REVEALING THE HISTORY OF ANATOLIAN SHEEP DOMESTICATION BY USING RETROVIRUS INTEGRATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İHSAN CİHAN AYANOĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

AUGUST 2013

Approval of the thesis:

REVEALING THE HISTORY OF ANATOLIAN SHEEP DOMESTICATION BY USING RETROVIRUS INTEGRATIONS

Submitted by **İHSAN CİHAN AYANOĞLU** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Gülay Özcengiz _____
Head of Department, **Biology**

Prof. Dr. İnci Togan _____
Supervisor, **Biology Dept., METU**

Examining Committee Members:

Assoc Prof. Dr. Can Bilgin _____
Biology Dept., METU

Prof. Dr. İnci Togan _____
Biology Dept., METU

Prof. Dr. Okan Ertuğrul _____
Genetics Dept., Fac. of Vet. Med., Ankara University

Assoc. Prof. Dr. Mayda Gürsel _____
Biology Dept., METU

Assist. Prof Dr. Emel Özkan Ünal _____
Animal Science Dept. Namık Kemal University

Date: 29.08.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: İHSAN CİHAN AYANOĞLU

Signature : _____

ABSTRACT

REVEALING THE HISTORY OF ANATOLIAN SHEEP DOMESTICATION BY USING RETROVIRUS INTEGRATIONS

Ayanoğlu, İhsan Cihan

M.Sc. Department of Biology

Supervisor: Prof. Dr. İnci Togan

August 2013, 91 pages

In the present study, it was aimed to contribute to the understanding of evolutionary history of sheep by using endogenous Jaagsiekte sheep retrovirus (enJSRV) integrations in 220 samples from 11 Turkish sheep breeds (Karayaka, Dağlıç, Kıvırcık, Sakız, Akkaraman, İvesi, Norduz, Karagül, Hemşin, Gökçeada and Morkaraman) and also in 30 samples from Anatolian wild sheep (*Ovis gmelinii anatolica*). Retrotypes of individuals were determined based on enJSRV integrations. The genetic distances between breeds were calculated and spatial distribution of retrotypes in Turkish sheep breeds have been determined by construction of synthetic maps.

In the study it has been shown that all previously observed retrotypes for Turkish sheep breeds (R2, R4 and R6) were also present among the individuals of the present study. In addition to these retrotypes, previously non-observed retrotypes for Turkish breeds (R0, R1 and R7) were also observed in Turkish sheep breeds which have been examined in this study. Different retrotype frequencies for Turkish sheep breeds have been observed when compared to previous studies. Spatial distribution of retrotypes showed that Turkey was separated into two broad regions where the regions were represented by fat-tailed and thin-tailed sheep breeds, generally. Seven of 30 Anatolian wild sheep samples exhibited retrotype of primitive breeds (R0) and also some individuals exhibited a new retrotype where there are no enJSRV integrations for enJSRV-6 integration site which was assumed to be "fixed insertion" previously. Thus, the present study brings new insight into evaluation of enJSRV-6 integration which is to use this integration and suggests that it can also be used as a polymorphic integration site. Also, comparative studies of enJSRV markers with sex dependent (mtDNA and Y-linked) markers indicated that possible second migration of the sheep may be mediated by males. The present study examines the enJSRV polymorphism

that is present in Turkey with a high resolution. The results contributes to the understanding of the evolutionary history of sheep and gives insights about the shaping of the genetic diversity observed among domestic sheep over Turkey on the basis of enJSRV integrations. Furthermore, polymorphism in enJSRV-6 especially high frequency of newly observed allele is may be indicating an interesting event related with the sheep domestication.

Keywords: Domestication, Turkish sheep breeds, Anatolian wild sheep, endogenous Jaagsiekte sheep retroviruses (enJSRVs), retrotype.

ÖZ

RETROVİRÜS ENTEGRASYON POLİMORFİZMİ KULLANARAK ANADOLU KOYUNLARINDAKİ EVCİLLEŞTİRME SÜRECİNİN ANLAŞILMASI

Ayanoğlu, İhsan Cihan

Yüksek Lisans, Biyoloji Bölümü

Tez Yöneticisi: Prof. Dr. İnci Togan

Ağustos 2013, 91 sayfa

Bu çalışmada, koyunların evrimsel tarihinin anlaşılmasına katkıda bulunmak amacıyla 11 Türk koyun ırkından (Karayaka, Dağlıç, Kıvırcık, Sakız, Akkaraman, İvesi, Norduz, Karagül, Hemşin, Gökçeada and Morkaraman) 220 bireydeki ve Anadolu yaban koyun popülasyonundan (*Ovis gmelinii anatolica*) 30 bireydeki içsel Jaagsiekte koyun retrovirüslerinin (enJSRV) entegrasyonu incelenmek üzere araştırılmıştır. Bu entegrasyonların değişik kombinasyonlarının varlığına göre bireylerin retrotipi belirlenmiştir. Koyun ırkları arasındaki genetik farklılıklar hesaplanmış olup, retrotiplerin Türk koyun ırklarındaki uzamsal dağılımları sentetik haritalar oluşturularak belirlenmiştir.

Bu çalışmada elde edilen sonuçlar daha önce Türk koyun ırklarında gözlemlenen bütün retrotip çeşitlerinin (R2, R4 ve R6) bu çalışmada kullanılan bireyler arasında da var olduğunu göstermektedir. Bu retrotiplere ek olarak Türk koyun ırklarında daha önce gözlemlenmemiş retrotiplerin (R0, R1 ve R7) de varlığı gözlemlenmiştir. Türk ırkları için retrotipler daha önceki çalışmalardan farklı frekanslarda gözlemlenmiş ve farklı bir retrotip dağılımı belirlenmiştir. Yapılan uzamsal dağılımlar sonucunda Türkiye'nin coğrafik olarak iki geniş bölgeye ayrıldığı ve bu iki bölgenin genelde yağlı kuyruklu ve yağsız kuyruklu Türk koyun ırkları tarafından temsil edildiği gözlemlenmiştir. İncelenen 30 Anadolu yaban koyunu örneğinden 7 tanesinde ilkel ırkların karakteristik özelliklerinden olan retrotip (R0) gözlemlenmiştir. Ayrıca Anadolu yaban koyunundan bazı bireylerde daha önce koyunlarda sabitlenmiş olarak düşünülen enJSRV-6 dahil hiç bir retrovirüs entegrasyonu gözlemlenmemiştir. Buna ilaveten bu çalışma daha önce koyun ırklarında sabitlendiği kabul edilen enJSRV-6 entegrasyonunu polimorfik bir belirteç olarak kullanma şeklinde de yeni bir bakış açısı getirmektedir. Son olarak nükleer bir belirteç olan enJSRV' nin kullanıldığı bu çalışmanın sonuçları annesel ve babasal belirteçlerin kullanıldığı diğer çalışmaların sonuçlarıyla kıyaslandığında, muhtemel ikincil bir koyun göçünün genelde erkekler tarafından gerçekleştiği söylenebilir.

Bu çalışma enJSRV entegrasyonlarını Anadolu yaban koyunu dahil olmak üzere Türk koyun ırklarında inceleyerek evcil koyunların evrimsel tarihini daha iyi anlamaya yardımcı olmakla beraber Türkiye'deki evcil koyun ırklarının genetik çeşitliliğinin nasıl şekillendiğinin anlaşılmasına katkıda bulunmaktadır. enJSRV-6 nın polymorfizmi ise koyunların evcilleştirme sürecinde yeni bir olayın işareti olabileceğinden dolayı merak uyarmaktadır.

Anahtar kelimeler: Evcilleştirme, Türk koyun ırkları, Anadolu yaban koyunu, içsel Jaagsiekte koyun retrovirüsleri (enJSRVs), retrotip.

To My Family

ACKNOWLEDGEMENTS

At first, I would like to thank to my supervisor Prof. Dr. İnci Togan for her valuable supervision, advices, guidance, useful critics and patience throughout this study. She always encouraged and motivated me in the best way at every single step of this study.

Also, I would like to thank to my thesis committee members Prof. Dr. Okan Ertuğrul, Assoc. Prof. Dr. C. Can Bilgin, Assoc. Prof. Dr. Mayda Gürsel and Asisst. Prof. Dr. Emel Özkan Ünal for their valuable criticism and advices.

I would like to thank to my lab mates Hande Acar and H. Alper Döm who helped me to get used to lab environment at the first place. Also I would like to present my deepest gratitude to Dr. Evren Koban Baştanlar for the both spiritual and material support she provided me especially in the first steps of this study. I would like to thank my lab mate Sevgin Demirci for introducing me statistical software which I have used in this study. Also, I would like to thank my lab mates: Dr. Füsun Özer, N. Dilşad Dağtaş, Eren Yüncü, Dr. Arzu Karahan, Dr. Arif Parmaksız, Begüm Uzun and Elçin Ekşi for their friendships and valuable contributions they have made to this study.

I wish to thank to The Scientific and Technological Research Council of Turkey (TÜBİTAK) for the financial support.

Especially I want to thank Neşe Çakmak and Melike Sever not only for their friendships but also for the technical support they provided me in the process of the preparation of this thesis. Also I am deeply grateful to İsmail Cem Yılmaz, Gökay Sedef and Özden Baltekin for their friendships, encouragement and support.

I am deeply thankful to İlknur Akbulut for her patience, care, love and the support she provided me in every way at every single stage of this study.

At last, I want to thank my family: M. Kemal Ayanoğlu, Nalan Ayanoğlu and N. Cihat Ayanoğlu for their unconditional infinite love and support. They were always there for me to motivate me in a way which I could not even imagine at the every step I stumbled in the process of this study.

Turkish sheep breed samples used in that study were provided by Scientific and Technical Research Council of Turkey (TÜBİTAK) as a part of the project In Vivo Conservation and Preliminary Molecular Identification of Some Turkish Domestic Animal Genetic Resources–I (TURKHAYGEN-I).

Anatolian wild sheep samples were provided by the General Directorate of Nature Conservation and National Parks, Ministry of Forestry and Hydraulic Works.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

%: Per Cent

°C: Degrees Celsius

AKK: Akkaraman

Arlequin: An Integrated Software Package for Population Genetics Data Analysis

BLAST: Basic Local Alignment Search Tool

bp: Base pair

BP: Before Present

DAG: Dağlıç

DNA: Deoxyribonucleic Acid

dNTP: Deoxyribonucleotide Triphosphate

enJSRV: Endogenous Jaagsiekte Sheep Retrovirus

env: Viral envelope gene

ERV: Endogenous Retrovirus

EtBr: Ethidium Bromide

$F_{ST:}$ The fixation index

GOK: Gökçeada

H: Y Chromosome Linked Haplogroup

HEM: Hemşin

HPG: Mitochondrial DNA Linked Haplogroup

IVE: İvesi

KCl: Potassium Chloride

KDE: Kernel Density Estimation

KIV: Kıvırcık

Km: Kilometer

KRG: Karagül

KRY: Karayaka

LTR: Long Terminal Repeat

M: Molar

MDS: Multidimensional Scaling

$MgCl_2$: Magnesium Chloride

mM: Millimolar

MRK: Morkaraman

MSA: Multiple Sequence Alignment

mtDNA: Mitochondrial DNA

NCBI: National Center for Biotechnology Information

ng: Nanogram

NH_4: Ammonium

NOR: Norduz

OGA: *Ovis gmelinii anatolica*

PCR: Polymerase Chain Reaction

pmol: Pico moles

R: Retrotype

r: The Standardized Mantel Statistic

RNA: Ribonucleic Acid

SAK: Sakız

SNP: Single Nucleotide Polymorphism

TAGEM: Turkish General Directorate of Agricultural Research

Taq: *Thermus aquaticus*

TURKHAYGEN-I: In Vivo Conservation and Preliminary Molecular Identification of Some Turkish Domestic Animal Genetic Resources–I

UV: Ultra Violet

V: Volt

µl: Microliter

CHAPTER 1

INTRODUCTION

Neolithic agricultural revolution can be defined as the transition from hunter-gatherer life style to animal husbandry and agricultural system. One of the main events of that transition is the human control over the reproduction and protection of animals, for the benefit of human kinds, which is called animal domestication. Animal domestication was initiated approximately 11,000 years ago in the Fertile Crescent which covers Southeastern Anatolia (Zeder, Emshwiller, Smith, & Bradley, 2006; Zeder, 2008). The present study contributes to the studies related with history of sheep domestication. In the study, molecular genetic markers and different statistical analyses were employed. The DNA samples of both Turkish native breeds and wild sheep (*Ovis gmelinii anatolica*) were used as the genetic material in this study.

1.1 Center of sheep domestication, first and second migration of domestic sheep

Classical definition of animal domestication has been stated by Bökönyi as: "… the capture and taming by man of animals of a species with particular behavioral characteristics, their removal from their natural living area and breeding community and their maintenance under controlled breeding conditions for profit." (Bökönyi, 1969). The term "for profit" may not be suitable for Neolithic context but it is possibly used by Bökönyi in order to exclude animals which have been domesticated as companions or pets.

More recent and inclusive definition of domesticated animal by Clutton-Brock was given as: "A domesticated animal is one that has been bred in captivity, for purposes of subsistence or profit, in a human community that maintains complete mastery over its breeding, organization of territory and food supply." (Brock, 1994)

There are different definitions of the term "breed". In the context of this study breed can be defined as: group of domestic livestock which share phenotypic traits which are well defined and identified so that they can be distinguished from other groups of animals within the same species. Turkey harbors at least 15 different sheep breeds all of which are defined

in detail at the catalogue (Native Animal Genetic Resource of Turkey) of Republic of Turkey Ministry of Food, Agriculture and Livestock (www.tarim.gov.tr). The breeds that were used in this study will be defined as a summary at section 1.2 of this chapter.

Another term which is used for some specific sheep breeds is called primitive sheep breed. Primitive sheep breeds possess some characteristics of their related wild sheep. It can change according to different breeds but presence of large horns, sometimes in both of the males and females, or short tails can be given as examples for the distinguishable characteristics of primitive breeds. In addition to their phenotypic resemblances to wild sheep, they are genetically closer to related wild sheep than other well defined and captivated domestic sheep breeds.

One of the primary questions about animal domestication is when and where wild animals were transformed into herded livestock. The interest in this question is partly because as well as the wild ancestors, early domestic herds and primitive sheep might be harboring unique genetic diversity which is not found in commercial domestic modern sheep breeds of today. These genetic sources, if they can be identified, may have a prime importance in conservation studies. To answer that question archaeozoologists used methods which rely on morphological changes until late 1990s (Zeder, Emshwiller, Smith, & Bradley, 2006). One of the most commonly accepted morphological marker for domestication was the drastic reduction in overall body size among populations composed of domesticated individuals as traced by archeological observations (Uerpmann, 1979; Meadow, 1989). According to this criterion established consensus was that animal domestication began with goats and then sheep approximately at 10,000 – 9,500 B.P. in the southern Levant (Uerpmann, 1979; Bar-Yosef & Meadow, 1995).

Then, it was realized that sex is the primary factor that affects body size of sheep and goat. There was a consistent difference between males and females where males are larger than females. In addition to sex difference, environmental conditions also affect body size: as the temperature and aridity increases smaller sized animals increase in number too (Zeder, Emshwiller, Smith, & Bradley, 2006).

The original interpretation that there is an association between body size reduction and initial domestication can be explained partly by the differences in the selection for culling of hunters and herders. In other words, archeological assemblages that have been generated by hunters are dominated by larger animals because hunters prefer large adult animals to get maximum outcome. On the other hand, in the archeological assemblages of herders are dominated by bones of smaller females which have been slaughtered after a few reproductive years. Furthermore, since excess males are not needed in herder populations in which males were harvested at young ages. For that reason, bones of adult males are not well represented in the archeological assemblages of herder populations (Zeder, 2008).

