DETERMINATION OF GENETIC DIVERSITY BETWEEN EGGPLANT AND ITS WILD RELATIVES

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
Yeliz TÜMBİLEN

July 2007
İZMİR

We approve the thesis of **Yeliz TÜMBİLEN**

Date of Signature

…… **18 July 2007**
Assoc. Prof. Dr. Sami DOĞANLAR
Supervisor
Department of Biology
İzmir Institute of Technology

…… **18 July 2007**
Assoc. Prof. Dr. Anne FRARY
Co-Supervisor
Department of Biology
İzmir Institute of Technology

…… **18 July 2007**
Assoc. Prof. Dr. Hülya İLBİ
Department of Horticulture
Ege University

…… **18 July 2007**
Assoc. Prof. Dr. Bahattin TANYOLAÇ
Department of Bioengineering
Ege University

…… **18 July 2007**
Assist. Prof. Dr. Abdullah KAHRAMAN
Department of Field Crops
Harran University

…… **18 July 2007**
Assoc. Prof. Dr. Sami DOĞANLAR
Head of Department
Izmir Institute of Technology

…….
Prof. Dr. Barış ÖZERDEM
Head of the Graduate School

ACKNOWLEDGEMENTS

I want to present my thanks to my advisor Assoc. Prof. Dr. Sami DOĞANLAR for his trust and all of his support to me.

I would like to also thank my co-supervisor Assoc. Prof. Dr. Anne FRARY for her great concern, valuable suggestions and kindness.

I am also grateful to Dr. Osman GÜLŞEN and Dr. Nedim MUTLU from Alata Bahçe Araştırma Enstitüsü Erdemli, Mersin and Alpaslan KİBAROĞLU from Beckman Coulter for all their kind and sincere support during my studies.

Also, I am grateful for the financial support of the TÜBİTAK (The Scientific and Technological Research Council of Turkey) during my education of Master of Science.

Finally, I present my gratitude to my dear parents, Birol TÜMBİLEN and Serpil TÜMBİLEN, for their endless support, encouragement and love.

In memory of my dear older sister Yeşim TÜMBİLEN, who will always be remembered.

ABSTRACT

DETERMINATION OF GENETIC DIVERSITY BETWEEN EGGPLANT AND ITS WILD RELATIVES

Eggplant (*Solanum melongena* L.) is an important crop and has a growing reputation and is now cultivated globally. It is a valuable member of the human diet in Asia, especially in India, which is a primary diversity center of the species. Turkey is the first in Europe and is in the first five countries around the world in terms of eggplant production. The Solanaceae family to which *S. melongena* belongs is an important family, too. Tomato, potato, tobacco and petunia are some example species of the Solanaceae family. This important family with 3000-4000 species shows a high level of morphological diversity which results in confusion about its systematics and this diversity is at the level of genera, species and cultivars.

The aims of the studies reported in this thesis were to analyze genetic diversity of Turkish eggplants and wild relatives in separate studies with different molecular tools. To reveal genetic diversity among eggplant cultivars grown in Turkey, the AFLP marker system was applied to the sample genotypes. For the investigation of genetic variation between *S. melongena* and its wild relatives, though, the SSR marker system was used.

For the AFLP data for Turkish eggplants, an r value of 0.97 was obtained which was in the best scale. Eigen values reported here were also informative. These results showed that the first component analysis explained 64.34% of the variation between samples. For three axes, though, a total of 72.21% variation was explained. According to the statistical results of SSR analysis, the r value of *Solanum* species' genotypic data was found to be 0.88. That means the correlation between sample genotypic data and dendrogram was found to be high. Due to the other statistical results which were Eigen values explained 46.12% of genotypes for first component analysis. With a total value of 55.28%, the 47 different genotypes were explained by the three principle component axes.

The results of AFLP studies showed that although a high similarity value was observed, diversity was detectable among the accessions. The results of SSR studies were also meaningful with their concordance with previous studies and observed diversity with a good fit to statistical results.

ÖZET

PATLICAN VE YABANİ TÜRLERİ ARASINDAKİ GENETİK ÇEŞİTLİLİĞİN BELİRLENMESİ

Patlıcan (*Solanum melongena* L.) gittikçe artan bir tanınma ile önemli ürünlerden bir tanesidir ve şu anda küresel olarak üretilmektedir. Özellikle Hindistan başta olmak üzere, birincil çeşitlilik merkezi olan Asya kıtasında, insan beslenmesinde önemli bir yeri vardır. Üretim miktarı bakımından Türkiye, Avrupa' da birinci ve dünya da da ilk beş ülke arasındadır. *S. melongena'*nın ait olduğu Solanaceae ailesi de önemlidir. 3000-4000 türün yer aldığı bu önemli aile, morfolojik olarak yüksek düzeyde çeşitlilik göstermektedir ve bu çeşitlilik sistematik açıdan çeşitler, türler ve cinsler seviyesindedir.

Bu tezde yapılan çalışmaların amacı, Türk patlıcanları ve yabani akrabaları arasındaki genetik çeşitliliği ayrı ayrı ve farklı moleküler teknikler kullanarak belirlemektir. Türkiye'de yetiştirilen patlıcan kültürleri arasındaki genetik çeşitliliği açığa çıkarmak üzere, AFLP işaretleyici sistemi örnek genotiplere uygulanmıştır. *S. melongena* ve yabani akrabaları arasındaki genetik varyasyonu araştırmak için ise SSR işaretleyici sistemi kullanılmıştır.

Türk patlıcanları AFLP verileri için, 0.97 r değeri bulunmuştur. Bu değer en iyi aralık dahilinde yer almıştır. Ayrıca, rapor edilen Eigen değerleri de oldukça açıklayıcı bulunmuştur. Bu sonuçlar, örnekler arasındaki çeşitliliğin temel bileşenler analizi ile ilk düzlemde %64.34 oranında açıklandığını göstermiştir. Üç düzlemde ise, toplam varyasyonun %72.21' i açıklanmıştır. SSR analizlerinin istatistiksel sonuçlarına göre, *Solanum* türleri genotipik verilerinin r değeri 0.88 bulunmuştur. Bu sonuç, örnek genotipik data ve dendrogram arasında bulunan ilişkinin yüksek olduğu anlamındadır. Diğer istatistiksel sonuçlara göre, Eigen değerleri, temel bileşenler analizi ile ilk düzlemde genotiplerin %46.12' sini açıklamıştır. Toplam %55.28' lik değer ile, 47 farklı genotip, temel bileşenler analizindeki ilk üç düzlemde açıklanmıştır.

AFLP çalışmalarının sonuçları, yüksek düzeyde benzerlik değeri gözlenmesine rağmen, tohum örnekleri arasında varyasyonun tespit edilebileceğini göstermiştir. SSR çalışmalarının sonuçları, önceki çalışmalarla uyumlu ve tespit edilen çeşitlilik ise istatistiksel olarak anlamlı bulunmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

1.1. Diversity and Systematics

There are presumably more than 30 million species in the world yet, a significant number of different organism types have not been identified. Human beings could not imagine the number of species until the possibility to travel to distant and remote places and closely examine them. However, the need for giving names to organisms is as old as human history. The basic aim was to be able to understand each other while explaining something in a short way. Nevertheless, the usage of different terminology in different languages or parts of the world or even in the same country but in a different region started to make it difficult to follow this basic aim. As a result, people started to consider an appropriate way of communication in a scientific way. The solution to this problem came in the early 18^{th} century. A Swedish naturalist, Carolus Linnaeus, proposed a system for naming organisms. In this binominal system of nomenclature, every organism has two names. The first name which represents the genus is written with a capital letter. The second name which is called an epithet is special for each organism. Both of these names are written in italics or underlined. This nomenclature is important because it is based on the usage of Latin words which serves as a common language overall and because of its easiness during the classification process.

Classification is basically the categorization of identified and named organisms in a hierarchical way. Identification and naming is then a necessity to be able to classify organisms. According to Linnaeus` system, which is the base of modern classification, two basic categories were constructed: species and genus. In terms of today's classification systems, organisms were grouped in seven categories which are species, genus, family, order, class, phylum for animals or division for plants and kingdom. Although there is some disagreement, most recently, all organisms started to be grouped into 3 different domains or superkingdoms: Bacteria, Archaea or Eukarya. In addition to these, there are also intermediate categories such as sub-, super- or infra- at each taxonomic level. The easiness of the binominal system does in fact rely upon the

availability to be broadened or become more complicated without any disorder. All these categories are the building blocks of the classification systems which start from the narrowest, most concordant groups and expand to the broadest, least similar taxa.

There are many classification systems based on different considerations. For example, people who are mostly interested in the geographical dispersal of organisms may classify them according to their habitats. Others may focus on the shape or color of a body part and propose a system. However, there are three fundamental systems. In the natural classification system, organisms are categorized based on their similarity and close relation. Members of higher taxa share less similarity and are less closely related and the opposite is true for lower taxa. The second system is an artificial classification system. In this system, a character or feature is taken into account and the grouping is done in terms of that trait. As an example, the classification of plants upon their medicinal properties or economic importance can be given. The last fundamental system is used to define fossil organisms and both natural and artificial systems provide data that support this definition. With respect to Linnaeus' classification idea, morphological and structural similarities were the major considerations and this method is still in usage. Today's classification of organisms is mainly carried out using the natural classification system.

Systematics is the science which defines diversity of organisms and includes their identification, naming and classification. As a result, it has a much more broadened aspect than all these proceedings which starts from first finding and then naming and place an organism into a taxon in systematics. This aspect also deals with the history of organisms on Earth, thus, it takes into account evolution. After being first elucidated by Lamarck although with an incorrect assumption and fully explained by Darwin using the theory of natural selection, evolutionary relationships of organisms have started to be considered within systematics more seriously. In this way, organisms can be more accurately grouped and new ones added to the systems in a well-suited way. This new way of systematics is called phylogenetics or phylogenetic systematics. There are three types of phylogenetic studies: phenetics, cladistics and evolutionary systematics (Westhead et al. 2002). The difference between phenetics and cladistics depends on the type of data chosen to study evolutionary relationships of organisms. All characters are used for phenetics while only shared characters are used for cladistics (Westhead et al. 2002). Hence, evolutionary systematics benefits from these two types of studies. As a general result for all three, though, phylogenetic trees or dendrograms

are drawn (Li 1997 and Raven et al. 1999b). Specifically phenograms and cladograms are terms that are used to refer to the trees drawn from these studies (Raven et al. 1999b and WEB_3 2007).

1.2. Genetic Diversity and Molecular Systematics

Darwin's theory which has real importance for systematics is in agreement with and supported by several sciences such as genetics, statistics and mathematics and has certain building blocks (Raven et al. 1999a and WEB_7 2007). These building blocks which are important in evolution are natural selection, genetic drift and founder effects. All three affect the same composition: genotype. Evolution can be defined as any changes in genotype due to several reasons but basically due to mutations.

Genetic diversity is different forms of genotype and occurs as a result of changes in genetic structure (WEB_2 2007). It potentially leads to speciation in the long term due to the process of evolution (Raven et al. 1999a and WEB_2 2007). Diversity in genetic composition is the basic feature which increases chance of survival for individuals and populations during natural selection (WEB_2 2007). The molecular components that form this genetic composition are DNA and proteins and their connection to classification is shown in Figure 1.1. (Li 1997, WEB_9 2007).

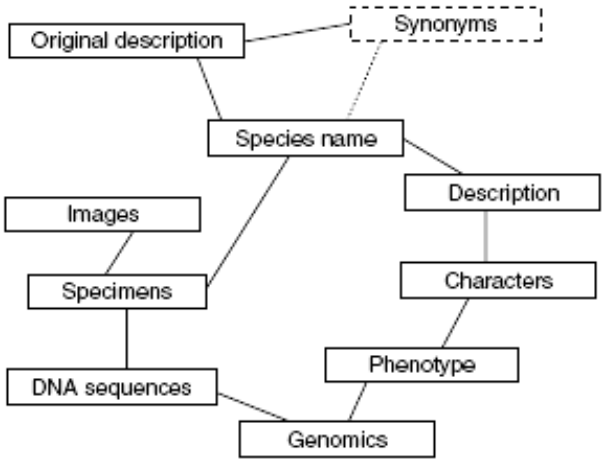

Figure 1.1. The relation of genomic data with taxonomy and taxonomic definition. (Source: Knapp et al. 2004)

Molecular systematics is the classification of organisms with the help of molecular techniques to detect diversity at the molecular level (Li 1997, Raven et al. 1999b and WEB_9 2007). There are several advantages of using molecular data compared to morphological data. Molecular data can be gathered via molecular techniques and are more abundant compared to morphological data (Li 1997 and Raven et al. 1999b). Even morphologically diverse species can be compared because of genotypic data's higher conservation compared to morphological data which is not only related with genotype but also affected by environment (Li 1997 and Raven et al. 1999b). Also, molecular techniques are open to improvement and every new method has increased easiness for application and fewer disadvantages in terms of obtained data (Li 1997 and Ranade 2003). As a result, an increased accumulation which is an important factor for molecular systematics is present (Li 1997).

There are three groups of statistical analysis systems used in phylogenetic studies (Li 1997, Ranade 2003 and Westhead et al. 2002). These are distance matrix methods, maximum likelihood methods and maximum parsimony methods (Li 1997, Ranade 2003 and Westhead et al. 2002). According to distance matrix methods, distances between any two taxa are calculated and clustering is organized using minimum distance (Li 1997, Ranade 2003 and Westhead et al. 2002). Neighbour-joining (NJ) and Unweighted Pairgroup Method with Arithmetic Averaging (UPGMA) are methods which depend on these kinds of distance values (Ranade 2003). UPGMA which is the simplest analysis method is specifically defined to construct phenograms while it is also used for phylogenetic trees (Li 1997 and Westhead et al. 2002). Maximum likelihood methods try to reach maximum likelihood value while constructing the final tree (Westhead et al. 2002 and Ranade 2003). The principle of maximum parsimony methods is based on using a minimum number of variables that exhibit phylogenetic differences between samples (Li 1997, Westhead et al. 2002 and Ranade 2003).

1.2.1. Molecular Techniques

There are various molecular techniques used experimentally for several purposes. Their usage for taxonomic studies is relatively new. Widespread application is reported as starting with protein analysis during the 1960s (Li 1997). This was followed

by DNA-based methods and recombinant technologies of which examples are shown in the Figure 1.2. (Li 1997).

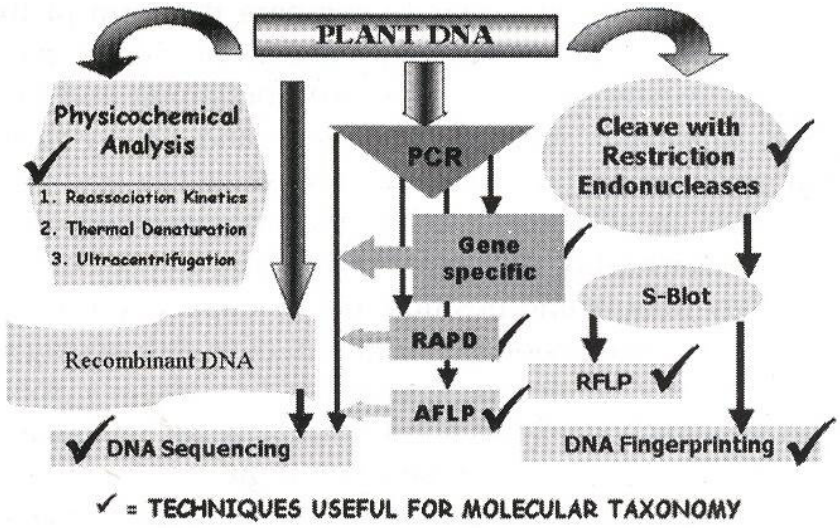

Figure 1.2. Several techniques that are used in molecular systematics studies. The methods for which usefulness has been proven are check-marked (Source: Ranade 2003).

Among these various techniques, AFLP and SSR technologies are defined in detail because they are of interest for this thesis.

1.2.1.1. AFLP

The amplified fragment length polymorphism (AFLP) marker technique was first developed by Vos et al. (Vos et al. 1995). It is a PCR-based technique that has the advantages of its easiness, speed and specificity (Mohan et al. 1997, Jones et al. 1997 Henry 1999 and Ranade 2003). The method also relies on restriction digestion of genomic DNA and, in fact, is a DNA fingerprinting process (Vos et al. 1995 and Ranade 2003).

AFLP consists of several steps. First, DNA is digested with two different enzymes. One of them cuts the DNA in several regions while the other cuts in few places: frequent and rare cutters (Vos et al.1995 and Jones et al. 1997). This provides an optimized number of DNA fragments, though, the number of fragments or bands, compared to other methods, is still high (Vos et al. 1995, Staub and Serquen 1996 and

Mohan et al. 1997). Usually between 50 and 100 bands per individual are obtained (Vos et al. 1995 and Staub and Serquen 1996). Enzyme pairs may be changed in experiments but the cutting patterns should be appropriate to the method. Some examples of these pairs are: *Eco*R I – *Mse* I, *Pst* I - *Mse* I, *Eco*R I – *Taq* I and *Hind*III - *Eco*R I (Mace et al. 1999 and WEB_6 2007). In this thesis study, specifically, *Eco*R I – *Mse* I restriction enzyme pairs were used. They have 6 bp and 4 bp recognition sites, respectively (Grandillo and Fulton 2002 and Invitrogen 2003). The next step of the protocol is related with these restriction sites. Small DNA pieces (adapters) specific for each restricted enzyme site are bound to the template DNA (Vos et al. 1995, Jones et al. 1997 and Ranade 2003). The template DNA plus adapter provides the binding sites for primers (Vos et al. 1995 and Jones et al. 1997). There are two PCR amplification steps in the protocol. These are called pre-selective and selective amplifications (Vos et al. 1995 and Invitrogen 2003). While pre-selective amplifications' primers have one extra nucleotide, selective amplification primers have two or three extra bases which may be selected differently (Vos et al. 1995, Mohan et al. 1997 and Salamini et al. 2004). These bases are responsible for selectivity and are used to increase specificity and decrease or increase number of bands (Vos et al. 1995 and Invitrogen 2003). The number of bands obtained from the AFLP technique is related with genome size and number of C and G bases in the extra nucleotides (Vos et al. 1995 and Invitrogen 2003). The last step is the running of samples on polyacrylamide gels and visualization of bands with either autoradiography or fluorescent detection (Vos et al. 1995 and Ranade 2003).

AFLP is a dominant type of marker (Henry 1999 and Ranade 2003). However, a large number of fragments and relatively high polymorphism make the technique favorable in molecular investigations (Vos et al. 1995, Staub and Serquen 1996, and Ranade 2003). Also, no prior sequence knowledge is needed for application of AFLP (Grandillo and Fulton 2002). Thus, it is a preferred method especially for taxonomic and mapping studies (Mohan et al. 1997 and Ranade 2003). Although it is a random process, selective amplifications increase the specificity (Vos et al. 1995 and Ranade 2003). The results are also reproducible (Mohan et al. 1997 and Henry 1999). Disadvantages are generally due to high expenses which are related with the detection steps (Staub and Serquen 1996 and Mohan et al. 1997). Automation is applicable, though, it is in fact another reason causing increase in expenses (Staub and Serquen 1996, Mohan et al. 1997 and Ranade 2003).

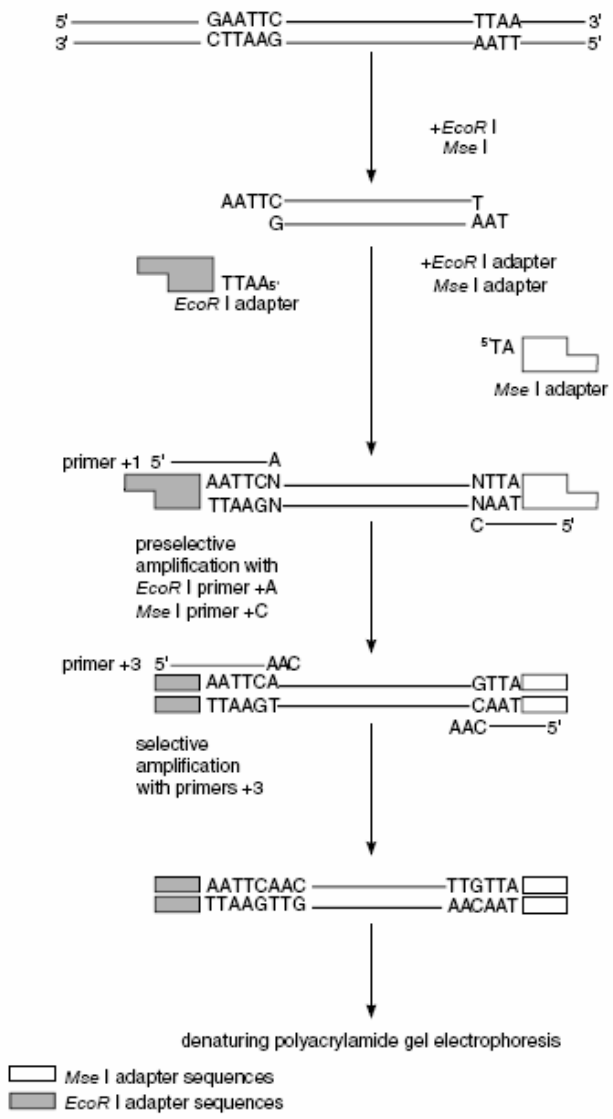

Figure 1.3. The protocol of AFLP technique.

(Source: Invitrogen 2003)

1.2.1.2. SSR

Simple sequence repeats or microsatellites are terms used to refer to tandemly repeated short nucleotide units between 1-5 bps in the genome (Staub and Serquen 1996, Powell et al. 1996, Jones et al. 1997 and Nunome et al. 2003a). These repeats show a genome-wide distribution and can be placed in either genes or non-coding regions of the nuclear genome or else in extranuclear genomes (Powell et al. 1996, Jones et al. 1997, Nunome et al. 2003b and Varshney et al. 2005). In the genome, this distribution is reported to be collected around particular regions of the chromosomes

such as centromeric areas (Nunome et al. 2003b). For the generation of SSR markers sequence data are needed (Jones et al. 1997, Nunome et al. 2003b and Varshney et al. 2005). These data are obtained in two ways. One way is by constructing a genomic library and screening using SSR probes (Jones et al. 1997 and Nunome et al. 2003a). The other way is based on sequence data supported by publicly available databases which comprises gene sequences and cDNA libraries (Nunome et al. 2003a, Nunome et al. 2003b and Varshney et al. 2005). ESTs (Expressed sequence tags) are included in the second way of accessing microsatellite-related sequence data (Rudd 2003 and Varshney et al. 2005). From all the resulting sequences, specific primers for SSRs can then be designed (Figure 1.4.). However, difference in the data type that is used classifies SSRs as genomic or genic SSRs (Rudd 2003 and Varshney et al. 2005). SSRs designed by using EST library data are one of the basic type of genic SSRs. Comparison of these SSRs in terms of their advantages and disadvantages was described in the studies of Rudd and Varshney et al. (Rudd 2003 and Varshney et al. 2005). Due to being a part of conserved regions of the genome, SSR primers designed from EST sequences, are expected to be suitable to apply to other related species (Varshney et al. 2005). This makes genic SSRs favored in comparison to genomic SSRs. SSRs identified by genomic library construction and search can be products of transcribed or non-transcribed regions (Varshney et al. 2005 and Nunome et al. 2003b). This feature while providing a high rate of polymorphism makes genomic SSRs less transferable among species (Varshney et al. 2005). One result of these interpretations is a disadvantage of genic SSRs such that less polymorphism is observed for genic SSRs (Varshney et al. 2005). Another disadvantage is related with the amount of sequence data which is publicly available as mentioned above. For example, eggplant was reported in SOL Genomics Network (http://sgn.cornell.edu) as having 3,181 ESTs in total (Mueller et al. 2005). That number was the lowest number within other species mentioned in the same study: tomato, potato, pepper and petunia (Mueller et al. 2005). However, sequencing studies continue with the continual addition of new data.

The production of ESTs is a sequencing event and can be directed from each end of cDNAs or from both ends (Rudd 2003). As a result, several ESTs are produced (Rudd 2003). These ESTs are then clustered to form contigs which include several sequence products that overlap and are included in the same region (Rudd 2003 and Krane and Raymer 2003). If these contigs include many members, then they are called a multi-member sequence assembly (Figure 1.5.). If a cluster includes small portions of a

cDNA, all of which in fact were synthesized from the initial complementary DNA, it is referred to as bridged sequence assembly (Figure 1.5.). Last, a cluster that consists of single ESTs or small clusters of ESTs are called singletons and small clusters, respectively (Figure 1.5.).

