

**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 Illumina 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 İSPANAK 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 göstermemiştir, 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. Ayrıca *S. oleracea L.* cv. Universal nuklear genomu Illumina 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 FIGURES	viii
LIST OF TABLES	ix
CHAPTER 1. INTRODUCTION	1
1.1 General Information about <i>Spinacia oleracea L.</i>	1
1.2 Origin and Production of Spinach.....	1
1.3 Nutrition of Spinach and Health Benefits	2
1.4 Genetic Diversity	2
1.5 Genetic Analysis with Marker Systems.....	3
1.5.1 Morphological Markers.....	3
1.5.2 Molecular Markers	3
1.5.2.1 Molecular Marker Systems in Spinach	3
1.5.2.2 Target Region Amplified Polymorphism Marker System	4
1.5.2.3 Simple Sequence Repeat (SSR) Marker Systems	4
1.6 Goals	5
CHAPTER 2. MATERIALS AND METHODS	6
2.1 Genetic Diversity Analysis	6
2.1.1 Materials.....	6
2.1.2 Methods.....	8
2.1.2.1 DNA Extraction.....	8
2.1.2.2 Marker Analysis	9
2.1.2.3 Diversity Analysis	10
2.1.2.4 Population Structure Analysis	10
2.2 New SSR marker Development.....	10
2.2.1 Materials.....	10
2.2.2 Methods.....	11
2.2.2.1 DNA Extraction.....	11
2.2.2.2 Next Generation Sequencing.....	11
2.2.2.3 Cleaning.....	11

2.2.2.4 Assembly and Assembly Evaluation	11
2.2.2.5 SSR Development.....	12
2.2.2.6 Annotation of SSR Motives.....	12
2.2.2.7 Primer Design.....	12
CHAPTER 3. RESULTS.....	13
3.1 Genetic Diversity Analysis.....	13
3.1.1 DNA extraction	13
3.1.2 Molecular Marker Analysis.....	13
3.1.2.1 SSR Marker Analysis	13
3.1.2.2 Diversity Analysis	14
3.1.2.2 Population Structure Analysis	17
3.2. SSR Marker Development.....	24
3.2.1. DNA Extraction	24
3.2.2. Cleaning	24
3.2.3. Assembly and Assembly Evaluation.....	24
3.2.4. SSR Development.....	25
3.2.5. Primer Development	26
3.2.6. Annotation of SSR motifs	26
CHAPTER 4. CONCLUSION	27
REFERENCES	28

LIST OF THE FIGURES

<u>Figure</u>	<u>Page</u>
Figure 1. Clustering of spinach world collection.....	15
Figure 2. Principle Coordinate Analysis.....	16
Figure 3. Distribution of clustering (K=2,20) ΔK values	18
Figure 4. Standard Deviation Values for (K=2,20)	18
Figure 5. Structure of spinach world collection germplasms	23

LIST OF THE TABLES

<u>Table</u>	<u>Page</u>
Table 1. Top 4 producers of spinach in 2013	2
Table 2. Plant individuals (<i>Spinacia oleracea</i>) used in this study.....	6
Table 3. Primer pairs information	9
Table 4. SSR markers that were used in genetic diversity analysis.....	13
Table 5. PIC values according to geographic areas	17
Table 6. Clustering probability of individuals	19
Table 7. Distribution of Contigs based on size range.....	24
Table 8. Distribution of SSRs by motif length.	25
Table 9. Most abundant simple sequence repeat (SSR) motifs	25

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 B₉), phyloquione (vitamin K₁), 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.