In recent years (Zeder, 2008), instead of the morphologic evidences, presence of distorted sex ratio (females outnumbering males) and ages of animals from different sexes (many old females but very few old males) were considered as the indicators of the herd management,

hence domestication. Furthermore, small sample atomic mass spectrometry (AMS) radiocarbon dating was used to identify the earliest site of first herd management. Thereby, center of domestication for an animal could be suggested. With the advancement of this new technology (AMS radiocarbon dating), the origins of domestication for sheep, goat, cattle and pigs are suggested to be residing mainly in south-eastern Anatolia (Zeder, 2008). Domestication centers for those four livestock species and their early dispersals in the Fertile Crescent are demonstrated in the map which is given in Figure 1.1

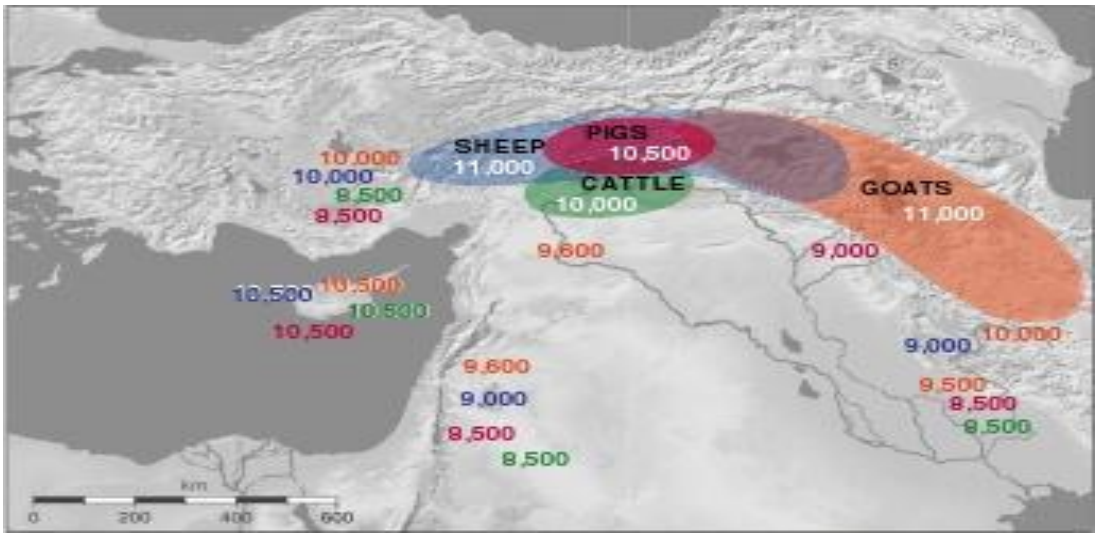

Figure 1.1 Domestication centers and early dispersal of domesticated livestock species in the Fertile Crescent. All dates are approximate times in B.P. Shaded areas indicates the suggested locations for initial domestication. Areas outside of the shaded areas demonstrate the approximate time for the first time domesticates that appeared in that specific location. Orange stands for goats, blue stands for sheep, green stands for cattle and fuscia stands for pigs. Map is retrieved from Zeder's (2008) study.

When compared to other livestock species (goats, pigs and cattle) sheep seems to be the initial focus for the transition from hunting to herding in Southeastern Anatolia approximately 10,500 B.P. (Peters, von den Driesch & Helmer, 2005). Goats show similar patterns and timing with sheep so it is suggested that both sheep and goats were domesticated together around between 11,000 B.P. and 10,500 B.P. in the region starting to from southern Anatolia and reaching to northern Zagros (Figure 1.1).

In addition to determining the location and time of domestication events, dispersion patterns of the domesticated animals are also one of major interests because it can help to understand

the similarities between different sheep breeds of which information can be assessed for conservative purpose. In addition to conservative studies dispersion patterns of domesticated animals may contribute to understanding of human migration events because migrations of animals and humans might have happened together. For the last 3 decades different models have been suggested to explain the Neolithic expansion across the Mediterranean Basin.

In early 1990s different excavations in Cyprus enabled us to have a new understanding about the Neolithic expansion in Mediterranean Basin (Guilaine, 2003). Before this discovery it was thought that Cyprus has been colonized around 8,500 B.P. (LeBrun *et al.*, 1987). Then, it was observed that these excavation sites in Cyprus date between 10,500 and 9,000 B.P. which suggest the arrival of early pioneer instead of fully established Neolithic mainland cultures. (Guilaine, 2003; Peltenburg, 2004). It is believed that these seafaring colonists transported all four major livestock species: sheep, goat, pig and cattle in their full wild size (but they must have been managed because they were travelling together with humans) to Cyprus because none of those species were present in Cyprus before this date. This colonization event provides valuable information about the dispersion of Neolithic package across the Mediterranean Basin.

In the light of this new discovery, Neolithic expansion to west is believed to be following two major routes: First one is the Maritime route which starts from Cyrus and continues all along the Mediterranean coast. It is believed that sheep has been carried to Western Europe by this route. For example sheep are thought to be first observed in Iberian Peninsula approximately 7,700-7,400 B.P. (Zilhão, 2001). Second migration route is thought to be through Danube Valley (Dobney and Larson, 2006; Zeder, 2008).

The direction of early migrations is not just from domestication centers to Europe. Also migrations to North Africa (Barker, 2002) and to Euroasia (Price, 2000) were documented. In the present study, migration events which are in the context of Neolithic expansion originated from domestication center and outspread into all directions are called as “the first migration of sheep”

It is believed that initially sheep were maintained primarily for meat but then specialization for secondary products presumably wool gained importance around 5,000 B.P. in Southwest Asia. It is suggested that selection and specialization for desired secondary products first took place in Southwest Asia and expanded to Europe, Africa and the rest of the Asia. (Chessa *et al.*, 2009). It was suggested that approximately five millenniums after the first domestication and expansion there was a second mass migration of sheep based on secondary products which is called “the second migration of sheep” whose extent of products is observed in nearly all of the modern domestic sheep of today.

As a summary, it may be anticipated that domestic sheep of Turkey might be mainly the product of the second migration with some remnants of the first migration of sheep. General characteristics of Turkish sheep breeds have been described in the catalogue of Turkish General Directorate of Agricultural research (TAGEM) (TAGEM, 2009). Distributions and

tail types of Turkish sheep breeds will be mentioned in Section 2.1. For detailed information refer to the catalogue.

1.2 Populations of wild sheep (*Ovis gmelinii*) in Anatolia

Ovis gmelinii exists as 2 different subspecies in modern-day Turkey: Anatolian wild sheep (*Ovis gmelini anatolica*) and Armenian Mouflon *(Ovis gmelinii gmelinii*). These two subspecies are geographically separated and there is no evidence that these two subspecies were connected for couple of hundred years. While Armenian Mouflon present in Eastern Turkey and northwestern Iran, Anatolian wild sheep located in Konya Bozdağ region which is in the central part of Turkey is an endemic subspecies to Turkey. Former and present distributions of these two subspecies are given in Figure 1.2 as reviewed by Arıhan (2000).

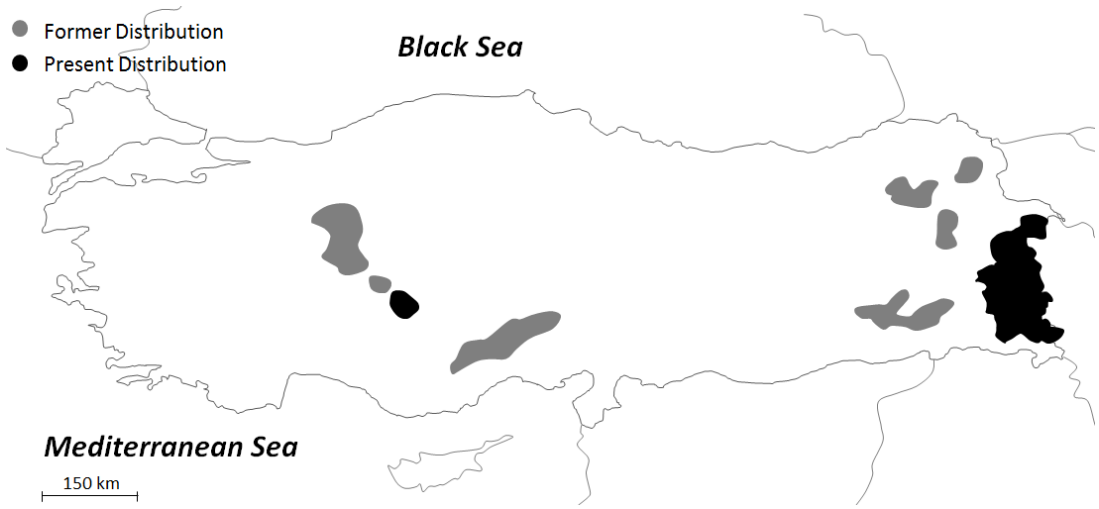

Figure 1.2 The former (shaded with grey) and the present (shaded with black) distribution of *Ovis gmelinii* subspecies. The map was taken from Arıhan's (2000) study.

Anatolian wild sheep is thought to had a wider range (approximately 50,000 sq. km.) before 20th century (Danford & Alston, 1877; Turan, 1984; Kaya, Bunch & Konuk, 2004) but at the beginning of the 20th century populations of Anatolian wild sheep started to decrease in number and this decrease led to extinction of Anatolian wild sheep populations between 1940-1970 except one population (with a size of as low as 35) which is located in Bozdağ, Konya (Arıhan, 2000; Sezen, 2000).The population in Bozdağ region has been under protection by a wild life protection program since 1966 (Sezen, 2000).

The distinctive feature of Anatolian wild sheep is that females have no horns like modern domesticated sheep. Due to this feature and because of the mtDNA haplogroup based studies it has been suggested that *Ovis gmelinii anatolica* may be one of the ancestors of the domesticated sheep (Hiendleder *et al.*, 2002). In this study 30 samples from Bozdağ population were used to determine retrotype frequencies (will be explained in the following section) by using endogenous retroviruses.

1.3 Endogenous Jaagsiekte sheep retroviruses (enJSRVs) as highly informative genetic markers.

Family of endogenous retroviruses (ERVs) has been examined to reveal history of domestic sheep in this study. The stable integration of provirus, which is retrovirus genome, in the genome of the host results in the formation of these ERVs (Boeke & Stoye, 1997). The insertion of the retrovirus genome into the host genome can be regarded as an insertional mutation. These ERVs are transmitted vertically from generation to generation in a Mendelian fashion.

At least 27 copies of ERVs related to exogenous and pathogenic Jaagsiekte retroviruses which are called endogenous Jaagsiekte retroviruses (enJSRVs) are present in sheep genome (Palmarini, Sharp, de las Heras, & Fan, 1999, Arnaud *et al.*, 2007; Arnaud, Varela, Spencer, & Palmarini, 2008). Some enJSRV loci are insertionally polymorphic while most of them are fixed in domestic sheep (Arnaud *et al.*, 2007). An irreversible single integration event of a retrovirus genome into host genome results in the presence of corresponding enJSRV in the host genome. Since these proviruses are vertically transmitted from generation to generation, it can be said that populations that harbor the same provirus at the same genetic location can be regarded as phylogenetically related. For those reasons enJSRVs can be used as highly informative genetic markers (Chessa *et al.*, 2009). Unlike many of the most commonly used markers for phylogenetic studies (e.g. mtDNA markers), enJSRVs are sex independent dominant nuclear markers.

Another important feature of these markers is about their inheritance. Insertionally polymorphic enJSRV loci which have been used for analyses in Chessa *et al.*'s (2009) study are on different chromosomes. These enJSRVs have been mapped by Fluorescent in situ hybridization (FISH) analysis. Only the location of enJSRV-7 locus could not be determined certainly because centromeric signal have been observed for this locus. In addition to enJSRV loci are present on different chromosomes; it has been shown that there is no sign of within population linkage ($p>0.858$) of these enJSRVs in the 65 groups which have been tested (Chessa *et al.*, 2009). Thus, it can be accepted that none of the possible insertionally polymorphic enJSRV pairs are significantly related with each other which assure that these polymorphic enJSRVs can be used as highly informative and independent markers. It can be inferred from these results that these insertionally polymorphic enJSRVs are independently inherited. Chromosomal locations of 5 insertionally polymorphic enJSRVs are given in Table 1.1.

Table 1.1 Chromosomal locations of 5 insertionally polymorphic enJSRVs (Chessa *et al.*, 2009)

enJSRV type	Chromosomal Location
enJSRV-8	3q21
enJSRV-15	6q13
enJSRV-16	10q24
enJSRV-18	11q17
enJS5F-16	15q23

1.4 enJSRVs in sheep genome, retrotypes assignments and their distribution in sheep breeds.

In Chessa *et al.*'s (2009) study 1362 animals from 133 domestic sheep breeds and the closest wild relatives were tested for the presence of the six insertionally polymorphic enJSRVs (enJSRV-7, enJSRV-8, enJSRV-15, enJSRV-16, enJSRV-18 and enJS5F16) which are inherited independently. The closest wild relatives of domestic sheep that used were Urial sheep (*Ovis vignei*), Asiatic Mouflon (*Ovis orientalis* or *Ovis* gmelinii) and Mediterranean Mouflon (feralized early domesticated *Ovis gmelinii* or *Ovis gmelinii*).

Among these 1362 animals frequencies of each insertionally polymorphic endogenous retroviruses are given in Table 1.2

Table 1.2 Frequencies of polymorphic enJSRVs among 1362 samples used in Chessa et al's (2009) study

Locus	Frequency
enJSRV-18	0.85
enJSRV-7	0.27
enJS5F16	0.30
enJSRV-15	0.03 – 0.05
enJSRV-16	
enJSRV-8	

While enJSRV-18 provirus has the highest frequency enJSRV-15, enJSRV-16 and enJSRV-8 seems to be very rare where enJSRV-7 and enJS5F-16 has a moderate frequency when compared to others among all 1362 samples. In addition to polymorphic retroviruses, enJSRV-6 also has been used in the study of Chessa et al (2009). This retrovirus is proposed to be fixed in all domestic sheep breeds so that its integration was used for DNA quality assessment in the study of Chessa et al (2009).

When retroviruses infect their hosts they duplicate their genome (RNA into DNA) by reverse transcription. Duplications of the 5' and 3' ends of the retrovirus genome lead to formation of two identical long terminal repeats (LTRs). Right after integration of retrovirus genome into the host genome 5' (proximal) and 3' (distal) LTRs of an endogenous retrovirus are expected to be identical but it can diverge by mutations over time at a similar rate of non-coding sequences.

In the light of this information, estimated time for the integration of a specific endogenous retrovirus can be determined by using the differences between proximal and distal LTRs of that endogenous retrovirus. The more differences between LTRs, the older the integration is.

In Chessa *et al.*'s (2009) study, it has been found that among all polymorphic enJSRVs, enJSRV-7 seems to be the oldest provirus because 5 nucleotides are different when proximal and distal LTRs of this provirus, which are 445 nucleotides long, are compared. On the other hand, all other insertionally polymorphic proviruses (enJSRV-8, enJSRV-15, enJSRV-16, enJSRV-18 and enJS5F16) have identical distal and proximal LTRs.

In addition to divergence of the LTRs of the endogenous retroviruses, presence of insertionally polymorphic retroviruses in Urial sheep, Asiatic and Mediterranean Mouflon have been used as an informative marker to predict the age of the provirus at the given genomic site.

Urial sheep is thought to be one of the closest living relatives of the domestic sheep and predicted to diverge from domestic sheep approximately 800,000 B.P. (Hernández Fernández *& Vrba*, 2005). On the basis of this information any enJSRV provirus that is shared by these two species may have integrated into the genome before the first domestication. Similarly Asiatic Mouflon which is accepted as the direct ancestor of the domestic sheep (Zohary, Tchernov & Horwitz, 1998; Hiendleder *et al.*, 2002) where Mediterranean Mouflon is believed to be small remained population of first domesticated sheep which readapted to feral life (Poplin, 1979; Hiendleder *et al.*, 2002)

Despite the fact that enJSRV-18 has the highest frequency (86%) in sheep, it has not been found in Urial sheep, the Mediterranean and the Asiatic Mouflon which are composed of 37 samples as total. Interestingly enJSRV-7 which has a lower frequency (27%) than enJSRV-18, it is found in 60% of Urial sheep, 86% of the Asiatic Mouflon and all of the Mediterranean Mouflon. These findings also confirm that enJSRV-7 is older than enJSRV-18 provirus and it might be the marker of first domestication. (Chessa *et al.*, 2009)

Based on the argument that among the above mentioned 6 enJSRVs in sheep genome, enJSRV-7 is the oldest polymorphic endogenous retrovirus, it was suggested that populations whose ancestors are the earliest domesticated sheep do not carry any polymorphic enJSRVs or only carry enJSRV-7.

Geographical dispersion of two major polymorphic loci can be summarized as: enJSRV-18 has a very high frequency in the old world with a very uniform distribution while enJSRV-7 has the highest frequency mostly in isolated island populations like Asiatic Mouflon (*Ovis orientalis* or *Ovis gmelinii*) and Cyprus Mouflon (*Ovis gmelini ophion*) which inhabits Cyprus. World-wide distributions of enJSRV-18 and enJSRV-7 are given in Figure 1.3 and 1.4 respectively. The distribution of other enJSRVs will be discussed later.