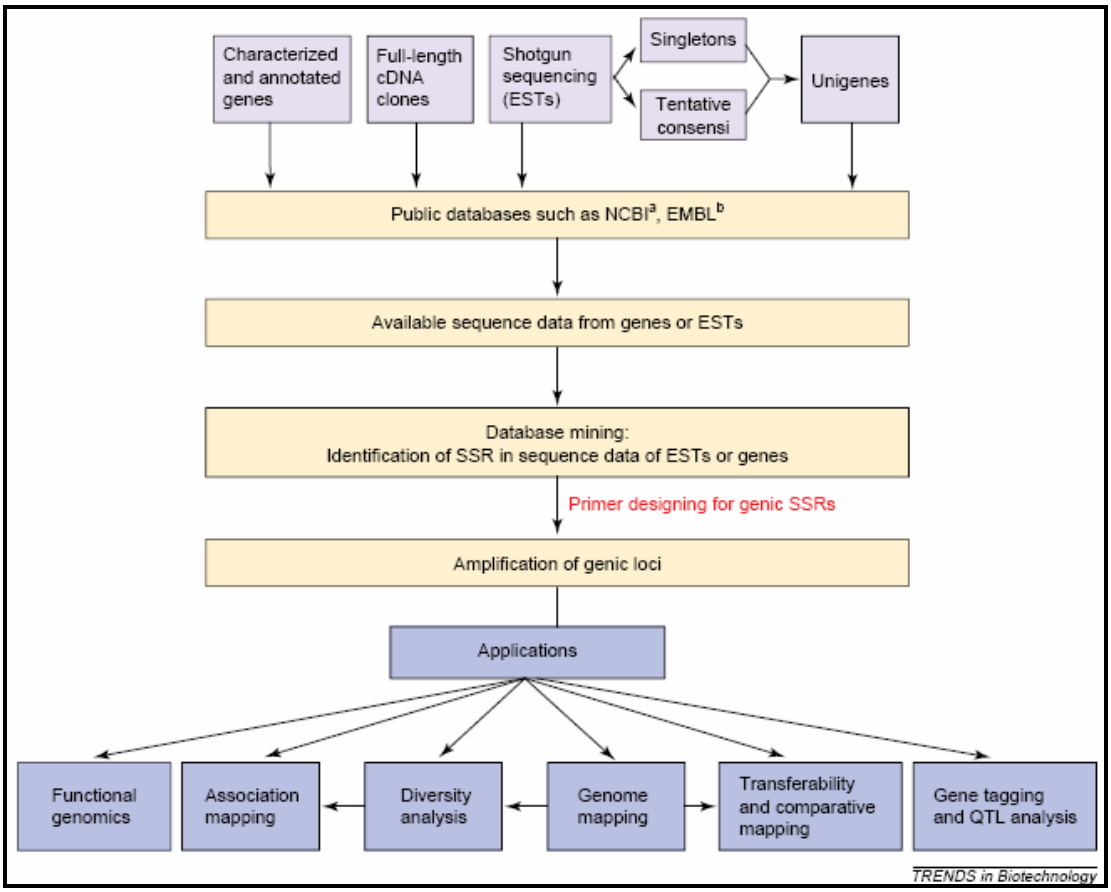

Figure 1.4. The design of genic SSR primers databases and their fields of use. (Source: Varshney et al. 2005)

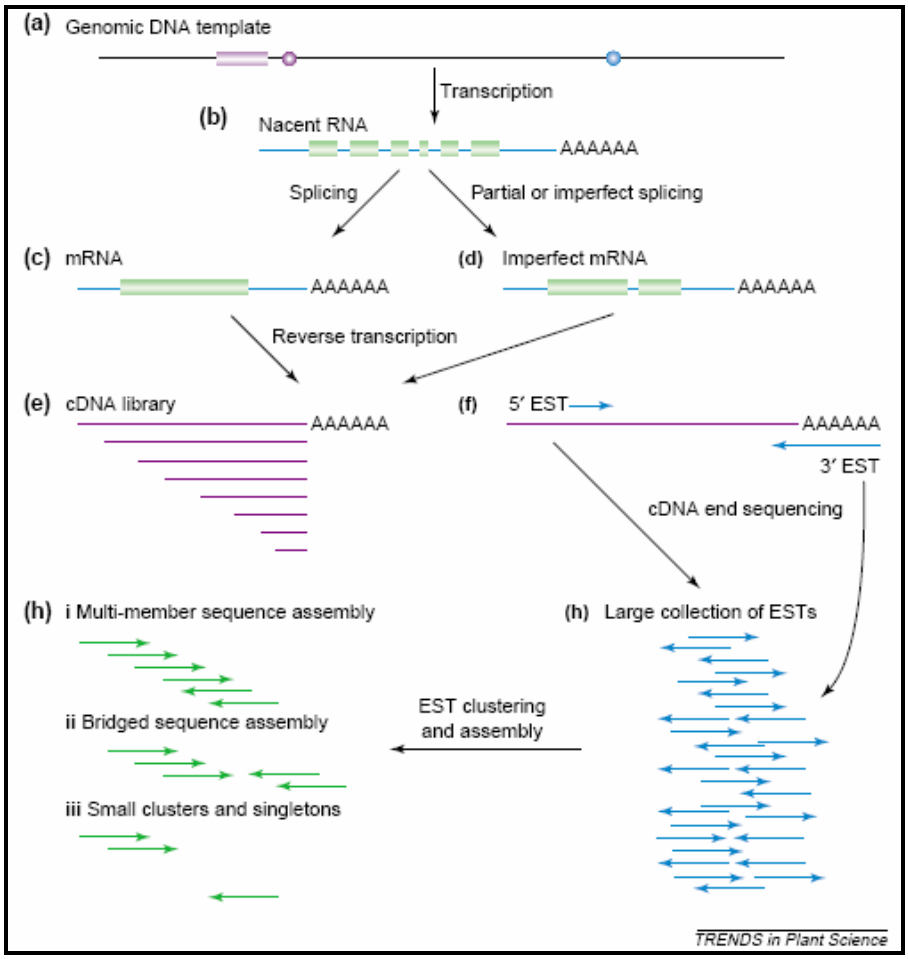

Figure 1.5. Production and clustering of ESTs from genomic DNA.
(Source: Rudd 2003)

SSR is a PCR based molecular method (Staub and Serquen 1996 and Jones et al. 1997). The principle is the detection of polymorphisms resulting from different numbers of repeat units in different individuals and is observed codominantly (Powell et al. 1996, Jones et al. 1997 and Henry 1999). The level of polymorphism is very high which makes SSR an ideal marker for mapping and diversity studies, fingerprinting and population genetics (Powell et al. 1996, Jones et al. 1997, Mohan et al. 1997 and Nunome et al. 2003a). It has an easy application procedure which is basic PCR amplification of the sample and then detection of the bands (Powell et al. 1996 and Jones et al. 1997). However, the major disadvantage is related with development of SSR primers which is defined above.

1.4. Solanaceae Family

The Solanaceae is a family in the plant kingdom and is a member of the Magnoliophyta division which is more generally referred to as angiosperms or flowering plants. The family is one of the five families of the Solanales order which is respectively a group within the 10 orders in Asterids (APG II 2003). The family includes 90 genera and estimated species number is between 3000-4000 (Knapp et al. 2004 and WEB_10 2007).

The Solanaceae family members are well adapted to different environments. They show a good dispersal to a wide region in the world and even to places that have harsh conditions such as deserts (Knapp et al. 2004, WEB_10 2007 and WEB_4 2007). As a result, the Solanaceae plants can be grown in several habitats and distributed worldwide (Figure 1.6.).

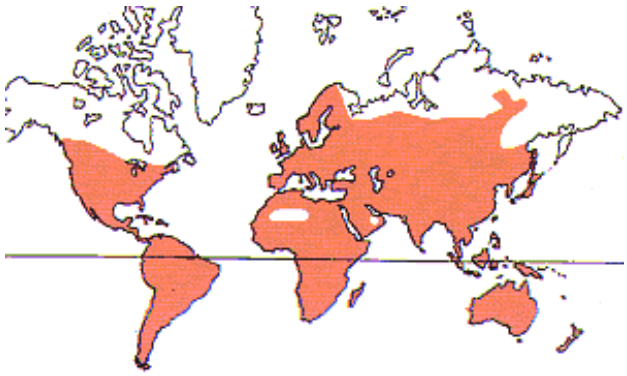

Figure 1.6. The dispersion of Solanaceae family members around the world.
(Source: WEB_10 2007)

However, the origin of diversity for most solanaceous plants is around tropical regions with accumulation around the Amazon and Andes parts of South America, thus they have a New World origin (Daunay et al. 2001, Knapp et al. 2004, WEB_10 2007 and WEB_4 2007). There are a few exceptions, though. For example, *Solanum melongena* (eggplant) and some of its related species are of Asia-Africa origin and are domesticated species, thus they are of Old World origin (Daunay et al. 2001, Doganlar et al. 2002a and Doganlar et al. 2002 b).

Owing to its inclusion of plant species that are important in relation to human diet, health issues and beauty and decorative needs, the family ranks third in economic importance (Mueller et al. 2005 and WEB_4 2007). This is also due to the Solanaceae consisting of a high number of domesticated species including tomato, pepper, potato, petunia, datura, tobacco, eggplant and others (Doganlar et al. 2002b, Mueller et al. 2005, WEB_10 2007).

In addition to distribution and usage diversity, morphological diversity among the Solanaceae family members which cover genus, species and cultivars is really noteworthy (Knapp et al. 2004 and WEB_10 2007). Flowers, fruits and leaves are important plant parts used commonly in taxonomy as they were targets of the domestication process (Doganlar et al. 2002b, Knapp et al. 2004 and WEB_10 2007). Recently, the huge variation for these traits is combined with molecular data for phylogenetic studies and an example was presented in the review of Knapp et al. (Knapp et al. 2004). In that study, the two forms of solanaceous fruits (berries or capsules), growth period (annual or perennial) and structure of the flower (actinomorphic or zygomorphic) were taken into account and the resulting tree is in Figure 1.7. (Knapp et al. 2004). There are also other morphological traits that show differences among the Solanaceae family such as prickles and hairs on several parts of the body, height and length of plant and plant organs (Doganlar et al. 2002b and Frary et al. 2003).

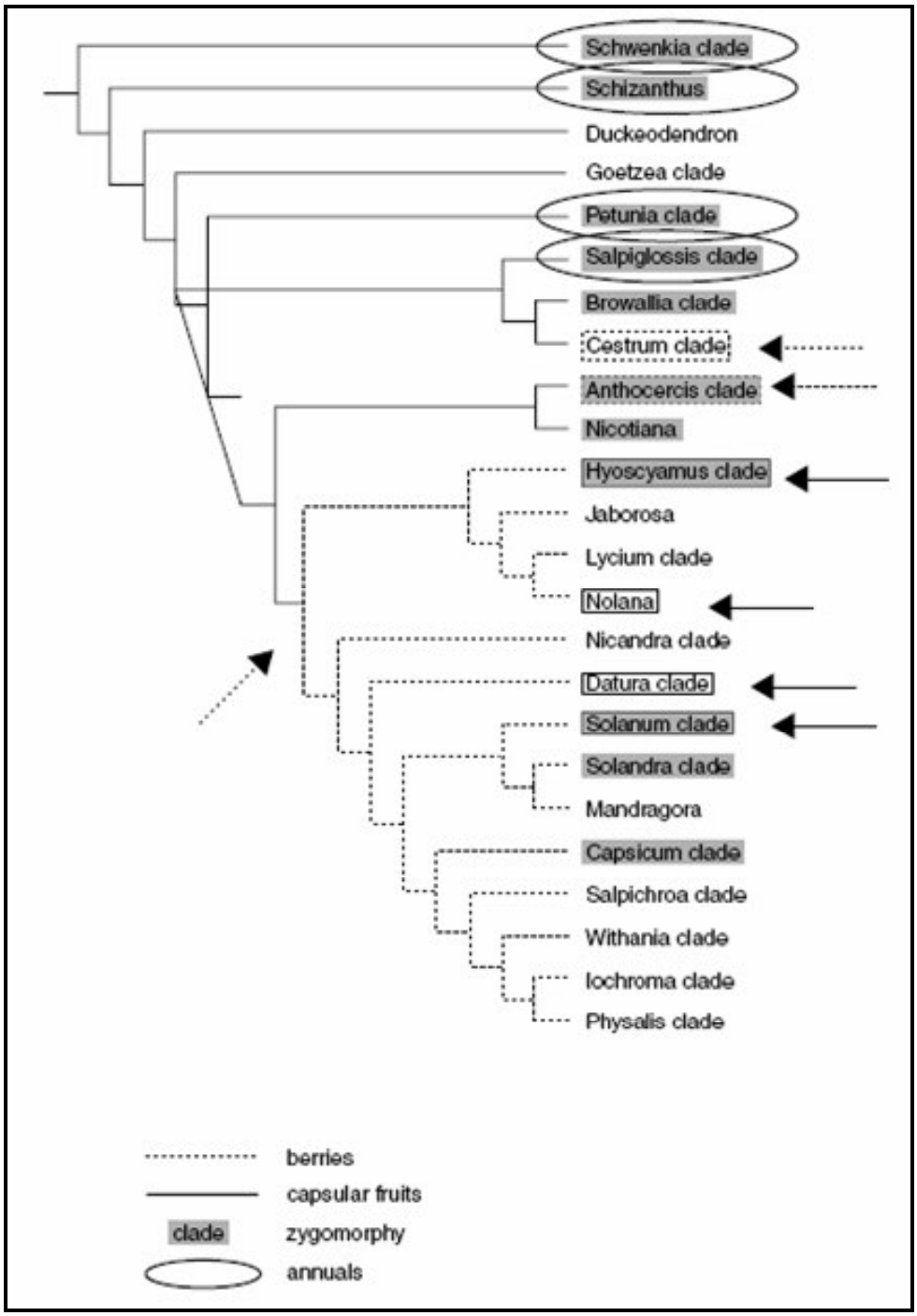

Figure 1.7. An example of a phylogenetic tree of the Solanaceae family indicating morphological traits (Source: Knapp et al. 2004).

1.4.1. Genus Solanum

The number of species in the Solanum genus is reported differently according to different sources as 1000-1400, 1500-2000 and 1000-2000 species (Sakata and Lester

1994, Isshiki et al. 1994c, Lester 1997, Daunay et al. 1998 and WEB_10 2007). This makes Solanum the most crowded genus of the family and almost half of solanaceous plants are members of this genus (Doganlar et al. 2002a, Knapp et al. 2004, WEB_10 2007 and WEB_11 2007). The confusion is not only about the exact number of species in the genus or family. It is also related with the number of names used for these species. These data are also not constant with over 3000, 3700 and close to 5000 names that are referred to in different papers (Sakata et al. 1994, Lester 1997, Daunay et al. 1998, Daunay et al. 1999, Furini and Wunder 2004). Due to all this indefiniteness and the importance of the genus for humans, there are considerable numbers of studies on the taxonomy, phylogeny and biotechnology of *Solanum* species. The importance for humans relies on the existence of several important crop plants in the genus.

1.4.2. Eggplant (*Solanum melongena*)

Due to confusion about use of the term eggplant, Lester indicated that eggplant may be used as nomenclature to refer to several *Solanum* species important for human diet and health such as *Solanum melongena, S. aethiopicum, S. macrocarpon, S. quitoense, S. sessiliforum* and related species (Lester 1998 and Daunay et al. 2001). However, another definition of eggplant only includes three cultivated species: *S. melongena, S. aethiopicum* and *S. macrocarpon* (Lester 1998, Daunay et al. 2001 and Doganlar et al. 2002a). Among these species, *S. melongena* is commonly referred to as eggplant and is of the most interest in published studies and as well as for this thesis (Lester 1998).

The name eggplant comes from the shape and color of the vegetable's fruit (Lester 1998, Economic Research Service, USDA 2006 and National Research Council 2006). Like an egg in shape and white-colored, this fruit led people to use this term in history (Economic Research Service USDA, 2006 and National Research Council, 2006). However, this fruit of African-origin was superseded by the Asian- domesticated species, *Solanum melongena* (Lester 1998, Daunay et al. 2001, Lester and Daunay 2003, National Research Council, 2006 and Frary et al. 2007). There are several terms used for *Solanum melongena.* Eggplant, brinjal eggplant, aubergine or guinea squash are examples of these terms (Nothmann 1986, Choudhury 1995, Lawande and Chavan

1998, Daunay et al. 1999 and Kashyap et al. 2003). However, Brinjal eggplant is the most common name used to refer to *Solanum melongena*.

The distribution and production of *S. melongena* differs according to countries and continents. The major production area is the continent Asia where the plant has real importance (Choudhury 1995, Lawande and Chavan 1998, Collonier et al. 2001, National Research Council 2006 and Frary et al. 2007). India and China are the two countries which are the primary cultivation centers and have the highest production (Lawande and Chavan 1998, Lester 1998, Daunay et al. 2001, Doganlar et al. 2002a, Doganlar et al. 2002b and Economic Research Service, USDA 2006). Thailand, Malaysia, Indonesia, and Philippines are the other important producers in Asia (Lawande and Chavan 1998, Daunay et al. 2001, Collonnier et al. 2001 and Doganlar et al. 2002b). In its own history, cultivation of brinjal eggplant spread to Japan after India and China (Frary et al. 2007). Japan is now an important eggplant producer in the world (Economic Research Service, USDA 2006). The introduction of eggplant to the west was primarily around the Mediterranean region which is the secondary "domestication region" and covers Turkey, Syria, and Persia (Nothmann 1986, Daunay et al. 2001, Kashyap et al. 2003, WEB_8 2007 and WEB_12 2007). Although later in history, the whole south Mediterranean region including countries such as Italy, Spain, France, and Greece became eggplant producers (Lawande and Chavan 1998, Daunay et al. 2001, Frary et al. 2007 and WEB_8 2007). Today, Turkey ranks the first in Europe in terms of total eggplant production (Economic Research Service, USDA 2006). Egypt is the most important brinjal eggplant producer in Africa (Economic Research Service, USDA 2006). America is far behind in terms of production and is reported as 20^{th} in the world and ranks first as an importer (Economic Research Service, USDA 2006). However, interest in eggplant has been increasing in the USA especially since the 1990s (Economic Research Service, USDA 2006). Overall, brinjal eggplant is now a globally cultivated plant species (Daunay et al. 2001, Doganlar et al. 2002a and Frary et al. 2007).

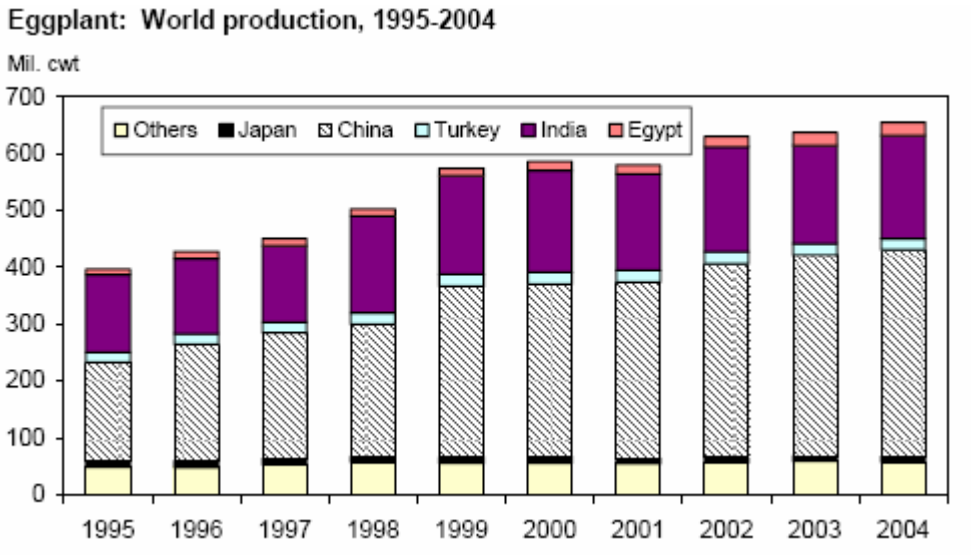

Figure 1.8. Worldwide production of eggplant between the years 1995 and 2004 (Source: Economic Research Service, USDA 2006).

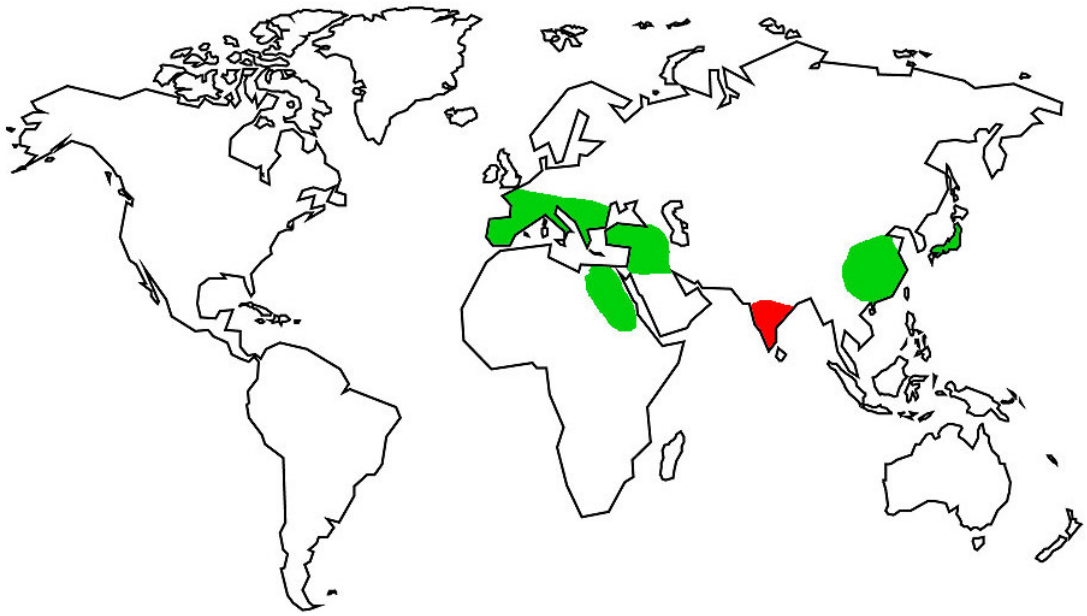

Figure 1.9. Primary and secondary diversity centers of eggplant (*Solanum melongena* L.). Red-colored region is the basic primary diversity center while green-colored regions are secondary diversity centers and major cultivation areas (Source: WEB_12 2007).

Eggplant or brinjal eggplant has three varieties differing from each other based on morphology affected by physiology and environment (Nothmann 1986, Lawande

and Chavan 1998 and Kashyap et al. 2003). These varieties are var. *esculentum*, var. *serpentinum* and var. *depressum* while the generally sold and consumed types are the offspring of these varieties (Nothmann 1986, Lawande and Chavan 1998 and Kashyap et al. 2003).

Eggplant (*S. melongena*) is a warm-loving plant with an ideal growing temperature between 22-30ºC, and has an erect and compact growth habit and large leaves and perfect flowers (Nothmann 1986 and Lawande and Chavan 1998). Autogamy or self-pollination is the usual way of fertilization although cross-pollination is also possible by insect (Nothmann 1986, Lawande and Chavan 1998, Daunay et al. 2001 and Frary et al. 2007). The plant is a biennial which is grown as an annual in general (Nothmann 1986). There is great morphological diversity among *S. melongena* varieties, cultivars, wild and weedy plants and between related species observed for several characters. Fruit color, size, shape and taste are the most noticeable traits that show differences among individuals (Collonnier et al. 2001, Kashyap et al. 2003, Nothmann 1986, Daunay et al. 2001 and Frary et al. 2007). The color differences of fruits are basically due to two color pigments' and their effects on appearance and are controlled by more than one gene (Nothmann 1986 and Frary et al. 2007). These pigments are chlorophyll a and b and anthocyanins which are in different amounts and in combination determine the exact color of the fruit (Nothmann 1986, Daunay et al. 2001 and Frary et al. 2007). As a result, eggplant fruits can be from white to black in color with a gradient of purple, yellow and green (Nothmann 1986, Lawande and Chavan 1998 and Daunay et al. 2001). In addition to the skin color uniformity of plants, striped or spotted color configurations are possible (Nothmann 1986, Lawande and Chavan 1998 and Daunay et al. 2001). The size of eggplant fruits may vary from grams to a kilo and vary greatly in length (Nothmann 1986 and Daunay et al. 2001). Another variable morphological character is the shape of the eggplant fruits. Round, egg shaped, oblong, pear shaped, long and curved are some examples of different forms of the fruits (Nothmann 1986, Lawande and Chavan 1998 and Daunay et al. 2001). Figure 1.10. shows some examples of different sized, shaped and colored Turkish eggplants.

