CHARACTERIZATION OF WORLD SPINACH GENETIC COLLECTION BY USING MOLECULAR MARKERS

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

MASTER OF SCIENCE

in Molecular Biology and Genetics

by
Mehmet GÖKTAY

December 2015
İZMİR

We approve the thesis of **Mehmet GÖKTAY**

Examining Committee Members:

_____
Prof. Dr. Anne FRARY
Department of Molecular Biology and Genetics, İzmir Institute of Technology

_____
Assoc. Prof. Dr. Jens ALLMER
Department of Molecular Biology and Genetics, İzmir Institute of Technology

_____
Assoc. Prof. Dr. Fevziye Çelebi TOPRAK
Department of Biology, Pamukkale University

25 December 2015

_____
Prof. Dr. Anne FRARY
Supervisor, Department of Molecular Biology and Genetics
İzmir Institute of Technology

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

_____
Prof. Dr. Bilge KARAÇALI
Dean of the Graduate School of Engineering and Sciences

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors Prof. Dr. Anne FRARY, Prof. Dr. Sami DOĞANLAR and Assoc. Prof. Dr. Jens ALLMER for their guidance, patience and unique support and contributions during my master period.

I would like to thank to all my friends in JLab and The Doganlar/Frary Lab for their kindness and help.

I would like to thank to TUBITAK-(113E326) project providing financial support through my master degree.

Finally, I would like to give special thanks and love to my family for their excellent emotional support, understanding, love and patience during all my stressful and hard times.

ABSTRACT

CHARACTERIZATION OF WORLD SPINACH GENETIC COLLECTION BY USING MOLECULAR MARKERS

Spinach (*Spinacia oleracea L.*) belongs to the *Amaranthaceae* family and is a popular and nutritious vegetable. It is believed that this green leafy plant originated from Persia which is now modern Iran and neighboring countries. In this study we screened 176 spinach world collection germplasm accessions with 15 known SSR markers. The markers produced 58 bands with 57 identified as polymorphic. PIC values of the markers ranged between 0.01 and 0.44. Average PIC value was 0.28. Based on allele analysis with Darwin5 and STRUCTURE tools, 176 individual were clustered into three groups. The first cluster mostly consisted of accessions from Europe and USA and contained 69 samples. The second cluster mostly consisted of lines from Asia and neighboring countries and contained 89 samples. The third cluster did not represent any characteristic according to geographic region, thus it was called an intermixed cluster which contained 18 samples. The maximum genetic dissimilarity of spinach accessions was 0.551 and minimum was 0.019. Average genetic dissimilarity was 0.217. Moreover we sequenced *S. oleracea L.* cv. Universal nuclear genome via IIIumina MiSeq technology and genome assembly was performed to develop new spinach–specific SSR markers. As a result, 3853 SSRs were identified in the assembled genome and we successfully designed 3275 primer pairs for these identified SSR motifs. These newly developed SSR markers will be helpful to scientists who are interested in spinach genome diversity and breeding.

ÖZET

DÜNYA ISPANAK GENETİK KOLLEKSİYONLARININ MOLEKÜLER MARKÖRLERLE KARAKTERİZASYONU

Ispanak (*Spinacia oleracea L.*) *Amaranthaceae* ailesine ait olan popüler ve besleyici bir sebzedir. Bu yeşil yapraklı bitkinin Persia (İran ve komşu ülkelerden) kökenlendiğine inanılır. Bu çalışmada Dünya Koleksiyonuna ait 176 ıspanak çeşidi, bilinen 15 SSR markır ile taranmıştır. Markırlar 57'si polimorfik olmak üzere 58 bant üretmiştir. Markırların PIC değerleri 0.01 ile 0.44 arasında değişmektedir. Ortalama PIC değeri ise 0.28'dir. Darwin5 ve STRUCTURE programlarının allel analizlerine göre 176 birey üç grupta toplanmıştır. Birinci grupta çoğunlukla Amerika ve Avrupa çeşitleri olmak üzere 69 örnek birey vardır. İkinci grup ise çoğunlukla Asya ve komşu ülkelerden olmak üzere 89 birey vardır. Üçüncü grup coğrafi alana göre hiç bir karakteristik göstermemiştr, bu nedenle 18 birey içeren bu grup intermixed grup olarak isimlendirilmiştir. Ispanak çeşitlerinin genetik farklılığı maksimum 0.051 ve minum 0.019'dur. Ortalama genetik farklılık ise 0.217'dir. Ayrica *S. oleracea L*. cv. Universal nuklear genomu IIIumina MiSeq teknolojisi kullanılarak dizilenip, ıspanak-spesifik SSR markır dizaynı için genom montajı yapıldı. Sonuç olarak, montajlanan genomda 3852 SSRs tespit edildi ve 3275 primer çifti, tespit edilen SSR motiflerine başarılı olarak dizayn edildi. Bu yeni geliştirilen SSR markırlar, ıspanak genom çeşitliliği ve ıslah konusu ile ilgilenen bilim adamlarına yardımcı olacaktır.

TABLE OF CONTENTS

LIST OF THE FIGURES

Figure **Page**

LIST OF THE TABLES

CHAPTER 1

INTRODUCTION

1.1. General Information about *Spinacia oleracea L.*

Spinach (*Spinacia oleracea L.*) is a leafy, flowering, edible, cross-pollinated plant which is in the *Amaranthaceae* family. It is a dioecious plant and both female and male plant leaves are alternate and simple. Spinach is a winter season plant and matures very quickly. Seed germination temperature between 7 and 25 °C is optimal. Although this green leafy plant grows from 5 to 30 °C degrees, it grows fastest at cool temperatures between 15 and 18 °C [1] [2]. Due to its highly nutritious nature, spinach is cultivated in most regions of the world [3].The chromosome number of spinach is 2n=12 and its estimated genome size is around 980 MB [3] [4]. Although spinach is not a model organism, its diploidy and average sized genome make it easy to study.

1.2. Origin and Production of Spinach

Spinach originated in Persia (modern Iran) and neighboring countries. It is probable that spinach was unfamiliar to other countries except its native land until the beginning of the Christian Era. Even later, it was unknown to Greeks and Romans. The first sign that spinach had spread to other countries was evidence of the crop in China in 657 A.D. According to old documents it was conveyed to Spain about 1100 A.D. from North Africa by the Moors [4] [5].Spinach was unknown in 13th century Germany but by the 14th century, it was commonly grown in Europe. The question of when spinach was transported to America remains unanswered, but it was probably in early colonial times [5] [6].

According to FAO statistics (Table 1), China is the largest producer of spinach worldwide. The United States of America, Japan and Turkey follow China in terms of production (Food and Agriculture Organization of the United Nations, 2013).

Table 1. Top 4 producers of spinach in 2013

Producers	**Production (tons)**
China	21.067.800
United States of America	336.200
Japan	258.427
Turkey	220.274
Worldwide Production	23.231.898

1.3. Nutrition of Spinach and Health Benefits

Spinach production and consumption may be linked with its very rich nutritional value. Spinach is a tremendous source of flavonoids with a number of different flavonoids identified by independent studies.[2] In addition, spinach contains high concentrations of vitamins A, E, C and K, folic acid and oxalic acid. For example, a 100 gram serving of fresh spinach provides 20% of the recommended daily intake of ascorbic acid (Vitamin C), B-carotene (provitamin A), lutein, folate (vitamin B_9), phylloquione (vitamin K_1), and alpha-tocopherol (vitamin E) [7]. This vegetable also contains carotenoids like lutein, violaxanthin and 9'-(Z)-neoxhantin. Numerous minerals are found in spinach including magnesium, manganese, calcium, phosphorus, iron, zinc, copper and potash [2].

From the pharmacological perspective, spinach has antioxidant, antiproliferative, antiinflammatory, antihistaminic and CNS depressant characteristics and is also hepatoprotective and protective against gamma radiation [2].

1.4. Genetic Diversity

Genetic diversity is the sum of all the genetic characteristic of a species. Diversity is determined by a pool of genes which affects the ability of the species to adapt to new environments. If this pool holds comparatively more variety than other species, it is more likely to survive in new or changing environmental conditions. In plant science, species diversity is used by breeders to develop new varieties which are resistant to salinity, drought, heat, pathogen attack or to develop new varieties which provide better quality and/or high yield. Plant diversity can be studied using various types of marker systems.