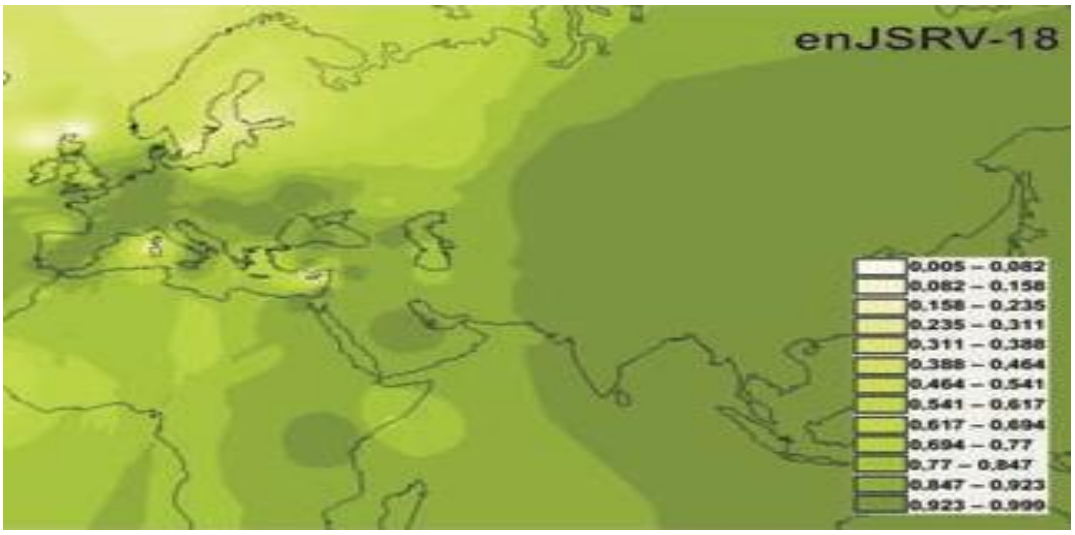

Figure 1.3 Spatial distribution of estimated enJSRV-18 frequencies for 133 breeds, which are divided into 65 groups, analyzed in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al.*'s (2009) study.

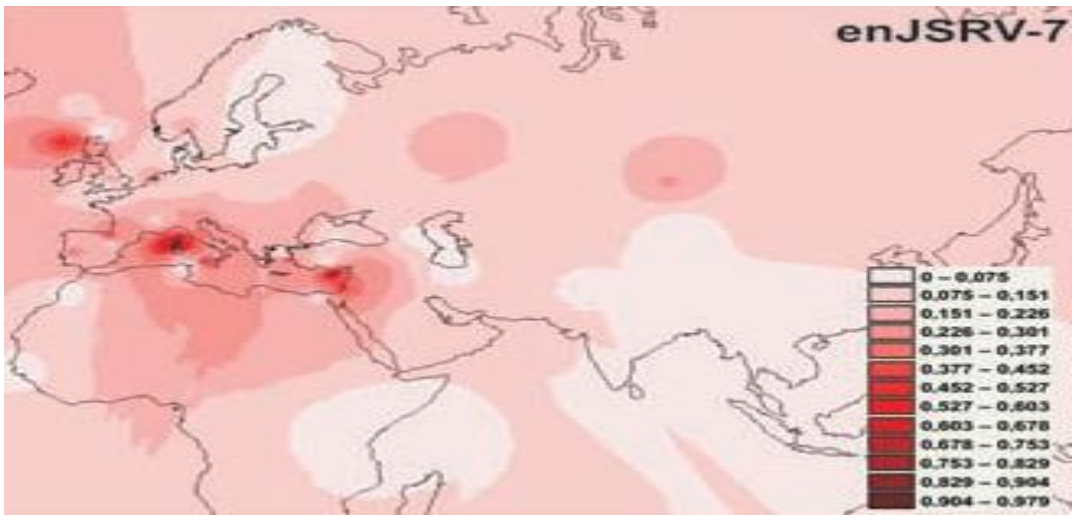

Figure 1.4 Spatial distribution of estimated enJSRV-7 frequencies for 133 breeds, which are divided into 65 groups, analyzed in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al.*'s (2009) study.

In Chessa *et al.*'s (2009) study, 65 groups which were formed from 133 different breeds have been analyzed according to frequencies of retrotypes. Retrotypes (R) are determined according to which enJSRVs are present in the genome of a specified sheep. There can be none or only one enJSRV in a sample as well as different combinations of enJSRVs can be present in one genome because they are independently inherited. With respect to presence of the enJSRVs, individuals were separated into 15 different retrotypes (R0-R14). Detailed list of all retrotypes are given in Table 1.4.

Table 1.3 Definition of retrotypes in terms of presence of enJSRVs (Chessa *et al.*, 2009)

Retrotype	Existing insertionally polymorphic enJSRVs
R0	None
R1	enJSRV-7
R2	enJSRV-18
R3	enJS5F-16
R4	enJSRV-7 + enJSRV-18
R5	enJSRV-7 + enJSRV-16
R6	enJSRV-18 + enJS5F16
R7	enJSRV-7 + enJSRV-18 + enJS5F16
R8	enJSRV-8
R9	enJS5F16 + enJSRV-8
R10	enJSRV-7 + enJS5F16 + enJSRV-8
R11	enJSRV-18 + enJSRV-8
R12	enJSRV-18 + enJS5F16 + enJSRV-8
R13	enJSRV-7 + enJSRV-18 + enJSRV-8
R14	enJSRV-7 + enJSRV-18 + enJS5F16 + enJSRV-8

Only eight of these retrotypes will be the concern of the present study but in order to understand the world-wide distribution of all retrotypes, all of them are defined (Table 1.3). World-wide distribution of retrotypes in the all samples (n=1362) used in Chessa *et al.*'s (2009) study are given in Figure 1.5.

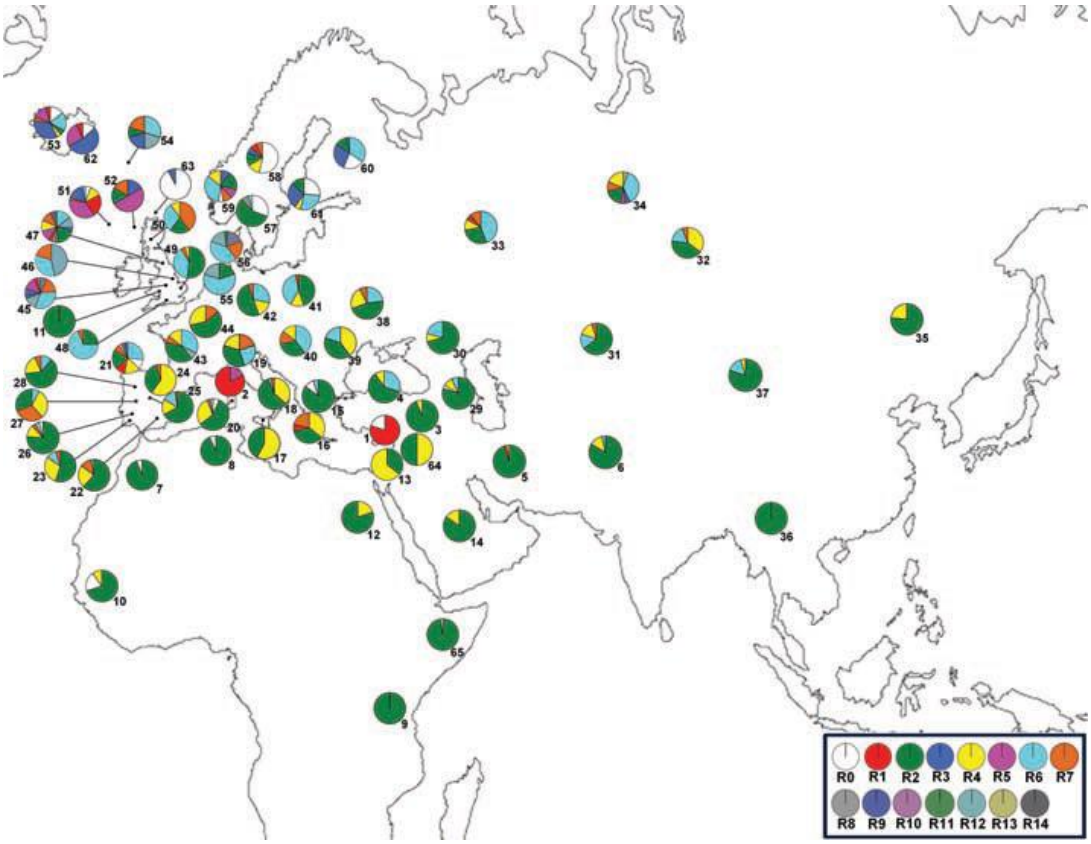

Figure 1.5 Frequencies of each retrotype for 133 breeds, which are presented in 65 groups, in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al.*'s (2009) study.

As it can be inferred from Figure 1.5, R2 (enJSRV-18 only) is the most dominant retrotype in the world especially in Asia and Africa. R4 (enJRV-7 + enJSRV-18) was another abundant retrotype especially in Iberian Peninsula, Southeastern Europe and Mediterranean Basin. R0 and R1 whose presences are the characteristic features of primitive breeds: Cyprus Mouflon and European Mouflon as well as in Scandinavian populations. These are

believed to be product of the first domestication and migration event. There are some interesting retrotypes like R3 (enJS5F16 only) is named as “Nordic” retrotype because it is only present in northern Europe. Other retrotypes will be discussed in coming chapters in more detail.

When the frequencies of modern Turkish breeds were examined R2 (enJSRV-18 only), R4 (enJSRV-7 + enJSRV-18) and R6 retrotypes (enJSRV-18 + enJS5F16) are present. Since Anatolia is known to be harboring the first sheep domestication center, furthermore, it is on the way of the second migration of sheep to Europe and again close to the origin of the second migration (Middle-East), it can be expected that there should be high retrotype diversity. However, there were very few retrotypes observed in the previous study (Chessa et al, 2009)

When a closer look is given to the samples examined from Turkey, it can be seen that only 2 groups of samples were taken into consideration. They are mainly located in middle part of the Turkey with a total sample size of 42. First group is the group number 3 in Figure 1.5 and consists of only 10 Karayaka samples. Second group (group number 2 in Figure 1.5) is a combination of samples from four different breeds: Akkaraman, Morkaraman, Hemşin and Tuj with a total sample size of 32. Thus, both the sample sizes and the number of breeds examined do seem to be far from representing the Turkish breeds.

1.5 Objectives of the study

Since the putative center of first sheep domestication center is near/in Southeastern Anatolia (Zeder, Emshwiller, Smith, & Bradley, 2006; Zeder, 2008), Turkish native breeds might be harboring the extents of first domesticated sheep. Furthermore, there might be a pattern in relation to first domestication within Turkey which might be useful in understanding migration routes of the first migration. In addition to that, proximity of Turkish native breeds to the predicted origin of second migration of domestic sheep makes this research interesting. Besides using native sheep, including a wild sheep (*Ovis gmelinii anatolica*) population for the purpose of analyzing the characteristics of the sheep populations using enJSRVs as a dominant genetic marker will further contribute to understanding of sheep domestication history and history of domestic sheep.

Summary of the objectives of this study can be stated as follows:

In this study 220 samples from 11 Turkish native sheep breeds and 30 samples from wild sheep (*Ovis gmelinii anatolica*) has been examined with respect to presence of insertionally polymorphic enJSRVs. Results obtained about the enJSRV integrations will be employed for the following purposes:

i. To determine the genotypes of the tested individuals on the basis of one or more insertionally polymorphic enJSRVs.
ii. To understand spatial distribution of retrotypes for Turkish native breeds and a wild population on a large scale which is expected to contribute to the understanding of sheep domestication in terms of first and second migration events by using data from a location which is very close to domestication center.
iii. To compare the results of nuclear dominant genetic marker (enJSRV) with sex dependent markers (mtDNA and Y-linked). The data for the latter are available. Comparative study is expected to reveal new insights about the sex of the migrated individuals.
iv. In general, to contribute to the understanding the evolutionary history of sheep in particular migration events.

CHAPTER 2

MATERIALS AND METHODS

2.1 Sheep and their samples

Within the framework of TURKHAYGEN-1 project (www.turkhaygen.gov.tr; Project no: 106G115) blood samples of 628 sheep from 13 different breeds were collected. Samples from 11 of these breeds (Karayaka, Dağlıç, Kıvırcık, Sakız, Akkaraman, İvesi, Norduz, Karagül, Hemşin, Gökçeada and Morkaraman) were employed in this study. Number of sheep individuals used in the present study their breed names and tail types are given in Table 2.1. The collection sites of the breed samples are shown in Figure 2.1

Table 2.1 Sample sizes of breeds from which blood samples were used

Breed	Tail Type	Sample size (n)
KARAYAKA	Thin & Long	20
DAĞLIÇ	Fat	20
KIVIRCIK	Thin & Long	20
SAKIZ	Semi-fat	20
AKKARAMAN	Fat	20
İVESİ	Fat	20
NORDUZ	Fat	20
KARAGÜL	Fat	20
HEMŞİN	Semi-fat	20
GÖKÇEADA	Thin & Long	20
MORKARAMAN	Fat	20
TOTAL		220

In addition to domestic sheep breeds, 30 samples from Anatolian wild sheep (*Ovis gmelinii anatolica*), which are located in Bozdağ region of Konya in Turkey were used for molecular

studies in the present study. Not all of these samples were informative in terms of enJSRV integrations of which reasons will be discussed in coming chapters.

Further information about breeds can be found in the catalogue of Turkish General Directorate of Agricultural research (TAGEM) (TAGEM, 2009).

Anatolian wild sheep blood samples were collected by the General Directorate of Nature Conservation and National Parks, Turkish Ministry of Forestry and Hydraulic works. DNA isolation from these samples was done mostly by Demirci (2012).

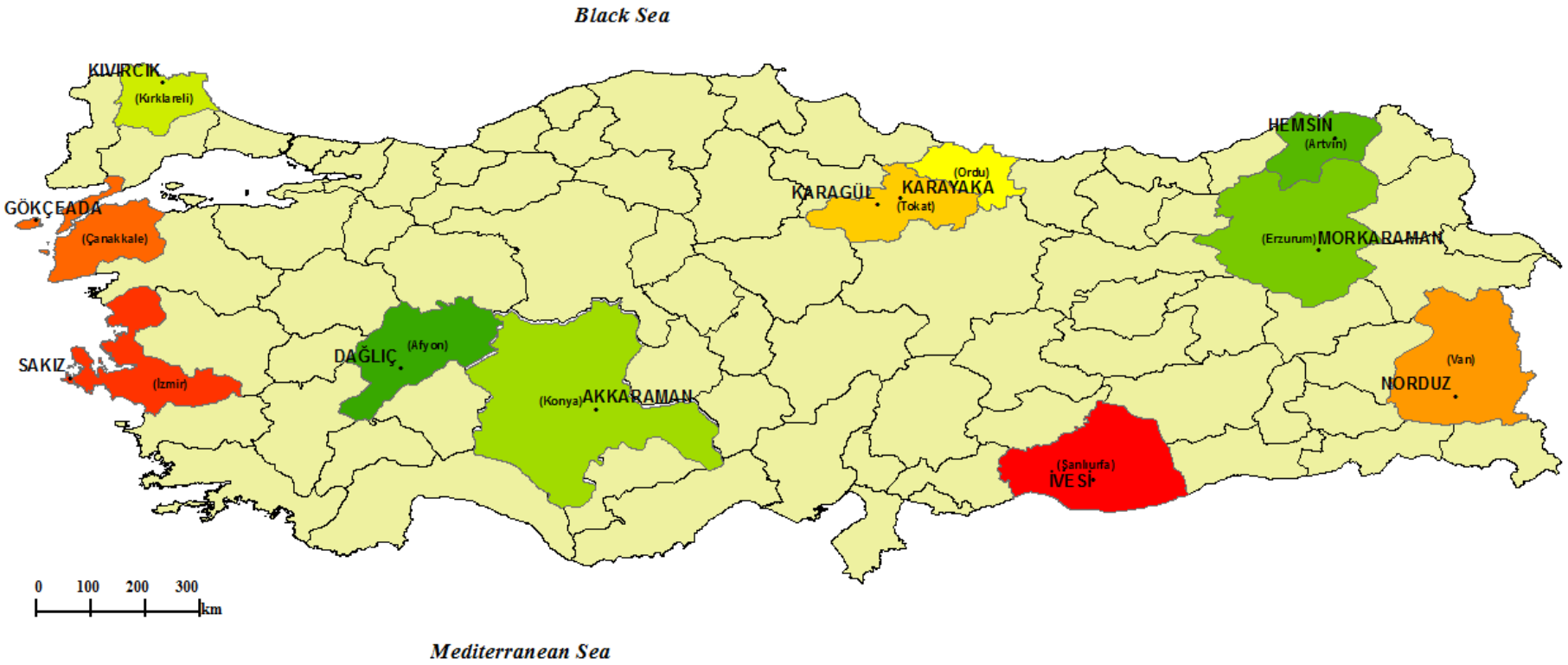

Figure 2.1 Sampling locations of breeds with respect to provinces on the map of Turkey. Map was constructed via ArcGIS 10 (ESRI) software

2.2 Laboratory experiments

2.2.1 DNA isolation from blood samples

DNA isolations were done by Doğan (2009), Yüncü (2009), Acar (2010) and Demirci (2012) in their studies. DNA isolations were performed by the modified version of standard phenol:chloroform isolation protocol (Sambrook, Foritsch & Maniatis, 1989) as given by (Koban, 2004). The modified isolation procedure is explained in detail in Acar's (2010) study.

The isolated DNA samples were checked both qualitatively and quantitatively by using Thermo Scientific NanoDrop 2000c spectrophotometer.

2.2.2 Amplification by Polymerase Chain Reaction (PCR)

2.2.2.1 General PCR strategy

The presence of enJSRV-6, -7, -18 and enJS5F-16 in each DNA sample was assessed by PCR at specific conditions which will be explained in the following sections.