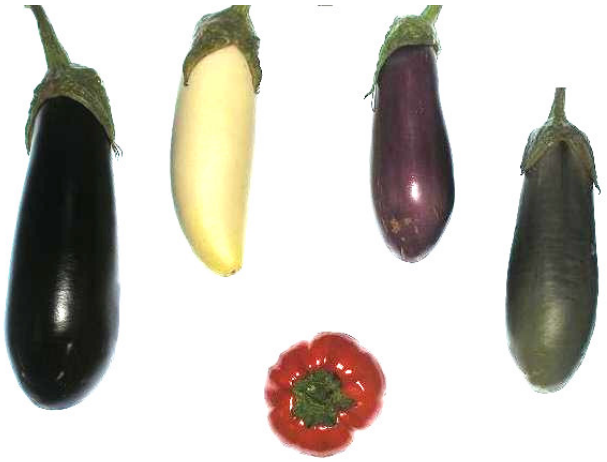

Figure 1.10. Examples of fruit diversity in eggplant.

When the taste of eggplant is considered, bitterness is the main concern and arises from the accumulation of a chemical in different amounts: glycoalkaloids (Lawande and Chavan 1998). The relation between bitterness and glycoalkaloid accumulation is directly proportional (Lawande and Chavan 1998). Consumption of the fruit is related with the ripening process (Nothmann 1986 and Lawande and Chavan 1998). Recommended time is before full maturity at which stage seed formation dominates (Nothmann 1986 and Lawande and Chavan 1998).

In addition to those features and morphological differences, there are also other important traits that exhibit a wide range of variety in eggplant. Flower color, hairiness, leaf shape, parthenocarpy, spines, resistance to pest and diseases are some examples (Nothmann 1986, Lawande and Chavan 1998 and Daunay et al. 2001). Spines are the physiological structures on several body parts and are common for not only eggplant but also the subgenus (Levin et al. 2006). For this reason, subgenus Leptostemonum was referred to as Spiny Solanums in the study of Levin et al. (Levin et al 2006). Resistance to pest and diseases is really important due to general susceptibility to these agents in eggplant which results in serious effects on production and yield (Lawande and Chavan 1998, Daunay et al. 2001 and Collonnier et al. 2001).

As reviewed in the study of Lawande and Chavan, the nutrient composition of eggplant changes according to cultivars and varieties. However, in general, the chemistry of eggplant is mostly composed of moisture: 92.7% (Lawande and Chavan 1998 and Collonnier et al. 2001). Carbohydrates, proteins, fiber and fat come after moisture as 4.0%, 1.4%, 1.3% and 0.3% respectively (Lawande and Chavan 1998 and

Collonnier et al. 2001). Chlorine, phosphorus and sulfur and, with respectively lower amounts, calcium and magnesium are abundant in eggplant (Lawande and Chavan 1998). Vitamin A and C are also important components of this chemical composition (Lawande and Chavan 1998 and Collonnier et al. 2001). Glutamic acid and aspartic acid are two amino acids that are in the highest quantity among the others for different varieties (Lawande and Chavan 1998).

Although eggplant is mainly considered as a food product, it also has medicinal effects (Daunay et al. 2001, Kashyap et al. 2003 and WEB_4 2007). Cholera, diabetes, asthma, bronchitis, dysuria, tooth ache and decrease in cholesterol are examples of health disorders on which eggplant has positive effects (Daunay et al. 2001 and Kashyap et al. 2003). As a family though, Solanaceae includes important genera that have pharmacological properties such as Datura, Belladona, Capsicum and Nicotiana (WEB_4 2007 and WEB_5 2007). Atropine, nicotine and capsaicin are alkaloid derivatives that have impacts on the neural system and epithelium (WEB_5 2007). All these chemicals including eggplant glycoalkaloids have a toxic effect in excessive amounts while serving as therapeutics in small amounts (WEB_5 2007).

Related with the chemical composition, it is known that allergy to Solanaceae members such as potato, tomato and bell pepper are possible for some individuals (Pramod and Venkatesh 2004 and WEB_5 2007). Recently, it has been reported that such allergies to eggplant are rare (Pramod and Venkatesh 2004). In that study, three allergens were detected by electrophoresis and immunoblotting assays and three different people were sampled for their complaints (Pramod and Venkatesh 2004).

1.5. Eggplant Genetic Diversity

As mentioned in the third section, the three varieties of *S. melongena*, the other two eggplant species and most of the species belonging to the Leptostemonum subgenus are diploids and have a haploid chromosome number of 12 (Choudhury 1995, Daunay et al. 2001 and Kashyap et al. 2003). The genus *Solanum* has not yet been properly identified. There is great morphological diversity observed in the genus both at the intra and interspecific levels (Furini and Wunder 2004, Karihaloo and Gottlieb 1995). As reported in different studies, morphological diversity is also detected between individuals of cultivars and between weedy and wild forms of the species (Isshiki et al.

1994b and Karihaloo and Gottlieb 1995). The distribution of the species in a wide region in the world and the existence of different origins of diversity and cultivation areas makes classification much more complicated (Lester 1997, Daunay et al. 2001, Lester and Daunay 2003, Levin et al. 2006 and National Research Council 2006). In addition to these problems, classification attempts were generally based on these morphological data which were in fact assisting the confusion about systematics (Mace et al. 1999b, Daunay et al. 2001, Furini and Wunder 2004 and Doganlar et al. 2002b). However, with new technology and as a result of accumulated molecular knowledge, genotypic information has started to be integrated into eggplant systematics and classification attempts (Daunay et al. 2001).

The Solanaceae as a family has also importance in genetic studies as well as great economic importance (Daunay et al. 2001, Frary et al. 2003 and WEB_4 2007). Three of the model systems used in plant genetics today: tomato, potato and tobacco, are members of the Solanaceae family. Two of these crops, tomato and potato, belong to genus *Solanum* (WEB_4 2007). Figure 1.11. shows the phylogenetic relationships among species used in biological studies.

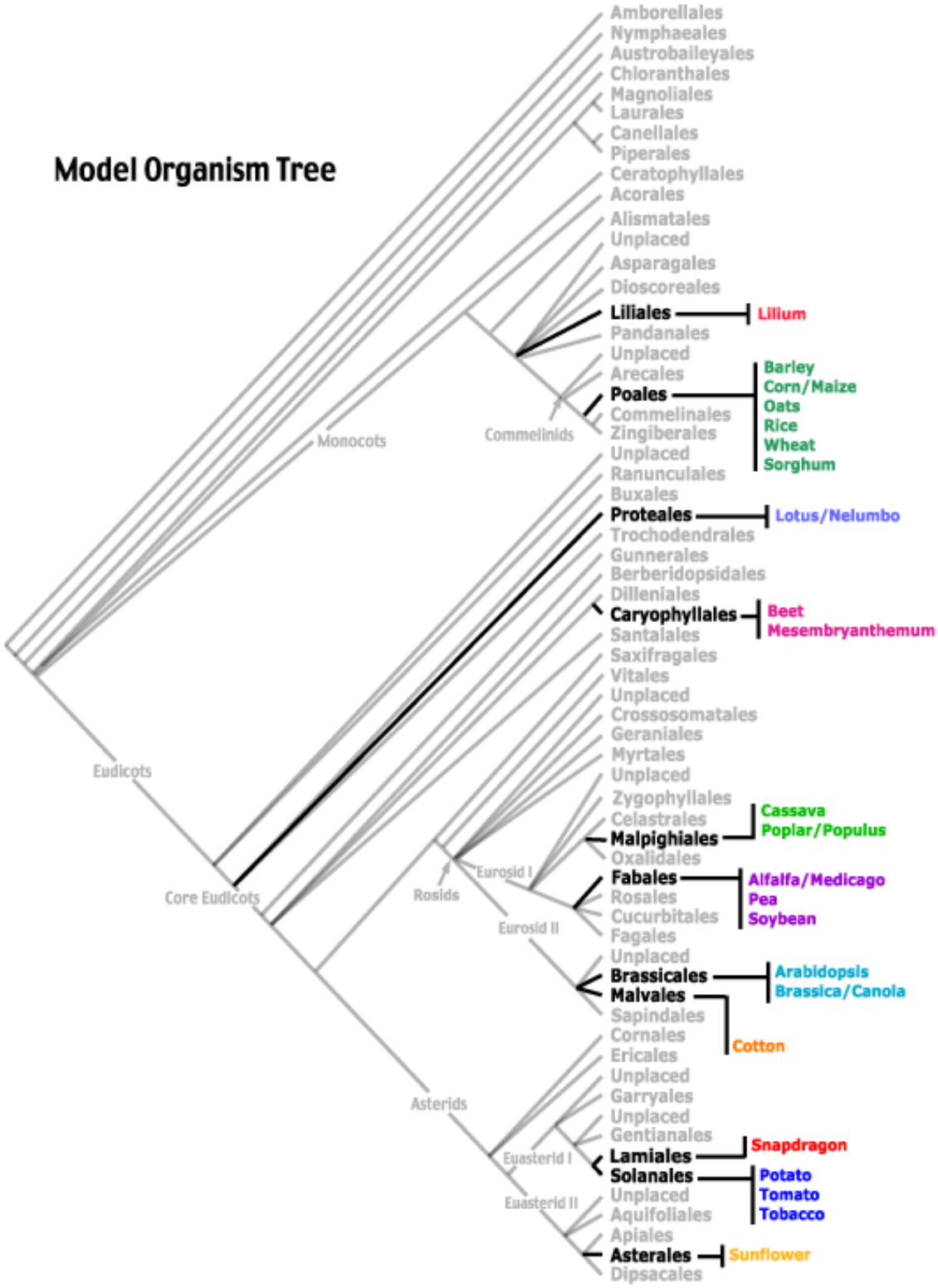

Figure 1.11. Plant species used in biological studies are depicted at the order level while species belonging to these orders are similarly colored. Order Solanales has three species: potato, tomato and tobacco and are in purple (Source: WEB_14 2007).

Recently with intense interest in tomato and other species of the family, there is increased accumulation of genetic data (Mueller et al. 2005). Much of the recent data are expressed sequence tags (ESTs) and can be reached online from a network called SGN (Mueller et al. 2005). Tomato, potato, pepper, petunia and eggplant all have EST libraries available with tomato having the most and eggplant the least ESTs (Mueller et al. 2005). The International Sol Project basically aims to produce data for comparative studies in the family via first sequencing the whole genome of tomato as a reference genome (Mueller et al. 2005 and WEB_4 2007). Sequencing of the tomato genome started in 2004 with collaboration between 10 countries (Mueller et al. 2005). Eggplant has been largely ignored in these studies end and progress is still behind other species in the Solanaceae family.

The systematics of flowering plants of which the Solanaceae family is a member has recently been organized by the Angiosperm Phylogeny Group (APG II 2003 and WEB_1 2007). The idea of the group was to obviate the questions and contradictions about classification (APG II 2003). The latest phylogenetic classification of angiosperms (flowering plants) and a detailed classification are shown in Figure 1.12. and Figure 1.13.

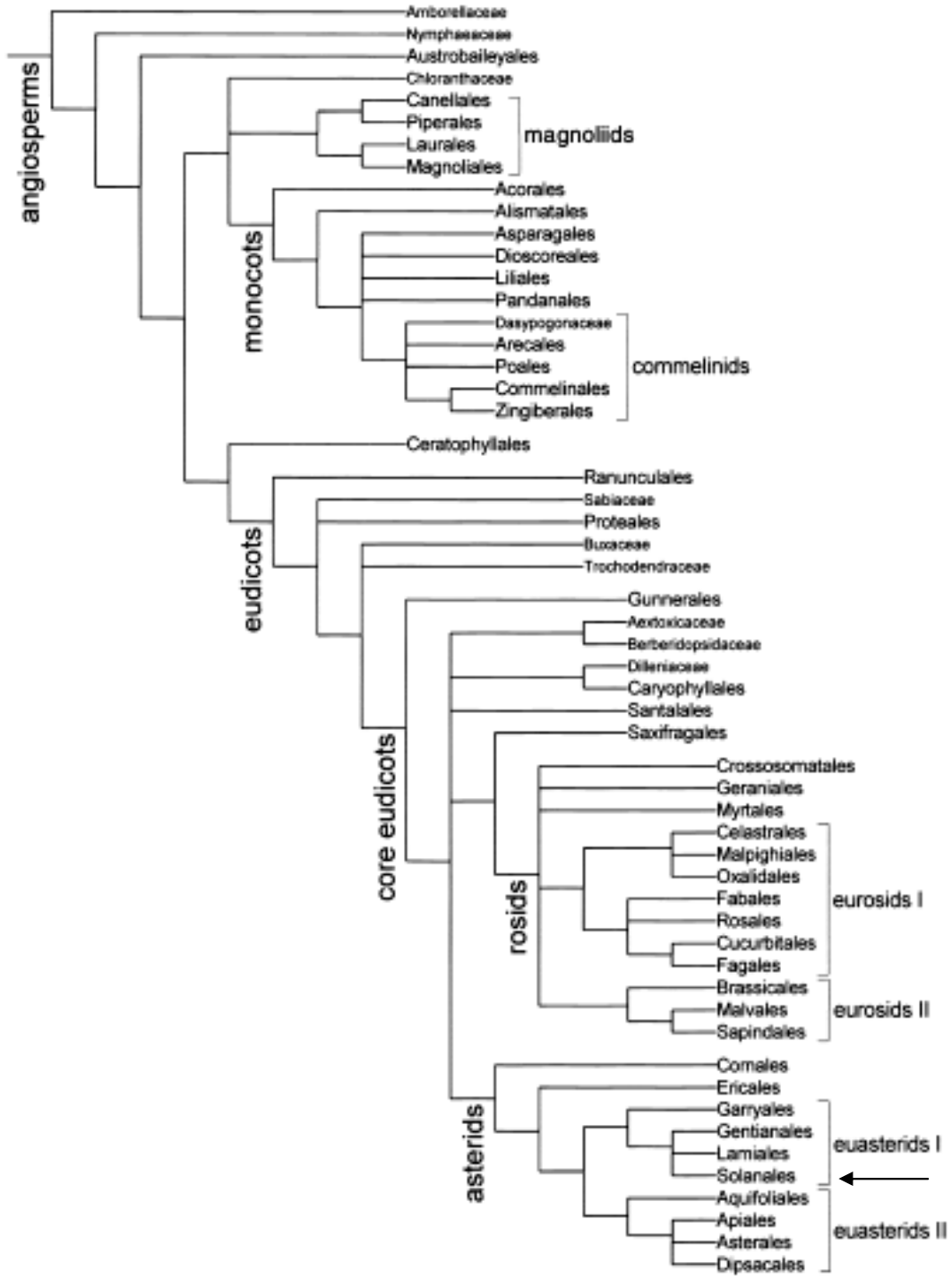

Figure 1.12. Phylogenetic classification of angiosperms. Solanales order is indicated by the arrow (Source: APG II 2003).

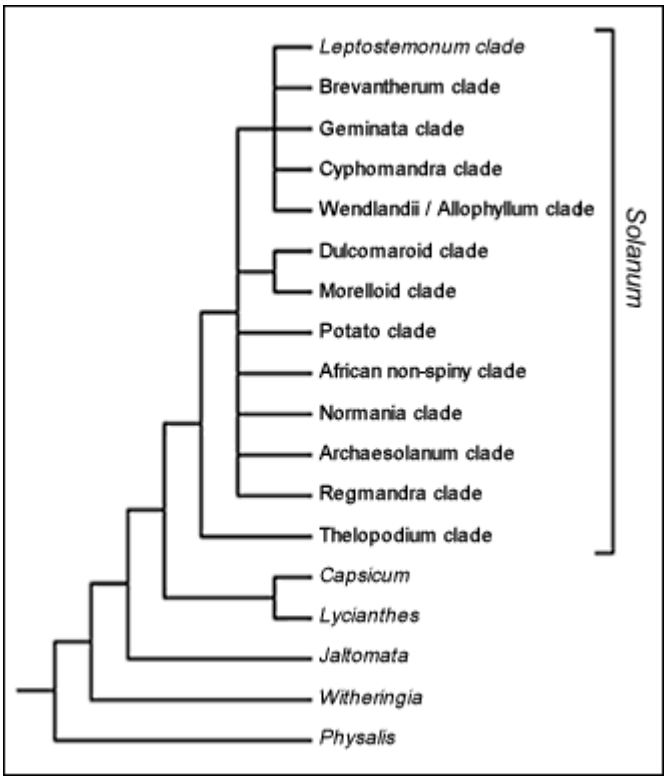

Figure 1.13. A detailed view of *Solanum* genus' classification. The Leptostemonum clade includes eggplant and its close relatives (Source: WEB_11 2007).

1.6. Studies of Genetic Diversity in Eggplant

In addition to morphological data, which has been used classically in taxonomy, other fields of science such as embryology, chemistry and anatomy have been used for revised classification of organisms (Daunay et al. 2001 and Singh et al. 2006). The usage of molecular data in taxonomy and systematics of organisms is the newest strategy for increasing accuracy in relation to evolutionary history. Molecular and other types of data are also being applied for classification in the Solanaceae family. Experiments based on molecular investigations started just three decades ago for this family (Daunay et al. 2001). In this view, some of the first studies were carried out at the protein level and examined differences in allozyme and isozyme patterns between individuals (Isshiki et al. 1994a, Isshiki et al. 1994b, Isshiki et al. 1994c, Karihaloo and Gottlieb 1995 and Kaur et al. 2004). Basically in these studies, *Solanum melongena*, commonly known as eggplant was compared with its weedy and wild forms and close relatives (Isshiki et al. 1994a, Isshiki et al. 1994b, Isshiki et al. 1994c, Karihaloo and

Gottlieb 1995 and Kaur et al. 2004). The purpose of these studies was to measure genetic diversity in those organisms and these types of markers were identified as being especially advantageous for cultivar studies (Isshiki et al. 1994a and Isshiki et al. 1994b). Although having important features that all markers should have such as codominancy, stability and also concordance with the previous classification attempts, limitations about the number of isozymes available and their possibility of further modifications at the cellular level, such as post-translational modification, resulted in declined interest in isozymes and allozymes (Staub and Serquen 1996, Daunay et al. 2001, Kaur et al. 2004 and Isshiki et al. 1994a). Thus, there are not much data accumulated from enzyme studies for eggplant (Kaur et al. 2004). However, studies based on enzymatic patterns are still being done. Recently, the technique was applied with an increased collection of isozymes and plant material and was found to be concordant with previous diversity results (Kaur et al. 2004).

Another way which is used to determine genetic diversity is concentrated on one of the extrachromosomal DNAs: the chloroplast genome. So far, several studies were done using chloroplast DNA analysis. In the early studies on *Solanum melongena* and its relatives, non-radioactively labeled total chloroplast DNA (cpDNA) was used as probe for detection in total DNA (Sakata et al. 1991, Sakata and Lester 1994 and Sakata and Lester 1997). The results were satisfactory in terms of data accumulation and agreement with previous studies. In addition to this, some interesting results were obtained. In the study of Sakata et al. (1997), it was found that morphological diversity was not related to cpDNA diversity and that morphologically similar lines could have quite different cpDNA patterns (Sakata and Lester 1997). A similar situation for the family Solanaceae was reviewed by Knapp et al. (2004). Although supporting important data in the taxonomy of Solanum, cpDNA studies are probably more effective for higher taxonomic levels due to cpDNAs maternal inheritance pattern and conservation (Sakata and Lester 1997, Olmstead et al. 1999 and Daunay et al. 2001). In the review of Daunay et al. (2001) it is proposed that the Leptostemonum subgenus which includes Melongena section is relatively suitable for cpDNA analysis.

Recently, studies about cpDNA are mostly concentrated on sequence data. The experiments are designed with a combination of different data just from cpDNA or a combination of nuclear DNA and cpDNA. In the study of Olmstead et al., restriction fragment length differences in the nuclear genome and the sequence of two chloroplast genes were used to construct cladograms reflecting phylogeny of the Solanaceae family

(Olmstead et al. 1999). In another study, Levin et al. studied only sequence data for phylogenic systematics of subgenus Leptostemonum of the genus *Solanum* (Levin et al. 2006). At a higher taxonomic level, Bremer et al. worked with 3 coding and 3 non-coding cpDNA regions to observe phylogeny of asterids to which the order Solanales belongs (Bremer et al. 2002). Also in this study, they checked the feasibility of using non-coding cpDNA regions to study phylogeny in higher taxa (Bremer et al. 2002). Today, there are many phylogeny studies using cpDNA genes or non-coding regions in different genera. Even in a recent study, it was proposed that as a comparison criterion the sequencing of the chloroplast genome is important (Martin et al. 2005). This approach represents the same idea that, in general, increased numbers of genes, individuals, species or markers give more accurate results (Martin et al. 2005). The Solanaceae family is then one step further than other plant families because the cpDNA of a member of the family, *Nicotiana tabacum*, was the first completely sequenced chloroplast genome (Olmstead et al. 1999).

Other detection techniques used for revealing plant diversity depend on nuclear genome analysis and use molecular markers (DNA-based markers) (Mohan et al. 1997 and Jones et al. 1997). Compared to morphological markers, molecular markers are noteworthy because they are unaffected by environmental changes and do not change the morphology of plants (Mohan et al. 1997, Jones et al. 1997 and Singh et al. 2006). Additionally, there are many more molecular markers in comparison to morphological markers (Jones et al. 1997). Molecular markers can be observed at any growth stage which is one of the most advantageous properties for breeders. The use of molecular markers leads to a new application field: marker assisted selection (MAS) (Staub and Serquen 1996, Mohan et al. 1997 and Kashyap et al. 2003). Via MAS, breeders can benefit from the early detection of traits of interest that have economic and agronomic importance (Staub and Serquen 1996 and Mohan et al. 1997). Some of these economically important traits are controlled by single genes (Staub and Serquen 1996). However, many important traits such as yield are under the control of several genes. In such cases, MAS has the most benefits (Staub and Serquen 1996). To summarize, molecular markers such as RFLP, RAPD and AFLP are important markers for not only eggplant but also for other plant species because they provide data for MAS and diversity studies. (Mohan et al. 1997 and Kashyap et al. 2003).

RFLP is one type of molecular marker. The basic principle of this marker is the difference in length of digested pieces of DNA segments (Staub and Serquen 1996 and

Jones et al. 1997). Digestion points are the restriction sites which are recognized by restriction enzymes (Staub and Serquen 1996 and Jones et al. 1997). Variation in length (polymorphism) is the result of a mutation affecting that restriction site (Jones et al. 1997). Despite its codominant nature, ability to define unique loci and reliability, RFLP has a time consuming protocol with additional steps to visualize bands via labeled probes (Staub and Serquen 1996, Mohan et al. 1997 and Jones et al. 1997). However, this technique has an important place in molecular markers in that it is the first marker type that was used in mapping studies for humans and, later, plants (Mohan et al. 1997). The first interspesific genetic linkage map for eggplant was constructed by using RFLP marker system (Doganlar et al. 2002). Two examples of diversity studies using RFLP are the papers of Isshiki et al. (Isshiki et al. 1998 and Isshiki et al. 2003). In these two different studies, they worked on mitochondrial and PCR amplified chloroplast DNA (Isshiki et al. 1998 and Isshiki et al. 2003). The aim for both studies was to look for complementation of DNA regions (mtDNA and cpDNA) with extracted and digested total DNA (Isshiki et al. 1998 and Isshiki et al. 2003). A different and impressive thing for the study of Isshiki et al. (1998) was the amplification of specific cpDNA fragments which reduced some of the labor during the process (Isshiki et al. 1998). Despite this, the results showed that mtDNA and cpDNA were not suitable materials to study diversity in *Solanum melongena* because of low variability (Isshiki et al. 2003).