1.5. Genetic Analysis with Marker Systems

Genetic markers are found throughout the genome and are used as flags for genome analysis experiments. There are two types of genetic markers: morphological markers and molecular (DNA) markers.

1.5.1. Morphological Markers

The inheritance of morphological markers can be traced visually without sophisticated biochemical or molecular techniques [8]. Examples of plant morphological markers include height or weight of the plant and color of its fruits or leaves. Morphological traits which are controlled by a single locus can be used as genetic markers but the gene's expression level must be reproducible in a wide-range of environments [8]. Unfortunately heterozygous and homozygous individuals will have the same phenotype for dominant traits, therefore in certain cases, morphological markers are not applicable. The number of morphological markers in a given individual is also inadequate for most types of diversity studies [8].

1.5.2. Molecular Markers

Molecular markers include both DNA markers and biochemical markers. Unlike biochemical markers which represent polymorphism at the protein level, DNA markers represent polymorphism at the DNA level[8]. DNA markers are very popular among breeders because these markers help them to select plants using DNA polymorphisms which can be used as flags for traits. Unlike morphological markers, these markers are stable, easily detectable and are not affected by environmental conditions [8].

1.5.2.1. Molecular Marker Systems in Spinach

There are several studies which studied the genetic diversity of spinach accessions/germplasms and tried to pinpoint the origin of this crop. A literature search up to the year 2015 shows that previous researchers used TRAP [9] and SSR markers in

their investigations [10] [11]. In our study, SSR markers were used but all related publications and marker systems will be reviewed in the following sections.

1.5.2.2. Target Region Amplified Polymorphism Marker System

TRAP is a technique that uses a fixed primer with a known sequence in combination with arbitrary primers to generate polymorphic markers. In 2007, Hu et al. showed that TRAP markers are an appropriate tool to study the genetic diversity of spinach. They used 38 accessions and 10 commercial hybrids to investigate genetic diversity of spinach and its correlation with geographical origin. Although their results revealed that there is a high level of polymorphism at the DNA level both within and among accessions, they could not find any correlation between genetic diversity and geographical origin. This was one of the first studies which attempted to reveal the correlation between geographical origin and genetic diversity among spinach accessions [9].

1.5.2.3. Simple Sequence Repeat (SSR) Marker Systems

SSRs are a group of repetitive DNA sequences that are 1 to 6 bases long. These repeat sequences can be found throughout most eukaryotic genomes [12]. Polymerase chain reaction (PCR) with two flanking primers is a very useful method to identify Simple Sequence Repeats Polymorphisms (SSRPs) [12]. SSR development was first applied to various mammalian species [12]. The first time that SSRs were identified in plants was in phage libraries of tropical tree genomes via hybridization of oligonucleotide probes of poly (GT) and poly (AG) [13]. After the presence of SSRs was demonstrated in plants, other studies showed the abundance of SSRs in plant genomes [14] [15]. SSR markers are co-dominantly inherited, reproducible, simple, PCR-based and extremely polymorphic. Consequently SSR markers have been widely used by plant scientists especially those who are working on plant genetics and ecology.

In 2007, Khattak et al. used 33 spinach hybrids from seven different breeding stations and performed PCR with 35 genic (exonic or intronic) SSR markers. They determined that 13 of 35 SSR markers were polymorphic and they could group the

individuals into three clusters. While two clusters contained European spinach cultivars, the third cluster was a mixture of European and Asian spinach types.

In 2013, Kuwahara et al. used 50 accessions collected from geographically diverse regions and they performed PCR with 6 SSR markers which were previously published by Khattak et al. [10]. They showed that west Asian accessions (Afganistan, Iran, Iraq, Syria) had the highest level of gene diversity. This outcome is consistent with the commonly accepted theory about the Persian origin of spinach [11].

1.6. Goals

The primary aim of this study was to reveal genetic diversity and population structure in a spinach world collection germplasms. For this purpose 176 spinach germplasm accessions (world collection) were used and PCR was performed with 15 SSR markers.

Another objective of this study was to develop new spinach-specific SSR markers. Until this investigation there were not enough SSR markers available for genetic diversity studies in spinach. Next Generation Sequencing Technology was applied to sequence *Spinacia oleracea* cv. Universal nuclear genome using IIIumina MiSeq technology. Raw read sequences were used to re-construct the spinach genome sequence itself. Using the assembled genome, 3853 new SSR markers were developed for future genetic diversity studies by the scientific community.

CHAPTER 2

MATERIALS AND METHODS

2.1. Genetic Diversity Analysis

2.1.1. Materials

A total of 176 spinach germplasm accessions were used in this study (Table 2). These accessions vary according to the area where they were collected. Seeds were provided by United States Department of Agriculture (USDA, Beltsville, MD, USA).

Table 2. Plant individuals (*Spinacia oleracea*) used in this study.

Accession	No	Sample Origin	Accession	No	Sample Origin
NSL6782	1	Netherlands	PI206474	89	Turkey
NSL6693	2	United States, Michigan	PI206473	90	Turkey
NSL6557	3	United States, Washington	PI206007	91	Sweden
NSL6099	4	United States, Pennsylvania	PI647854	92	Georgia
NSL22003	5	United States, California	PI647855	93	Georgia
PI193619	6	Ethiopia	PI647856	94	Georgia
PI169690	7	Turkey, Kocaeli	PI491261	95	Greece
PI169674	8	Turkey, Istanbul	PI491262	96	Greece
PI169675	9	Turkey, Manisa	PI499372	97	Former Soviet Union
PI169678	10	Turkey ,Hatay	PI508504	98	Korea, South
PI175924	11	Turkey, Samsun	PI527332	99	China
PI175923	12	Turkey, Balıkesir	PI169682	100	Turkey, Edirne
PI175595	13	Turkey	PI169684	101	Turkey, Kırklareli
PI175313	14	India	PI169686	102	Turkey, Balıkesir
PI175312	15	India	PI169683	103	Turkey, Kırklareli
PI648951	16	United States, Maryland	PI648944	104	China, Beijing
PI648955	17	United States, Maryland	PI648941	105	China, Beijing
PI648952	18	United States, Maryland	PI648939	106	Turkey, Mugla
PI648953	19	United States, Maryland	PI254565	107	Afghanistan
PI648954	20	United States, Maryland	PI249920	108	Spain
PI604784	21	Afghanistan	PI229792	109	Iran
PI604786	22	Nepal	PI229731	110	Iran
PI604787	23	Afghanistan, Jowzjan	NSL37031	111	United States, Georgia

(Cont. on the next page)

Table 2. (cont.)

PI604777	24	Japan, Hokkaido	NSL32678	112	United States, NY
PI604778	25	Japan, Hokkaido	NSL32629	113	United States, Lousiana
PI604780	26	Japan, Hokkaido	NSL31384	114	United States, Oregon
PI604783	27	Afghanistan	NSL40592	115	United States, Michigan
NSL6098	28	United States, Virginia	PI433210	116	China
NSL6097	29	United States, Minnesota	PI433211	117	France
NSL6096	30	United States, Missouri	PI379551	118	Former Serbia and Montenegro
NSL6095	31	United States, Missouri	PI379552	119	Former Serbia and Montenegro
NSL6094	32	United States, Pennsylvania	PI419004	120	China
PI604790	33	Afghanistan	PI286435	121	Nepal
PI604791	34	Afghanistan	PI296393	122	Iran
PI606707	35	Netherlands, North Holland	PI303138	123	Netherlands
PI608762	36	Thailand	PI319220	124	Egypt
NSL81329	37	United States, Maryland	PI220686	125	Afghanistan
NSL42771	38	United States, California	PI222270	126	Iran
PI478393	39	China	PI266926	127	Germany
PI445783	40	Syria	PI263873	128	Greece
PI445784	41	Syria	PI176773	129	Turkey, Konya
PI224959	42	Iran	PI212328	130	Afghanistan
PI227230	43	Japan	PI212120	131	Afghanistan
PI227383	44	Iran	PI220121	132	Afghanistan
PI200882	45	Afghanistan	PI176372	133	Italy
PI204733	46	Turkey	PI358259	134	Former Serbia and Montenegro
PI204632	47	Turkey, Sivas	PI358269	135	Former Serbia and Montenegro
PI531448	48	Hungary	PI350710	136	France
PI531449	49	Hungary	PI360894	137	Netherlands
NSL184380	50	United States, California	PI360895	138	Netherlands
NSL184379	51	United States, California	PI183246	139	Egypt
NSL184378	52	United States, California	PI192945	140	China
PI169673	53	Turkey, Aydin	PI193618	141	Ethiopia
PI169671	54	Turkey, Burdur	PI179597	142	Belgium
PI169670	55	Turkey, Antalya	PI181086	143	India
PI209645	56	Iran, Fars	PI181923	144	Syria
PI209646	57	Turkey	PI648950	145	United States, Maryland
PI209647	58	Iran, Fars	PI648947	146	China, Beijing
PI177082	59	Turkey, Kayseri	PI648937	147	Syria
PI176779	60	Turkey, Bilecik	PI648938	148	Turkey, Mugla
PI177557	61	Turkey, Ankara	PI163309	149	India
PI175929	62	Turkey, Erzincan	PI103063	150	China, Beijing
PI175928	63	Turkey, Yozgat	PI174389	151	Turkey, Elazig