Two sets of PCRs called 5'-PCR and 3'-PCR were performed for each DNA sample with primers which are specific to enJSRV integrations. All the primers which were used in this study were designed by Chessa *et al.* (2009).

The 5' PCR primers are named as 5'FlankF and ProvR. 5'FlankF primers were used as a forward primer and designed as complementary to the 5' flanking region which is adjacent to 5'LTR of corresponding provirus. As a reverse primer ProvR which is complementary to untranslated *gag* region was used

The 3' PCR primers are called ProvF and 3'FlankR. ProvF primer was used as a forward primer and it was designed at 3' end of *env* of the provirus. Reverse primer is 3'FlankR and it was designed at flanking region which is adjacent to 3'LTR of each specific retrovirus.

The two sets of primers used for the PCR amplifications and their relative locations on their integration sites are shown in Figure 2.2

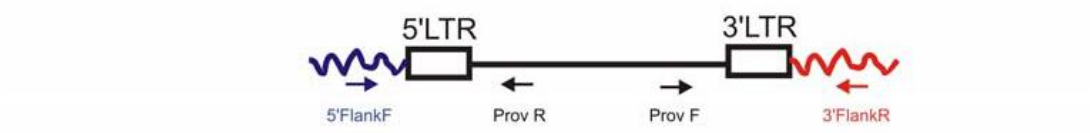

Figure 2.2 The two sets of primers used for the PCR amplification and their relative locations on the integration sites. Figure retrieved from Chessa *et al.*'s (2009) study.

2.2.2.2 PCR conditions for the amplification of integration sites

2.2.2.2.1 Amplification of enJSRV-6 locus

Parts of the enJSRV-6 locus in domestic sheep (EF680319), which is 5914 base pairs (bp) long, were amplified by two sets of PCRs (5'- and 3'- PCRs).

5' PCR was performed to amplify approximately 800-900 bp long part of the locus which starts from flanking region adjacent to 5' LTR and ends at 824th position of the locus. Sequences of the primers used for 5'PCR amplification are given below.

5'Flank Forward (5'FlankF): 5'-CCAGTTCCAGAAGGGAAGGGAG-3'

Provirus Reverse (ProvR): 5'-AGCCCTACAACCTGGGTTGCCCA-3'

Table 2.2 shows the components of the PCR mixture and Table 2.3 shows the 5'PCR conditions for amplification of the specified part of the enJSRV-6 locus.

Table 2.2 PCR mixture for 5'PCR of enJSRV-6

PCR buffer (KCl)	1.5 X
$MgCl_2$	2mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.3 PCR conditions for 5'PCR of enJSRV-6

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	35
Annealing	61^0C	45 seconds	
Extension	72^0C	50 seconds	
Final Extension	72^0C	10 minutes	1

3' PCR was performed to amplify approximately 600-700 bp long part of the locus which starts from the 5278^{th} position of the locus and ends at flanking region adjacent to 3' LTR. Sequences of the primers used for 3'PCR amplification are given below.

Provirus Forward (ProvF): 5'-AGCCTTCATTCACTGTGGCGA-3'

3'Flank Reverse (3'FlankR): 5'-CAGGGGAATAACTGGTGCTACCT-3'

Table 2.4 shows the components of the PCR mixture and Table 2.5 shows the 3'PCR conditions for amplification of the specified part of the enJSRV-6 locus.

Table 2.4 PCR mixture for 3'PCR of enJSRV-6

PCR buffer (KCl)	1 X
$MgCl_2$	2.5mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.5 PCR conditions for 3'PCR of enJSRV-6

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	30
Annealing	61^0C	35 seconds	
Extension	72^0C	45 seconds	
Final Extension	72^0C	10 minutes	1

2.2.2.2.2 Amplification of enJSRV-7 locus

Parts of the enJSRV-7 locus in domestic sheep (EF680298), which is 7941 base pairs (bp) long, were amplified by two sets of PCRs (5'- and 3'- PCRs).

5' PCR was performed to amplify approximately 500-600 long part of the locus which starts from flanking region adjacent to 5' LTR and ends at 493rd position of the locus. Sequences of the primers used for 5'PCR amplification are given below.

5'Flank Forward (5'FlankF): 5'-TGTGCACACGTGGTGGGAGTC-3'

Provirus Reverse (ProvR): 5'-GTAGTGGCGAGGAAAACTGTCGAGC-3'

Table 2.6 shows the components of the PCR mixture and Table 2.7 shows the 5'PCR conditions for amplification of the specified part of the enJSRV-7 locus.

Table 2.6 PCR mixture for 5'PCR of enJSRV-7

PCR buffer (KCl)	1.5 X
$MgCl_2$	2mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.7 PCR conditions for 5'PCR of enJSRV-7

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	20 seconds	35
Annealing	57^0C	30 seconds	
Extension	72^0C	45 seconds	
Final Extension	72^0C	10 minutes	1

3’ PCR was performed to amplify approximately 650-750 bp long part of the locus which starts from the 7302nd position of the locus and ends at flanking region adjacent to 3’ LTR. Sequences of the primers used for 3’PCR amplification are given below.

Provirus Forward (ProvF): 5’-AGCCTTCATTCACTGTGGCGA-3’

3’Flank Reverse (3’FlankR): 5’-AGGAACTCCAGGGTCGTGCCA-3’

Table 2.8 shows the components of the PCR mixture and Table 2.9 shows the 3’PCR conditions for amplification of the specified part of the enJSRV-7 locus.

Table 2.8 PCR mixture for 3’PCR of enJSRV-7

PCR buffer (KCl)	1 X
$MgCl_2$	2.5mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.9 PCR conditions for 3’PCR of enJSRV-7

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	30
Annealing	59^0C	35 seconds	
Extension	72^0C	45 seconds	
Final Extension	72^0C	10 minutes	1

2.2.2.2.3 Amplification of enJSRV-18 locus

Parts of the enJSRV-18 locus in domestic sheep (EF680301), which is 7941 base pairs (bp) long, were amplified by two sets of PCRs (5’- and 3’- PCRs).

5’ PCR was performed to amplify approximately 500-600 long part of the locus which starts from flanking region adjacent to 5’ LTR and ends at 493rd position of the locus. Sequences of the primers used for 5’PCR amplification are given below.

5’Flank Forward (5’FlankF): 5’-GGGAAGATTCGTTTCTTAGGCGCTC-3’

Provirus Reverse (ProvR): 5’-GTAGTGGCGAGGAAAACTGTCGAGC-3’

Table 2.10 shows the components of the PCR mixture and Table 2.11 shows the 5’PCR conditions for amplification of the specified part of the enJSRV-18 locus.

Table 2.10 PCR mixture for 5’PCR of enJSRV-18

PCR buffer (NH_4)	1.5 X
$MgCl_2$	2mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.11 PCR conditions for 5’PCR of enJSRV-18

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	35
Annealing	62^0C	45 seconds	
Extension	72^0C	55 seconds	
Final Extension	72^0C	15 minutes	1

3’ PCR was performed to amplify approximately 650-750 bp long part of the locus which starts from the 7305th position of the locus and ends at flanking region adjacent to 3’ LTR. Sequences of the primers used for 3’PCR amplification are given below.

Provirus Forward (ProvF): 5’-AGCCTTCATTCACTGTGGCGA-3’

3’Flank Reverse (3’FlankR): 5’-AGGAACTCCAGGGTCGTGCCA-3’

Table 2.12 shows the components of the PCR mixture and Table 2.13 shows the 3'PCR conditions for amplification of the specified part of the enJSRV-18 locus.

Table 2.12 PCR mixture for 3'PCR of enJSRV-18

PCR buffer (KCl)	1 X
$MgCl_2$	2.5mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.13 PCR conditions for 3'PCR of enJSRV-18

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	30
Annealing	60^0C	30 seconds	
Extension	72^0C	40 seconds	
Final Extension	72^0C	10 minutes	1

2.2.2.2.4 Amplification of enJS5F-16 clone

Parts of the enJS5F-16 enJSRV-like clone in domestic sheep (AF136224), which is 6916 base pairs (bp) long, were amplified by two sets of PCRs (5'- and 3'- PCRs).

5' PCR was performed to amplify approximately 500-600 long part of the clone which starts from flanking region adjacent to 5' LTR and ends at 494^{th} position of the clone. Sequences of the primers used for 5'PCR amplification are given below.

5'Flank Forward (5'FlankF): 5'-GGATAAGCTACACTATAAAACCAAAG-3'

Provirus Reverse (ProvR): 5'-GTAGTGGCGAGGAAAACTGTCGAGC-3'

3’ PCR was performed to amplify approximately 650-750 bp long part of the clone which starts from the 6279th position of the clone and ends at flanking region adjacent to 3’ LTR. Sequences of the primers used for 3’PCR amplification are given below.

Provirus Forward (ProvF): 5’-AGCCTTCATTCACTGTGGCGA-3’

3’Flank Reverse (3’FlankR): 5’-CCATATGTAGGGATTGGGGGGTG-3’

The PCR mixture and PCR conditions of 5’ and 3’ PCRs for enJS5F-16 were identical. Table 2.14 shows the components of the PCR mixture and Table 2.15 shows the PCR conditions for amplifications of the specified parts of the enJS5F-16.

Table 2.14 PCR mixture for enJS5F-16

PCR buffer (NH_4)	1.5 X
$MgCl_2$	2mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.15 PCR conditions for enJS5RF-16

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	25 seconds	30
Annealing	57^0C	45 seconds	
Extension	72^0C	55 seconds	
Final Extension	72^0C	10 minutes	1

2.2.3 Visualization of amplified PCR products

Amplified PCR products were run on 2% agarose gel which contains EtBr at 120 Volts for 35 minutes. After agarose gel electrophoresis procedure, visualization of agarose gels was performed by using Vilber Lourmat CN-3000WL instrument.

While running samples on agarose gel, positive controls were also run for enJSRV-7 and enJS5F-16 integrations because these integrations are rare for our breeds. When there are few or none amplified DNA products for those integration sites, the presence of positive controls ensures that this result is not due to PCR itself. To ensure that the amplified DNA fragments are specific, DNA ladders (Gene Ruler 100bp) were used at all gels to determine the size of amplified DNA

2.3 The procedure for data interpretation

2.3.1 Determination of the presence of a specific enJSRV

The presence of specific enJSRV integration was validated by the positive results for both of the PCRs (5'- and 3'- PCRs) for corresponding specific enJSRV. If there is no amplification or there is amplification only at one of the PCRs, it is regarded as there is no integration of that specific enJSRV for that individual (Chessa *et al.,* 2009). Only one exception to that was done for enJSRV-6 case: Some individuals have amplified product for only one of the PCRs either 5'- or 3'- PCR. The reason will be discussed in 4th Chapter.

2.3.2 Retrotype determination by using the combinations of different enJSRV integrations

In this study, we are interested in 8 retrotypes namely: R0, R1, R2, R3, R4, R5, R6 and R7. These retrotypes contain the enJSRVs that has been investigated in this study either as a single integration or as specific combinations of the integrations with the exception of R0. Retrotypes and their formations were discussed in Section 1.5 and summarized in Table 1.3.

2.4 Statistical Analyses

In order to represent the landscape of the retrotype distribution of Turkish native sheep breeds, pie charts representing the retrotype frequencies of the breeds were constructed. These pie charts were located on the centroids of the sampling sites of the breeds and shown on the map of Turkey via ArcGIS 10 (ESRI) software (Figure 3.5).

2.4.1 Construction of synthetic maps

Genetic trends of populations on geographical maps can be shown by synthetic maps. In order to construct synthetic maps, first pairwise genetic distances between populations are needed to be calculated. Then, this pairwise matrix should be reduced into a vector for each population. Finally by using these vectors synthetic maps can be constructed by density estimation or interpolation methods according to nature of the data.

2.4.1.1 Calculation of pairwise genetic distances between Turkish sheep breeds

As a first step of synthetic map construction, F_{ST} values (Wright 1965) were used to measure the pair-wise genetic differences between the breeds. It is a widely used measure developed to express the genetic differences between the populations (Allendorf & Luikard, 2007), It varies between 0 (when units, here breeds, can be considered as panmictic) and 1(when there is no gene flow between the breeds). Breeds are described with respect to their retrotype frequencies.

The F_{ST} values can be calculated by the formula which is given below:

$$F_{ST} = \frac{(H_T - H_S)}{H_T}$$

Equation 2.1

where H_T is the calculated heterozygosity from the overall retrotype frequencies based on the total population composed by the pair of breeds (Wright, 1951) and H_S is the average observed heterozygosity in breeds under construction.

In this study pairwise F_{ST} values were calculated from retrotype frequencies by using Arlequin 3.11 (Excoffier, Laval & Schneider, 2005) software.

2.4.1.2 Multidimensional scaling (MDS) analysis

As a second step multidimensional scaling (MDS) analysis was used to summarize pairwise genetic distances in one dimension (Gower, 1966). First the space constructed by

independent compound axes describing the genetic differences between the breeds is formed. Positions of the breeds on the first axis which covers the maximum portion of the variance that exist between breeds is used in the Kernel Density Estimation (KDE).

In this study metric MDS analysis were performed from pairwise F_{ST} values between populations by using R 2.14 software (http://www.r-project.org/).

2.4.1.3 Kernel Density Estimation (KDE)

At the last step Kernel Density Estimation (KDE) method was used to construct synthetic maps. Spatial patterns can be inferred by KDE. KDE calculates a circular area for a specified location "s" with a determined bandwidth (τ) by using quadratic kernel function which is called k value.

For a specified location (s), the center of the event has the highest value in kernel function because it is most contributed point in a location. The event is designated by s_i.

$\hat{\lambda}_\tau(s)$ which is the density for location s can be estimated by the formula (Gatrell *et al.*, 1996) given below.

$$\hat{\lambda}_\tau(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{(s-s_i)}{\tau}\right) \qquad \textbf{Equation 2.2}$$

Where k stands for the kernel function, s stands for the specified location, τ stands for the bandwidth, n stands for the sample size and s_i stands for the event.

In this study, centers of the collection sites of the breeds were used as s and values of the first dimension of the MDS analysis was used to estimate kernel densities.

Synthetic maps by Kernel density estimation method was constructed by ArcGIS 10 software. To observe the density distribution of Turkish breeds in terms of their genetic differences two different maps were constructed where the whole area were asked to be divided into either 2 or 3 classes. For both of the maps k value is 2.15.

In synthetic maps kernel surfaces with same density is shown with the same color so breeds which are genetically similar is expected to have the same density color on the map.

2.4.2 Mantel's test

Mantel's (1967) test is an approach that is used to evaluate correlation between dissimilarity (distance) matrices. In Mantel's test variables are presented in dissimilarity matrices in which pairwise genetic similarities among sample locations were summarized or they are the geographic distances themselves. Aim of the Mantel's test is to determine if correlation in randomized matrices is similar to observed correlation.

Mantel's statistic is based on a simple cross-product term (Z) where the non-redundant elements of the two matrices are considered (excluding the diagonal) and this term can be calculated by the formula given below:

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} y_{ij}$$

Equation 2.3

This equation is normalized by using standardized equation where variable of different measurement units can be considered within the same framework. The standardized Mantel statistic (r) is calculated by this standardized and normalized equation like Pearson correlation coefficient between the two matrices.

The normalized equation is given below:

$$r = \frac{1}{(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(x_{ij} - \bar{x})}{s_x} \cdot \frac{(y_{ij} - \bar{y})}{s_y}$$

Equation 2.4

where x and y are variables (or sets of variables) measured at locations i and j and n is the number of elements in the distance matrices, and the s_x and s_y are standard deviations for variable x and y.

The elements of a distance matrix are dependent. For that reason permutation procedures are used for evaluation of the test of significance for Mantel's test statistics. For those permuted matrices Mantel statistics are recomputed. By using high numbers of iterations the distribution of values for the statistic are generated. Generally ~1000 permutations used for α=0.05, ~5000 permutations used for α=0.01, ~10000 permutations used for greater precision (Manly, 1991)

In order to determine if there is a significant relationship between genetic distance and geographic distance, Mantel's test was performed by using R 2.14 (http://www.r-project.org/). Two different mantel tests were performed: 1) Mantel test statistics was calculated to see if pairwise F_{ST} values are significantly related to geographical distances and 2) Mantel test statistics was calculated to see if pairwise F_{ST} values are significantly related to the logarithm of pairwise geographical distance. Both of the Mantel's tests are based on 999 permutations.