RAPD is another molecular marker. It has a PCR (Polymerase Chain Reaction) based principle which was firstly defined by two different groups: Welsh and McClelland (1990) and Williams et al. (1990) (Staub and Serquen 1996, Mohan et al. 1997 and Jones et al. 1997). During the assay, just a single primer randomly binds to and allows amplification of several DNA regions. Thus, a banding pattern with 5 to 10 bands is obtained (Staub and Serquen 1996 and Jones et al. 1997). The advantages of this marker system rely on its easiness of application which results in reduced cost and time (Staub and Serquen 1996 and Jones et al. 1997). However, it usually shows a dominant character and generally they are specific to species (Staub and Serquen 1996 and Jones et al. 1997). In addition to these disadvantages, RAPD markers do not carry two of the most important features that a marker should exhibit: reliability and reproducibility (Jones et al. 1997). There are several diversity and mapping studies about eggplant using RAPD. For example; as reviewed in Kashyap et al., an eggplant molecular linkage map was constructed by Nunome et al. and fruit shape and color were mapped with RAPD and AFLP markers (Kashyap et al. 2003 and Nunome et al. 2001).

In another study, Karihaloo et al. looked at diversity between *Solanum melongena* and the weedy form *insanum* (Karihaloo et al. 1995). As a result of this study, it was reported that there is no need to define them as different species due to very high genetic similarity (Karihaloo et al. 1995). More recently, a study was designed upon 5 different species of eggplant to determine their diversity by RAPD analysis (Singh et al. 2006). The genetic differences observed in this study were high and were the result of the fact that sampling was commonly from India which is one of the most important diversity regions in the world (Singh et al. 2006).

AFLP is one of the most favorite markers with its many advantages and fewer disadvantages as described in Section 1.2.1.1. There are several studies about eggplant AFLP. These studies are generally concentrated on diversity of eggplant *(Solanum melongena)* while the results support the suitability of AFLP for that kind of analysis as firstly indicated by the study of Mace et al. (Mace et al. 1999b, Furini and Wunder 2004). In another study of Mace et al., they used AFLP markers to reveal phylogenetic relations of Datureae which is a member of the Solanaceae family (Mace et al. 1999a). This study showed that AFLP analysis was more informative than isozyme, morphological and ITS-1 markers when the same accessions were compared (Mace et al. 1999a). Furini and Wunder also studied eggplant and related species. However in addition to AFLP data, morphological data were evaluated as a verification tool especially for such a diverse and complicated genus (Furini and Wunder 2004). It was also emphasized in the study that the way of deciding which plants should be saved in the seed banks should be revised by the addition of molecular data (Tanksley and McCouch 1997 and Furini and Wunder 2004). AFLP as a technique is also used for mapping studies for eggplant as with other plant species (Mohan et al. 1997, Kashyap et al. 2003 and Frary et al. 2007). With combinations of other molecular tools, AFLP was used for constructing several genetic linkage maps of eggplant (Kashyap et al. 2003 and Frary et al. 2007).

Like AFLP, SSR is an important molecular marker type owing to its significant properties which were described in Section 1.2.1.2. Based on these features, SSRs can be used in breeding, MAS, mapping, fingerprinting, population genetics and phylogenetic studies (Staub and Serquen 1996, Powell et al. 1996, Jones et al. 1997, Nunome et al. 2003a, Nunome et al. 2003b and Varshney et al. 2005). Named differently for plants and vertebrates as SSRs (Simple Sequence Repeats) and STRs (Simple Tandem Repeats), respectively, repeated sequences as markers are really

informative for both plants and vertebrates (Staub and Serquen 1996). The commonly observed types of repeats are different between humans and plants and also among plant species and it was estimated that 10 fold fewer SSRs are found in plants than humans and diagramed as shown in the Figure 1.14. (Powell et al. 1996, Mohan et al. 1997 and Nunome et al. 2003a).

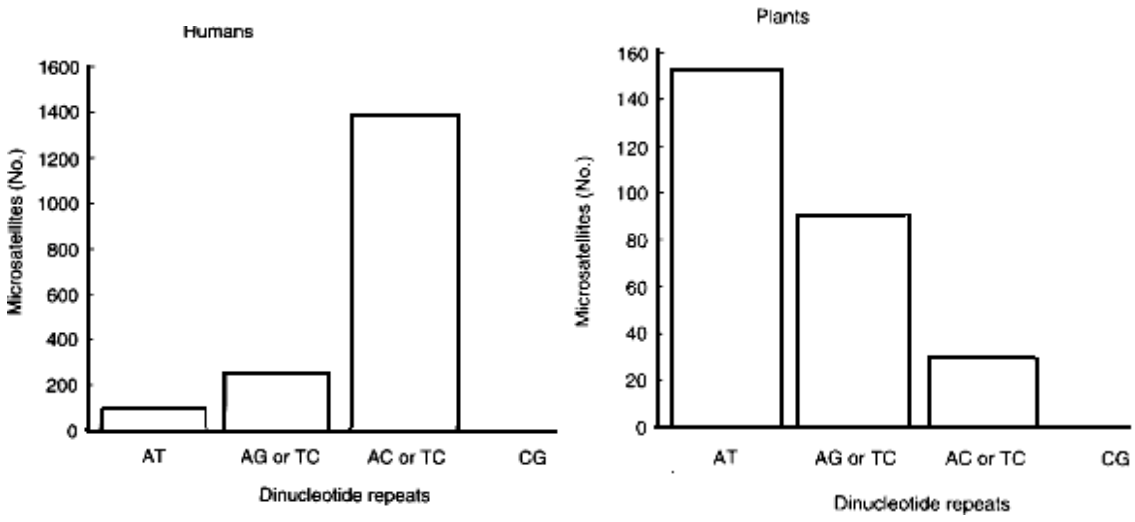

Figure 1.14. Number of different types of dinucleotide repeats in humans and plants. (Source: Powell et al. 1996)

The first study about SSRs in eggplant concentrated on their suitability as a marker system for molecular analysis of this plant (Nunome et al. 2003b). In that study, Nunome et al. built a linkage map of eggplant that had SSR, AFLP and RAPD markers (Nunome et al. 2003b). In another study, Nunome et al. specifically examined trinucleotide repeats in eggplant (Nunome et al. 2003a). The reason trinucleotides were used was because of their greater suitability for allele differentiation (Nunome et al. 2003a).

In this thesis, two separate assays upon two different sample set were applied. For the first one, AFLP technique was used to reveal genetic diversity among several Turkish local varieties. For the second assay, SSR molecular marker technique was used to identify genetic similarity between *Solanum melongena* and related species. For both studies, materials used in the experiments were kit-extracted DNAs of greenhouse-grown samples.

CHAPTER 2

MATERIALS AND METHODS

2.1. Materials

2.1.1. Plant Material

The plant material used in this thesis' studies can be categorized into two groups in terms of two different experimental designs: AFLP and SSR. One group of material was Turkish eggplants. The seeds were supplied by Dr. Ayfer Tan; Aegean Agricultural Research Institute (AARI), Turkey (Ege Tarımsal Araştırma Enstitüsü Menemen, Türkiye) and these accessions are listed in Table 2.1.

Table 2.1. Turkish eggplants characterized by AFLP.

Given Genotype Numbers	Pedigree Number	Accession number	Cultivar Name
1	06T53	TR 66688	Burdur Yerli Patlıcan
2	06T54	TR 66667	Isparta Patlıcan
3	06T55	TR 66572	Uşak Patlıcan
4	06T56	TR 43010	Çanakkale Kır Patlıcan
5	06T57	TR 40300	Gaziantep Mor Dolmalik
6	06T58	TR 37266	Kastamonu Uzun Patlıcan
7	06T59	TR 66013	Bursa Topan Patlıcan
8	06T60	TR 43306	Edirne Kırmızı Patlıcan
9	06T61	TR 66017	Bilecik Kemer Patlıcan
10	06T62	TR 66012	Eskişehir Tombul Ak
11	06T63	TR 66559	Kütahya Patlıcan
12	06T64	TR 62668	Manisa Uzun Patlıcan
13	06T65	TR 68530	Zonguldak Patlıcan
14	06T67	TR 70633	Kemer-27
15	06T68	TR 50591	İzmir Patlıcan
16	06T74	TR 70635	Topan-374
17	06T75	TR 62004	
18	06T76	TR 52348	
19	06T77	TR 62430	
20	06T78	TR 62423	
21	06T79	TR 62491	
22	06T80	TR 62525	
23	06T82	TR 62581	
24	06T84	TR 62667	
25	06T85	TR 62736	
26	06T86	TR 62776	
27	06T87	TR 62385	
28	06T89	TR 61593	
29	06T91	TR 61563	
30	06T92	TR 61564	
31	06T93	TR 61706	
32	06T94	TR 61518	
33	06T95	TR 61493	
34	06T96	TR 61766	
35	06T97	TR 61856	
36	06T99	TR 62049	
37	06T100	TR 62043	
38	06T102	TR 61985	
39	06T103	TR 62073	
40	06T104	TR 62072	
41	06T105	TR 62139	
42	06T106	TR 62101	
43	06T107	TR 62100	
44	06T108	TR 61956	
45	06T111	TR 66009	
46	06T112	TR 66014	
47	06T113	TR 66011	

Table 2.1. Turkish eggplants characterized by AFLP (Cont.).

Given Genotype Numbers	Pedigree Number	Accession number	Cultivar Name
48	06T114	TR 66018	
49	06T115	TR 66334	
50	06T116	TR 66331	
51	06T117	TR 55852	
52	06T118	TR 52522	
53	06T120	TR 43134	
54	06T121	TR 43919	
55	06T122	TR 66579	
56	06T123	TR 66584	
57	06T124	TR 66587	
58	06T125	TR 66589	
59	06T126	TR 66597	
60	06T127	TR 66672	
61	06T128	TR 66667	
62	06T129	TR 66687	
63	06T130	TR 66695	
64	06T131	TR 66698	
65	06T132	TR 66701	
66	06T134	TR 66709	
67	06T135	TR 66720	
68	06T136	TR 66728	
69	06T137	TR 66730	
70	06T138	TR 43768	
71	06T139	TR 55862	
72	06T140	TR 55976	
73	06T141	TR 56029	
74	06T142	TR 61540	
75	06T143	TR 61620	
76	06T144	TR 61892	
77	06T146		Black Beauty
78	06T147		MM738
79	06T149		Çamlıca
80	06T148		Dusky
81	06T875	MM 0195	*S. linnaeanum*
82	06T877	MM 0232	*S. aethiopicum* group Gilo
83	06T874	MM 0150	*S. macrocarpon*

Seeds for 77 different Turkish lines, three non-Turkish cultivars (Black Beauty, MM738 and Dusky) and three wild types as outgroups were sown and grown in the greenhouse with 10 seeds planted per line. The second group consisted of wild relatives of eggplant seeds of which were obtained from Dr. Marie- Christine Daunay; French National Research Institute (INRA), France and are listed in Table 2.2.

Table 2.2. Eggplant and its wild relatives tested with SSR markers.

Given Genotype Numbers	Pedigree Number	Accession Numbers	Species Names
1	06T860	MM 0661	*S. incanum* group A
2	06T861	MM 0574	*S. aethiopicum* group Kumba
3	06T862	MM 0497	*S. violaceum*
4	06T863	MM 0374	*S. viarum*
5	06T865	MM 0577	*S. incanum* group C
6	06T866	MM 0498	*S. melongena* group E
7	06T867	MM 0376	*S. capsicoides*
8	06T868	MM 0373	*S. scabrum*
9	06T870	BIRM/S. 2458	*S. melongena* group H
10	06T871	LF3.24	*S. melongena* group H
11	06T872	MM 0132	*S. macrocarpon*
12	06T873	MM 0134	*S. aethiopicum* group Aculeatum
13	06T874	MM 0150	*S. macrocarpon*
14	06T875	MM 0195	*S. linnaeanum*
15	06T876	MM 0210	*S. campylacanthum*
16	06T877	MM 0232	*S. aethiopicum* group Gilo
17	06T878	MM 0284	*S. sisymbrifolium*
18	06T879	MM 1248	*S. incanum* group D
19	06T880	MM 1259	*S. anguivi*
20	06T881	MM 1269	*S. semilistellatum*
21	06T882	MM 1350	*S. melanospermum*
22	06T883	MM 1426	*S. incanum* group B
23	06T884	MM 0337	*S. incanum* group D
24	06T885	MM 0700	*S. incanum* group A
25	06T886	MM 0702	*S. incanum*
26	06T887	MM 0707	*S. incanum* group A
27	06T889	MM 0712	*S. incanum* group A
28	06T890	MM 0713	*S. incanum* group D
29	06T891	MM 0715	*S. incanum* group C
30	06T892	MM 0738	*S. melongena* group H
31	06T893	MM 0824	*S. marginatum*
32	06T895	MM 0982	*S. anguivi*
33	06T896	MM 1005	*S. lidii*
34	06T897	MM 1007	*S. macrocarpon*
35	06T899	MM 1010	*S. melongena* group G
36	06T900	MM 1129	*S. macrocarpon*
37	06T901	MM 1137	*S. dasyphyllum*
38	06T902	MM 1169	*S. aculeantrum*
39	06T903	MM 1235	*S. lurchellii*
40	06T904	MM 1244	*S. incanum* group B
41	06T906	MM 0669	*S. melongena* group E
42	06T907	MM 0672	*S. incanum* group C
43	06T908	MM 0674	*S. lichtensteinii*
44	06T909	MM 0675	*S. melongena* group E
45	06T910	MM 0676	*S. incanum* group D
46	06T911	MM 0677	*S. incanum* group C
47	06T913	MM 0686	*S. melongena* group F

Eggplant and its wild relatives were represented by total 47 different individuals that are encompassed by 20 different species. Within these 20 species; *S. incanum*, *S. melongena* and *S. aethiopicum* had individual groups which were represented by several accessions. Species with the number of accessions for each group and total number of accessions for each species are listed in Table 2.3. Same as Turkish eggplants, wild eggplants were grown in the greenhouse and each species were represented by 10 individuals.

Table 2.3. List of eggplant and its wild relatives with number of accessions tested.

	Species Names	**Number of accessions**	**Total number of accessions**
1	*S. incanum*	1	
	Group A	4	
	Group B	2	
	Group C	4	
	Group D	4	15
2	*S. melongena*	3	
	Group E	1	
	Group F	1	
	Group G	3	8
	Group H		
3	*S. aethiopicum*	1	
	Group Aculeatum	1	
	Group Gilo		
	Group Kumba	1	3
4	*S. violaceum*		1
5	*S viarum*		1
6	*S. capsicoides*		1
7	*S. scabrum*		1
8	*S. macrocarpon*		4
9	*S. linnaeanum*		1
10	*S. campylacanthum*		1
11	*S. sisymbrifolium*		1
12	*S. anguivi*		2
13	*S. semilistellatum*		1
14	*S. melanospermum*		1
15	*S. marginatum*		1
16	*S. lidii*		1
17	*S. dasyphyllum*		1
18	*S. aculeantrum*		1
19	*S. lurchellii*		1
20	*S. lichtensteinii*		1

2.1.2.Sample DNAs

2.1.2.1. Extraction

For the extraction process, Wizard Genomic DNA Purification Kit, (Promega, Madison, WI, USA), was utilized. The protocol was applied with a few modifications. Genomic DNA from the fresh and youngest leaves of 10 plants representing each individual was extracted separately. Instead of directly using 600 µl Nuclei Lysis Solution to each tube, amount was added in two steps. At first step, 250 µl of solution was used for grinding. Then remaining 350 µl was added to each tube and ground tissue was mixed several times for better homogeneity. Another modification was about centrifugation. Instead of 3 min. at 13.000 – 16.000 g, samples were spun at 10.000 g for 5 min. at 6^{th} step in the protocol and 10.000 g for 2 min. at 9^{th} step. At the 10^{th} step, ethanol washed samples were spun again at 10.000 g for 2 min. After rehydration of DNA with DNA rehydration solution, 5 µl DNA per each individual was taken and combined in a new tube with the DNAs of the other individuals of the same accession.

2.1.2.2. Quantity Checking

Mixed samples were checked in NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA) to determine the quantity of the DNAs. The NanoDrop values of Turkish and Wild eggplants are in Table 2.4. and Table 2.5. According to these values, the amount of DNA that was used in the experiments was adjusted as described in Section 2.2.1. for AFLP experiments and Section 2.2.2.2. for SSR experiments.

Table 2.4. Turkish Eggplants Nanodrop Results.

Pedigree Number	ng/ul
06T53	663.61
06T54	513.21
06T55	802.84
06T56	806.55
06T57	701.71
06T58	706.64
06T59	411.32
06T60	1224.08
06T61	804.88
06T62	514.05
06T63	563.4
06T64	299.96
06T65	800.29
06T66	1638.53
06T67	598.48
06T68	989.15
06T74	992.28
06T75	1014.87
06T76	191.73
06T77	831.69
06T78	741.69
06T79	348.19
06T80	102.28
06T81	782.35
06T82	1255.79
06T83	70.08
06T84	889.27
06T85	506.98

Pedigree Number	ng/ul
06T86	488.46
06T87	672.35
06T89	715.87
06T91	583.79
06T92	456.02
06T93	913.89
06T94	687.44
06T95	1113.16
06T96	779.57
06T97	695.55
06T99	478.93
06T100	2031.05
06T102	738.43
06T103	689.2
06T104	197.12
06T105	917.37
06T106	162.26
06T107	4410.06
06T108	1074.34
06T111	124.78
06T112	908.53
06T113	811.18
06T114	1047.41
06T115	1798.6
06T116	1450.27
06T117	1476.57
06T118	116.44
06T119	167.45

Pedigree Number	ng/ul
06T120	1156.8
06T121	1004.07
06T122	1048.06
06T123	972.8
06T124	1959.06
06T125	110.35
06T126	120.99
06T127	162.17
06T128	47.77
06T129	54.64
06T130	843.13
06T131	188.59
06T132	936.93
06T134	1373.68
06T135	1106.52
06T136	239.36
06T137	98.67
06T138	2619.18
06T139	1210.54
06T140	1148.92
06T141	98.54
06T142	1236.51
06T143	3085.77
06T144	239.18
06T146	352.57
06T147	182.34
06T149	839.13

Table 2.5. Wild Eggplants Nanodrop Results.

Pedigree Number	ng/ul
06T860	647.04
06T861	960.96
06T862	338.08
06T863	429.92
06T865	966.16
06T866	160.46
06T867	1151.32
06T868	484.65
06T870	944.99
06T871	1140.79
06T872	877.97
06T873	398.95
06T874	929.56
06T875	628.03
06T876	610.16
06T877	385.55
06T878	320.47

Pedigree Number	ng/ul
06T879	233.24
06T880	596.69
06T881	274.93
06T882	414.5
06T883	317.42
06T884	858.41
06T885	225.5
06T886	156.09
06T887	243.62
06T888	101.97
06T889	217.64
06T890	244.27
06T891	446.42
06T892	414.13
06T893	683.53
06T895	591.43
06T896	331.32

Pedigree Number	ng/ul
06T897	447.13
06T899	1185.94
06T900	389.76
06T901	896.81
06T902	332.64
06T903	118.43
06T904	88.21
06T905	698.1
06T906	751.88
06T907	440.85
06T908	465.78
06T909	507.66
06T910	331.29
06T911	476.01
06T913	361.75

2.2. Methods

2.2.1. AFLP

For AFLP experiments, two different kits were used: Invitrogen AFLP Core Reagent Kit and Invitrogen AFLP Starter Primer Kit. Several pre-experiments were done to optimize the protocol. The final protocol was the one defined in the user manual of AFLP Analysis System I, AFLP Starter Primer Kit, Version B, 2003 with a few modifications as described below. Water used during the whole process was either that provided with the kit or Sigma Water (Sigma-Aldrich Company, LTD Irvine, Ayrshire KA12 8NB, UK). As Taq polymerase, Promega GoTaq DNA Polymerase (Promega, Madison, WI, USA) was used in amplification reactions. Prepared samples were analyzed with the CEQ™ 8800 Genetic Analysis System (Beckman Coulter, Inc., Fullerton, CA, USA). The chemicals used during this analysis were Sample Loading Solution, Size Standard-600, Mineral Oil, Separation Buffer, Separation Gel and Separation Capillary Array all of which were Beckman Coulter products.

The protocol consisted of several steps. Firstly, isolated sample DNAs were restricted with two enzymes the properties of which were defined in Section 1.2.1.1. These two enzymes were *Eco*R I and *Mse* I and were supplied in the kit as a mixture. Modification in this step was the adjustment of each sample DNA concentration to ~ 700 ng/µl.

At the second step, adapters specific to *Eco*R I and *Mse* I restriction sites were bound to those regions. For this step, no changes were applied.

Next step included the first PCR reaction. In this step instead of 1:50 dilution, PCR products were diluted 1:40: 1 µl sample DNA and 39 µl sample loading solution (SLS).

The second PCR reaction was based on trinucleotide extension of the previous step's samples. In this step, fluorescent labeled primers were used for detection in the CEQ 8800 Genetic Analysis System. Different *Eco*R I primers with different triplets were labeled by Sigma-Proligo (Sigma-Aldrich Company, LTD Irvine, Ayrshire, UK). These primers were the same as the kit primers and were diluted before use to 1 pmol/µl. Because there are no defined and recommended primer combinations for eggplant for this product, information about suitable primer combinations for other related Solanaceous species was used (Invitrogen 2003). Thus, combinations that worked well for tomato, pepper and potato were defined. Of the total 22 primer combinations, 10 were selected and applied to Turkish eggplants (Table 2.6.).

Table 2.6. Selective primer combinations that were applied to Turkish eggplants.

*Eco*R I / *Mse* I combinations					
	M - CAC	M - CAT	M - CAG	M - CAA	M - CTA
E - ACA	√ (3.pri.com)				
E - ACC		√ (5.pri.com)			
E - ACT	√ (11.pri.com)	√ (13.pri.com)	√ (6.pri.com)		
E - AAC			√ (7.pri.com)		
E - AGC				√ (16.pri.com)	√ (19.pri.com)
E - AGG			√ (18.pri.com)	√ (17.pri.com)	

Other modifications in this step were using labeled *Eco*R I primers instead of labeling them as described in the user manual. For Mix 1 and for each sample, 2.5 µl *Eco*R I, 1.5 µl *Mse* I primers were used. 1µl dH_2O was added per sample to complete the total volume of Mix 1 to 5 µl. Mix 2 was prepared as the manual described and 5 µl from Mix 1, 10 µl from Mix 2 and 5 µl DNA were mixed. First PCR profile for this selective amplification was chosen due to defined PCR machine properties.

The last step was the preparation of the samples, for the machine. Selective PCR products were first diluted 1:5 with dH_2O: 2 µl DNA and 8 µl SLS. Then, a second dilution with SLS was done: 3 µl DNA was mixed with 30 µl SLS and 0.5 µl size standard-600.

As appropriate for the size standard that was used, (Size Standard-600, GenomeLab, Beckman Coulter, Inc., Fullerton, CA, USA), Frag 4 method in the system was chosen. The profile of the method was: capillary temperature 50ºC, denaturation temperature 90ºC for 120 sec., injection voltage 2.0kV for 30 sec. and with a separation voltage 4.8 kV for 60.0 min. After definition of the plate and method, system was started to be run.

2.2.2. SSR

2.2.2.1. Design and Checking of the SSR Primers

For design of SSR primers, an EST library of *Solanum melongena* with 3181 sequences was accessed from Sol Genomics Network (WEB_4 2007). SSRs in the sequences were found using the SSR Discovery Input program from PBC Public Databases (WEB_15 2006). Among these designed primers, the primers that had certain repeat numbers were selected for synthesis by Integrated DNA Technologies, Inc. IA, USA. In the next step, synthesized primers were checked for amplification in PCR reactions. PCR conditions were a preliminary denaturation for 5 min. at 94ºC; 35 cycles at 94ºC for 30 s., 50ºC for 1 min., 72ºC for 1 min.; final extension for 5 min at 72ºC and hold at 4ºC. For annealing temperature, a general estimation was done with 5ºC less then the melting temperature of the SSR primers. Generally, 50ºC was applied to all primers. PCR reaction was 25 µl per sample: 2.5 µl 10x PCR Buffer; 0.5 µl dNTP, 0.5 µl of F primer and R primer; 0.25 µl Taq Polymerase; 18.75 µl dH_2O and 2 µl sample

DNA. The products were prepared for gel electrophoresis by adding 2 µl blue juice to each sample and the gel was 3% agarose 1xTAE. Samples were electrophoresed for at least 4 hours at 120 mA. For visualizing of DNA bands in the gel, ethidium bromide was used either by adding it directly to the gel or by staining the gel with an ethidium bromide solution after electrophoresis. At last step, these gels were viewed using the AlphaImager Gel Documentation System (Alpha Innotech, San Leandro, CA, USA).