(Cont. on the next page)

Table 2. (cont.)

PI175925	64	Turkey, Canakkale	PI174388	152	Turkey, Elazig
PI169690	65	Turkey, Kocaeli	PI174387	153	Turkey, Gaziantep
PI358256	66	Former Serbia and Montenegro	PI358250	154	Former Serbia and Montenegro
PI358255	67	Former Serbia and Montenegro	PI648956	155	United States, Maryland
PI358254	68	Former Serbia and Montenegro	PI648960	156	United States, Maryland
PI179044	69	Turkey, Cankiri	PI174386	157	Turkey, Mardin
PI179042	70	Turkey, Kutahya	PI174385	158	Turkey Diyarbakir
PI179043	71	Turkey, Eskisehir	PI173972	159	India
PI179041	72	Turkey, Zonguldak	NSL6092	160	United States, New York
PI179591	73	Belgium	NSL6093	161	United States, IIIinois
PI179593	74	Belgium	PI648964	162	United States, California
PI179594	75	Belgium	PI648965	163	United States, Maryland
PI179595	76	Belgium	Ames20170	164	China
PI531454	77	Hungary	NSL6082	165	United States, New York
PI531455	78	Hungary	PI165994	166	India
PI531456	79	Hungary	PI165504	167	India
PI531457	80	Hungary	PI261789	168	France
NSL26513	81	United States, Michigan	PI256080	169	Afghanistan
NSL28216	82	United States, Wyoming	PI179508	170	Iraq
NSL28218	83	Sweden	PI179588	171	Belgium
PI205235	84	Turkey	PI179590	172	Belgium
PI205234	85	Turkey	PI167195	173	Turkey, Gaziantep
PI204736	86	Turkey	PI370602	174	Former Serbia and Montenegro
PI204735	87	Turkey	PI368824	175	Former Serbia and Montenegro
PI207518	88	Afghanistan	PI361127	176	United Kingdom, England

2.1.2. Methods

2.1.2.1. DNA Extraction

The youngest and lightest green leaves (roughly 3-4) were collected and genomic DNA extraction was performed. DNA extraction was carried out manually with a CTAB DNA isolation procedure [16]. Extracted DNA was suspended with TE buffer. Quality control was performed using ND-1000 spectrophotometer and all DNA samples were stored at -20 °C for further analysis.

2.1.2.2. Marker Analysis

A total of 15 SSR primer pairs were used for analysis of the spinach world collection. Spinach SSR markers are listed in Table 3**.** [17]

Table 3. Primer pairs information [17].

Primer name	Forward Primer Sequences(5'-3')	Reverse Primer Sequences(3'-5')	Annealing Temperature
SO1	GCGTTTCTAATTGCACCATATCA	TTTGGCGGTTGGTAGGTTTG	55 °C
SO3	ACTAGTGAGGGGGCCAGTTTACA	CAGCTGAGGCTCTTCTTCTTCTTC	55 °C
SO6	AGGGATAGTTTCGACACGGAGAGA	TTTAAGGGCAAAGGGAGCATCA	55 °C
SO7	TGCTTGGGTTCTATTGGTC	TTTTTTTTACAGAAGTGAATGCAA	55 °C
SO10	TGAAATCAACAACGATTACGC	AAGAACAAAGCAAGATAAGGTTCC	55 °C
SO13	GGTACTCCATCCGTCCCAACC	TCCAAAAGAAGCGGCCATGT	55 °C
SO14	CATCTCTACATCCGCCATTTCCA	GTGAAAAAACTCCGACGAAACACC	55 °C
SO16	AACATTTCAGGGGATTTCGTTTCA	GCCCTCGGTAACGCGACATA	55 °C
SO17	CCTGAGGTGCAGAACAATAACACG	GAGGGGATTTCACCTTAACCTTGC	55 °C
SO19	AAACTCTTTCTGATGGAGAGC	TTTGGAGGAGAGAGAGTGG	55 °C
SO28	GGCATCGTACATAGTTGTCC	TTTAATATGCAACACTTTTATCC	55 °C
SO29	AACATTTCAGGGGATTTCGTTTCA	GCCCTCGGTAACGCGACATA	55 °C
SO42	CCTCTAGGACCAATAATAATGC	CTCTCAACTTTGCTATCAACC	55 °C
SO48	AAGAGATCCAAATGCAAAGGAAG	GCAACACTAAAAATACCCTAATCG	55 °C
SO51	AGAGAAAAACCACCCAATCTCA	AAGAGGTGAGAGACGAGTGGAG	55 °C

Polymerase Chain Reaction was performed with the following components in total volume of 25 µl: 2.5 µl 10X Buffer (50 mM KCl, 10 mM Tris-HCl, 1.5 mM $MgCl_2$, pH: 8.3), 0.5 µl dNTP (0.2 mM), 0.5 µl Forward and Reverse primers (10 pmol), 0.25 µl *Taq* polymerase, 1.5 µl $MgCl_2$,18.25 µl sterile double-distilled water and 1 µl DNA. Amplification reaction conditions were: 5 minutes DNA denaturation at 94 °C, 35 cycles with 30 seconds at 94 °C, 45 seconds at 55 °C annealing, 45 seconds at 72 °C extension and final extension 10 minutes at 72 °C with using AB Applied Biosystems Veriti 96 well Thermal Cycler.

Amplicons were analyzed with Fragment Analyzer Automated CE System (Advanced Analytical Technologies, Inc, USA). This system can read amplified fragments in range 35 to 500 nucleotides and is able to identify 3 nucleotide differences between fragments.

2.1.2.3 .Diversity Analysis

According to the absence and presence of bands, allelic data were scored with "1" present or "0" absent and this allelic information was used to create a matrix. This matrix was used for genetic diversity analysis with DARwin5 (Dissimilarity Analysis Representation for Windows) tool [18]. Thus, allelic data from marker analysis was used for calculation of a distance matrix using Dice's coefficient and clustering analysis with Unweighted Neighbor-Joining algorithm.

2.1.2.4. Population Structure Analysis

STRUCTURE 2.2.3 was used with the allelic data to identify the correct number of subpopulations which best explain population structure by application of clustering [25]. This program was run with parameters of burn-in period of 100,000 and 500,000 MCMC replications and ad hoc statistics introduced by Evanno et al. [19].

STRUCTURE HARVESTER program was further used to identify the correct number of (K) subpopulations. K was tested from 2 to 10 with 20 iterations for each group. ΔK is a measurement which is used to determine correctness of clustering. Higher ΔK value is desired and indicates the correct number of subpopulations.

2.2. New SSR marker Development

2.2.1. Materials

A spinach cultivar *S. oleracea* Universal was used for SSR marker development.