2.5 Experiments and analyses for enJSRV-6 negative samples

2.5.1 Amplification of enJSRV-6 negative samples with enJSRV-6 empty locus primers

12 samples (5 from OGA population and 7 from different native sheep breeds) in which there are no enJSRV-6 integrations called "enJSRV-6 negative" were selected to be amplified with enJSRV-6 empty locus primers. These samples were discarded from the data previously because of being negative for enJSRV-6 integration. Selected samples for amplification of empty enJSRV-6 locus are given in Table 2.16

Table 2.16 Selected samples for the amplification of empty enJSRV-6 locus

Samples
OGA 1
OGA 2
OGA 3
OGA 4
OGA 11
KIV 30
DAG 14
KRG 2
AKK 11
SAK 14
SAK 20
NOR 13

A set of primers used for the PCR amplification of empty enJSRV-6 locus. The amplified region with empty locus primers are shown in Figure 2.3

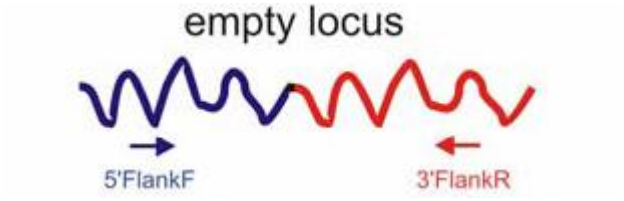

Figure 2.3 Set of primers which is used for the PCR amplification of empty enJSRV-6 locus. Figure retrieved from Chessa *et al.*'s (2009) study.

Primers that were used for this amplification are 5' Flank Forward (5'FlankF) and 3' Flank Reverse (3'FlankR) whose sequences are given in section 2.2.2.2.1.

2.5.2 PCR conditions for the amplifications of enJSRV-6 empty locus

Flanking regions (both 5' and 3') of the enJSRV-6 locus (EF680319) which is 5914 base pairs (bp) long were amplified by a set of PCR. The amplified product is expected to be around 700-800 bp long.

Table 2.17 shows the com components of the PCR mixture and Table 2.18 shows the PCR conditions for amplification of the empty enJSRV-6 locus.

Table 2.17 PCR mixture for PCR of empty enJSRV-6 locus

PCR buffer (NH_4)	1.5 X
$MgCl_2$	2.5mM
dNTP	0.5mM
Primer	10 pmol
DNA	30ng-60ng
Taq Polymerase	0.5 unit
Total volume	15µl

Table 2.18 PCR conditions for PCR of empty enJSRV-6 locus

Step	Temperature	Duration	Number of Cycles
Initial Denaturation	94^0C	3 minutes	1
Denaturation	94^0C	20 seconds	30
Annealing	60^0C	35 seconds	
Extension	72^0C	45 seconds	
Final Extension	72^0C	10 minutes	1

In those amplifications by PCR, 2 two samples (HEM 35 and MRK33) which were identified as enJSRV6 positive used as negative controls like distilled water (dH_2O).

2.5.3 Sequencing of empty enJSRV-6 locus

In order to get rid of remaining components of the PCR mix (excess primers, salts and genomic DNA), amplified products for empty enJSRV-6 locus were purified via Roche purification kit for PCR product with the provided protocol by the manufacturer.

The purified amplified products were sequenced by using chain termination technique (Sanger, Donelson, Coulson, Kossel, & Fischer, 1973). Forward (5'FlankF) and reverse (3'FlankR) primers were used for sequencing reactions. The reactions of sequencing were performed by the biotechnology company named as RefGen ((http://refgen.com.tr/). For the reactions ABI Prism BigDye Terminator™ device was used and sequencing data were collected via ABI Prism 3100 ™ DNA analyzer.

The sequences of the amplified products were retrieved from chromatograms and converted to FASTA format by using the software named Chromas Pro v. 1.5 (http:\\www.technelysium.com.au/ChromasPro.html/).

Exported sequences were aligned by using ClustalW algorithm (Thompson *et al.*, 1994) which is embedded in Bioedit software v.7.1.3 (Hall, 1997-2011). Any mismatches in the multiple sequence alignment (MSA) were corrected manually based on chromatograms of the sequences.

From 24 sequences for 12 samples (2 sequences for each sample) a consensus sequence was obtained for the amplified products for empty enJSRV-6 locus. Obtained sequence was analyzed by using Basic Local Alignment Search Tool (BLAST) (Altschul, Gish, Miller, Myers, & Lipman, 1990) embedded in the website of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/).

2.5.4 Amplification of polymorphic enJSRV integrations for enJSRV-6 negative samples from Ivesi and Karayaka breeds.

20 enJSRV-6 negative samples (10 from Karayaka and 10 from İvesi breed) were amplified by PCR for empty enJSRV-6 locus with the procedure which is explained in sections 2.5.1 and 2.5.2.

Samples in which positive results were obtained for the amplification of empty enJSRV-6 locus were amplified for enJSRV-7, enJSRV-18 and enJS5F-16 integrations by PCR according to conditions explained in section 2.2.2.2.2, 2.2.2.2.3 and 2.2.2.2.4 respectively.

CHAPTER 3

RESULTS

3.1 DNA isolation and PCR amplification

DNA isolation from blood samples was performed by standard phenol:chloroform method (Sambrook, Foritsch & Maniatis, 1989). The isolated DNA samples were checked both qualitatively and quantitatively via Nanodrop instrument, which is a micro-volume spectrophotometer.

The isolated DNA samples were amplified by Polymerase Chain Reaction (PCR) for four specific endogenous retrovirus integration sites separately. Optimization of the PCRs was difficult even for the most frequent integration (enJSRV-6). For each of the integrations breed specific modifications were done. For example, 5' primers for enJSRV-6 integration were amplifying the DNA samples from Karayaka breed successfully at 61^0C annealing temperature with 1.5X KCl Taq Buffer but DNA samples from İvesi breed were just been able to be amplified at 64^0C annealing temperature with 1X NH_4 Taq buffer.

Amplified PCR products were run on 2% agarose gel which contains EtBr and visualization of agarose gels were performed by using Vilber Lourmat CN-3000WL device. For enJSRV-7 and enJS5F-16 integrations positive controls were used because these integrations are relatively rare (but they were more common than enJSRV-8 and enJSRV-15) for our breeds. When there are few or none amplified DNA products for those integration sites, these positive controls ensures that this result is not due to PCR itself but it is because there are no integration at all for those no-product samples. To ensure that the amplified DNA fragments are specific, DNA ladders were used to determine the size of amplified DNA. Figure 3.1, Figure 3.2, Figure 3.3 and Figure 3.4 demonstrate the example gel photos for enJSRV-6, enJSRV-7, enJSRV-18 and enJS5F-16 integrations respectively.

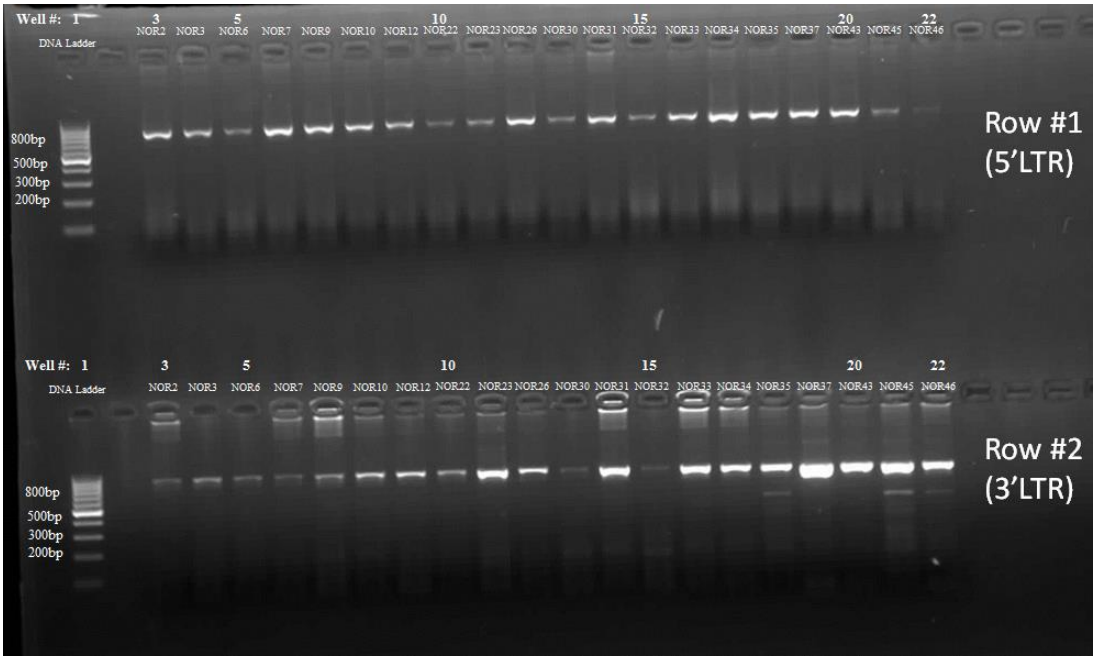

Figure 3.1 Visualization of enJSRV-6 integrations on the agarose gel. Details of the gel picture are explained in the text.

The first row contains the amplified products for the 5' LTR region of enJSRV-6 locus and the second one contains the amplified products for the 3' LTR region of enJSRV-6 locus. For each of the row: the first well contains the DNA ladder (Gene Ruler 100bp) and the next 20 wells (from well 3rd to 22nd) contain 20 samples from Norduz breed.

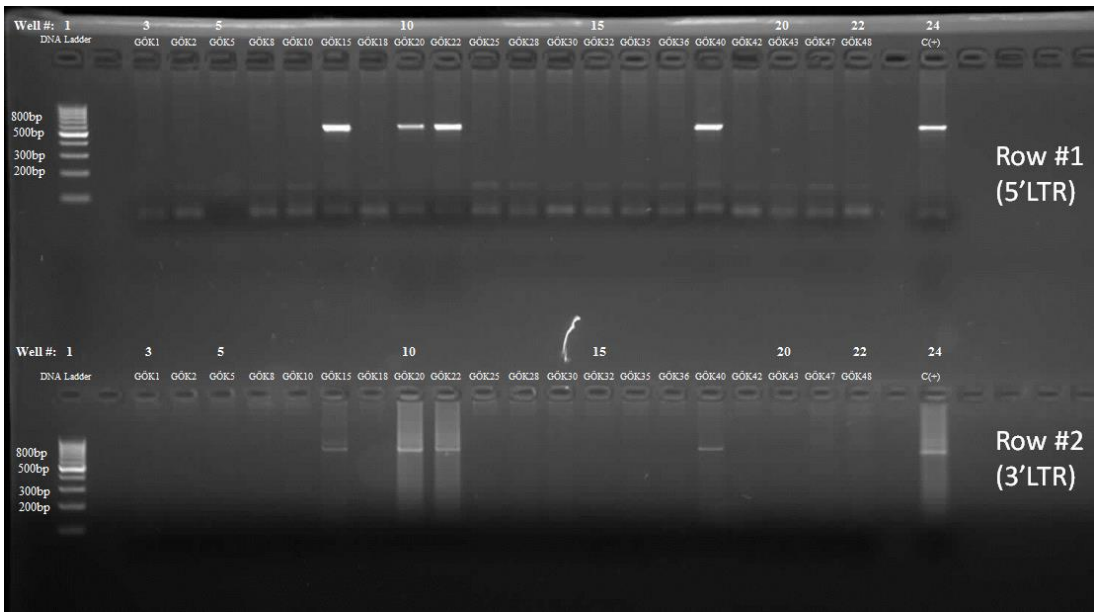

Figure 3.2 Visualization of enJSRV-7 integrations on the agarose gel. Details of the gel picture are explained in the text.

The first row contains the amplified products for the 5' LTR region of enJSRV-7 locus and the second one contains the amplified products for the 3' LTR region of enJSRV-7 locus. For each of the row: the first well contains the DNA ladder (Gene Ruler 100bp) and the next 20 wells (from well 3rd to 22nd) contain 20 samples from Gökçeada breed. The last well (24th) at the end is the positive control and contains an amplified product of a sample which was identified as positive for enJSRV-7 before.

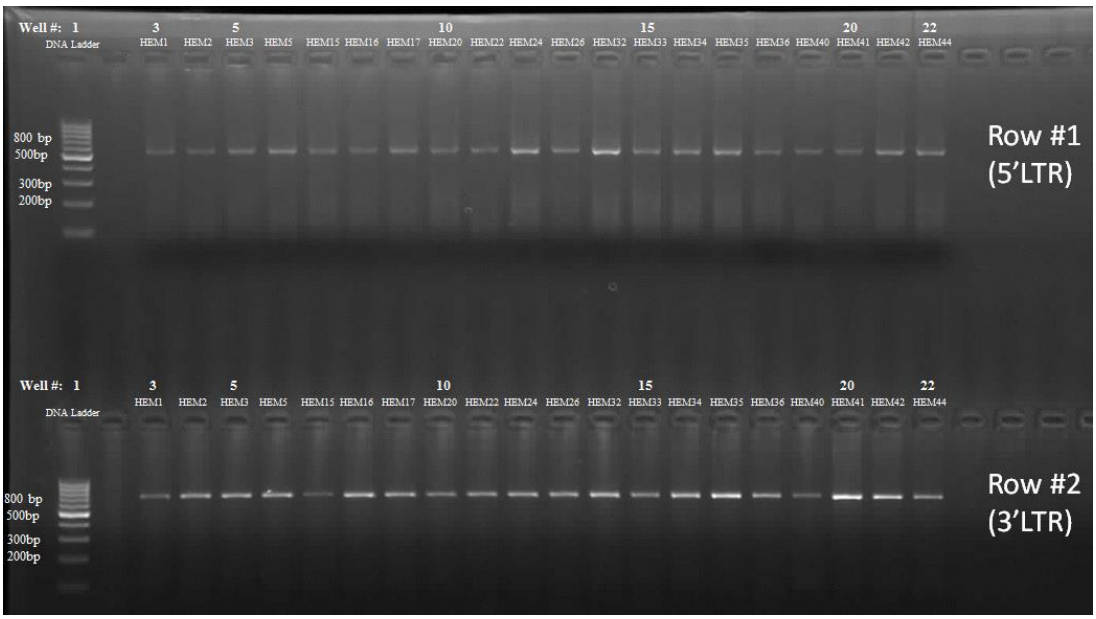

Figure 3.3 Visualization of enJSRV-18 integrations on the agarose gel. Details of the gel picture are explained in the text.

The first row contains the amplified products for the 5' LTR region of enJSRV-18 locus and the second one contains the amplified products for the 3' LTR region of enJSRV-18 locus. For each of the row: the first well contains the DNA ladder (Gene Ruler 100bp) and the next 20 wells (from well 3rd to 22nd) contain 20 samples from Hemşin breed.

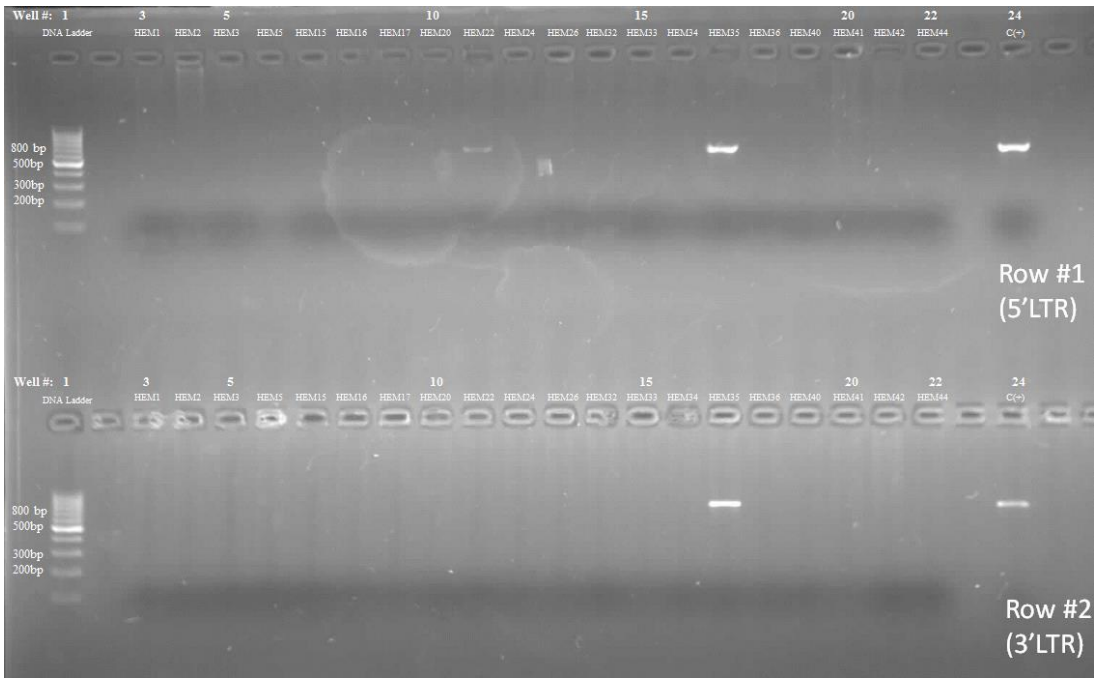

Figure 3.4 Visualization of enJS5F-16 integrations on the agarose gel. Details of the gel picture are explained in the text.