2.2.2.2. SSR Protocol

As a result of gel electrophoresis, primers giving polymorphic bands were detected. These primer combinations' forward pairs were extended by adding M13 sequence. M13 sequence was added to the 5' end of the forward primer whereas the reverse primer remained same as previously designed (Table 2.7.). These newly designed forward primers and separate fluorescent M13 primers were synthesized by Sigma-Proligo (Sigma-Aldrich Company, LTD Irvine, Ayrshire, UK).

The best PCR conditions and the amounts of components in the experiments were determined after several preliminary experiments. PCR reactions were 20 µl total for each sample and were composed of 13.15 µl dH_2O, 2 µl 10x PCR buffer, 0.4 µl dNTP, 0.2 µl Taq Polymerase, 0.75 µl of each primer (F and R primers and M13) and 2 µl sample DNA. Sample DNA concentrations were adjusted ~10 ng/µl by dilution with dH_2O. The profile of the PCR was: 94ºC for 5 min.; 94ºC for 30 s., 56ºC for 45 s., 72ºC for 45 s. for 27 cycles; 94ºC for 30 s., 53ºC for 45 s., 72ºC for 45 s. for eight cycles; 72ºC for 10 min, hold at 4ºC.

Before loading the samples for analysis in the CEQ™ 8800 Genetic Analysis System, PCR products were diluted 1:10 with sample loading solution (SLS). For each sample, 3 µl PCR products were diluted with 27 µl SLS and 0.5 µl size standard-600.

Suitable with the used size standard, Frag 4 method in the system was used. The profile of the method was: capillary temperature 50ºC, denaturation temperature 90ºC for 120 sec., injection voltage 2.0kV for 30 sec. and with a separation voltage 4.8 kV for 60.0 min. After definition of the plate and method, system was started to be run.

Table 2.7. Repeat motifs and sequences for the SSR primers. M13 sequence was added to the forward sequence.

Given Code	Repeat Motif and Number	Forward Sequence	Reverse Sequence
smSSR01	(ATT)21	TGTAAAACGACGGCCAGTGTGACTACGGTTTCACTGGT	GATGACGACGACGATAATAGA
smSSR02	(TA)9 (GA)8	TGTAAAACGACGGCCAGTATTGAAAGTTGCTCTGCTTC	GAAAGAGGAGATCCAGGAGT
smSSR03	(TA)9 (GA)8	TGTAAAACGACGGCCAGTATTGAAAGTTGCTCTGCTTC	GATCGAACCCACATCATC
smSSR04	(TA)9 (GA)8	TGTAAAACGACGGCCAGTCTCTGCTTCACCTCTGTGTT	CCATGAAAGAGAAGATCGAG
smSSR05	(TA)9 (GA)8	TGTAAAACGACGGCCAGTTCTGCTTCACCTCTGTTCTT	AGTAGAGCAACGACGACAAT
smSSR06	(TA)9 (GA)8	TGTAAAACGACGGCCAGTTCTGCTTCACCTCTGTTCTT	GAAAGAGGAGATCGAGGAGT
smSSR07	(TAA)20 (CGA)8	TGTAAAACGACGGCCAGTTGAATGGAATTACACAAGCA	ATTCTCTAAACCTCAGCCAA
smSSR08	(TAA)20 (CGA)4 - (TAA)22	TGTAAAACGACGGCCAGTAATGCAAACAATTATCAGGG	ACAACTCAGCCAGTCGTAAT
smSSR09	(TTTGC)3	TGTAAAACGACGGCCAGTCACATGGGAACCTACTTACC	GACGACCATCAAACAAGAAT
smSSR10	(TTTGC)3	TGTAAAACGACGGCCAGTAAGCTTCGGAGGAAGATAAG	GGGAGATGGAATAAGTCACA
smSSR11	(AGC)6	TGTAAAACGACGGCCAGTAAACAAACTGAAACCCATGT	AAGTTTGCTGTTGCTGCT
smSSR12	(ACCAA)3	TGTAAAACGACGGCCAGTAAACAGAAACCAGAGTACTTCA	CAGAAGAAGGTTCAGTTTGC
smSSR13	(AT)9	TGTAAAACGACGGCCAGTAGGAATTAACATGGTTCAACA	TTCCTCTTACAACCACATCC
smSSR14	(ATTA)4	TGTAAAACGACGGCCAGTATACCACATCAATCCAAAGC	CATCATCATCTTCACAGTGG
smSSR15	(CCTTT)3	TGTAAAACGACGGCCAGTCTGTGGTTGCCTTATCAGTA	TAGTCCAAGGGTTTGATGAC
smSSR16	(AGA)7	TGTAAAACGACGGCCAGTAAGAATTTGATGTTGAACCG	CTTTATCAGCCAATTTCTGG
smSSR17	(ATAC)4	TGTAAAACGACGGCCAGTTCTTGCCATTTAATTTCCTC	CTATGTCCCTATTATGCCCA
smSSR18	(TAAT)4	TGTAAAACGACGGCCAGTTTAGGCATTTGATTTAGCCT	TATGTCCCTAAGCATAACGG
smSSR19	(GAA)6	TGTAAAACGACGGCCAGTGAACAATGATTCATCGGATT	AGTTGATGTTGAATTTCCCA
smSSR20	(AGA)5	TGTAAAACGACGGCCAGTACAAGGAAGGACACAAACAC	ATCTAATCACTGTCGCTGCT
smSSR21	(TAC)5	TGTAAAACGACGGCCAGTAAGTTTACATGACAGCACCA	TTGCCATCATCAATACCATA
smSSR22	(GCC)5	TGTAAAACGACGGCCAGTCTCCGTCAAATTCCTATCAA	GGGAGTCCACATAGAGCATA
smSSR23	(AAG)5	TGTAAAACGACGGCCAGTAGAGAAGAAGCCAGCAGAA	TCTGAATCTCCCGAGAAGTA
smSSR24	(TCA)5	TGTAAAACGACGGCCAGTGATTTATGGCTTCTGATGGA	TCCTAACCCACTTGATGAAC
smSSR25	(TGA)5	TGTAAAACGACGGCCAGTTCCTAACCCACTTGATGAAC	GATTTATGGCTTCTGATGGA
smSSR26	(AAG)5	TGTAAAACGACGGCCAGTCAACTTCGATCTTCAATTCC	TCTGAATCTCCCGAGAAGTA
smSSR27	(TGT)5	TGTAAAACGACGGCCAGTATACATTTGAGCCGAGAGTG	TAAATCTGAGAAGGTCGCAT
smSSR28	(TCA)5	TGTAAAACGACGGCCAGTCACACTCCTCAGAACTCCAT	CAGCAGTACCTCTTGGTCAT
smSSR29	(CTT)5	TGTAAAACGACGGCCAGTTCCACTTCAATTTCCAAGTC	GATCGCTTAGCAGAAGCC
smSSR30	(GAA)5	TGTAAAACGACGGCCAGTGATCGCTTAGCAGAAGCC	TCCACTTCAATTTCCAAGTC

Table 2.7. Repeat motifs and sequences for the SSR primers. M13 sequence was added to the forward sequence (Cont.).

Given Code	Repeat Motif and Number	Forward Sequence	Reverse Sequence
smSSR31	(TCC)5	TGTAAAACGACGGCCAGTCTTCCTACCCACACTTCATC	TAGGCCGGAGATAGTTGTAA
smSSR32	(GAA)5	TGTAAAACGACGGCCAGTCCCACTGATCAGAAGAAGTT	TAGCACACATCCATACCAAA
smSSR33	(TCA)5	TGTAAAACGACGGCCAGTTTGCTAGAAATAGCAAAGGG	CGTGGTGTGTATGATGCTTA
smSSR34	(AGA)5	TGTAAAACGACGGCCAGTACAAGGAAGGACACAAACAC	ATCTAATCACTGTCGCTGCT
smSSR35	(ATG)5	TGTAAAACGACGGCCAGTCACCACCAAAGAATTCCTAA	TTGCTAGAAATAGCAAAGGG
smSSR36	(CTG)5	TGTAAAACGACGGCCAGTAGCACCAGGACAATGAATAC	CCATTTCTTTCTCGACCTTA
smSSR37	(AAG)5	TGTAAAACGACGGCCAGTAAAGAAGCTTCCGACGAA	CACTTGTTTCAGCACTTTGA
smSSR38	(GCT)5	TGTAAAACGACGGCCAGTGCCATAGATGAAAGGTCAGA	GGATTTATGGACAAGGTGAA
smSSR39	(TCA)5	TGTAAAACGACGGCCAGTTTGCTAGAAATAGCAAAGGG	CGTGGTGTGTATGATGCTTA
smSSR40	(AAG)5	TGTAAAACGACGGCCAGTTTCTTTGATCTTCAATTCCAA	ATGAAGCTGTTCATGATTCC
smSSR41	(TCA)5	TGTAAAACGACGGCCAGTCTCCTCCTGGTAAGGAGTCT	GCAGTATAGAGACGCGAAAT
smSSR42	(CAC)5	TGTAAAACGACGGCCAGTACAGTACACCAGAAACGGAA	GTTACAATGACGGTGGATCT
smSSR43	(GCT)5	TGTAAAACGACGGCCAGTACACCTAAACAACAACCAGG	GGTGGTGTTCAGTCATCTTT
smSSR44	(CCA)5	TGTAAAACGACGGCCAGTTGCATTTCATACAGAAACCA	GCAAGGATATCACTGAGCTG
smSSR45	(TTC)5	TGTAAAACGACGGCCAGTTTTCTCAACCCAAACTGAAC	GCAGCTCTCGCATAGATAGT
smSSR46	(CAC)5	TGTAAAACGACGGCCAGTGGAAACCTTCATTCACTTCA	AGGTCACCGTTACAATTACG
smSSR47	(AGA)5	TGTAAAACGACGGCCAGTACACGATGATCATAAGGGAG	ATCTAATCACTGTCGCTGCT
smSSR48	(GCT)5	TGTAAAACGACGGCCAGTGCCATAGATGAAAGGTCAGA	GGATGGAAAGGATAAGAAGG
smSSR49	(ATG)5	TGTAAAACGACGGCCAGTTAGTCAACTGCATCACCAGA	CCACTCCCACTACTGTCACT
smSSR50	(ATG)5	TGTAAAACGACGGCCAGTTATCAGTCAACTGCATCACC	TGCATTTACGTGAGCTCTAA

CHAPTER 3

RESULTS AND DISCUSSION OF AFLP DATA

3.1. Results

3.1.1. Pre-Experiments and Their Results

To determine the final form of the protocol described in Section 2.2.1. that gave the best results for eggplant samples and their relatives, several preliminary experiments were done. These attempts were changes in the amounts of the components or dilution ratios. Also, because no primer combinations for the selective PCR amplification step were specified for eggplant in the kit protocol, various combinations were tested.

At first, sample DNA amounts of 0.5 µl, 1 µl, and 1.5 µl were tested. Then, in accordance with the recommended amount (250 ng), DNA concentrations were fixed to ~ 100 ng/ µl. From these dilutions 2.5 µl was taken for each sample and used in the restriction digestion step. However, the best results for eggplant samples were ~ 700 ng/ µl for restriction digestion. For selective PCR, 5 µl and 7 µl DNA from the previous step were tried and 5 µl was determined to be better.

In the second PCR, although all primers had the same triplet extensions, the selective primers for both *Eco*R I and *Mse* I were not the kit primers. An important point was the need for addition of dNTP which was an extra variable. Different dNTP amounts applied in the experiments were 0.4 µl and 0.6 µl. However, due to discordant results for both samples and amounts and to eliminate a variable, it was decided that *Mse* I primers would be used from the kit. This was basically because dNTP were included with the kit primers. In this step, also, different primers amounts were tested. Different from the user manual, 2.5 µl *Eco*R I and 1.5 µl *Mse* I were decided to be best in the end.

As mentioned previously in Section 2.2.1., due to a lack of defined data about selective primer combinations in the kit manual (Invitrogen 2003), combinations that worked best in related species were determined. Twenty two such combinations were applied to two different Turkish eggplants and the 10 giving the best results were selected (Table 2.6.). Of these 10 combinations, two of them (E-ACT/M-CAG and E-

AAC/M-CAG) were the same as previously published AFLP primer combinations used in eggplant (Mace et al. 1999a and Mace et al. 1999b).

One of the most challenging parts of the experiment which was not mentioned in the protocol was the dilution of samples. Two different dilution steps were of concern: dilution with water before sample loading solution (SLS) and SLS dilution. Dilutions of 1:5, 1:10, 1:20 and 1:40 in water after selective PCR were tested. In those tests, 1:20 was the best resulting dilution ratio. For SLS dilution, three different dilutions were tested: 1:5, 1:10 and 1:20. The results of these dilution ratios were related with the amount of sample DNA and kit primer used during the experiments. Generally when sample DNA volume (water diluted) was high, a higher dilution ratio with SLS and less selective kit primer gave better results. If less DNA was taken, more kit primer and less SLS worked better. There are also TE (supplied by the kit) dilutions according to protocol. For these steps, several attempts were made to find out the best resulting one for eggplant samples. The only change was after the first PCR: instead of 1:50, a 1:40 dilution was applied as described in Section 2.2.1.

In summary, each component in the experiment and their concentrations were sensitive. Less than should be or excess amount of DNA, dNTP, Taq polymerase and primers affect the results and sometimes no result may be obtained. The most important thing that is emphasized in the manual (Invitrogen 2003) was the purity of DNA. For such an importance, extraction of sample DNAs was done with DNA purification kit and quantities of DNAs were measured for each sample as described in detailed in Section 2.1.2. Results from an AFLP experiment and an expanded view of that figure are shown in Figure 3.1. and Figure 3.2.

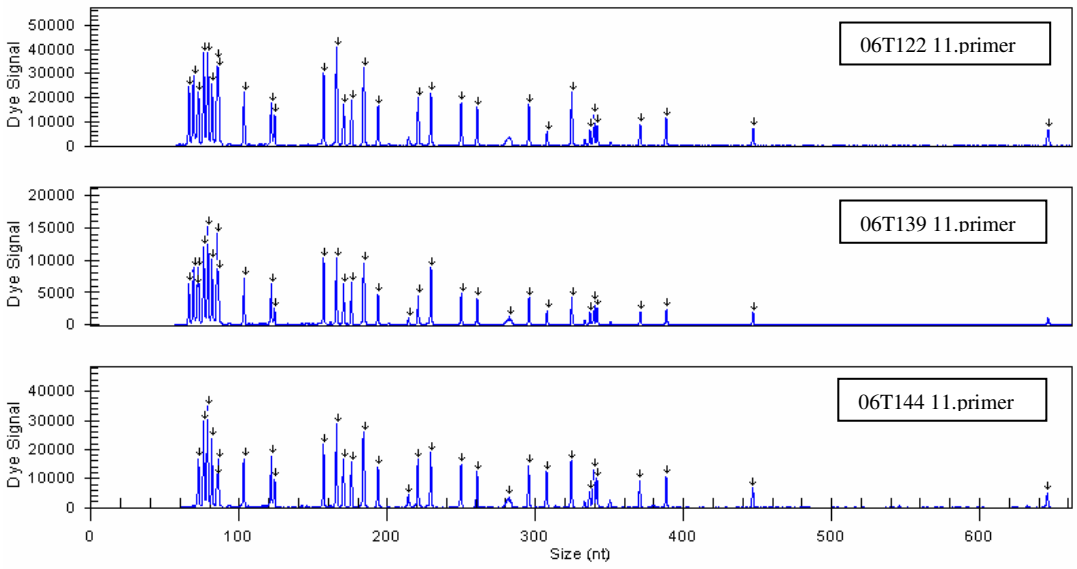

Figure 3.1. An example of AFLP study results for three different samples (06T122, 06T139 and 06T144) with one primer combination (11. primer com). Size standards are not shown in the figure.

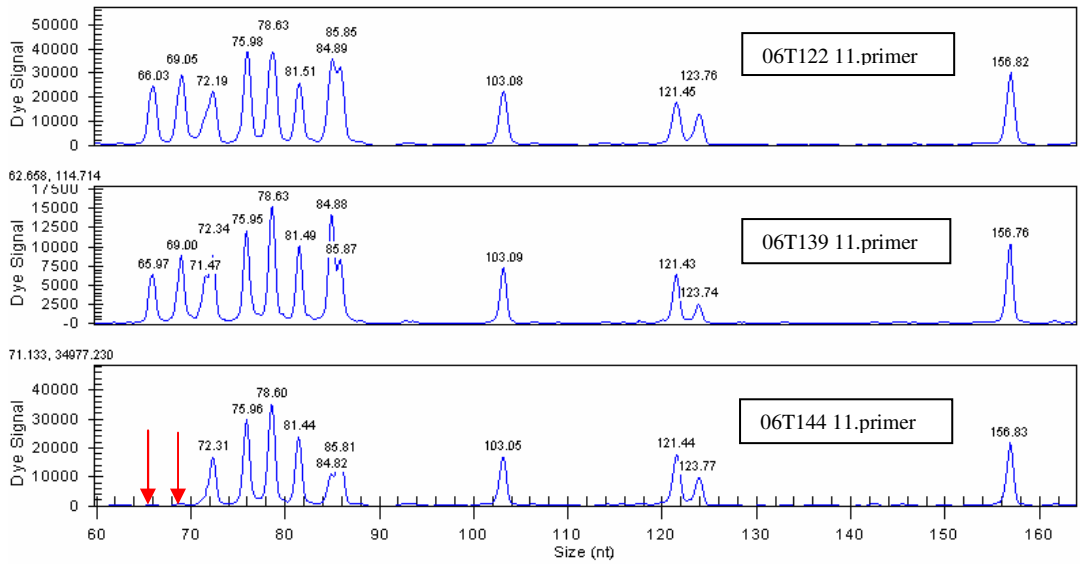

Figure 3.2. A closer view of the same results shown in Figure 3.1. Polymorphisms detected within the three samples are indicated by arrows. Fragment sizes are indicated above each peak.

3.1.2. Analysis and results of the AFLP Data

For analysis of AFLP data, the results of the experiments were genotyped based on the presence and absence of peaks (bands) as 1 and 0, respectively. The preliminary analyzed data were then used to draw a dendrogram. In this study, to draw the dendrogram of Turkish eggplants, NTSYS-pc version 2.2j, (Applied Biostatistics Inc, Setauket, New York, USA), was used. This software consists of several clustering methods including UPGMA (unweighted-pair group method arithmetic average) and enables the correlation of data and construction of dendrograms with two and three-dimensional plots of the components.

To draw the dendrogram, qualitative data were used to generate a matrix determining similarity and dissimilarity among samples. The chosen method was Dice's method (Dice 1945) which is one of the coefficients evaluating similarity of the samples (Mohammadi and Prasanna 2003 and Gulsen et al. 2007). The defined similarity matrix was then used to draw a dendrogram with the clustering method UPGMA via the SHAN module in the software. To decide the efficiency of clustering, the cophenetic correlation coefficient was calculated with the Mantel method (1967) (Mohammadi and Prasanna 2003).

As a second step, Principle Component Analysis (PCA) was done to form two-dimensional and three-dimensional plots representing samples organization in multiple planes (Mohammadi and Prasanna 2003). To do that, a correlation matrix of the data was calculated with SIMINT module in the software. Then, Eigen values were calculated with Eigen module in the software of which values are listed in Table 3.1.

Table 3.1. Eigen values representing principal components of the study AFLP Turkish eggplants at three dimensions are listed in order.

	Eigenvalue	Percent	Cumulative
1	53.40756115	64.3465	64.3465
2	4.66385657	5.6191	69.9656
3	1.86849208	2.2512	72.2168

At the last step, the acquired tree and plots were arranged for the final form while samples were labeled and graphs were organized as shown in Figure 3.3., Figure 3.4. and Figure 3.5.

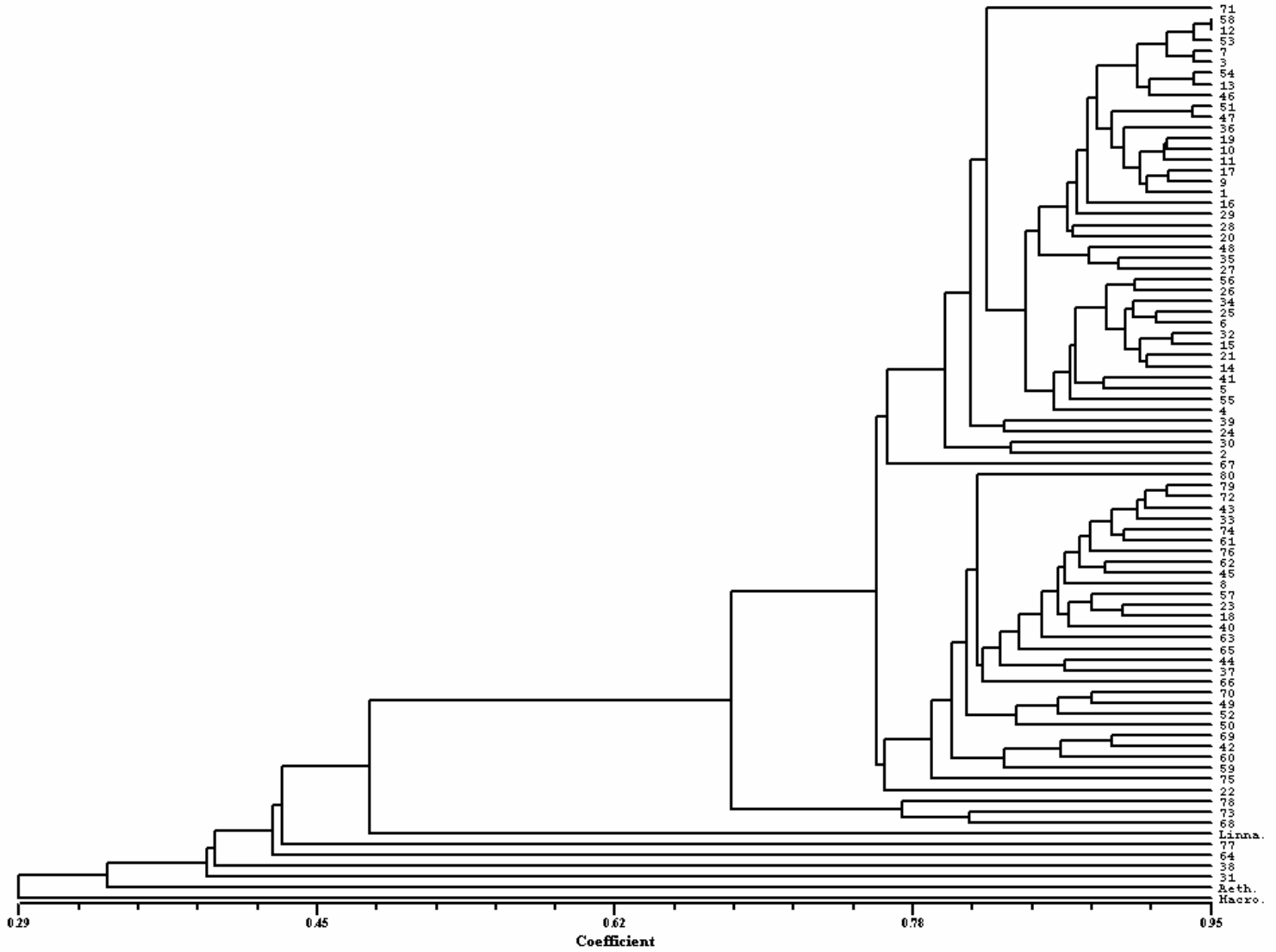

Figure 3.3. Dendrogram showing coefficient of similarity among Turkish eggplants and three outgroups.

