaliaceae). *Journal of Biotechnology* 6, no. March: 539-543.

Matthews, David E, Gerard R Lazo, Victoria L Carollo, and Olin D Anderson. 2004. Chapter 14 Bioinformatics and triticeae genomics : resources and future developments. *Victoria*: 425-446.

Meudt, Heidi M, and Andrew C Clarke. 2007. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in plant science* 12, no. 3 (March): 106-17. doi:10.1016/j.tplants.2007.02.001. http://www.ncbi.nlm.nih.gov/pubmed/17303467.

Mohan, M., Suresh N., A Bhagwat, T G Krishna, Masahiro Yano, C R Bhatia, and Takuji Sasaki. 1997. Genome mapping , molecular markers and marker-assisted selection in crop plants. *Genome Research*: 87-103.

Mondal, N., K. V. Bhat, and P. S. Srivastava. 2010. Variation in Fatty Acid Composition in Indian Germplasm of Sesame. *Journal of the American Oil Chemists' Society* 87, no. 11 (June): 1263-1269. doi:10.1007/s11746-010-1615-9. International, and Agricultural Sciences. 2007. Genetic diversity in sesame (Sesamum http://www.springerlink.com/index/10.1007/s11746-010-1615-9.

Ram, S. Ganesh, K. Sundaravelpandian, M. Kumar, K. K. Vinod, J. R. Kannan Bapu, and T. S. Raveendran. 2006. Pollen–pistil interaction in the inter-specific crosses of Sesamum sp. *Euphytica* 152, no. 3 (August): 379-385. doi:10.1007/s10681-006-9225-1. http://www.springerlink.com/index/10.1007/s10681-006-9225-1.

Rold, I. 2000. AFLP markers reveal high polymorphic rates in ryegrasses (Lolium spp .): 125-134.

Sharmila, Vijayarajan, Siruthaiyur Krishnamoorthy Ganesh, and Mahalingam Gunasekaran. 2007. Generation mean analysis for quantitative traits in sesame (Sesamum indicum L.) crosses. *Genetics and Molecular Biology* 30, no. 1: 80-84. doi:10.1590/S1415-47572007000100015. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572007000100015&lng=en&nrm=iso&tlng=en.

Somers, Daryl J. 2004. Chapter 2 Molecular Marker Systems and Their Evaluation for Cereal Genetics. *Agriculture*: 19-34.

Uzun, B., and M Cagirgan. 2006. Comparison of determinate and indeterminate lines of sesame for agronomic traits. *Field Crops Research* 96, no. 1 (March): 13-18. doi:10.1016/j.fcr.2005.04.017. http://linkinghub.elsevier.com/retrieve/pii/S0378429005001474.

Uzun, B., D. Lee, P. Donini, and M. I. Cagirgan. 2003. Identification of a molecular marker linked to the closed capsule mutant trait in sesame using AFLP. *Plant Breeding* 122, no. 1 (February): 95-97. doi:10.1046/j.1439-0523.2003.00787.x. http://doi.wiley.com/10.1046/j.1439-0523.2003.00787.x.

Kumar M., S., Annapurna Kishore and S. C. Hiremath 2008. Cytological Analysis of Interspecific Hybrid between Sesamum indicum L X S . *Analysis* 21, no. 4.

Varshney, Rajeev K, Andreas Graner, and Mark E Sorrells. 2005. Genic microsatellite markers in plants: features and applications. *Trends in biotechnology* 23, no. 1 (January): 48-55. doi:10.1016/j.tibtech.2004.11.005. http://www.ncbi.nlm.nih.gov/pubmed/15629858.

Wei, Li-Bin, Hai-Yang Zhang, Yong-Zhan Zheng, Wang-Zhen Guo, and Tian-Zhen Zhang. 2008. Developing EST-Derived Microsatellites in Sesame (Sesamum indicum L.). *Acta Agronomica Sinica* 34, no. 12 (December): 2077-2084. doi:10.1016/S1875-2780(09)60019-5. http://linkinghub.elsevier.com/retrieve/pii/S1875278009600195.

Wei, Li-bin, Hai-yang Zhang, Yong-zhan Zheng, and Hong-mei Miao. 2009. A Genetic Linkage Map Construction for Sesame (Sesamum indicum L .). *Genomics* 31, no. April: 199-208.

Were, B, a Onkware, S Gudu, M Welander, and a Carlsson. 2006. Seed oil content and fatty acid composition in East African sesame (Sesamum indicum L.) accessions evaluated over 3 years. *Field Crops Research* 97, no. 2-3 (June): 254-260. doi:10.1016/j.fcr.2005.10.009. http://linkinghub.elsevier.com/retrieve/pii/S037842900500273X.

Wold, S., Esbensen K., Geladi P. 2011. Principle Component Analysis. *Scence Direct.*