The first row contains the amplified products for the 5' LTR region of enJS5F-16 locus and the second one contains the amplified products for the 3' LTR region of enJS5F-16 locus. For each of the row: the first well contains the DNA ladder (Gene Ruler 100bp) and the next 20 wells (from well 3rd to 22nd) contain 20 samples from Hemşin breed. The last well (24th) at the end is the positive control and contains an amplified product of a sample which was identified as positive for enJSRV-7 before.

3.2 Endogenous Jaagsiekte sheep retrovirus (enJSRV) integration polymorphism in Turkish sheep breeds

3.2.1 enJSRV integrations in Turkish native sheep breeds

The presence of specific enJSRV integration was validated by the positive results for both of the primer pairs for corresponding specific enJSRV. If there is no amplification or amplification only for one pair, it is regarded as there is no integration of enJSRV which was under examination (Chessa *et al.,* 2009). More than one enJSRV integrations can be present in one sample.

11 native Turkish sheep breed were examined for the presence of four different enJSRV integrations which are enJSRV-6, enJSRV-7, enJSRV-18 and enJS5F-16. The distribution of enJSRV integrations in eleven breeds is represented in Table 3.1. The frequencies of each type of the integration for eleven breeds are represented in the Table 3.2.

Table 3.1 Distribution of enJSRV integrations in Turkish native sheep breeds

Name of breed	Sample size	enJSRV-6	enJSRV-7	enJSRV-18	enJS5F-16
KARAYAKA	20	20	3	19	2
DAĞLIÇ	20	20	2	19	1
KIVIRCIK	20	20	5	19	3
SAKIZ	20	20	5	17	1
AKKARAMAN	20	20	2	20	2
İVESİ	20	20	2	20	1
NORDUZ	20	20	2	18	1
KARAGÜL	20	19	1	18	0
HEMŞİN	20	20	4	20	1
GÖKÇEADA	20	20	4	18	0
MORKARAMAN	20	20	2	20	2
TOTAL	220	219	33	208	14

Table 3.2 Frequencies of enJSRV integrations for Turkish native breeds

Name of breed	Sample size	enJSRV-6	enJSRV-7	enJSRV-18	enJS5F-16
KARAYAKA	20	1	0.15	0.95	0.1
DAĞLIÇ	20	1	0.1	0.95	0.05
KIVIRCIK	20	1	0.25	0.95	0.15
SAKIZ	20	1	0.25	0.85	0.05
AKKARAMAN	20	1	0.1	1	0.1
İVESİ	20	1	0.1	1	0.05
NORDUZ	20	1	0,1	0.9	0.05
KARAGÜL	20	0.95	0.05	0.9	0
HEMŞİN	20	1	0.2	1	0.05
GÖKÇEADA	20	1	0.2	0.9	0
MORKARAMAN	20	1	0.1	1	0.1

In these experiments we did not use samples if they were not amplified for enJSRV-6 integration. The criterion was set by Chessa *et al.* (2009). Therefore, it is present in all 220

samples except one sample from Karagül breed as it can be observed from the distribution and frequencies of enJSRV integrations. Among the three insertionally polymorphic enJSRVs, enJSRV-18 integration is the most abundant one for Turkish breeds. enJSRV-7 and enJS5F-16 are less common when compared to other two.

enJSRV integrations for each sample with their sampling locations are given in Appendix A.

3.2.2 enJSRV integrations in Anatolian wild sheep (*Ovis gmelinii anatolica)* population

30 samples from Anatolian wild sheep (*Ovis gmelinii anatolica)* were amplified for the enJSRV-6 integration. Only 7 of 30 samples were positive for the enJSRV-6 integration. Afterwards, these 7 enJSRV-6 positive samples were examined for the presence of polymorphic enJSRV integrations (enJSRV-7, enJSRV-18 and enJS5F-16) like native breeds but no positive results were observed for those integrations. Therefore 7 samples have R0 (no insertionally polymorphic enJSRV integrations) retrotype while other 23 remained as unidentified when presence of enJSRV-6 integration is set as the criterion for DNA quality.

3.3 Spatial distribution of the retrotypes

3.3.1 Spatial Distribution of the retrotypes among Turkish breeds

Retrotype frequencies were calculated from enJSRV integration in 219 samples excluding 1 sample which was negative for enJSRV-6 integration and demonstrated in Table 3.3 Frequencies of retrotypes for all breeds were summarized by pie charts on the map of Turkey according to center of collection sites of breeds (Table 3.5). Figure 3.5 shows the retrotype profile of Turkey breeds. Pie charts on the map of Turkey were generated by using ArcGIS 10 (ESRI) software.

Table 3.3 Retrotype frequencies of Turkish sheep breeds

	R0	R1	R2	R3	R4	R5	R6	R7
KARAYAKA	0.0500	0.0000	0.7000	0.0000	0.1500	0.0000	0.1000	0.0000
DAĞLIÇ	0.0500	0.0000	0.8500	0.0000	0.0500	0.0000	0.0000	0.0500
KIVIRCIK	0.0000	0.0500	0.6500	0.0000	0.1500	0.0000	0.1000	0.0500
SAKIZ	0.1000	0.0500	0.6000	0.0000	0.2000	0.0000	0.0500	0.0000
AKKARAMAN	0.0000	0.0000	0.8000	0.0000	0.1000	0.0000	0.1000	0.0000
İVESİ	0.0000	0.0000	0.8500	0.0000	0.1000	0.0000	0.0500	0.0000
NORDUZ	0.1000	0.0000	0.7500	0.0000	0.1000	0.0000	0.0500	0.0000
KARAGÜL	0.0526	0.0000	0.8947	0.0000	0.0526	0.0000	0.0000	0.0000
HEMŞİN	0.0000	0.0000	0.7500	0.0000	0.2000	0.0000	0.0500	0.0000
GÖKÇEADA	0.0000	0.1000	0.8000	0.0000	0.1000	0.0000	0.0000	0.0000
MORKARAMAN	0.0000	0.0000	0.8000	0.0000	0.1000	0.0000	0.1000	0.0000

Table 3.4 Locations of breeds as centers of the site of collection

Breed	Latitude	Longitude
AKKARAMAN	37.916625	32.86665
DAGLIC	38.41444444	30.44777778
HEMSIN	41.19721364	42.03051364
IVESI	37.07	39.02833333
KARAGUL	40.394	36.349125
GOKCEADA	40.20446	25.93656
KARAYAKA	40.46729286	36.63696429
KIVIRCIK	41.86985	27.50695
MORKARAMAN	39.841484	41.827072
NORDUZ	38.075	43.51333333
SAKIZ	38.2956625	26.3508125

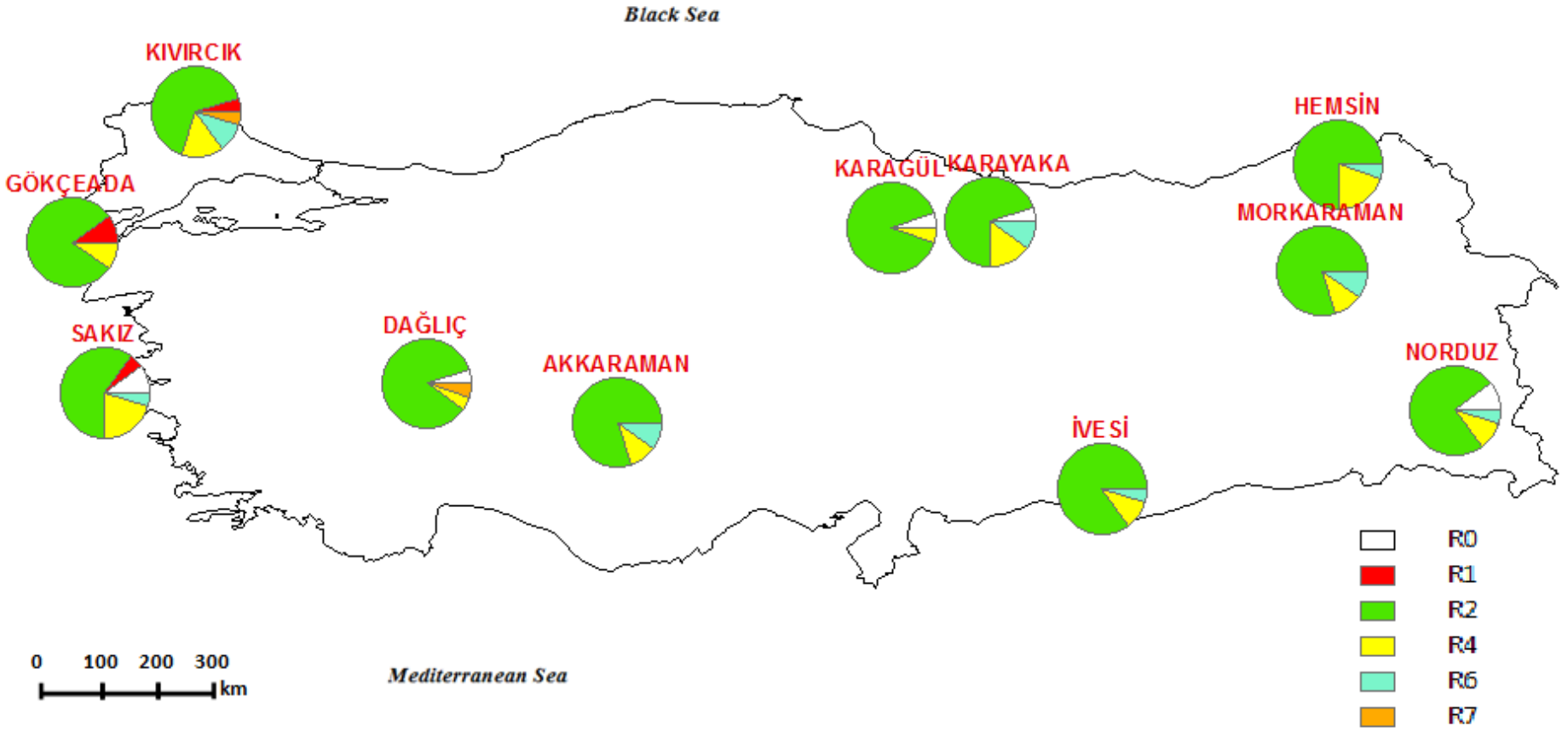

Figure 3.5 Distribution of retrotype frequencies of Turkish sheep breeds on the map of Turkey

In the Figure 3.5, R2 (enJSRV-18 only) is the predominant retrotype in Turkey. However, at the west part of the Turkey, proportion of the R2 retrotype is decreasing and this decrease is not corresponding to the increase of a single retrotype. R4 (enJSRV-7 + enJSRV-18) frequency is increasing in varying frequencies in different breeds. Also presence of R1 (enJSRV-7 only) and R7 (enJSRV-7 + enJSRV-18 + enJS5F-16) only in western breeds is noteworthy.

3.3.2 Synthetic Maps

In order to construct synthetic maps, first pairwise F_{ST} values were calculated from retrotype frequencies by using Arlequin 3.11 (Excoffier, Laval & Schneider, 2005). F_{ST} values (Table 3.5) has been used in metric multidimensional scaling (MDS) analysis which was performed by using R 2.14 (http://www.r-project.org/). First dimension of MDS analysis results (Table 3.6) were used to construct synthetic maps by Kernel density estimation method via ArcGIS 10 software. The centers of collection sites for each breed were used in synthetic map which was constructed. The constructed synthetic maps are shown in Figure 3.6 and Figure 3.7.

Table 3.5 Pairwise F_{ST} distances of Turkish breeds with respect to retrotypes

	KRY	AKK	GÖK	DAĞ	KIV	İVE	KRG	HEM	SAK	NOR	MRK
KRY	0.00000										
AKK	0.00000	0.00000									
GÖK	0.00000	0.00000	0.00000								
DAĞ	0.00733	0.00000	0.00000	0.00000							
KIV	0.00000	0.00000	0.00000	0.02556	0.00000						
İVE	0.00000	0.00000	0.00000	0.00000	0.00893	0.00000					
KRG	0.02841	0.00000	0.00000	0.00000	0.05660	0.00000	0.00000				
HEM	0.00000	0.00000	0.00000	0.00709	0.00000	0.00000	0.02462	0.00000			
SAK	0.00000	0.01648	0.01648	0.05263	0.00000	0.04306	0.08086	0.00000	0.00000		
NOR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	
MRK	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.01648	0.00000	0.00000

Breeds used for analysis are: Kıvırcık (KIV), Gökçeada (GÖK), Sakız (SAK), Dağlıç (DAĞ), Akkaraman (AKK), Karagül (KRG), Karayaka (KRY), İvesi (İVE), Hemşin (HEM), Morkaraman (MRK), Norduz (NOR).

Table 3.6 Results of multi-dimensional scaling (MDS) analysis for Turkish sheep breeds

	Z1	Z2	Z3
KARAYAKA	-0.004665437	-0.000129204	-0.003306429
AKKARAMAN	0.000911039	0.00300086	0.001169488
GOKCEADA	0.000911039	0.00300086	0.001169488
DAGLIC	0.01516713	-0.006247413	-0.003359896
KIVIRCIK	-0.018031086	-0.010526579	0.001463594
IVESI	0.009158894	-0.008276549	0.002399539
KARAGUL	0.040783738	0.004625053	0.00060703
HEMSIN	-0.003640351	0.001284472	-0.001877622
SAKIZ	-0.041032268	0.005107128	-0.000133792
NORDUZ	-0.000473738	0.005160514	0.000699111
MORKARAMAN	0.000911039	0.00300086	0.001169488

Z1, Z2 and Z3 stand for first, second and third dimensional values of multidimensional scaling which was calculated by using pairwise F_{ST} values.

Figure 3.6 Synthetic map with 2 classes was constructed by first dimensional values of MDS. Kernel density estimation (KDE) was performed via ArcGIS 10 (ESRI) software

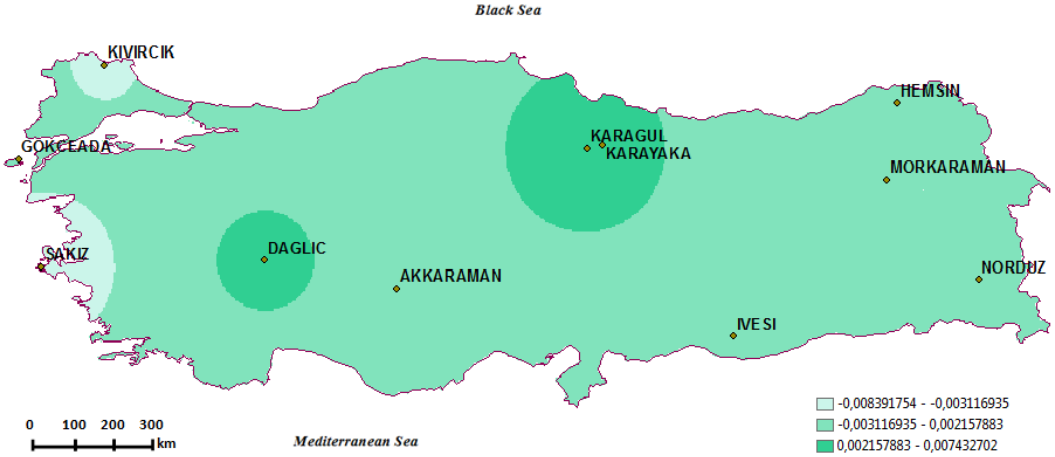

Figure 3.7 Synthetic map with 3 classes was constructed by first dimensional values of MDS. Kernel density estimation (KDE) was performed via ArcGIS 10 (ESRI) software.

For the synthetic maps breeds are divided into 2 classes (Figure 3.6) and then into 3 classes (Figure 3.7). The border near the west coast of Turkey is present with a slight difference for both of the cases. Generally Central Eastern and Eastern part of the Turkey seems to be uniform. Thus it can be said that the breeds in those regions are genetically similar to each other.

3.3.3 Relationship between pair-wise genetic distances of breeds and geographical distance

In order to determine if there is a significant relationship between genetic distance and geographic distance within Turkey, Mantel's test was performed by using R 2.14 (http://www.r-project.org/). Two different Mantel tests were performed, in those tests pairwise F_{ST} values are tested for the presence significant relation to geographical distance and logarithm of geographical distance separately.