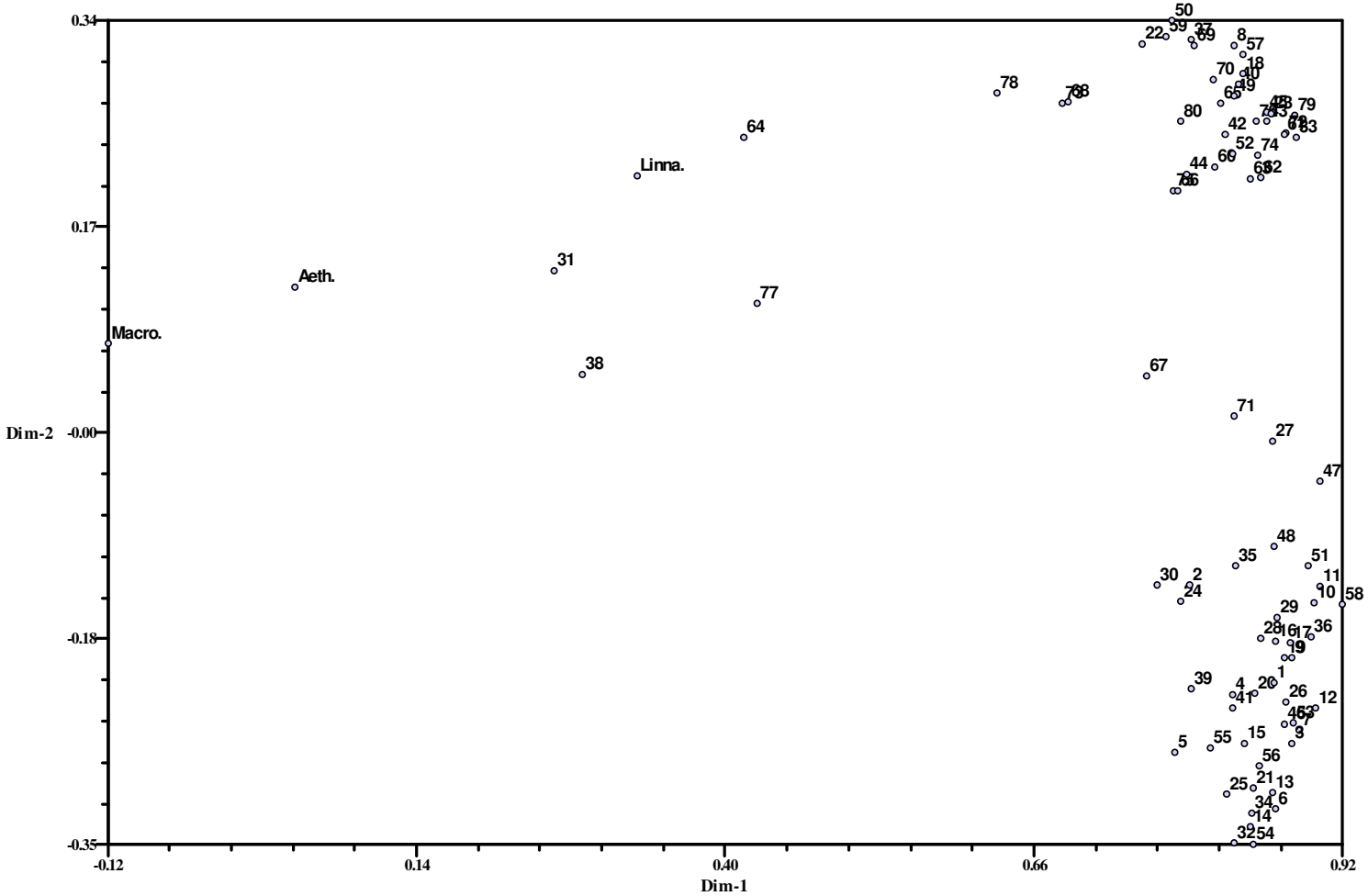

Figure 3.4. Two-dimensional plot of Turkish eggplant accessions.

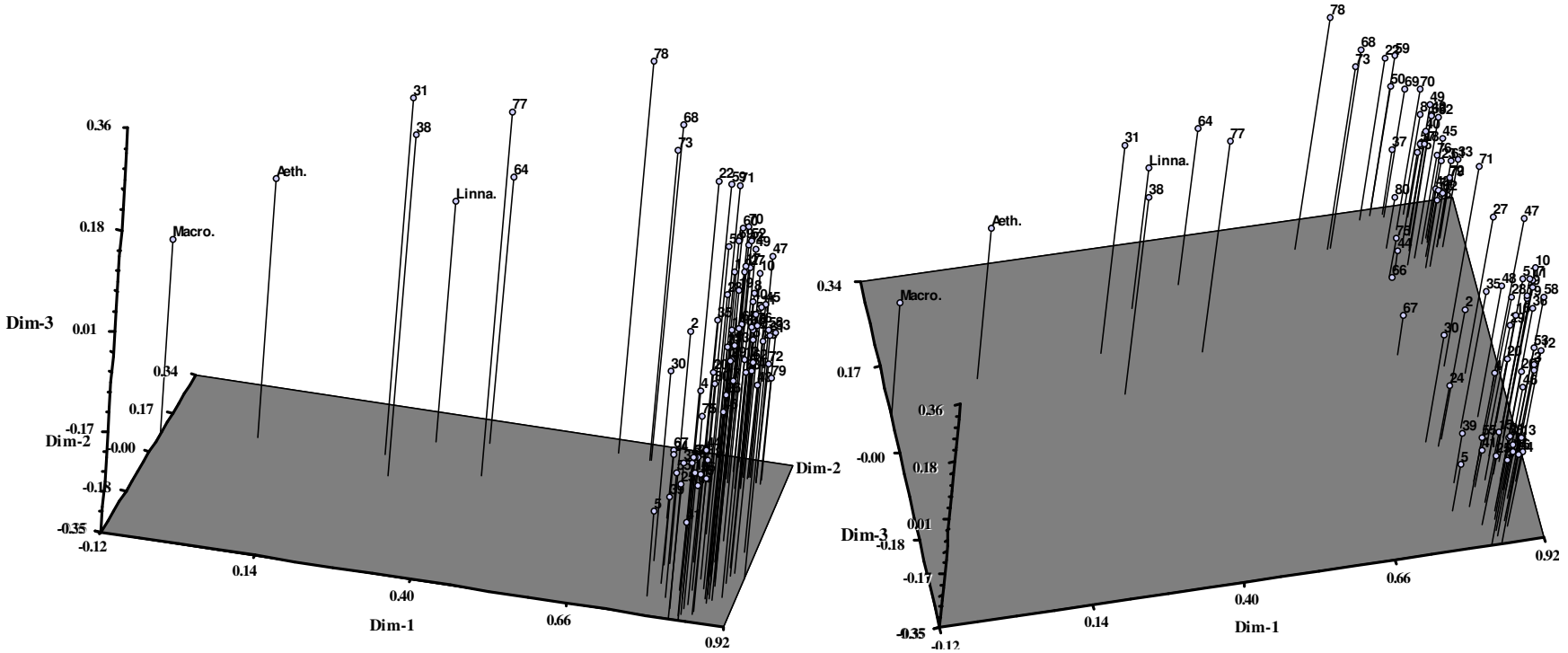

Figure 3.5. Three-dimensional graphs of Turkish eggplants AFLP results.

3.2. Discussion

One of the statistical results of AFLP analysis was the correlation matrix result: r value. In the review of Mohammadi and Prasanna, an r value of more than 0.9 is defined as a very good correlation (Mohammadi and Prasanna 2003). For the AFLP data for Turkish eggplants, an r value of 0.97 was obtained and indicates that the correlation coefficient between the similarity matrix of data and dendrogram was in the best scale (Mohammadi and Prasanna 2003). Reported Eigen values were also informative. These results showed that the first component explained 64.34% of the variation between samples (Table 3.1.). For three axes, a total of 72.21% variation was explained.

The similarity within the total of 83 samples including three outgroups was between 0.29 and 0.95, with a mean value of 0.62 (Figure 3.3.). There were just two samples identical with 0.95 similarity, genotype numbers 58 and 12. Among the outgroups, *Solanum linnaeanum* was the closest sample to Turkish eggplants which were representative of *Solanum melongena*. At the point where *S. linnaeanum* joined to most of the genotypes the similarity coefficient was 0.48. Though, from that point, Turkish eggplants were separated into two groups with a 0.68 cophenetic correlation coefficient. One group which was relatively small consisted of three samples with genotype numbers 78, 73 and 68. The other group was a very large one represented by 73 different genotypes out of 83 (88%) in total (Figure 3.3.). That large group also separated into two big groups of 43 and 30 samples in each group (Figure 3.3.). The correlation coefficient of these two groups was 0.75 (Figure 3.3.). The least similar samples for each of these two groups were genotypes 67 and 22 while 67 belonged to the first, and 22 belonged to the second group.

S. macrocarpon was the least similar outgroup of the three outgroups and had 0.29 similarity coefficient to all remaining genotypes. *S. aethiopicum*, though, was relatively more similar with a 0.33 coefficient value. Samples which were relatively distant from most of the Turkish eggplants even from *S. linnaeanum* outgroup were *S. melongena* genotypes 77, 64, 38 and 31 (Figure 3.3.).

The concordance of the results with previous studies was dependent on the outgroups similarity to *S. melongena*. This was because in previous studies, generally, the variation between *S. melongena* and *Solanum* species was investigated (Sakata et al. 1991, Isshiki et al. 1998, Mace et al. 1999b, Furini and Wunder 2004, Singh et al. 2006

and Levin et al. 2006). Intraspecies diversity examples, though, were again concentrated on *S. melongena* and its closest relatives such as *S. incanum* (Sakata and Lester 1994, Karihaloo and Gottlieb 1995 and Karihaloo et al. 1995). AFLP studies about *S. melongena* were the studies of Furini and Wunder and Mace et al. (Furini and Wunder 2004 and Mace et al. 1999b). In these studies, genetic variation or similarity were investigated within the Solanaceae family. As a consequence of AFLP results, the outgroups' relatedness to Turkish eggplants (*S. melongena*) was similar in the results of previous studies. Among the species *S. linnaeanum*, *S. aethiopicum* and *S. macrocarpon*, *S. linnaeanum* was the closest relative of *S. melongena* (Furini and Wunder 2004 and Levin et al. 2006). In fact, *S. melongena* and these outgroups were reported as in the same subgenus (Leptostemonum) and closer than the other subgenus species (Furini and Wunder 2004). The closeness of *S. linnaeanum* was a result of belonging to the same section, Melongena, while *S. aethiopicum* and *S. macrocarpon* were reported to belong to Oliganthes section (Furini and Wunder 2004).

CHAPTER 4

RESULTS AND DISCUSSION OF SSR DATA

4.1. Results

4.1.1 Pre-Experiments and Their Results

As a result of the database search mentioned in Section 2.2.2.1., 158 different SGN sequences were identified as having at least one SSR. When the SSRs were counted individually, the total number of SSRs increased to 168 as nine of the sequences had two SSRs and one had three SSRs (Table 4.1.). In total, seven compound repeats were identified meaning that two SSRs followed each other (Table 4.1.). Overall, the AT repeat was the most common repeat representing 8.33 % of the total (Table 4.1). The longest simple SSR was a TAA SSR with 22 repeat units (Table 4.1.). Based on total length, the longest SSR was the compound repeat $(TAA)_{20}$ $(CGA)_{8,}$ 84 nucleotides long (Table 4.1.).

When the repeat motifs were classified in terms of the number of the bases in the repeat, it was observed that the most common ones were trinucleotide repeats which represented 56.7% of the total (Table 4.1.). TCA and TTC/AAG were the two most frequently identified trinucleotide repeats with 8 SSRs identified for each (Table 4.1.).

Table 4.1. Repeat motifs, numbers of SSRs identified and average repeat numbers for the SSRs identified in the eggplant EST library.

Repeat Motif	Number of SSRs Identified	Average repeat #
Dinucleotide		
AT/ TA	14/ 2	5,78/ 6,5
AC	3	6
GA	1	7
TC/ AG	1/ 1	5/ 6
CA	1	8
Trinucleotide		
TCA	8	4,625
ATT/ TAA	1/ 1	21/ 22
ACC/ TGG	2/ 3	4/ 4
AGA/ TCT	6/ 2	5/ 4
AGC	4	5,25
TGC	6	4
TTC/ AAG	8/ 8	4,375/ 4,5
CTT/ GAA	4/ 3	4,25/ 5,33
TAC/ ATG	1/ 3	5/ 5
AAC/ TTG	2/ 1	4/ 4
GCA	1	6
GCC	1	5
ACT/ TGA	2/ 3	4/ 4,33
ACA/ TGT	4/ 3	4/ 4,33
CAT	4	4
CCG	4	4
AAT/ TTA	1/ 2	4/ 4
GTG/ CAC	1/ 4	4/ 4,5
ATA	2	4
CCA	2	4,5
GAT	2	4,5
TCC	1	5
GAG	1	4
CTG	1	5
GCT	4	4,75
CAG	1	4
CAA	3	4
Tetranucleotide		
CTGG	1	3
ATAG	2	3
TTTA	2	3
ATAC	1	4
TAAT/ ATTA	1/ 1	4/ 4
AAAC	1	3
TGAC	1	3
Pentanucleotide		
TTTGC	2	3
ATTTT	1	2

Table 4.1. Repeat motifs, numbers of SSRs identified and average repeat numbers for the SSRs identified in the eggplant EST library (Cont.).

Repeat Motif	**Number of SSRs Identified**	**Average repeat #**
AAATA	2	2
AAAAT	3	2
AATTG	1	2
AAAAG	1	2
ACCAA	1	3
ATAAA	4	2
CCTTT	1	3
CATGC	1	2
TTCCT	1	2
Compound Repeats Dinucleotide		
$(TA)_9$ $(GA)_8$	5	17
Compound Repeats Trinucleotide		
$(TAA)_{20}$ $(CGA)_8$	1	28
$(TAA)_{20}$ $(CGA)_4$	1	24

The longest SSRs were selected for primer design. The criterion taken into account was the number of repeat motifs. Thus, only SSRs containing dinucleotides greater than 8, trinucleotides greater than 4, and tetranucleotides greater than 3 units long were used for primer design. A total of 50 SSR primer pairs were designed (Table 4.2.).

In the next step, the EST sequences having SSRs were analyzed for their uniqueness. Thus, the 158 SSR-containing sequences were found to represent 110 unigenes (Table 4.3.). The remaining 48 sequences were members of these unigene families (Table 4.3.). In the table, unique ESTs are listed with (–) in the unigene status part (Table 4.3.). The ESTs that are in the same unigene family with other members are listed with their SGN EST identifier codes in the status part (Table 4.3.). SGN ESTs that in fact belong to an EST family but had no other SSR primers designed for them were listed having more than one unigene member but no ESTs in the family (Table 4.3.).

Table 4.2. SSR primers repeat motifs and sequences.

Given Code	Primer Code	Repeat Motif and Number	Forward Sequence	Reverse Sequence	Left TM	Right TM	Size
smSSR01	sgn|E513845	(ATT)21	GTGACTACGGTTTCACTGGT	GATGACGACGACGATAATAGA	55,041	55,346	310
smSSR02	sgn|E514583	(TA)9 (GA)8	ATTGAAAGTTGCTCTGCTTC	GAAAGAGGAGATCCAGGAGT	54,815	54,889	327
smSSR03	sgn|E514601	(TA)9 (GA)8	ATTGAAAGTTGCTCTGCTTC	GATCGAACCCACATCATC	54,815	54,264	145
smSSR04	sgn|E514602	(TA)9 (GA)8	CTCTGCTTCACCTCTGTGTT	CCATGAAAGAGAAGATCGAG	55,529	54,996	320
smSSR05	sgn|E514645	(TA)9 (GA)8	TCTGCTTCACCTCTGTTCTT	AGTAGAGCAACGACGACAAT	55,140	55,047	165
smSSR06	sgn|E514647	(TA)9 (GA)8	TCTGCTTCACCTCTGTTCTT	GAAAGAGGAGATCGAGGAGT	55,140	55,059	315
smSSR07	sgn|E519315	(TAA)20 (CGA)8	TGAATGGAATTACACAAGCA	ATTCTCTAAACCTCAGCCAA	55,129	54,183	240
smSSR08	sgn|E520555	(TAA)20 (CGA)4 - (TAA)22	AATGCAAACAATTATCAGGG	ACAACTCAGCCAGTCGTAAT	55,183	54,877	395
smSSR09	sgn|E513913	(TTTGC)3	CACATGGGAACCTACTTACC	GACGACCATCAAACAAGAAT	54,494	55,020	344
smSSR10	sgn|E513947	(TTTGC)3	AAGCTTCGGAGGAAGATAAG	GGGAGATGGAATAAGTCACA	55,452	54,946	248
smSSR11	sgn|E515884	(AGC)6	AAACAAACTGAAACCCATGT	AAGTTTGCTGTTGCTGCT	54,531	54,589	126
smSSR12	sgn|E516012	(ACCAA)3	AAACAGAAACCAGAGTACTTCA	CAGAAGAAGGTTCAGTTTGC	53,397	55,156	313
smSSR13	sgn|E517027	(AT)9	AGGAATTAACATGGTTCAACA	TTCCTCTTACAACCACATCC	54,667	55,033	263
smSSR14	sgn|E517698	(ATTA)4	ATACCACATCAATCCAAAGC	CATCATCATCTTCACAGTGG	54,991	54,721	241
smSSR15	sgn|E518171	(CCTTT)3	CTGTGGTTGCCTTATCAGTA	TAGTCCAAGGGTTTGATGAC	53,832	55,033	116
smSSR16	sgn|E518867	(AGA)7	AAGAATTTGATGTTGAACCG	CTTTATCAGCCAATTTCTGG	55,217	55,070	390
smSSR17	sgn|E519219	(ATAC)4	TCTTGCCATTTAATTTCCTC	CTATGTCCCTATTATGCCCA	54,553	55,149	115
smSSR18	sgn|E519312	(TAAT)4	TTAGGCATTTGATTTAGCCT	TATGTCCCTAAGCATAACGG	54,376	55,387	342
smSSR19	sgn|E520513	(GAA)6	GAACAATGATTCATCGGATT	AGTTGATGTTGAATTTCCCA	54,868	55,468	241
smSSR20	sgn|E513907	(AGA)5	ACAAGGAAGGACACAAACAC	ATCTAATCACTGTCGCTGCT	55,003	55,131	205
smSSR21	sgn|E514329	(TAC)5	AAGTTTACATGACAGCACCA	TTGCCATCATCAATACCATA	54,132	54,840	249
smSSR22	sgn|E514434	(GCC)5	CTCCGTCAAATTCCTATCAA	GGGAGTCCACATAGAGCATA	55,310	55,154	276
smSSR23	sgn|E515341	(AAG)5	AGAGAAGAAGCCAGCAGAA	TCTGAATCTCCCGAGAAGTA	55,388	54,996	338
smSSR24	sgn|E515827	(TCA)5	GATTTATGGCTTCTGATGGA	TCCTAACCCACTTGATGAAC	55,216	55,033	229
smSSR25	sgn|E515828	(TGA)5	TCCTAACCCACTTGATGAAC	GATTTATGGCTTCTGATGGA	55,033	55,216	228
smSSR26	sgn|E516013	(AAG)5	CAACTTCGATCTTCAATTCC	TCTGAATCTCCCGAGAAGTA	54,836	54,996	373

Table 4.2. SSR primers repeat motifs and sequences (Cont.).

Given Code	Primer Code	Repeat Motif and Number	Forward Sequence	Reverse Sequence	Left TM	Right TM	Size
smSSR27	sgnlE516784	(TGT)5	ATACATTTGAGCCGAGAGTG	TAAATCTGAGAAGGTCGCAT	55,408	55,040	184
smSSR28	sgnlE517072	(TCA)5	CACACTCCTCAGAACTCCAT	CAGCAGTACCTCTTGGTCAT	55,084	55,313	301
smSSR29	sgnlE517168	(CTT)5	TCCACTTCAATTTCCAAGTC	GATCGCTTAGCAGAAGCC	55,167	56,235	188
smSSR30	sgnlE517192	(GAA)5	GATCGCTTAGCAGAAGCC	TCCACTTCAATTTCCAAGTC	56,235	55,167	188
smSSR31	sgnlE517356	(TCC)5	CTTCCTACCCACACTTCATC	TAGGCCGGAGATAGTTGTAA	54,592	55,104	225
smSSR32	sgnlE517618	(GAA)5	CCCACTGATCAGAAGAAGTT	TAGCACACATCCATACCAAA	54,280	54,994	317
smSSR33	sgnlE517678	(TCA)5	TTGCTAGAAATAGCAAAGGG	CGTGGTGTGTATGATGCTTA	54,998	55,550	191
smSSR34	sgnlE517743	(AGA)5	ACAAGGAAGGACACAAACAC	ATCTAATCACTGTCGCTGCT	55,003	55,131	205
smSSR35	sgnlE517795	(ATG)5	CACCACCAAAGAATTCCTAA	TTGCTAGAAATAGCAAAGGG	55,229	54,998	269
smSSR36	sgnlE517835	(CTG)5	AGCACCAGGACAATGAATAC	CCATTTCTTTCTCGACCTTA	55,057	54,620	231
smSSR37	sgnlE517892	(AAG)5	AAAGAAGCTTCCGACGAA	CACTTGTTTCAGCACTTTGA	56,119	54,976	115
smSSR38	sgnlE517980	(GCT)5	GCCATAGATGAAAGGTCAGA	GGATTTATGGACAAGGTGAA	55,288	54,967	211
smSSR39	sgnlE518064	(TCA)5	TTGCTAGAAATAGCAAAGGG	CGTGGTGTGTATGATGCTTA	54,998	55,550	191
smSSR40	sgnlE518161	(AAG)5	TTCTTTGATCTTCAATTCCAA	ATGAAGCTGTTCATGATTCC	55,012	55,105	283
smSSR41	sgnlE518430	(TCA)5	CTCCTCCTGGTAAGGAGTCT	GCAGTATAGAGACGCGAAAT	55,026	54,827	267
smSSR42	sgnlE518630	(CAC)5	ACAGTACACCAGAAACGGAA	GTTACAATGACGGTGGATCT	55,666	54,886	160
smSSR43	sgnlE519141	(GCT)5	ACACCTAAACAACAACCAGG	GGTGGTGTTCAGTCATCTTT	55,073	54,913	333
smSSR44	sgnlE519591	(CCA)5	TGCATTTCATACAGAAACCA	GCAAGGATATCACTGAGCTG	55,129	56,011	233
smSSR45	sgnlE519680	(TTC)5	TTTCTCAACCCAAACTGAAC	GCAGCTCTCGCATAGATAGT	55,252	54,969	172
smSSR46	sgnlE519853	(CAC)5	GGAAACCTTCATTCACTTCA	AGGTCACCGTTACAATTACG	55,167	55,206	272
smSSR47	sgnlE520160	(AGA)5	ACACGATGATCATAAGGGAG	ATCTAATCACTGTCGCTGCT	54,983	55,131	189
smSSR48	sgnlE520161	(GCT)5	GCCATAGATGAAAGGTCAGA	GGATGGAAAGGATAAGAAGG	55,288	55,308	152
smSSR49	sgnlE520192	(ATG)5	TAGTCAACTGCATCACCAGA	CCACTCCCACTACTGTCACT	55,187	55,035	317
smSSR50	sgnlE520238	(ATG)5	TATCAGTCAACTGCATCACC	TGCATTTACGTGAGCTCTAA	54,452	54,788	255

Table 4.3. SGN ESTs and their unigene status.

SGN EST Identifier	Number of Unigene Members	Unigene Status ESTs in the Unigene Family
SGN-E513833	1	-
SGN-E513845	3	SGN-E519315 - 520555
SGN-E513876	6	SGN-E513909 - 513941 - 514099 - 515598 - 520441
SGN-E513913	2	SGN-E513947
SGN-E513907	5	SGN-E517743 - 517980 - 520160 - 520161
SGN-E513915	2	SGN-E513916
SGN-E513959	1	-
SGN-E513954	4	SGN-E517716 - 517947 - 518931
SGN-E514038	1	-
SGN-E514161	1	-
SGN-E514249	3	SGN-E514250 - 514252
SGN-E514275	1	-
SGN-E514279	1	-
SGN-E514329	1	-
SGN-E514364	>1	-
SGN-E514405	1	-
SGN-E514434	>1	-
SGN-E514583	4	SGN-E514602 - 514645 - 514647
SGN-E514589	1	-
SGN-E514599	1	-
SGN-E514601	1	-
SGN-E514796	4	SGN-E516027 - 519337 - 519339
SGN-E514812	2	SGN-E520465
SGN-E514885	2	SGN-E516490
SGN-E515218	2	SGN-E515220
SGN-E515228	1	-
SGN-E515249	>1	-
SGN-E515280	3	SGN-E520089 - 520115
SGN-E515318	1	-
SGN-E515331	>1	-
SGN-E515341	3	SGN-E516013 - 518161
SGN-E515531	1	-
SGN-E515767	1	-

SGN EST Identifier	Number of Unigene Members	Unigene Status ESTs in the Unigene Family
SGN-E517356	1	-
SGN-E517380	>1	-
SGN-E517385	1	-
SGN-E517618	1	-
SGN-E517645	>1	-
SGN-E517670	4	SGN-E517903 - 518057 - 519243
SGN-E517672	1	-
SGN-E517678	3	SGN-E517795 - 518064
SGN-E517698	>1	-
SGN-E517702	1	-
SGN-E517712	>1	-
SGN-E517804	>1	-
SGN-E517835	1	-
SGN-E517846	1	-
SGN-E517892	1	-
SGN-E518073	1	-
SGN-E518083	1	-
SGN-E518171	>1	-
SGN-E518430	1	-
SGN-E518441	2	SGN-E519135
SGN-E518630	>1	-
SGN-E518715	1	-
SGN-E518750	1	-
SGN-E518751	1	-
SGN-E518838	1	-
SGN-E518850	1	-
SGN-E518867	1	-
SGN-E518869	1	-
SGN-E518919	1	-
SGN-E519141	1	-
SGN-E519202	1	-
SGN-E519219	>1	-
SGN-E519312	1	-

Table 4.3. SGN ESTs and their unigene status (Cont.).