2.2.2. Methods

2.2.2.1. DNA Extraction

The youngest and lightest green leaves (roughly 3-4) were collected and Genomic DNA extraction was performed. DNA extraction was carried out manually with a CTAB DNA isolation procedure [16]. Extracted DNA was suspended with TE buffer. Quality control was performed using ND-1000 spectrophotometer and all DNA samples were stored at -20 °C for further analysis.

2.2.2.2. Next Generation Sequencing

The nuclear genome was sequenced by Illumina Mi-Seq Next Generation Sequencing technology at San Diego, CA, USA. Raw data were supplied to us that consisted of approximately 15 million paired-end reads and read length was 300 nucleotides.

2.2.2.3. Cleaning

Adapter and linker sequences were removed from reads with cutadapt2 (v.1.8.3.) software [20] using default settings with minimum-length switch (-m=50). Any trimmed reads smaller than 50 nucleotides were removed from dataset since they disrupt mapping and assembly process. Reads were then mapped to human genome with bowtie (v.2.1.0) software [21] using default settings against human genome to remove possible human contaminants.

2.2.2.4. Assembly and Assembly Evaluation

Abyss (3.8.1), a *de novo*, parallel, paired end sequence assembler [22], was employed for sequence assembly. More than 100 runs were performed with different settings such as changing kmer (all possible substrings of length k contained in reads) and required number of reads to make a contig. Assembly quality was based on various parameters, such as the weighted median of contig lengths (N50), a commonly used

measure. The best assembly was identified according to the N50 value, assembly nucleotide length (closeness to the estimated size of the *S.oleracea* genome), length of largest contig and contig number. The settings that were finally chosen were: kmer (k=175) to create contigs.

2.2.2.5. SSR Development

Only contigs larger than 1000 nucleotides were analyzed for SSR detection using our in-house tool SiSeer [23]. The minimum number of repeats required to identify perfect SSRs were: 10 for mononucleotides, four for dinucleotides and three for motifs comprised of three or more nucleotides.

2.2.2.6. Annotation of SSR Motives

Identified SSR sequences were extracted with their genomic contex (padded with 100 nucleotides) and were converted to FASTA formatted sequences. These queries were searched against the Uniprot non–redundant plant protein database (Taxonomy = Viridiplantae) with blastx [24].

2.2.2.7. Primer Design

Primer design was performed with Primer3 [25] (primer core v.2.3.6) using default parameters with (PRIMER TASK = generic, PRIMER_OPT=SIZE=20, PRIMER_MIN_SIZE=18,PRIMER_MAX_SIZE=24,PRIMER_PRODUCT_SIZE=100 -300, PRIMER_MAX_TM=60, PRIMER_MIN_TM=50, PRIMER_OPT_TM=55).

CHAPTER 3

RESULTS

3.1. Genetic Diversity Analysis

3.1.1. DNA extraction

All germplasm nuclear DNA was extracted and DNA analyzed by ND-1000 spectrophotometer for its quality and quantity.

3.1.2. Molecular Marker Analysis

3.1.2.1. SSR Marker Analysis

The spinach world collection which included 176 germplasm accessions was analyzed using 15 SSR markers (Table 3). A total of 58 polymorphic SSR alleles were used to calculate genetic diversity and to perform clustering analysis among spinach germplasm. PIC (Polymorphism Information Content) value was calculated for each SSR marker [26]. In this study the SSR markers produced a total of 58 PCR bands, 57 of which polymorphic. PIC values of the SSR markers (Table 4) are ranged between 0.01 ± 0.00 (SO7) and 0.44 ± 0.03 (SO48) and average PIC value was 0.28.

Table 4. SSR markers that were used in genetic diversity analysis

Primer Name	Polymorphic Bands	Amplified Bands	PIC	Standard Error
SO1	4	4	0.39	0.08
SO3	5	5	0.34	0.06
SO6	7	7	0.29	0.09
SO7	2	3	0.01	0.00
SO10	3	3	0.19	0.10
SO13	2	2	0.36	0.12
SO14	3	3	0.21	0.11

(Cont. on the next page)

Table 4. (cont.)

SO16	3	3	0.17	0.05
SO17	6	6	0.34	0.04
SO19	4	4	0.26	0.07
SO28	2	2	0.20	0.12
SO29	3	3	0.26	0.12
SO42	4	4	0.32	0.05
SO48	6	6	0.44	0.03
SO51	3	3	0.41	0.05
Total Bands	**57**	**58**		

3.1.2.2. Diversity Analysis

Genetic diversity of the world spinach collection was calculated using Dice coefficient. Clustering analysis was performed with Darwin 5 program using Unweighted Neighbor joining algorithm. According to clustering analysis, two main groups and one outgroup were produced (Figure 1). Among 15400 dissimilarity values, maximum genetic dissimilarity of spinach accessions was 0.551 and minimum was 0.019. Average genetic diversity was 0.217.

The two main groups were labeled as A and B, respectively, outgroup was labeled as C. While group A contains 69 individuals group B contains 89 and group C contains 18 individuals. Although there is not exact clustering among individuals, group A mostly consisted of germplasm from Europe and USA (57.7%), group B included germplasm mostly from Asia and neighboring countries (69.8%). Group C was clearly intermixed with accessions from Asia, Europe and United States.