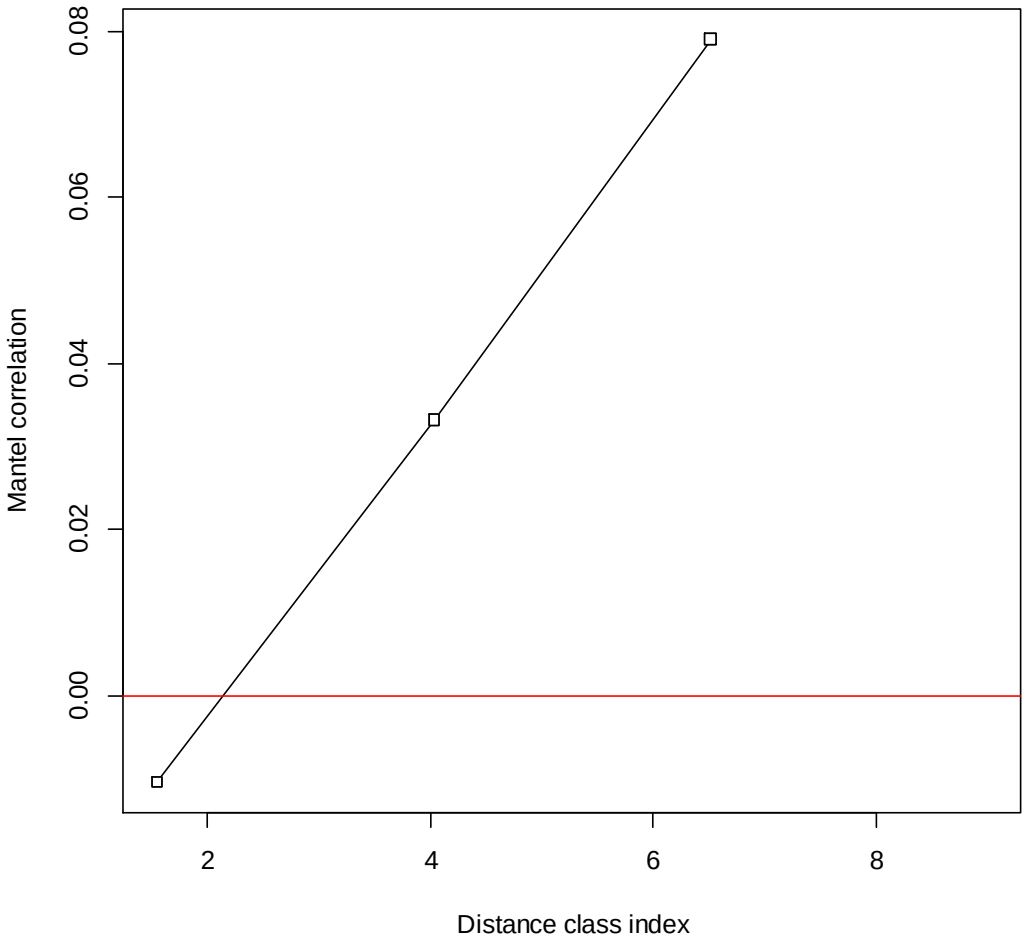

Figure 3.8 Mantel's test for the relationship between pairwise F_{ST} values and geographical distance.

The Mantel statistics r, which is based on Pearson's product-moment correlation, is calculated as "-0.04405". Significance value is 0.615 and this test is based on 999 permutations.

Upper limits of r can be listed for different confidence levels as: 0.188 for 90%, 0.263 for 95%, 0.355 for 97.5% and 0.404 for 99%.

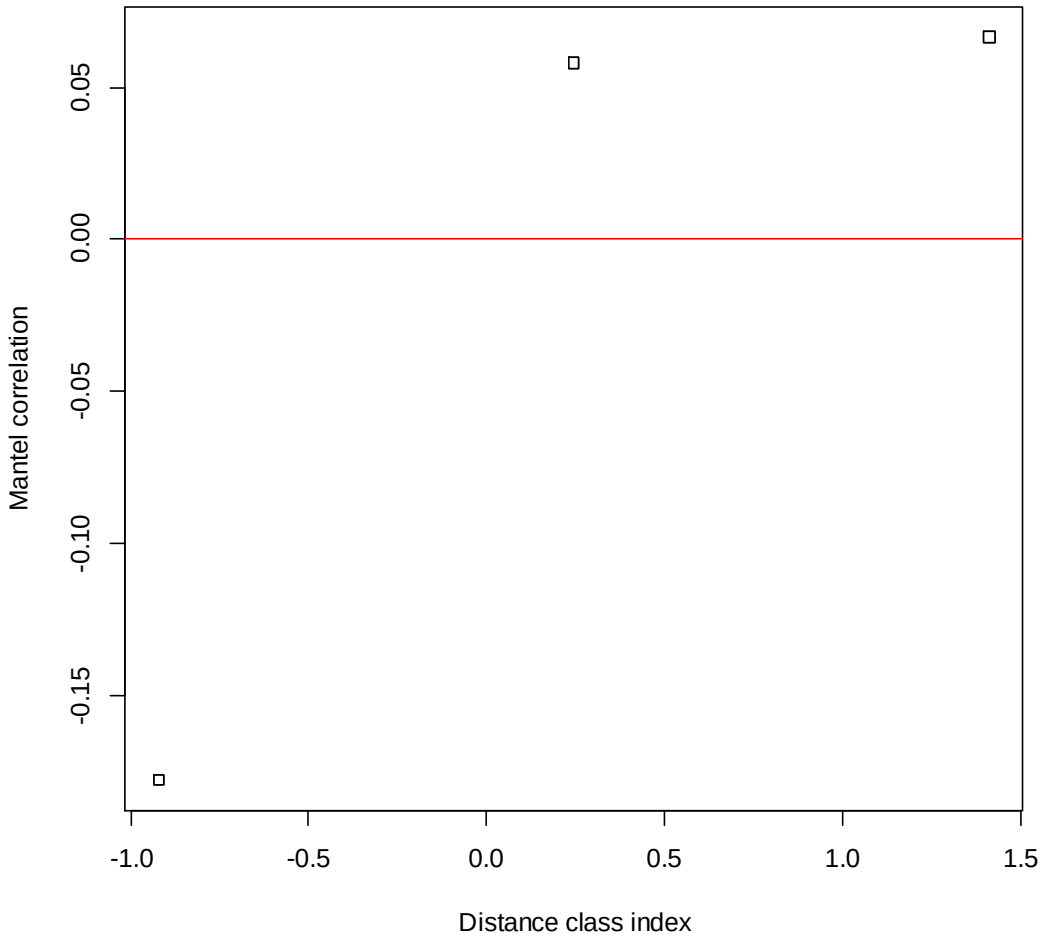

Figure 3.9 Mantel's test for the relationship between pairwise F_{ST} values and logarithm of geographical distance.

The Mantel statistics r, which is based on Pearson's product-moment correlation, is calculated as "-0.07308". Significance value is 0.755 and this test is based on 999 permutations.

Upper limits of r can be listed for different confidence levels as: 0.156 for 90%, 0.216 for 95%, 0.251 for 97.5% and 0.293 for 99%.

For both of the test results, there is no significant association between genetic distance and geographical distance.

3.4 Results of the experiments for enJSRV-6 negative samples

3.4.1 Amplification of enJSRV-6 negative samples with empty locus primers for enJSRV-6

The gel image for the amplifications of 5 enJSRV-6 negative *Ovis gmelinii anatolica* (OGA) samples with enJSRV-6 empty locus primers is given in Figure 3.10 and the gel image for the same amplification of 7 enJSRV-6 negative Turkish domestic breed samples is given in Figure 3.11.

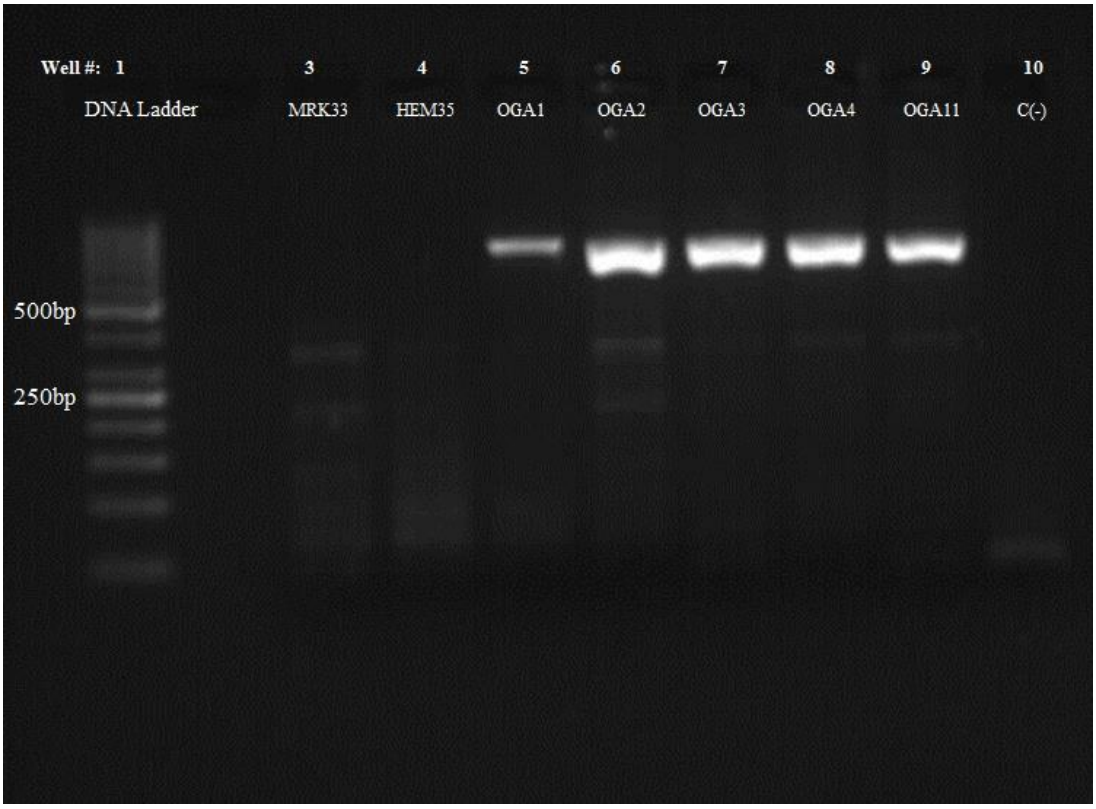

Figure 3.10 Visualization of empty locus (enJSRV-6) amplifications on the agarose gel. Details of the gel picture are explained in the text.

The first well contains the DNA ladder (Gene Ruler 50bp plus), 3rd and 4th well contains 2 samples from native breeds which were identified as enJSRV-6 positive. The next 5 wells (from well 5 to 9) contain 5 enJSRV6 negative OGA samples. The last well (10th well) contains the negative control (dH_2O instead of DNA).

MRK33 and HEM35 samples were identified as enJSRV-6 positive there should not be any amplification with empty locus primers because if there was enJSRV-6 integration, the region will be too long to be amplified by the empty locus primers. In other words, in addition to simple negative control (dH_2O), these two samples were used as specific negative controls.

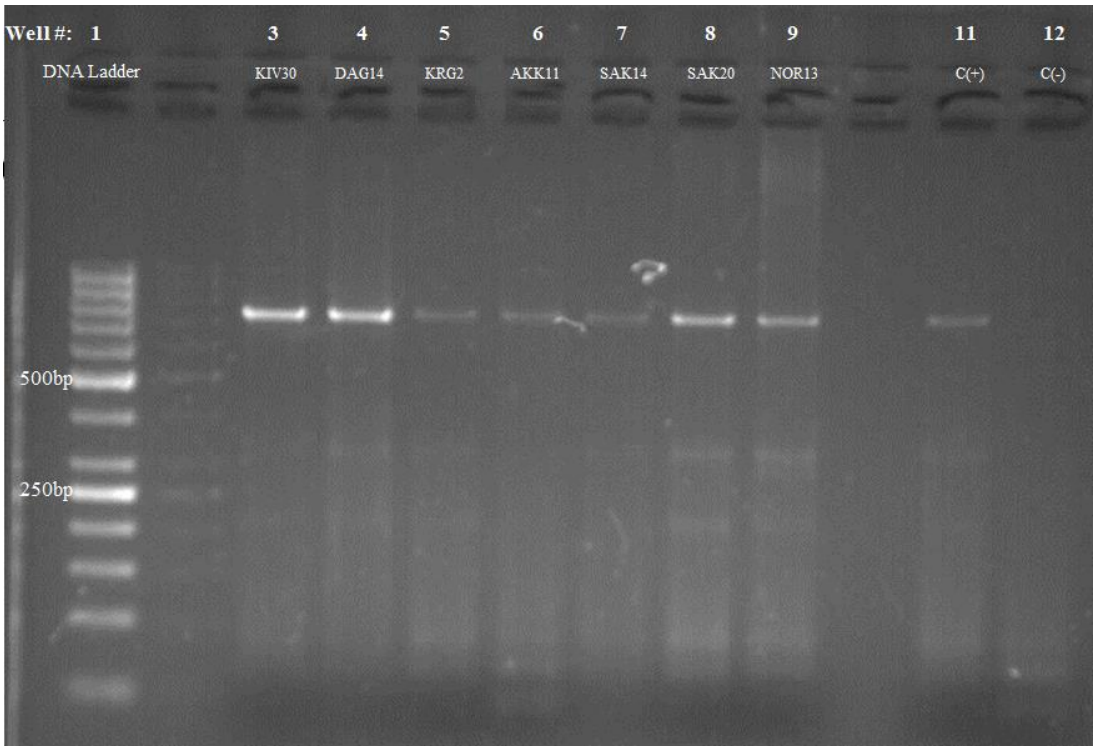

Figure 3.11 Visualization of empty locus (enJSRV-6) amplifications on the agarose gel. Details of the gel picture are explained in the text.

The first well contains the DNA ladder (Gene Ruler 50bp plus). The next 7 wells (from well 3 to 9) contain 7 enJSRV6 negative native samples. The 11th well contains OGA1 as positive control and the 12th well contains HEM35 as negative control.

It is explained in the Figure 3.10 why HEM35 sample was used as negative control and in the light of the results obtained from Figure 3.10 OGA1 was used as a positive control in Figure 3.11.

When both of the empty locus amplifications are considered, there were no amplifications for enJSRV-6 positive samples. For all of the 12 enJSRV-6 negative samples (5 OGA and 7 native) there were amplification with empty locus primers. These results indicate that for those samples that their DNA quality was not bad but they did not have enJSRV-6 integration at all.

3.4.2 Sequences of the amplified products with enJSRV-6 empty locus primers.

In order to be sure that these amplified products are empty enJSRV-6 locus, all of the amplified products (as total of 12 samples) were sequenced. Sequences that are obtained from enJSRV-6 negative samples with the use of both forward and reverse primers were aligned by using Multiple Sequence Alignment (MSA) with the algorithm of ClustalW (Thompson *et al.*, 1994) by using Bioedit v. 7.1.3 (Hall 1997-2011). The aligned sequences of 5 enJSRV-6 negative OGA samples are given in Figure 3.12.

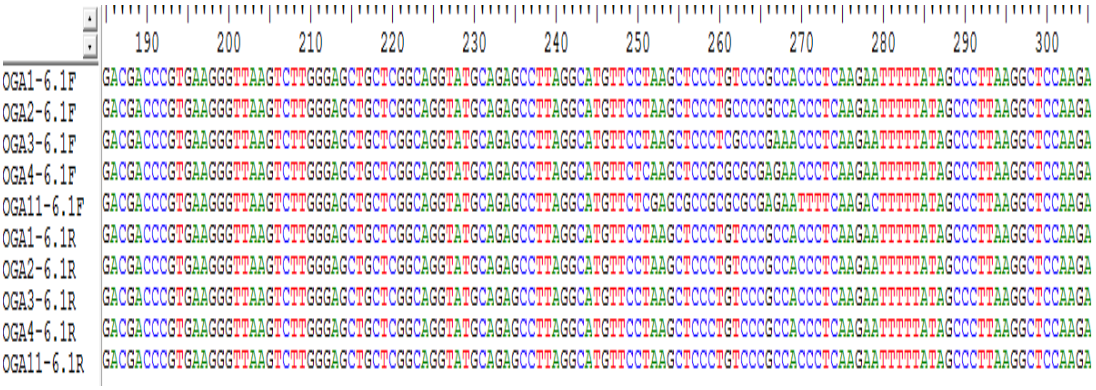

Figure 3.12 Part of the multiple sequence alignment (MSA) of the sequences of 5 enJSRV-6 negative OGA samples

Mismatches between sequences (like around 260th position) were corrected manually based on chromatograms. With the addition of 7 sequenced enJSRV-6 negative native sheep samples (shown in Figure 3.11) final sequence of the amplified products (731 bp long) with enJSRV-6 empty locus primers was determined as the following:

5'-
GGAAGGGAGATCACCCATCAGAAGACATACTTTACAATCTTATCCAAAATCTGA
ATTTTGGTAGATAGCTGTTACCAGTGTGAAGAAAGTAGATCTCACTCTTGAGAG
TTCTAACTACCCTAACATTTATTCTGATCCCTCTGGGCCCATGACATACTAGTTA
TTCTTTTGATTGTTGCGGGGGACGACCCGTGAAGGGTTAAGTCTTGGGAGCTGC
TCGGCAGGTATGCAGAGCCTTAGGCATGTTCCTAAGCTCCCTGTCCCGCCACCC
TCAAGAATTTTTATAGCCCTTAAGGCTCCAAGATGTCCAGTTCCTGCAACCTGTC
CTAGAAGATAGATTATCTTATCATGTATACTTCATAGAAGATAGATTATCTGATT
GTGTTCTGTATACAATGGTAAGGGTCTGGTGATTGTATCTTGAGATTAAAAACA
ACCTTGTGAATGTCATAAGTCACGTACTTTACCCTATATATACTGCAGCACAAT
AAAGCAAGGCATCAGCCATTTTGGGCTGATCCTCTCAACCCCATCTTTTGTCTCT
CTCTTATCTTCTTAGCGGGGACGCTCCGTTCTCTCCCTGTGCAGGTGCGACTCTT
GCTTGTGCTGGCTGCGGCAGATTGTCATTCCTGAATATAATCTGATGTTGAAAGT
AGTGAATTGTGTCAGATCCTTGGTTCTCAGCTCTATTACTCTATTCACTACTCCT
GATGGAGGTAGCACCAGTTATTCCCCTGA -3'

Obtained sequence was analyzed by using Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) embedded in the website of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The results obtained from BLAST run are given in Figure 3.13 and 3.14.