SGN EST Identifier	Number of Unigene Members	Unigene Status ESTs in the Unigene Family
SGN-E515782	1	-
SGN-E515827	2	SGN-E515828
SGN-E515838	4	SGN-E515840 - 517812 - 517813
SGN-E515884	1	-
SGN-E515985	>1	-
SGN-E516001	1	-
SGN-E516012	>1	-
SGN-E516287	2	SGN-E516310
SGN-E516412	>1	-
SGN-E516480	1	-
SGN-E516525	2	SGN-E516575
SGN-E516784	1	-
SGN-E516862	>1	-
SGN-E517027	1	-
SGN-E517041	>1	-
SGN-E517072	1	-
SGN-E517074	>1	-
SGN-E517168	2	SGN-E517192
SGN-E517174	1	-
SGN-E517185	1	-
SGN-E517317	1	-
SGN-E517318	1	-

SGN EST Identifier	Number of Unigene Members	Unigene Status ESTs in the Unigene Family
SGN-E519392	>1	-
SGN-E519431	1	-
SGN-E519467	1	-
SGN-E519591	>1	-
SGN-E519680	1	-
SGN-E519737	1	-
SGN-E519853	1	-
SGN-E520021	1	-
SGN-E520010	2	SGN-E520012
SGN-E520049	1	-
SGN-E520056	1	-
SGN-E520120	1	-
SGN-E520121	1	-
SGN-E520147	>1	-
SGN-E520154	2	SGN-E520155
SGN-E520192	2	SGN-E520238
SGN-E520221	2	SGN-E520223
SGN-E520230	1	-
SGN-E520254	1	-
SGN-E520454	>1	-
SGN-E520470	1	-
SGN-E520513	1	-

After design of the primers and their synthesis, they were checked for amplification as described in detail in Section 2.2.2.1. Verification was done using DNA from one accession only (Figure 4.1.). As exemplified in Figure 4.1., 19 smSSR primers were checked for amplification. Except for the eight primers which gave faint bands, all the primers worked successfully with the sample DNA. Generally, single bands were observed for the total 50 smSSR primers with a few exceptions as shown in Figure 4.2.

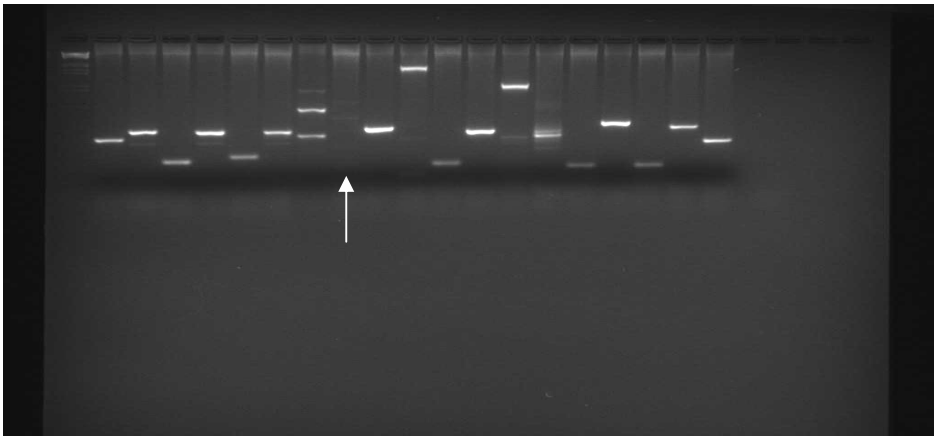

Figure 4.1. Amplification results for 19 smSSR primers checked with a single DNA. A weaker result is indicated by the arrow.

Primers that gave successful amplification were applied to whole DNA samples from individuals of wild species to identify the SSRs that revealed polymorphism. In Figure 4.2., an example of a polymorphic SSR, smSSR10, is shown.

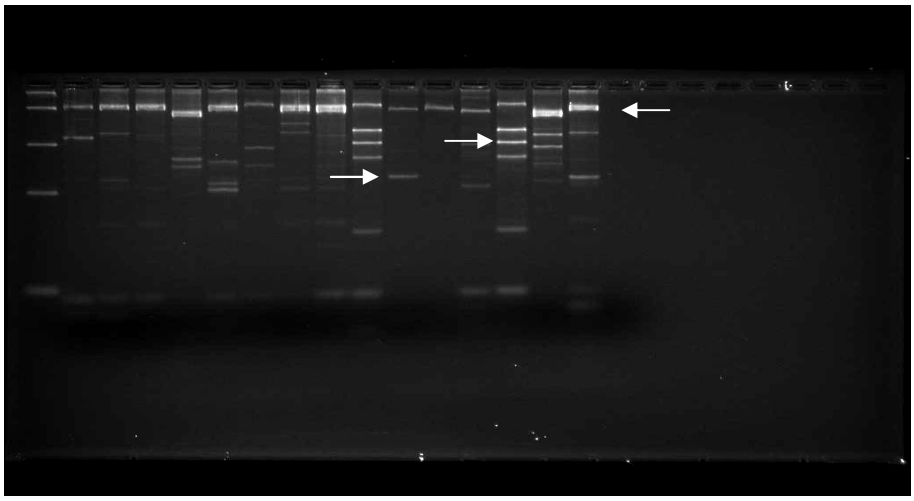

Figure 4.2. Amplification of DNA from 16 wild eggplant accessions with smSSR10. Some of the polymorphic bands are indicated by arrows.

Before the final protocol for SSR analysis was determined, several preliminary experiments were done. These preliminary experiments were based on changes of the PCR conditions, the amount of forward, reverse and M13 primers and dilution amount of PCR products with sample loading solution (SLS).

The challenging part about PCR conditions of the SSR experiments with M13 primer were annealing step cycles (Section 2.2.2.2.). In the study of Schuelke, which was primarily taken as reference, 30 cycles were applied in the experiments (Schuelke 2000). However, the 30 cycle PCR condition was not successful for eggplant samples. For this reason; 25, 26, 27 cycled PCR profiles were tried. Within these, 25 cycles also did not work for the samples. Also, because of giving weaker PCR products in comparison to 27 cycled PCR profile products, 26 cycled PCR profile was not selected.

Another variable was related to the amount of primers used in the experiments. According to the reference study, equal amount of forward, reverse and M13 primers were used: 1.0 µl (Schuelke 2000). However, no satisfactory results were obtained by these amounts. Decreasing the amount of each primer in different combinations, to 0.75 µl while the other 2 were stable, did not give different or better results in the end. Final volumes of primers which gave successful results were equal amount of each primer: 0.75 µl.

The other variant of the SSR experiments with M13 primer was about dilution of PCR products with sample loading solution (SLS) (Beckman Coulter, Inc., Fullerton,

CA, USA). 1:5, 1:10, 1:20, and 1:30 were the different SLS dilution ratios of the samples. 1:20 and 1:30 resulted in weaker peaks or even no peaks while differing from sample to sample. However, 1:5 and 1:10 dilutions gave best results with the 1:10 dilution applied to the PCR products.

In Figure 4.3., an example of SSR sample results analyzed by the CEQ 8800 Genetic Analysis System is shown.

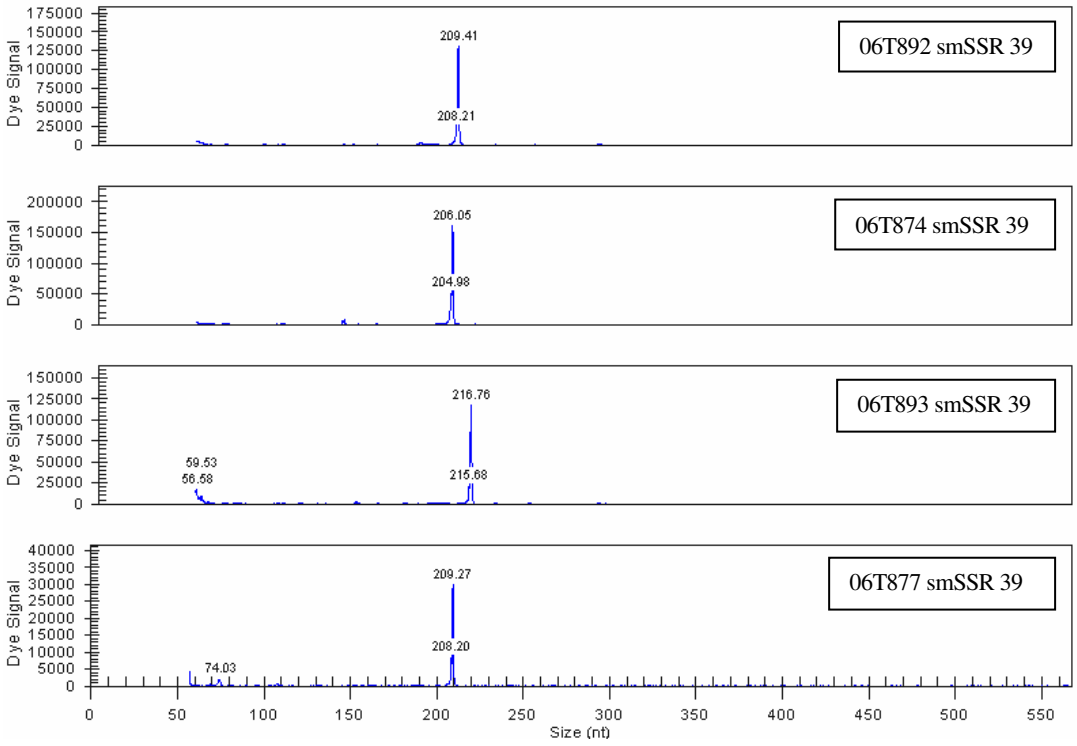

Figure 4.3. An example of SSR studies for 4 samples (Pedigree numbers: 06T892, 06T874, 06T893 and 06T877) with one primer pair (smSSR39) is shown in the figure. Polymorphism can be detected by examining the size of the peaks.

4.1.2. Analysis and Results of the SSR Data

Although a total of 50 primers were designed and applied to the samples (Table 4.2.), data for only 25 primers, listed in Table 4.4., were used for further analysis.

Table 4.4. SSR primers selected for use in the analysis and for drawing dendrogram.

Given Codes		
smSSR 09	smSSR 20	smSSR 39
smSSR 11	smSSR 21	smSSR 40
smSSR 12	smSSR 22	smSSR 42
smSSR 14	smSSR 24	smSSR 44
smSSR 15	smSSR 29	smSSR 45
smSSR 16	smSSR 31	smSSR 46
smSSR 17	smSSR 35	smSSR 47
smSSR 18	smSSR 36	
smSSR 19	smSSR 37	

Seventeen of these selected primers were previously defined as produced from unique ESTs. Four of the 25 SSRs share the same EST family with other SSRs that were not used in the dendrogram analysis. Therefore, for the purposes of this study, they can be considered as unique. The remaining four SSR primers were in two families with two SSRs used from each family. Detailed information is given in Table 4.3.

For the analysis of SSR data, the results of the experiments were grouped based on their presence and absence as 1 and 0. This was achieved using the software of the CEQ 8800 System. These analyzed data were then used to draw a dendrogram. In this study, NTSYS-pc version 2.2j was used to draw the dendrogram of eggplant species. The same matrix and dendrogram parameters were used for the wild eggplants as for the Turkish eggplant analysis. Eigen values are listed in Table 4.5. The derived tree and plots are shown in Figure 4.4., Figure 4.5., Figure 4.6. and Figure 4.7.

Table 4.5. Eigen values representing principal components of the study SSR wild eggplants at three dimensions are listed in order.

	Eigenvalue	Percent	Cumulative
1	21.67843519	46.1243	46.1243
2	2.17524931	4.6282	50.7525
3	2.12832677	4.5284	55.2809

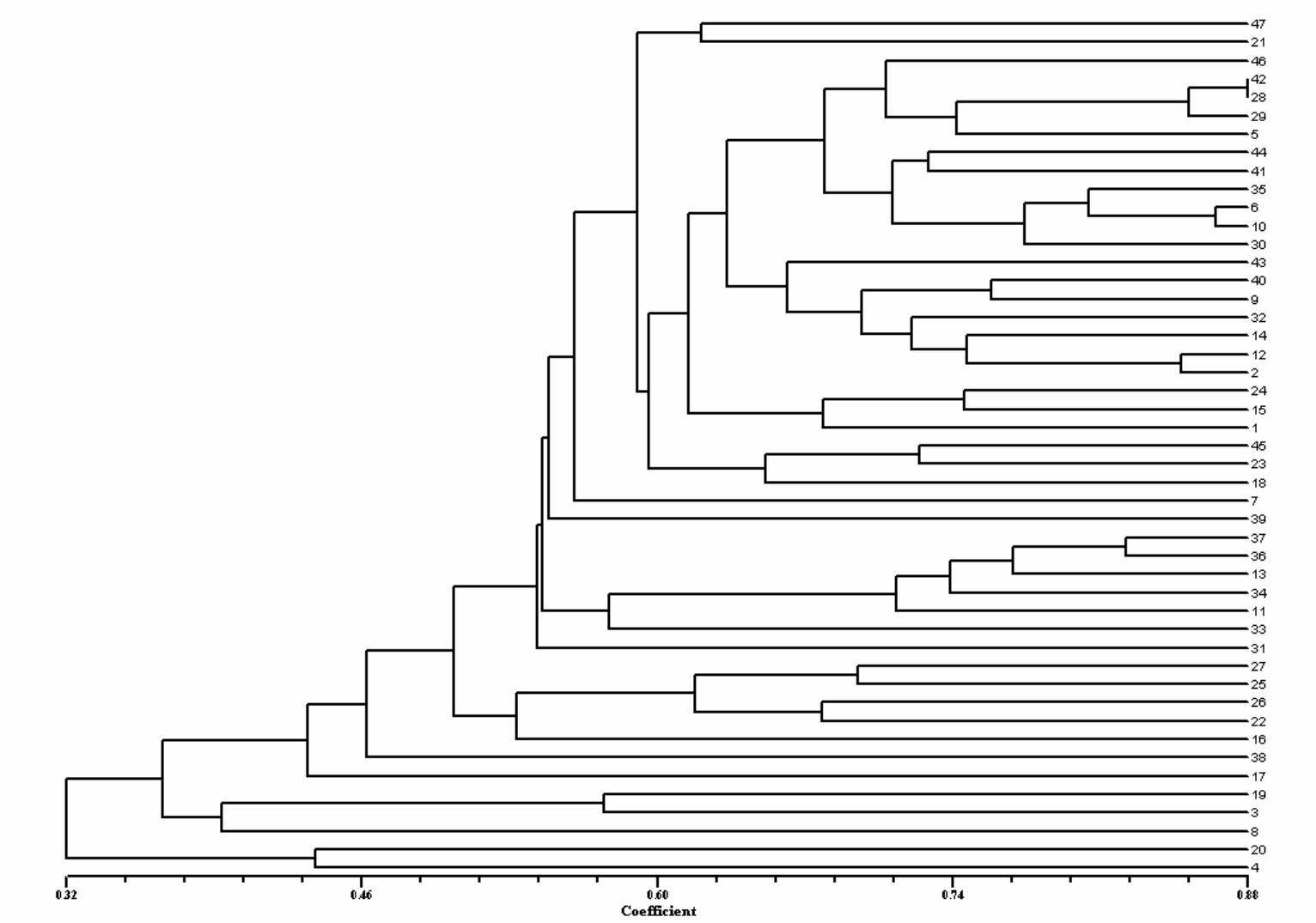

Figure.4.4. Dendrogram showing similarity among *S. melongena* and its wild relatives.

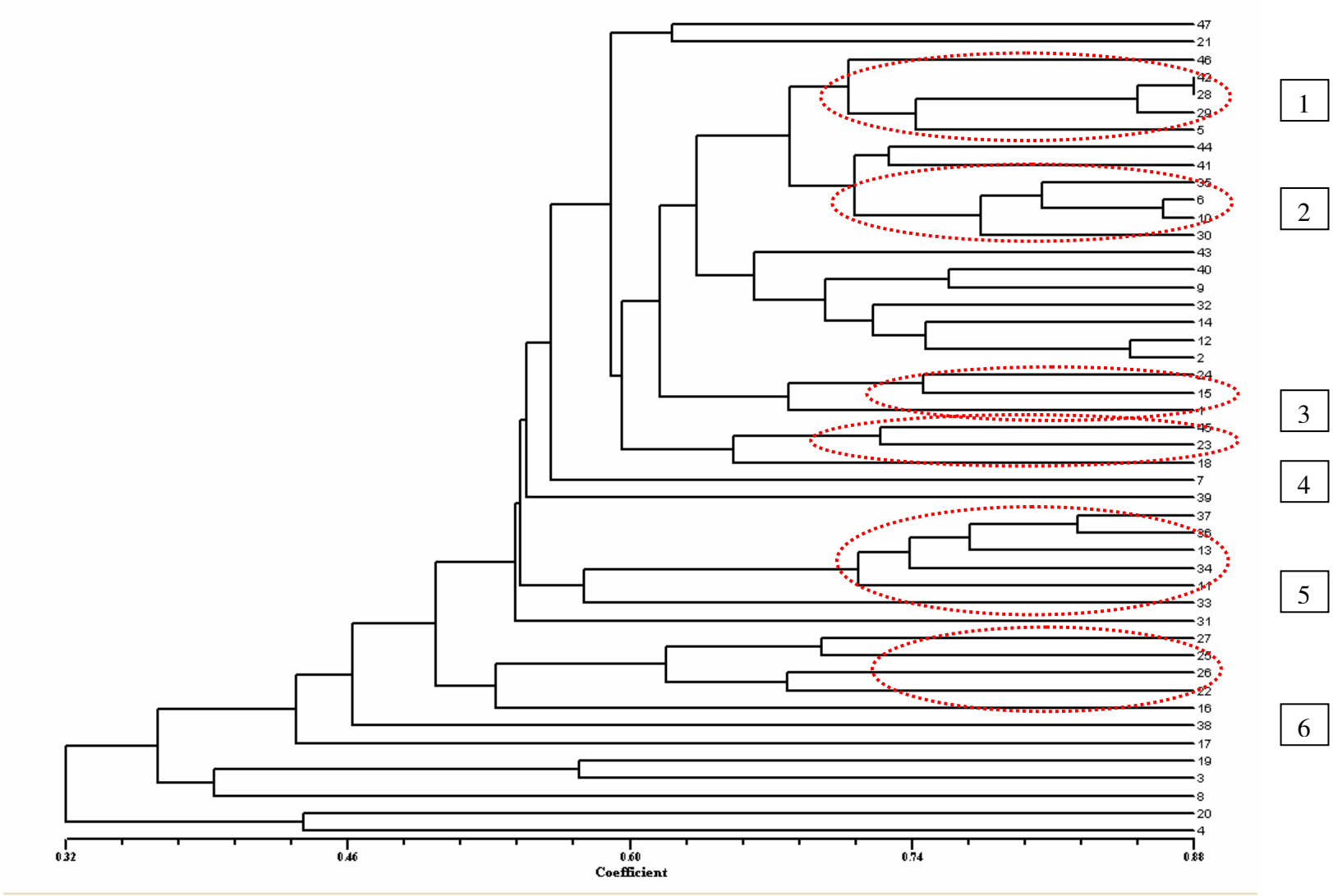

Figure 4.5. Dendrogram showing similarity among *S.melangena* and as wild relatives with clusters indicated.

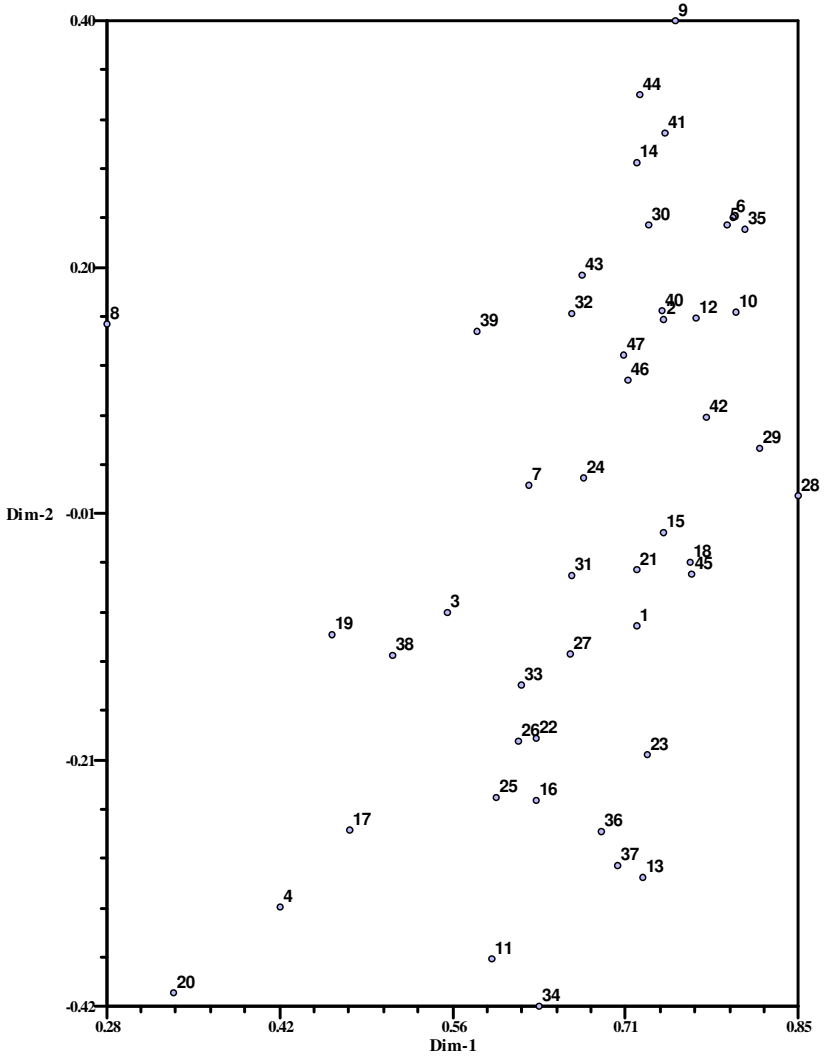

Figure 4.6. Two-dimensional plot of wild eggplants SSR data.