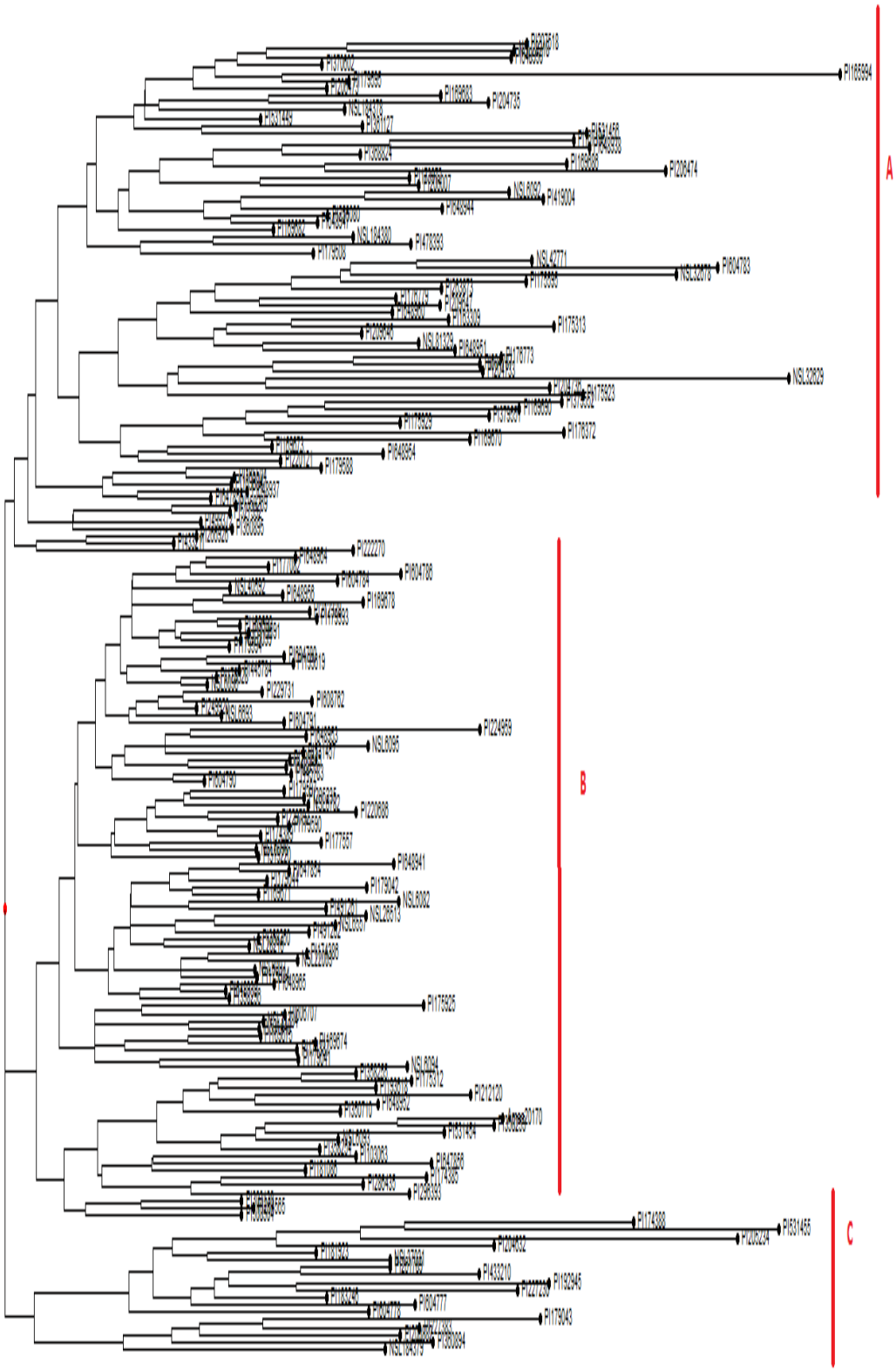

Figure 1. Clustering of spinach world collection

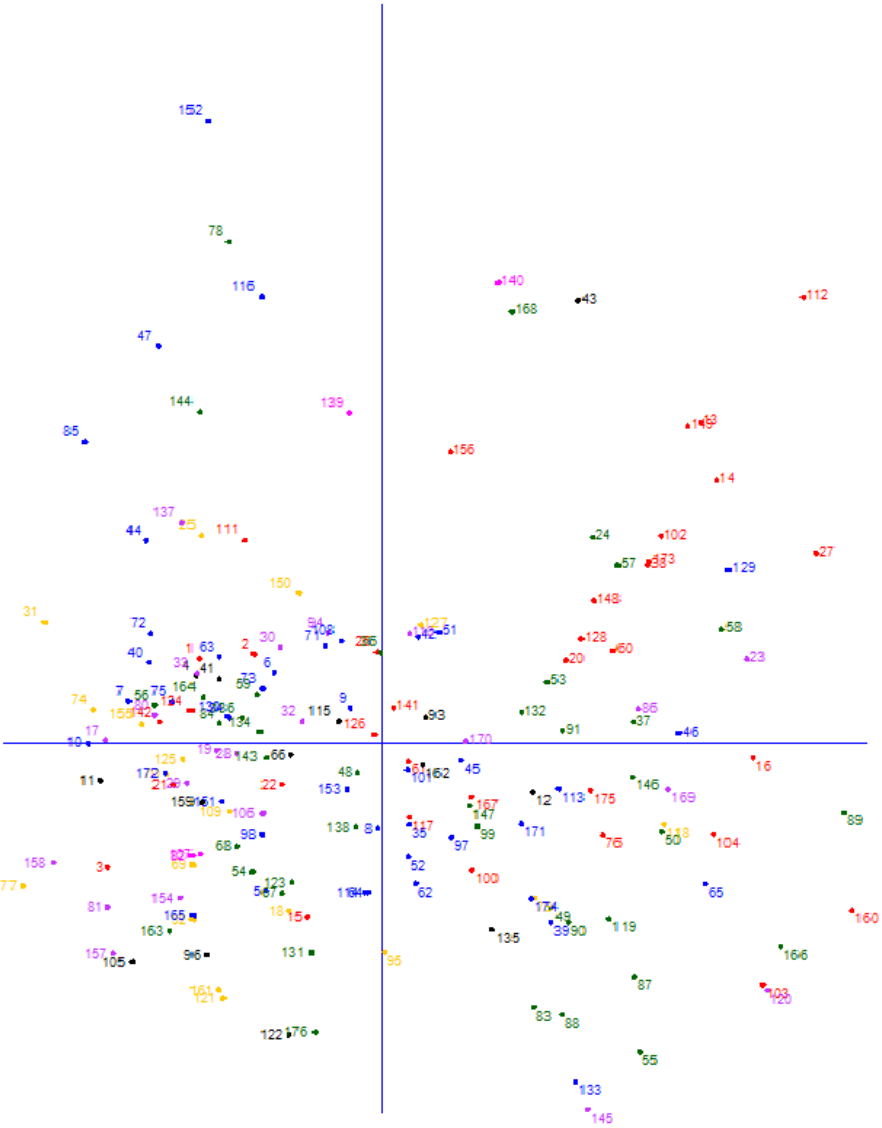

Figure 2. Principle Coordinate Analysis

In the Principle Coordinate Analysis (PCoA), accessions are colored based on their regions. Coloring was applied as follows; Turkey=Blue, USA=Red, Far East=Yellow, Europe =Green, Persia=Purple. There was no clear discrimination in terms of geographic origin of spinach accessions.

Furthermore, the spinach accessions were divided into groups based on geographic origin and PIC values were calculated to compare according to geographical regions; Far-East (Japan, Thailand, Korea South, China, Nepal), Europe (Netherlands, Hungary, Belgium, Greece, Germany, Italy, France), Persia (Afghanistan, Syria, Iran, Iraq, Egypt) (Table 5). These results show no significant differences in germplasm

diversity from different regions indicating that genetic diversity has been maintained throughout spinach's distribution.

In 2007, Hu et al. used the TRAP marker system to screen polymorphism of spinach accessions and studied correlation with geographic origin. Although they showed that there is high polymorphism between samples, they were unsuccessful infinding a relationship between genetic diversity and geographical origin. In 2007, Khattak et al. used 33 spinach hybrids from Europe and Asia. They clustered 33 spinach hybrids into three clusters. The first two clusters consisted of Europe hybrids and third cluster consisted of a mixture from Asia and Europe. They used very small and distinct population in terms of geographical origin. That is probably why they were successful in correlating genetic diversity with geographical origin.

In 2013, Kuwahara et al. used 50 accessions collected from geographically diverse regions and they screened this population with 6 SSR markers. They claimed that they were successful in correlating geographical origin and genetic diversity of geographically distant spinach samples. However, in terms of average genetic diversity distribution (He), their results did not show a clear distinction.

All of the previous studies showed high polymorphism between accessions, but the number of accessions used in previous work is very low comparatively with this study. Studies with larger populations point out the need for more primer pairs which target new SSR markers.

Table 5. PIC values according to geographic areas

Geographic Area	PIC (average)	Standard Deviation	Number of Individuals
USA	0.28	0.18	34
Turkey	0.29	0.16	43
Europe*	0.27	0.19	29
Far East*	0.24	0.17	18
Persia*	0.27	0.17	28

3.1.2.2. Population Structure Analysis

The highest ΔK (correctness of clustering) values were identified for K=2 (Figure 3) indicating that the spinach collection was best described as two subpopulations. In terms of deciding correct number of subpopulations, standard

deviation (SD) for each K is another important component besides ΔK which is shown in (Figure 4).

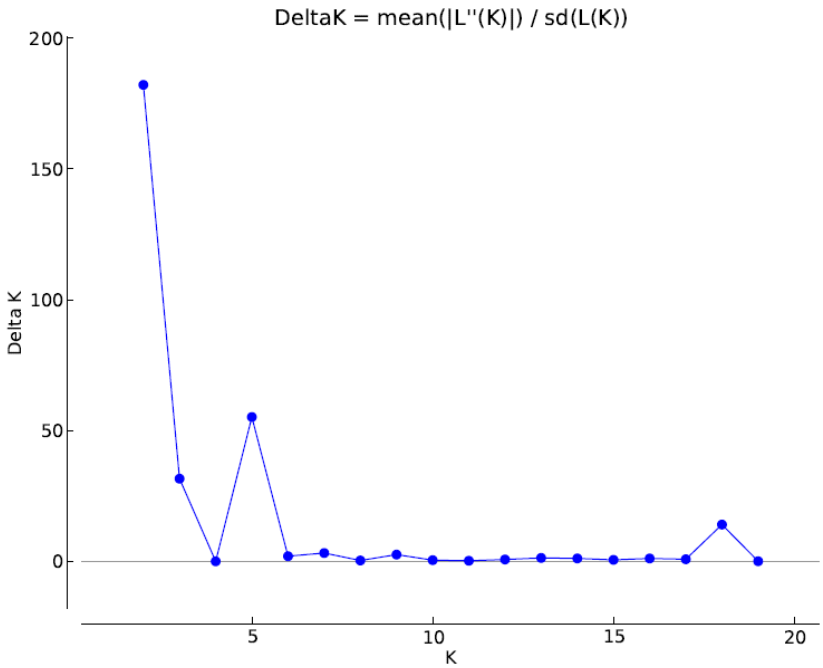

Figure 3. Distribution of clustering (K=2,20) ΔK values

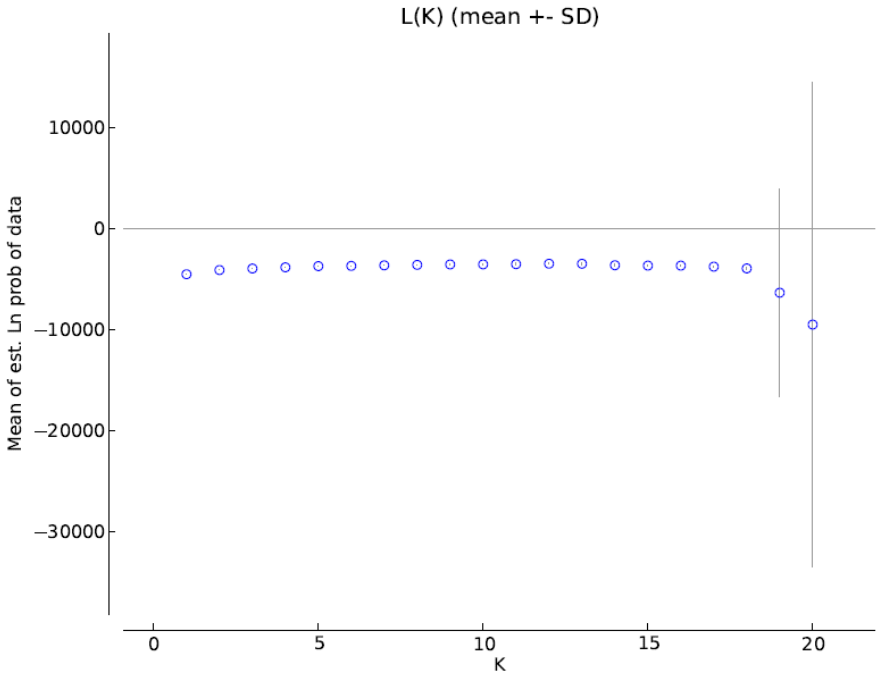

Figure 4. Standard Deviation Values for (K=2,20)

The structure of the population is represented the best when K value is 2. Thus each individual's probability of membership to each cluster was calculated. A threshold of P>0.7 was used for cluster assignment while cluster 1 contained 63 individuals, cluster 2 contained 64 individuals. In addition, 49 were found to be admixed (Table 6). According to Darwin5 and STRUCTURE program results, three groups (subpopulations) were created. Number of clusters which were created by Darwin5 and subpopulations created by STRUCTURE program are same but we do not see Asia or USA/Europe groups either cluster 1 or cluster 2 because high number of individuals were labeled as admixed.

Table 6. Clusering probability of individuals

ID	Genotype	Origin	Cluster 1	Cluster 2	Cluster Number
101	PI169684	Turkey,Mardin	0.7	0.3	1
83	NSL28218	Germany	0.704	0.296	1
29	NSL6097	Syria	0.717	0.283	1
67	PI358255	Greece	0.737	0.263	1
73	PI179591	Turkey,Kırklareli	0.753	0.247	1
148	PI648938	United States,IIIinois	0.795	0.205	1
158	PI174385	Afghanistan	0.798	0.202	1
28	NSL6098	Syria	0.809	0.191	1
129	PI176773	Turkey,Antalya	0.82	0.18	1
137	PI360894	Afghanistan	0.833	0.167	1
160	NSL6092	United States, Maryland	0.837	0.163	1
52	NSL184378	Turkey,Kutahya	0.839	0.161	1
49	PI531449	Former Serbia and Montenegro	0.842	0.158	1
69	PI179044	Korea,South	0.843	0.157	1
65	PI169690	Turkey	0.846	0.154	1
77	PI531454	China	0.859	0.141	1
134	PI358259	Sweden	0.864	0.136	1
1	NSL6782	United States, Michigan	0.866	0.134	1
15	PI175312	United States,Maryland	0.867	0.133	1
53	PI169673	Belgium	0.872	0.128	1
138	PI360895	Spain	0.875	0.125	1
75	PI179594	Turkey,Mugla	0.88	0.12	1
68	PI358254	Former Soviet Union	0.883	0.117	1
60	PI176779	United States,Michigan	0.891	0.109	1
114	NSL31384	Turkey	0.894	0.106	1
62	PI175929	Turkey	0.895	0.105	1

(Cont. on the next page)

Table 6. (cont.)

33	PI604790	Afghanistan	0.899	0.101	1
110	PI229731	Afghanistan, Jowzjan	0.9	0.1	1
81	NSL26513	Afghanistan	0.906	0.094	1
121	PI286435	China	0.906	0.094	1
100	PI169682	United States,Maryland	0.909	0.091	1
84	PI205235	Greece	0.91	0.09	1
20	PI648954	United States,Virginia	0.922	0.078	1
35	PI606707	Turkey,Sivas	0.923	0.077	1
132	PI220121	Sweden	0.923	0.077	1
123	PI303138	Former Serbia and Montenegro	0.925	0.075	1
127	PI266926	China	0.928	0.072	1
19	PI648953	Afghanistan	0.929	0.071	1
17	PI648955	Afghanistan	0.931	0.069	1
120	PI419004	Egypt	0.935	0.065	1
135	PI358269	Georgia	0.935	0.065	1
165	NSL6082	Turkey,Gaziantep	0.935	0.065	1
34	PI604791	Turkey	0.936	0.064	1
55	PI169670	Belgium	0.936	0.064	1
58	PI209647	Hungary	0.938	0.062	1
131	PI212120	Hungary	0.939	0.061	1
90	PI206473	Netherlands	0.94	0.06	1
133	PI176372	Turkey	0.941	0.059	1
119	PI379552	Netherlands	0.942	0.058	1
139	PI183246	Iran	0.943	0.057	1
50	NSL184380	Former Serbia and Montenegro	0.945	0.055	1
64	PI175925	Turkey	0.946	0.054	1
162	PI648964	India	0.946	0.054	1
161	NSL6093	China,Beijing	0.947	0.053	1
163	PI648965	France	0.947	0.053	1
79	PI531456	China	0.949	0.051	1
16	PI648951	United States,Maryland	0.953	0.047	1
145	PI648950	Afghanistan	0.956	0.044	1
155	PI648956	China,Beijing	0.957	0.043	1
136	PI350710	Greece	0.968	0.032	1
174	PI370602	Turkey,Diyarbakir	0.969	0.031	1
51	NSL184379	Turkey,Cankiri	0.97	0.03	1
38	NSL42771	United States,California	0.973	0.027	1
10	PI169678	Turkey	0.026	0.974	2
14	PI175313	United States,Maryland	0.028	0.972	2
45	PI200882	Turkey,Bilecik	0.029	0.971	2

(Cont. on the next page)

Table 6. (cont.)