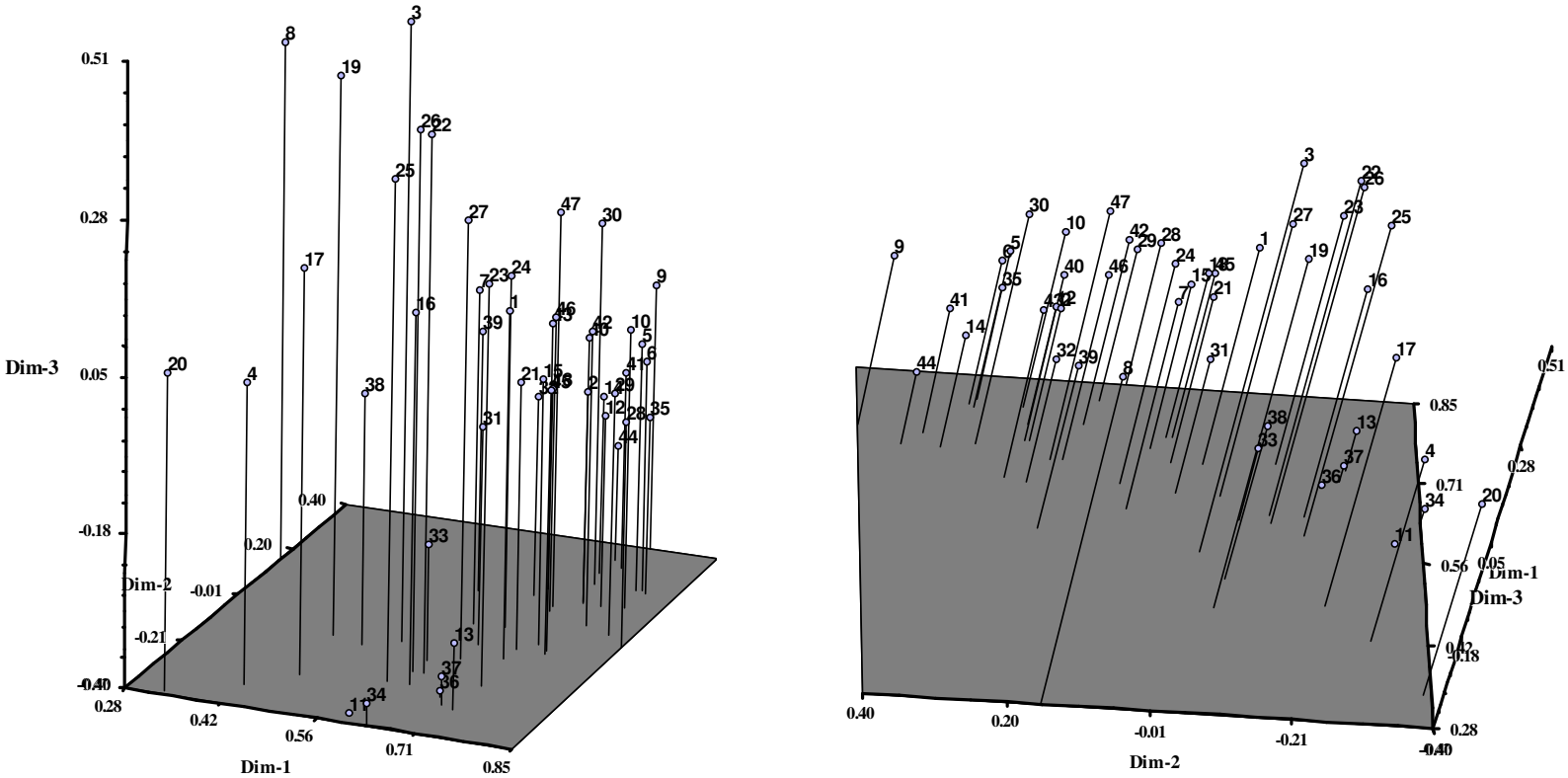

Figure 4.7. Three-dimensional plots of wild eggplants SSR data.

4.2. Discussion

According to the statistical results of SSR analysis, the r value of *Solanum* species' genotypic data was found to be 0.88. This value indicates a good fit as defined in the review of Mohammadi and Prasanna (Mohammadi and Prasanna 2003). That means the correlation between sample genotypic data and dendrogram was found to be high. Other statistical results which were Eigen values explained 46.12% of genotypes for first component analysis (Table 4.5.). With a total value of 55.28%, the 47 different genotypes were explained by the three component axes (Table 4.5.).

The scale of the dendrogram was between 0.32 and 0.88 with a mean value of 0.60 (Figure 4.4.). According to the least similarity value, genotypes were separated into 2 groups with 0.32 coefficient. One group consisted of two samples, genotypes 20 and 4 which were *S. semilistellatum* and *S. viarum*, respectively. All the other samples formed the other group in the large scale and with 0.36 similarity. There were just two genotypes which gave identical results but with a 0.88 correlation coefficient (Figure 4.4.). The second most similar samples were *Solanum melongena* group E and H members (genotype numbers 6 and 10) with a 0.86 cophenetic correlation value (Figure 4.5., Group-2).

In terms of the samples' clustering, the dendrogram was compatible with the expected results and with the previous studies. For example, *Solanum incanum*, *Solanum melongena*, *Solanum macrocarpon* members all formed separate clusters (Figure 4.5., Group-1, Figure 4.5., Group-2 and Figure 4.5., Group-5). An important grouping was the one which included *S. incanum* and *S. melongena* clusters with a 0.67 correlation value (Figure 4.5., Group-1 and Figure 4.5., Group-2). This result agreed with the interpretation of the relative closeness of these 2 species which was reported in several studies (Sakata and Lester 1994, Karihaloo et al. 1995, Furini and Wunder 2004, Mace et al. 1999b, Singh et al. 2006). One point which was quite noteworthy was the separation of different group members of *S. incanum* into different clusters (Figure 4.5., Group-1, Figure 4.5., Group-4 and Figure 4.5., Group-6). This result showed the diversity among different groups and within the same species. Another important grouping was *S. macrocarpon* which also included *S. dasyphyllum* (Figure 4.5., Group-5). In the recent phylogenetic study of the Leptostemonum clade by Levin et al., these two species were included in the same clade with a very high similarity. In the same study and in our SSR analysis of eggplant and its

wild relatives, *S. campylacanthum* and *S. incanum* were grouped together (Levin et al. 2006, Figure 4.5., Group-3).

As a marker system, SSR is accepted as a valuable molecular analysis tool (Powell et al. 1996). However, due to their conservative nature and expected low level of polymorphism, the usefulness of SSRs derived from ESTs for clustering analysis has been questioned (Rudd 2003 and Varshney et al. 2005). However, the overall results of the present study were satisfactory in terms of their statistical values and concordance with previously published data. The correlation coefficient 0.88 for the highest similarity between genotypes and the least 0.32 exhibited a good separation from a conserved region of the genome. However, increasing data and sample numbers may increase the accuracy of the clustering results.

CHAPTER 5

CONCLUSION AND FUTURE PERSPECTIVE

In this thesis, the research was separated into two parts. For both parts, the materials used in the experiments were DNAs from greenhouse-grown samples. The general aim for each part was to reveal genetic differences or similarities between the plant materials which were members of different accessions or species.

For the first part, materials were Turkish eggplants all of which belonged to different accessions of *Solanum melongena*. To reveal genetic diversity among Turkish eggplants and three outgroups, which were *Solanum linnaeanum*, *Solanum aethiopicum* and *Solanum macrocarpon*, the AFLP marker system was used because it has been proven to be an efficient molecular tool to reveal genetic diversity not only in other systems but also in the Solanaceae family (Mace et al. 1999b). Another reason for this method's selection was related with its potential to produce a high amount of genetic data. When it was considered that the possible genetic diversity at the intraspecies level was low in eggplant, it was determined that as much data as possible should be obtained and used in the analysis (Daunay et al. 2001). According to the results of the AFLP experiments in Turkish eggplants, this idea was shown to be correct. Statistical results were quite satisfactory and a dendrogram was drawn that was concordant with the data in these analyses with an r value of 0.97.

In the second part, the aim was to find genetic diversity between eggplant and its wild relatives. From 20 different species, three species clustered into individual groups, and a total of 47 different genotypes were tested with SSR marker system. The reason for the selection of this marker system was related with its highly polymorphic nature, easiness to study and its reliability. However, the design of that marker system is laborious work. For that reason, SSR primers for this study were designed from eggplant ESTs which were publicly available on SOL Genomics Network (http://sgn.cornell.edu). Due to the fact that the source of the primers was ESTs, these SSRs were expected to amplify more conserved regions in the genome. That led to a relatively low level of polymorphism and diversity within the materials. However, analysis of the results and statistical values obtained were in the good scale with a correlation coefficient value of 0.88. The concordance within the results and clusters in

the dendrogram also concluded it as a reliable work. However, increase in the amount of data may give a better separation of the samples and a better statistical result.

The overall importance of these studies is related with taxonomic issues and also breeding and preservation attempts. Because of confusion about eggplant or in the Solanaceae family, genotypic data serves as a powerful means and data source in determination of similarity between individuals. This similarity can be at any level in the organization of organisms. In addition to this, for the preservation and maintenance of crop plants which are important for human health and diet is an important subject. This is basically related with diversity conservation. Using the results of this study, the most diverse species or accessions can be selected for preservation without any time and money lost and species or accessions can be identified accurately in taxonomy. For future work, these new SSR primers can be used and integrated into the mapping studies of eggplant.

REFERENCES

Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R.G., Anderberg, A.A., Källersjö, M., and Barkhordarian, E., 2002. "Phylogenetics of asterids based on 3 conding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels", *Molecular Phylogenetics and Evolution*, Vol.24, pp.274-301.

Choudhury, B., 1995. *Eggplant* (Evolution in Crop Plants edited by Smartt, J., and Simmonds, N.W.), pp.464-465.

Collonnier, C., Fock, I., Kashyap, V., Rotino, G.L., Daunay, M.C., Lian, Y., Mariska, I.K., Rajam, M.V., Servaes, A., Ducreux, G., & Sihachakr, D., 2001. "Applications of Biotechnology in eggplant", *Plant Cell, Tissue and Organ Culture*, Vol.65, pp.91-107.

Daunay, M.C., Lester, R.N., Dalmon, A., Ferri, M., Kapilima, W., Poveda Aguilar, M.M., and Jullian, E., 1998. *The use of Wild Genetic Resources for Eggplant (Solanum melongena) Breeding. II: Crossibility and Fertility of Interspecific Hybrids* (X. EUCARPIA Meeting on Genetics and Breeding of Capsicum & Eggplant), pp.19-24.

Daunay, M.-C., Dalmon, A., & Lester, R.N., 1999. *Management of a collection of Solanum species for eggplant (Solanum melongena) breeding purposes* (Solanaceae IV edited by Nee, M., Symon, D.E., Lester, R.N., & Jessop, J.P., Royal Botanic Gardens, Kew), pp. 369-383.

Daunay, M.C., Lester, R.N., Gebhardt, C.H., Hennart, J.W., Jahn, M., Frary, A. and Doganlar, S., 2001. *Genetic Resources of Eggplant (Solanum melongena L.) and Allied Species: A New Challenge for Molecular Geneticists and Eggplant Breeders* (*Solanaceae V* edited by Van Den Berg, R.G., Barendse, G.W. and Mariani, Nijmegen University Press, Nijmegen, The Netherlands), pp. 251–274 .

Dice, L.R. 1945. "Measures of the amount of ecological association between species", *Ecology*, Vol.26, pp.297-302.

Doganlar, S., Frary, A., Daunay, C., Lester, N., and Tanskley, S., 2002a. "A Comparative Genetic Linkage Map of Eggplant *(Solanum melongena)* and Its Implication For Genome Evoluation in the Solanaceae", *Genetics*, Vol.161, pp.1697-1711.

Doganlar, S., Frary, A., Daunay, M.C., Lester, R.N., and Tanksley, S.D., 2002b. "Conservation of Gene Function in the Solanaceae as Revealed by Comparative Mapping of Domestication Traits in Eggplant", *Genetics*, Vol.161, pp.1713-1726.

Economic Research Service, USDA, 2006. "Commodity Highlight: Eggplant", *Vegetables and Melons Outlook*, VGS-318, pp.23-27.

Frary, A., Doğanlar, S., 2003. "Comparative Genetics of Crop Plant Domestication and Evolution", *Turk J Agric For*, Vol.27, pp.59-69.

Frary, A., Doganlar, S., Daunay, M.C., Tanksley, S.D., 2003. "QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species", *Theor Appl Genet*, Vol.107, pp.359-370.

Frary, A., Doganlar, S. and Daunay, M.C., 2007. "Eggplant", *Genome Mapping and Molecular Breeding in Plants*, Vol.5, pp.231-257.

Furini, A., Wunder, J., 2004. "Analysis of eggplant (*Solanum melongena*)-related germplasm: morphological and AFLP data contribute to phylogenetic interpretations and germplasm utilization", *Theor Appl Genet*, Vol.108, pp.197-208.

Grandillo, S., Fulton, T.M., 2002. *Approaches to gene mapping* (Molecular Plant Biology Volume One edited by Gilmartin, P.M., and Bowler, C.), pp. 101-137.

Gulsen, O., Karagul, S. & Abak, K., 2007. "Diversity and relationships among Turkish okra germplasm by SRAP and phenotypic marker polymorphism", *Biologia Bratislava*, Vol.62/1, pp.41-45.

Henry, R.J. 1999. *Molecular Techniques for the Identification of Plants* (Applied Plant technology edited by V.L. Chopra, V.L., Malik, V.S., & Bhat, S.R.), pp.269-283.

Invitrogen, 2003. *AFLP Analysis System I, AFLP Starter primer Kit, Version B, User Manual*.

Isshiki, S., Okuba, H., Fujieda, K., 1994a. "Phylogeny of eggplant and related Solanum species constructed by allozyme variation", *Scientia Horticulturae*, Vol.59, pp.171-176.

Isshiki, S., Okuba, H., & Fujieda, K., 1994b. "Genetic control of isozymes in eggplant and its wild relatives", *Euphytica*, Vol.80, pp.145-150.

Isshiki, S., Okuba, H., Oda, N., and Fujieda, K., 1994c. "Isozyme Variation in Eggplant (*Solanum melongena* L.), *J. Japan. Soc. Hort. Sci*., Vol.63(1), pp.115-120.

Isshiki, S., Uchiyama, T., Tashiro, Y., & Miyazaki, S., 1998. "RFLP analysis of a PCR amplified region of chloroplast DNA in eggplant and related *Solanum* species", *Euphytica*, Vol.102, pp.295-299.

Isshiki, S., Suzuki, S., and Yamashita, K, 2003. "RFLP analysis of mitochondrial DNA in eggplant and related *Solanum* species", *Genetic Resources and Crop Evolution*, Vol.50, pp.133-137.

Jones, N., Ougham, H., and Thomas, H., 1997. "Markers and mapping: we are all geneticists now", *New Phytol*., Vol.137, pp.165-177.

Karihaloo, J.L., Gottlieb, L.D., 1995. "Allozyme variation in the eggplant, *Solanum melongena* L. (Solanaceae)", *Theor Appl Genet*, Vol.90, pp.578-583.

Karihaloo, J.L., Brauner, S., Gottlieb, L.D., 1995. "Random amplified polymorphic DNA variation in the eggplant", *Theor Appl Genet*, Vol.90, pp.767-770.

Kashyap, V., Kumar, S.V., Collonnier, C., Fusari, F., Haicour, R., Rotino, G.L., Sihachakr, D., Rajam, M.V., 2003. "Biotechnology of eggplant", *Scientia Horticulturae*, Vol.97, pp.1-25.

Kaur, M., Singh, S., and Karihaloo, J.L., 2004. "Diversity of Enzyme Electrophoretic Patterns in the Eggplant Complex", *J. Plant Biochemistry & Biotechnology*, Vol.13, pp.69-72.

Knapp, S., Bohs, L., Nee, M., Spooner, D.M., 2004. "Solanaceae – a model for linking genomics with biodiversity", *Comp Funct Genom*, Vol.5, pp.285-291.

Krane, D.E. and Raymer, M.L., 2003. *Genomics and Gene Recognition* (Fundamental Concepts of Bioinformatics by Pearson Education/Benjamin Cummings), pp.117-153.

Lawande, K.E., and Chavan, J.K., 1998. *Eggplant (Brinjal)* (Handbook of Vegetable Science and Technology edited by Salunkhe, D.K., Kadam, S.S.), pp.225-243

Lester, R.N., 1997. "Typification of nineteen names of African *Solanum* species described by A. Richard and others, including *S. campylacanthum* and *S. panduriforme*", *Botanical Journal of the Linnean Society*, Vol.125, pp.273-293.

Lester, R.N., 1998. *Genetic Resources of Capsicum and Eggplants* (X. EUCARPIA Meeting on Genetics and Breeding of Capsicum & Eggplant), pp.25-30.

Lester, R.N. and Daunay, M.C., 2003. *Diversity of African Vegetable Solanum species and its implications for a better understanding of plant domestication* (Rudolf Mansfeld edited by Knüpffer, H. and Ochsmann, J.), Vol.22, pp.137-152.

Levin, R.A., Myers, N.R., and Bohs, L., 2006. "Phylogenetic relationships among the "Spiny Solanums" (*Solanum* subgenus *Leptostemonum*, Solanaceae)", *American Journal of Botany*, Vol.93(1), pp.157-169.

Li, W.-H., 1997. *Molecular Phylogenetics* (Methods in Molecular Evolution by Wen-Hsiung Li), pp.99-147.

Mace, E.S., Gebhardt, C.G., Lester, R.N, 1999a. "AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae)", *Theor Appl Genet*, Vol.99, pp.634-641.

Mace, E.S, Lester, R.N., Gebhardt, C.G., 1999b. "AFLP analysis of genetic relationships among the cultivated eggplant, Solanum melongena L., and wild relatives (Solanaceae)", *Theor Appl Genet*, Vol.99, pp.626-633.

Mantel, N., 1967. “The detection of disease clustering and a generalized regression approach”, *Cancer Res*., Vol.27, pp.209-220.

Martin, W., Deusch, O., Stawski, N., Grünheit, N., and Goremykin, V., 2005. “Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution”, *TRENDS in Plant Science*, Vol.10, No.5, pp.203-209.

Mohammadi, S.A. and Prasanna, B.M., 2003. “Analysis of Genetic Diversity in Crop Plants-Salient Statistical Tools and Considerartions”, *Crop Sci*., Vol.43, pp.1235-1248.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T.G., Yano, M., Bhatia, C.R., and Sasaki, T., 1997. “Genome mapping, molecular markers and marker-assisted selection in crop plants”, *Molecular Breeding*, Vol.3, pp.87-103.

Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., Lin, C., Wright, M.H., Ahrens, R., Wang, Y., Herbst, E.V., Keyder, E.R., Menda, N., Zamir, D., and Tanksley, S.D., 2005. “The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond”, *Plant Physiology*, Vol.138, pp. 1310-1317.

National Research Council, 2006. *Eggplant (Garden Egg)* (Lost Crops of Africa Volume II: Vegetables), pp.136-153.

Nothmann, J., 1986. *Eggplant* (CRC Handbook of Fruit Set and Development edited by Monselise, S.P.), pp. 145-152.

Nunome, T., Suwabe, K., Ohyama, A., and Fukuoka, H., 2003a. “Characterization of Trinucleotide Microsatellites in Eggplant”, *Breeding Science*, Vol.53, pp.77-83.

Nunome, T., Suwabe, K., Iketani, H., and Hirai, M., 2003b. “Identification and characterization of microsatellites in eggplant”, *Plant Breeding*, Vol.122, pp. 256-262.

Olmstead, R.G., Sweere, J.A., Spangler, R.E., Bohs, L., & Palmer, J.D., 1999. *Phylogeny and Provisional Classification of the Solanaceae Based on Chloroplast DNA* (Solanaceae IV edited by Nee, M., Symon, D.E., Jessop, J.P., Hawkes, J.P.), pp.111-137. Royal Botanic Gardens, Kew.

Powell, W., Machray, G.C., Provan, J., 1996. “Polymorphism revealed by simple sequence repeats”, *TRENDS in Plant Science*, Vol.1, No.1, pp.215-222.

Pramod, S.N., Venkatesh, Y.P., 2004. “Allergy to eggplant (*Solanum melongena*)”, *J Allergy Clin Immunol*, Vol.113, No.1, pp.171-173. [Letters to the editor].

Ranade, S.A., 2003. *Application of molecular sequences and DNA profiles for taxonomy and phylogenetics of plants* (Molecular Insight in Plant Biology edited by Nath, P., Mattoo, A.K., Ranade, S.A., Weil, J.H.), pp.49-66.

Raven, P.H., Evert, R.F., Eichhorn, S.E., 1999a. *The Process of Evolution* (Biology of Plants, 6th Edition, Freeman & Worth, New York, NY), pp.235-259.

Raven, P.H., Evert, R.F., Eichhorn, S.E., 1999b. *Systematics: The Science of Biological Diversity* (Biology of Plants, 6th Edition, Freeman & Worth, New York, NY), pp.261-280.

Rudd, S., 2003. "Expressed sequence tags: alternative or complement to whole genome sequences",*TRENDS in Plant Science*, Vol.8, No.7, pp.321-329.

Salamini, F., Heun, M., Brandolini, A., Özkan, H., and Wunder, J., 2004. "Comment on "AFLP data and the origins of domesticated crops"", *Genome*, Vol.47, pp.615-620.

Sakata, Y., Nishio, T., & Matthews, P.J., 1991. "Chloroplast DNA analysis of eggplant (*Solanum melongena*) and related species for their taxonomic affinity", *Euphytica*, Vol.55, pp.21-26.

Sakata, Y., Lester, R.N., 1994. "Chloroplast DNA diversity in eggplant (*Solanum melongena*) and its related species S. *incanum* and S. *marginatum*", *Euphytica*, Vol.80, pp.1-4.

Sakata, Y., Lester, R.N., 1997. "Chloroplast diversity in brinjal eggplant (*Solanum melongena* L.) and related species", *Euphytica*, Vol.97, pp.295-301.

Schuelke, M. 2000. "An economic method for the fluorescent labeling of PCR fragments", *Nature Biotechnology*, Vol.18, pp.233-234.

Singh, A.K., Singh, M., Singh, A.K., Singh, R., Kumar, S., and Kalloo, G., 2006. "Genetic Diversity within the genus *Solanum* (Solanaceae) as revealed by RAPD markers", *CURRENT SCIENCE*, Vol.90, No.5, pp.711-716.

Staub, J.E., Serquen, F.C., 1996. "Genetic Markers, Map Construction, and Their Application in Plant Breeding", *HORTSCIENCE*, Vol.31(5), pp.729-740.

Tanksley, S.D., McCouch, S.R., 1997. "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild", *SCIENCE*, Vol.277, pp.1063-1066.

The Angiosperm Phylogeny Group, 2003. "An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II", *Botanical Journal of the Linnean Society*, Vol.141, pp.399-436.

Varshney, R.K., Graner, A., and Sorrells, M.E., 2005. "Genic microsatellite markers in plants: features and applications", *TRENDS in Biotechnology*, Vol.23, No.1, pp.48-55.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M., 1995. "AFLP: a new technique for DNA fingerprinting", *Nucleic Acids Research*, Vol.23, No.21, pp.4407-4414.

Westhead, D.R., Parish, J.H. and Twyman, R.M., 2002. *Phylogenetics, Cladistics and Ontology* (Instant Notes of Bioinformatics edited by Homes, B.D.), pp.97-108.

WEB_1 2007. Wikipedia 03/18/2007. http://en.wikipedia.org/wiki/Flowering_plant

WEB_2 2007. Wikipedia 04/10/2007. http://en.wikipedia.org/wiki/Genetic_diversity

WEB_3 2007. Wikipedia 03/27/2007. http://en.wikipedia.org/wiki/Cladistics

WEB_4 2007. Sol Genomics Network 03/17/2007. http://www.sgn.cornell.edu/about/ about_solanaceae.pl

WEB_5 2007. Wikipedia 03/17/2007. http://en.wikipedia.org/wiki/Solanaceae

WEB_6 2007. AFLP 10/19/2006. http://biologi.uio.no/fellesavdelinger/DNA-KAFFE/ Kaffe_Resources/AFLP.html

WEB_7 2007. NCBI A Science Primer 03/17/2007. http://www.ncbi.nlm.nih.gov/ About/primer/phylo.html

WEB_8 2007. Wikipedia 03/18/2007. http://en.wikipedia.org/wiki/Eggplant

WEB_9 2007. Wikipedia 04/10/2007. http://en.wikipedia.org/wiki/ Molecular_systematics

WEB_10 2007. Solanaceae Source 03/17/2007. http://www.nhm.ac.uk/research-curation/projects/solanaceaesource/solanaceae/

WEB_11 2007. Solanaceae Source 03/17/2007. http://www.nhm.ac.uk/research-curation/projects/solanaceaesource/solanum/

WEB_12 2007. Aubergina (*Solanum melongena* L.) 04/10/2007. http://www.biologie.uni-hamburg.de/b-online/schaugarten/SolanummelongenaL/ SolanummelongenaL.jpg

WEB_13 2007. SOL Genomics Network 03/17/2007. http://sgn.cornell.edu

WEB_14 2007. Model Organism Tree 04/10/2007. http://www.mobot.org/ mobot/research/APweb/trees/modeltreemap.html

WEB_15 2006. PBC Public Databases 10/19/2006. http://hornbill.cspp.latrobe.edu.au/ cgi-binpub/ssrprimer/indexssr.pl