106	PI648939	Afghanistan	0.03	0.97	2
104	PI648944	United States,New York	0.031	0.969	2
87	PI204735	Italy	0.035	0.965	2
6	PI193619	Turkey,Istanbul	0.036	0.964	2
98	PI508504	Turkey,Gaziantep	0.04	0.96	2
40	PI445783	Turkey,Burdur	0.041	0.959	2
56	PI209645	Belgium	0.041	0.959	2
168	PI261789	Netherlands	0.042	0.958	2
41	PI445784	Iran,Fars	0.043	0.957	2
140	PI192945	Iran	0.043	0.957	2
105	PI648941	India	0.049	0.951	2
21	PI604784	United States,Missouri	0.05	0.95	2
54	PI169671	Belgium	0.05	0.95	2
169	PI256080	Afghanistan	0.05	0.95	2
88	PI207518	Former Serbia and Montenegro	0.051	0.949	2
126	PI222270	United States,Minnesota	0.054	0.946	2
170	PI179508	Syria	0.055	0.945	2
124	PI319220	United States,California	0.058	0.942	2
3	NSL6557	United States,Pennsylvania	0.059	0.941	2
103	PI169683	United States,Maryland	0.062	0.938	2
46	PI204733	Turkey,Erzincan	0.063	0.937	2
74	PI179593	China,Beijing	0.063	0.937	2
61	PI177557	United States,Wyoming	0.066	0.934	2
42	PI224959	Turkey	0.072	0.928	2
151	PI174389	Turkey,Ankara	0.073	0.927	2
5	NSL22003	Turkey,Kocaeli	0.079	0.921	2
2	NSL6693	United States, Washington	0.083	0.917	2
149	PI163309	United States,Maryland	0.088	0.912	2
13	PI175595	United States,Maryland	0.089	0.911	2
93	PI647855	Ethiopia	0.093	0.907	2
4	NSL6099	Ethiopia	0.096	0.904	2
125	PI220686	Japan,Hokkaido	0.098	0.902	2
22	PI604786	United States,Pnnsylvania	0.102	0.898	2
115	NSL40592	Georgia	0.112	0.888	2
9	PI169675	Turkey,Balıkesir	0.113	0.887	2
43	PI227230	Iran,Fars	0.113	0.887	2
76	PI179595	United States,Oregon	0.113	0.887	2
122	PI296393	India	0.114	0.886	2
128	PI263873	United States, California	0.123	0.877	2
80	PI531457	Iran	0.125	0.875	2

(Cont. on the next page)

Table 6. (cont.)

94	PI647856	Syria	0.128	0.872	2
25	PI604778	Thailand	0.132	0.868	2
30	NSL6096	Iran	0.134	0.866	2
150	PI103063	Japan,Hokkaido	0.157	0.843	2
7	PI169690	Turkey,Hatay	0.168	0.832	2
175	PI368824	United States,New York	0.169	0.831	2
71	PI179043	Turkey,Edirne	0.172	0.828	2
86	PI204736	Afghanistan	0.174	0.826	2
153	PI174387	Turkey,Zonguldak	0.177	0.823	2
173	PI167195	United States,Maryland	0.177	0.823	2
102	PI169686	United States,California	0.178	0.822	2
167	PI165504	United Kingdom,England	0.179	0.821	2
66	PI358256	Georgia	0.188	0.812	2
12	PI175923	India	0.208	0.792	2
99	PI527332	Former Serbia and Montenegro	0.238	0.762	2
8	PI169674	Turkey,Samsun	0.241	0.759	2
27	PI604783	United States, California	0.251	0.749	2
141	PI193618	United States,New York	0.261	0.739	2
108	PI249920	Turkey,Manisa	0.271	0.729	2
11	PI175924	India	0.283	0.717	2
72	PI179041	Turkey,Balıkesir	0.292	0.708	2
37	NSL81329	Hungary	0.323	0.677	admixed
113	NSL32629	Turkey,Canakkale	0.34	0.66	admixed
117	PI433211	United States,Lousiana	0.344	0.656	admixed
97	PI499372	Turkey,Elazig	0.347	0.653	admixed
96	PI491262	India	0.364	0.636	admixed
152	PI174388	Turkey,Kocaeli	0.379	0.621	admixed
146	PI648947	Former Serbia and Montenegro	0.384	0.616	admixed
48	PI531448	Former Serbia and Montenegro	0.395	0.605	admixed
32	NSL6094	Iran	0.402	0.598	admixed
82	NSL28216	Iran	0.404	0.596	admixed
157	PI174386	Egypt	0.406	0.594	admixed
23	PI604787	Afghanistan	0.409	0.591	admixed
142	PI179597	United States,Michigan	0.411	0.589	admixed
36	PI608762	Hungary	0.416	0.584	admixed
116	PI433210	Turkey,Kırklareli	0.417	0.583	admixed
118	PI379551	Nepal	0.418	0.582	admixed
47	PI204632	Turkey,Yozgat	0.42	0.58	admixed
159	PI173972	India	0.42	0.58	admixed
164	Ames20170	Belgium	0.431	0.569	admixed

(Cont. on the next page)

Table 6. (cont.)

24	PI604777	Netherlands	0.44	0.56	admixed
26	PI604780	United States,Maryland	0.454	0.546	admixed
130	PI212328	Turkey,Eskisehir	0.457	0.543	admixed
143	PI181086	France	0.458	0.542	admixed
95	PI491261	China,Beijing	0.47	0.53	admixed
166	PI165994	Former Serbia and Montenegro	0.484	0.516	admixed
39	PI478393	Turkey,Aydin	0.493	0.507	admixed
171	PI179588	Turkey,Mugla	0.52	0.48	admixed
44	PI227383	Turkey,Kayseri	0.523	0.477	admixed
18	PI648952	Japan,Hokkaido	0.543	0.457	admixed
78	PI531455	Former Serbia and Montenegro	0.545	0.455	admixed
92	PI647854	China	0.551	0.449	admixed
156	PI648960	United States,Georgia	0.562	0.438	admixed
31	NSL6095	Japan	0.563	0.437	admixed
154	PI358250	Afghanistan	0.567	0.433	admixed
59	PI177082	Hungary	0.576	0.424	admixed
112	NSL32678	United States,California	0.577	0.423	admixed
176	PI361127	Belgium	0.579	0.421	admixed
172	PI179590	Turkey,Elazig	0.58	0.42	admixed
109	PI229792	Nepal	0.607	0.393	admixed
89	PI206474	France	0.61	0.39	admixed
91	PI206007	Netherlands	0.622	0.378	admixed
70	PI179042	China	0.625	0.375	admixed
107	PI254565	Iraq	0.636	0.364	admixed
63	PI175928	Turkey	0.644	0.356	admixed
144	PI181923	Former Serbia and Montenegro	0.653	0.347	admixed
147	PI648937	Belgium	0.662	0.338	admixed
85	PI205234	Turkey,Konya	0.664	0.336	admixed
111	NSL37031	United States,Missouri	0.67	0.33	admixed
57	PI209646	Hungary	0.689	0.311	admixed

Figure 5. Structure of spinach world collection germplasms

In the Figure 5, X axis demonstrates all individuals and Y axis demonstrates probability calculation for Cluster 1 and Cluster2.

3.2. SSR Marker Development

3.2.1. DNA Extraction

S. oleracea (Universal) cultivar DNA were extracted. Extracted DNA quality and quantity was analyzed using ND-1000 spectrophotometer.

3.2.2. Cleaning

First we applied the Cutadapt2 program to remove adapter sequences from reads and to remove any trimmed reads smaller than 50 nucleotides. Later, those trimmed reads were mapped to the human genome with bowtie2 to remove possible human contaminants. Finally after these two steps were performed, 0.16 % of the original data was discarded.

3.2.3. Assembly and Assembly Evaluation

Cleaned reads were assembled with different K-mers ranging between 55 and 195. The best assembly was identified when K is 175 according to largest contig and K50 value. The largest contig of this assembly was 72,530 nucleotides. Distribution of contig lengths is shown in Table 7.

Table 7. Distribution of Contigs based on size range.

Range	**Number of Contigs**
1000-2000	17.792
2000-3000	622
3000-4000	90
4000-5000	28
6000-10.000	4
10.000-15.000	1
15.000-20.000	2
20.000-50.000	1
50.000-100.000	4

3.2.4. SSR Development

A total of 3853 SSR markers were identified via using SiSeer tool [23]. Maximum length of SSRs was 98 nt, minimum length 6 nt and average length was 15 nt. Distributions of SSR markers according to motif length is shown in Table 8. Among 3852 SSR motifs the most frequent mononucleotide repeats, representing 54.33% of all SSRs. Second common SSR type is dinucleotides, representing 15.68%. The percentage of abundance of remaining repeat types are 11.86% (hexanucleotide), 8.46% (trinucleotide), 5.24% (heptanucleotide), 2.18 (tetranucleotide), 1.22% (octonucleotide) and 1.01% (pentanucleotide). The most frequent SSR motifs are shown at Table 9 . A/T the most frequent mononucleotides repeat (51.20%) followed by AT/TA dinucleotide repeats (23.92%, 22.61%) respectively.

Table 8. Distribution of SSRs by motif length.

Pattern	**Number of SSRs**	**Frequency (%)**
Mononucleotide	2093	54.33
Dinucleotide	604	15.68
Trinucleotide	326	8.46
Tetranucleotide	84	2.18
Pentanucleotide	39	1.01
Hexanucletoide	457	11.86
Heptanucleotide	202	5.24
Octanucleotide	47	1.22
Total	3852	100

Table 9. Most abundant simple sequence repeat (SSR) motifs

SSR Motif	**Number of SSRs**	***Motif Frequency (%)**
T	874	51.20
A	823	48.21
AT	128	23.92
TA	121	22.61
TC	44	8.22
AG	42	7.85
CT	38	7.10
GA	33	6.16
AC	28	5.23
AAT	41	14.28
TTA	21	7.31
TAT	20	6.96
TTG	16	5.57
ATA	15	5.22
ATT	15	5.22
TAAT	8	11.2

(Cont. on the next page)

Table 9 (cont.)

AATA	6	8.45
TTAT	6	8.45
TTTA	6	8.45
AAAT	5	7.04
TTAA	5	7.04
ATTT	4	5.63
CTTAT	5	13.51
AATAA	2	5.40
CAGTT	2	5.40
CTGAA	2	5.40
CTTTT	2	5.40
TTTCT	2	5.40

3.2.5. Primer Development

Among 3853 newly identified SSRs, 3275 primer pairs were developed successfully through Primer3 tool. Thus, a low proportion of SSRs (15%) was not suitable to develop primer pairs in terms of defined criterions. A table of these primers will be available at http://plantmolgen.iyte.edu.tr/data/. once the results are published.

3.2.6. Annotation of SSR motifs

Among the 3275 SSR markers, 1419 SSR markers were successfully annotated through Blastx program against NR-Plant Protein database (Taxonomy = Viridiplantae). Annotation of SSR motifs will be available at http://plantmolgen.iyte.edu.tr/data/. once the results are published.

CHAPTER 4

CONCLUSION

Spinach is a green leafy plant which is grown almost all over the world because of its nutrient content. Unfortunately there are not enough studies in terms of genetic diversity of spinach. Today climate change, global warming and hunger are important concerns and highlight the necessity to learn more about vegetable genomes such as spinach.

In this study, 15 known SSR markers were used to screen 176 spinach accessions from all over the world to study its genomic diversity. This is one of the first studies in terms of high number of samples (accessions) analyzed comparative to previously published papers examining spinach genome diversity. Unlike previous studies which showed correlation between geographical origin and genetic diversity of spinach germplasms, we were unable to make similar conclusion because comparatively high number of individuals. Another reason might be a lack of available spinach-specific molecular markers. Although 57 of 58 bands were polymorphic, more bands with new SSR markers might help to get better clustering according to regions.

In order to address the problem of a lack of sufficient spinach-specific markers, we sequenced *Spinacia oleracea* (Universal) cultivar nuclear genome and assembled its genome for SSR marker development. As a result we identified 3853 SSR markers and we successfully designed 3275 primer pairs for these identified SSR motives. These newly designed primer pairs can be used for future genetic diversity studies in spinach and most probably for closely related species.

REFERENCES

[1] S. T. Koike, U. F. Advisor, M. County, M. Cahn, U. F. Advisor, M. County, S. Fennimore, U. W. Specialist, and U. S. Agricultural, "SPINACH PRODUCTION IN CALIFORNIA University of California," *UC Veg. Res. Inf. Cent.*, vol. 2006, 2008.

[2] T. Tetrafluoroborate and E. Nh, "Spinacia oleracea Linn: A pharmacognostic and pharmacological overview," vol. 3, no. 9, pp. 2280–2283, 2010.

[3] M. Ito, "Characterization of Spinach Chromosomes by Condensation Patterns and Physical Mapping of 5 s and 45s rDNAs by FISH," *Sci. Technol.*, vol. 125, no. 1, pp. 59–62, 2000.

[4] K. Arumuganathan and E. D. Earle, "Nuclear DNA content of some important plant species," *Plant Mol. Biol. Report.*, vol. 9, no. 3, pp. 208–218, 1991.

[5] R. Boswell, "Garden Peas and Spinach from the Middle East," *Natl. Geogr. Mag.*, 1949.

[6] R. J. and S. Carol, *The food Encyclopedia : Over 8.000 Ingredients, Tools, Techniques and People*. 2006.

[7] G. E. Lester, D. J. Makus, D. M. Hodges, and J. L. Jifon, "Summer (Subarctic) versus winter (Subtropic) production affects spinach (Spinacia oleracea L.) leaf bionutrients: Vitamins (C, E, Folate, K1, provitamin A), lutein, phenolics, and antioxidants," *J. Agric. Food Chem.*, vol. 61, no. 29, pp. 7019–7027, 2013.

[8] L. S. Kumar, "DNA markers in plant improvement: An overview," *Biotechnol. Adv.*, vol. 17, no. 2–3, pp. 143–182, 1999.

[9] J. Hu, B. Mou, and B. a. Vick, "Genetic diversity of 38 spinach (Spinacia oleracea L.) germplasm accessions and 10 commercial hybrids assessed by TRAP markers," *Genet. Resour. Crop Evol.*, vol. 54, no. 8, pp. 1667–1674, 2007.

[10] J. Z. K. Khattak, J. L. Christiansen, a. M. Torp, and S. B. Andersen, "Genic microsatellite markers for discrimination of spinach cultivars," *Plant Breed.*, vol. 126, no. 4, pp. 454–456, 2007.

[11] K. Kuwahara, R. Suzuki, Y. Ito, T. Mikami, and Y. Onodera, "An analysis of genetic differentiation and geographical variation of spinach germplasm using SSR markers," *Plant Genet. Resour.*, vol. 12, no. 02, pp. 185–190, 2014.

[12] Y.-J. Park, J. K. Lee, and N.-S. Kim, "Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crops," *Molecules*, vol. 14, no. 11, pp. 4546–4569, 2009.

[13] S. . Condit, R., Hubbel, “Abundance and DNA sequence of two-base repeat regions in tropical tree genomes,” *Genome*, vol. 34, pp. 66–71, 1991.

[14] M. Morgante and a M. Olivieri, “PCR-amplified microsatellites as markers in plant genetics.,” *Plant J.*, vol. 3, no. 1, pp. 175–182, 1993.

[15] S. . Wang, Z., Weber, J., L Zhong, G., Tanksley, “Survey of plant short tandem DNA repeats,” *Theor. Appl.*, vol. 88, pp. 1–6, 1994.

[16] D. J. . Doyle, J.J., “A rapid DNA isolation procedure for small quantities of fresh leaf tissue,” *Phytochem. Bull. 19*, 1987.

[17] “www.pgdbj.jp.” . (access time : Feb 2014)

[18] X. P. and J. C. G. Perla Hamon, Marc Sequin, *Genetic Diversity of cultivated tropical plants*. 2003.

[19] J. Evanno, Regnaut, S., Goudet, “Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study,” *Mol. Ecol.*, vol. 14, no. 8, pp. 2611–2620, 2005.

[20] M. Martin, “Cutadapt removes adapter sequences from high-thorughput sequencing reads,” *EMBnet.journal*, vol. 17, pp. 10–12, 2011.

[21] B. Langmead and S. L. Salzberg, “Fast gapped-read alignment with Bowtie 2.,” *Nat. Methods*, vol. 9, no. 4, pp. 357–9, 2012.

[22] B. L. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, “AbySS: A paralel assembler for short read sequence data,” *Genome Res.*, 2009.

[23] “http://bioinformatics.iyte.edu.tr/index.php?n=Softwares.SiSeeR.” .

[24] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool.,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–10, 1990.

[25] R. M. Koressaar T., “Enhancement and modifications of primer design program Primer3,” *Bioinformatics*, vol. 23, no. 10, pp. 1289–91, 2007.

[26] A. Kumar, P. Mishra, S. C. Singh, and V. Sundaresan, “Efficiency of ISSR and RAPD markers in genetic divergence analysis and conservation management of Justicia adhatoda L., a medicinal plant,” *Plant Syst. Evol.*, vol. 300, no. 6, pp. 1409–1420, 2014.