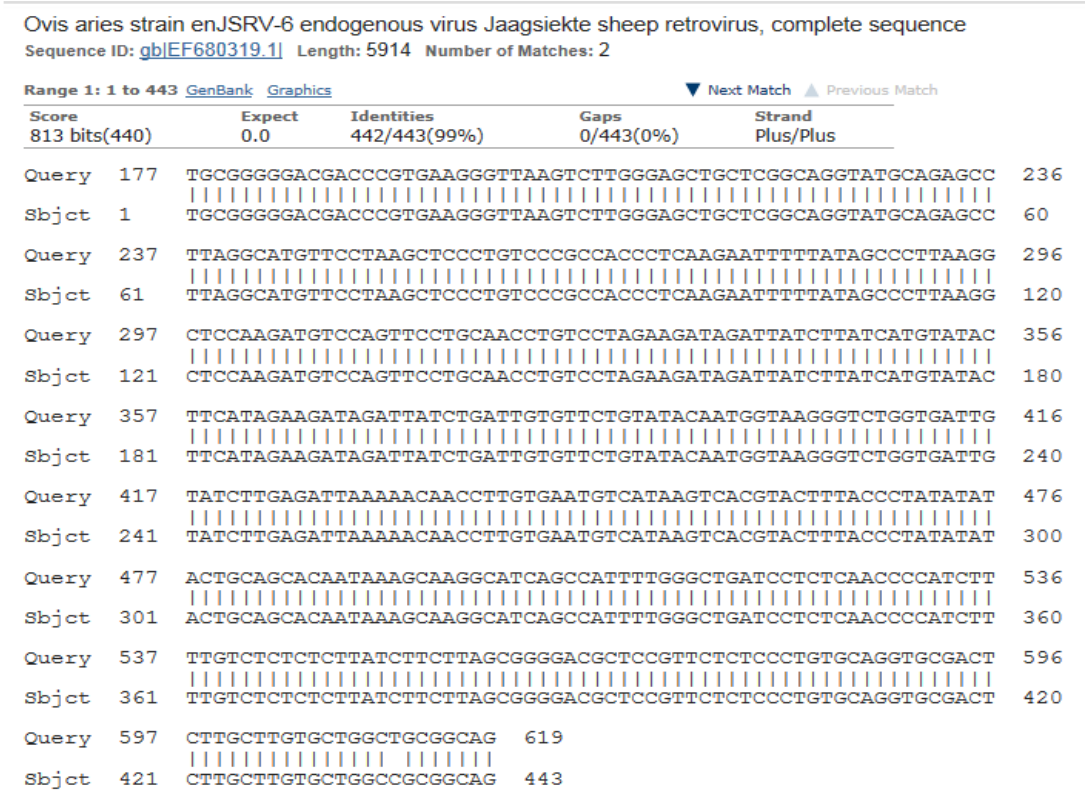

Figure 3.13 The first part of the snapshot of BLAST hit of the sequence of amplified products by empty locus primers

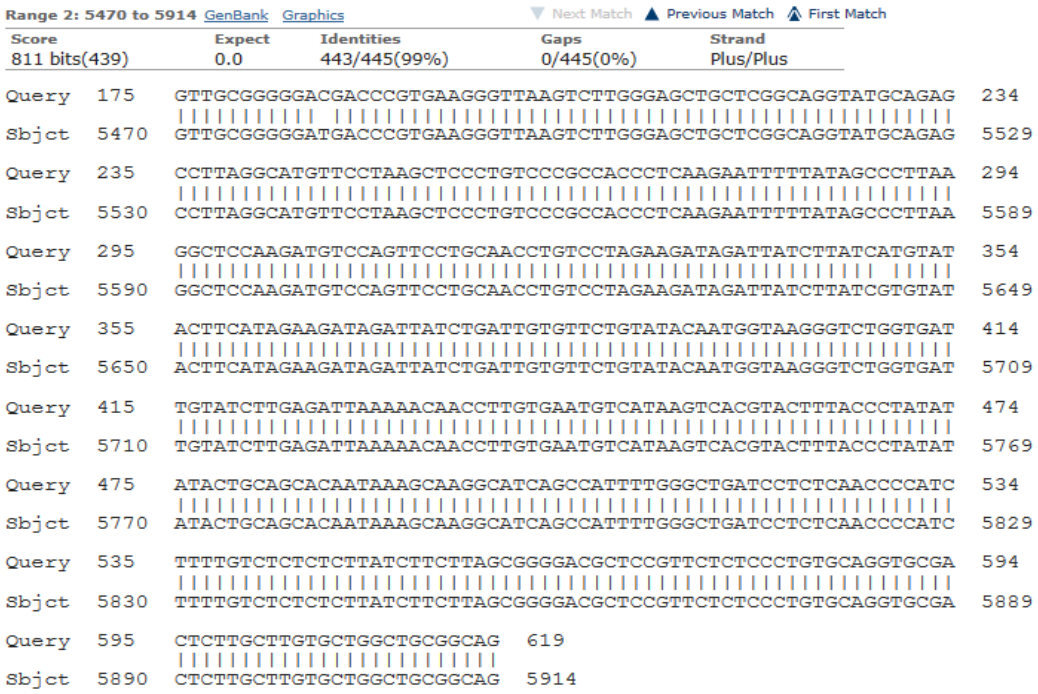

Figure 3.14 The second part of the snapshot of BLAST hit of the sequence of amplified products by empty locus primers

BLAST algorithm gave hit to enJSRV-6 sequence (EF680319.1) (5914 bp long) from two different sites 5' region of the sequence (Figure 3.13) and 3' region of the sequence (Figure 3.14). In other words, our sequence contains parts from the flanking regions of enJSRV-6 site. It means that the sequences we obtained are from the region where enJSRV-6 integrates itself into the host but it is not from the provirus region of enJSRV-6.

This proves that at least these 12 enJSRV-6 samples mentioned before are negative for enJSRV-6 but this is not due to DNA quality, it is because there is no integration of enJSRV-6 at all.

3.4.3 Polymorphic enJSRV integrations in enJSRV-6 negative samples

10 samples from Karayaka and 10 samples from İvesi breeds which had been discarded from original data because they were negative for enJSRV-6 integration were amplified with empty locus primers for enJSRV-6. The results for those amplifications are given in Figure 3.15 and summarized in Table 3.7.

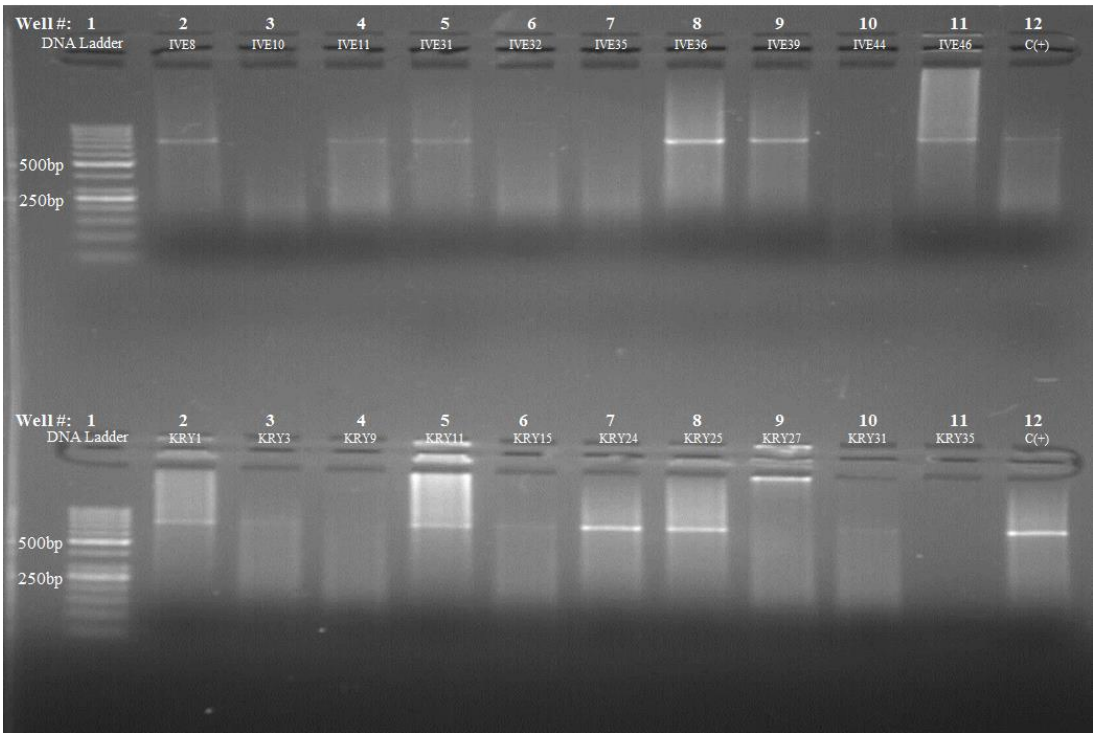

Figure 3.15 Visualization of empty locus (enJSRV-6) amplifications on the agarose gel. Details of the gel picture are explained in the text.

For both of the rows the first well contains DNA ladder (Gene Ruler 50 bp plus) and the last well contains OGA1 sample as a positive control. For the upper row: Wells from 2 to 11 contain 10 enJSRV-6 negative samples from İvesi breed. For the lower row: Wells from 2 to 11 contain 10 enJSRV-6 negative samples from Karayaka breed.

Table 3.7 Amplifications of enJSRV-6 negative samples for empty enJSRV-6 locus

Sample ID	enJSRV-6 empty locus	Sample ID	enJSRV-6 empty locus
IVE 8	+	**KRY 1**	+
IVE 10	-	**KRY 3**	-
IVE 11	+	**KRY 9**	-
IVE 31	+	**KRY 11**	+
IVE 32	-	**KRY 15**	-
IVE 35	-	**KRY 24**	+
IVE 36	+	**KRY 25**	+
IVE 39	+	**KRY 27**	-
IVE 44	-	**KRY 31**	-
IVE 46	+	**KRY 35**	-

As it can be observed from Figure 3.15 and Table 3.7; 6 of 10 İvesi samples and 4 of 10 Karayaka samples shows positive results with the empty locus primers. As previously mentioned, we may say that for those 10 samples being enJSRV-6 negative is not due to DNA quality so these 10 samples (6 from İvesi and 4 from Karayaka) were examined for the presence of insertionally polymorphic enJSRV. Integrations of the polymorphic enJSRVs for those 10 samples are given in Table 3.8

Table 3.8 Polymorphic enJSRV integrations for enJSRV-6 negative native samples

Sample ID	enJSRV-7	enJSRV-18	enJS5F-16
IVE 8	-	+	-
IVE 11	-	-	-
IVE 31	+	+	-
IVE 36	+	+	-
IVE 39	-	+	-
IVE 46	-	-	+
KRY 1	-	+	-
KRY 11	-	-	-
KRY 21	+	+	-
KRY 25	-	+	-

IVE 11 and KRY 11 samples show no positive results for any polymorphic enJSRVs so they can be defined as no enJSRV insertions including enJSRV-6. In other words, they are in R0 retrotype as defined by Chessa *et al.* (2009) but they are also negative for enJSRV-6 integration.

For Karayaka and İvesi breeds, differences between the frequencies of polymorphic enJSRV integrations when only enJSRV positive samples are considered (Table 3.2) and when enJSRV-6 negative breeds are also included are summarized in Table 3.9

Table 3.9 Comparison of the frequencies of polymorphic enJSRVs according to presence of enJSRV-6

	Only enJSRV-6 positive samples			enJSRV-6 negative samples included		
	enJSRV-7	enJSRV-18	enJS5F-16	enJSRV-7	enJSRV-18	enJS5F-16
Karayaka	0.15	0.95	0.1	0.167	0.917	0.083
İvesi	0.1	1	0.05	0.115	0.923	0.077

It is observed that frequencies of polymorphic enJSRVs changes differently. It is noteworthy that decrease in the frequency of enJSRV-18 integration and increase in the frequency of enJSRV-7 integration is consistent for both of the breeds when the enJSRV-6 negative samples are considered.

CHAPTER 4

DISCUSSION

4.1 Using enJSRVs as molecular genetic markers

In understanding the evolutionary history of sheep, parallel to the studies in livestock, first mtDNA based studies were used (Hiendleder, Mainz, Plante, & Lewalski, 1998; Bruford, Bradley, & Luikart, 2003; Meadows *et al.*, 2005; Pedrosa *et al.*, 2005; Periera *et al.*, 2006*;* Meadows, Cemal, Karaca, Gootwine, & Kijas, 2007). Then markers from male specific non-recombining Y chromosome region were developed and used (Meadows, Hawken, & Kijas, 2004; Meadows *et al.*, 2006; Meadows, Hiendleder, & Kijas, 2011). mtDNA and Y chromosome markers were providing the evolutionary history of sheep from maternal and paternal sites respectively. On the other hand, studies harboring information captured by microsatellite (Uzun *et al.*, 2006; Peter *et al.*, *2007*) and single nucleotide polymorphism (SNP) (Kijas *et al.*, 2012) revealed the evolutionary history of sheep on the basis of both sexes.

The recently employed markers (enJSRVs) are only present in sheep and goats. enJSRV integrations in these species began approximately 5-7 million years ago even before the split between the genus *Capra* and the genus *Ovis*. The integration sites (enJSRV-18, enJSRV-7, enJSRV-8, enJSRV-15, enJSRV-16, and enJS5F16) used by Chessa *et al*. (2009) are observed in only sheep particularly in domestic sheep. Thus the integration sites used in the study have emerged presumably after the first sheep domestication 12,000 years ago (Zeder, 2008) and continued to evolve since then (Arnaud *et al.*, 2007; Chessa *et al.*, 2009). At least 27 copies of enJSRVs are present in the sheep genome (Arnaud *et al.*, 2008). Most enJSRV loci are fixed in domestic sheep but some are insertionally polymorphic (Arnaud *et al.*, 2007). These insertionally polymorphic enJSRVs (enJSRV-18, enJSRV-7, enJSRV-8, enJSRV-15, enJSRV-16, and enJS5F16) have been used as molecular genetic markers (Chessa *et al.*, 2009).

enJSRV integration sites are just like ALU (Berkman *et al*., 2008) and if they are integrated they will be detected but if they are removed still they will leave a trace at the site. Thus unlike point mutations (which can occur recurrently on the same site and we can have information about only the latest mutation), enJSRV integrations are irreversible. Because of their irreversibility sequential changes at their long terminal repeats (LTRs) can be reliably traced. Retrotypes defined by the combinations of irreversible enJSRV integrations which are molecular markers differing from each other by the unique events occurring both spatially and temporarily are suitable markers for the evolutionary history of sheep. The two retrotypes (R0 and R1) are thought to be old enough to occur even before (or approximately at the down of) the first domestication. Thus geographic distribution patterns of sheep retrotypes defined on the basis of enJSRV integrations, being on the nuclear chromosomes, are expected to reveal information about possible migration centers, routes and even events corresponding to two different time periods in a way both sexes are represented.

In this study 219 samples from 11 breeds (n=20 for each breed) and also 7 samples (in the first stage of the study) from *Ovis gmelinii anatolica* (Anatolian wild sheep) have been tested for the presence of the 3 insertionally polymorphic enJSRVs (enJSRV-7, enJSRV-18 and enJS5F-16). To contribute to the understanding of evolutionary history of sheep in particular in the lands of the first domestication center of sheep in Anatolia. As recommended all the samples were tested for the presence of enJSRV-6 to assess the DNA quality of samples because enJSRV6 integration is regarded as fixed in all sheep breeds (Chessa *et al.*, 2009). Previously it was stated that 30 samples from Anatolian wild sheep were used for analysis. Only 7 of them have been analyzed for the presence of polymorphic enJSRV integrations because 23 of the individuals were negative for enJSRV-6 integration.

In order to check presence of 3 relatively common insertionally polymorphic enJSRVs and fixed enJSRV-6 the primers which have been designed by Chessa *et al* (2009) was used in this study. Optimization for each pair of the primers (for both 5' and 3' PCRs) has been a major difficulty. Although the PCR conditions that have been optimized by us for each enJSRV have been given in a detailed way at tables (from table 2.2 to 2.15), these conditions provided are only the starting points for the establishment of tediously long PCR optimization processes for each of the integrations sites. These start points are the results of the first optimization efforts for integration sites by using a few individuals from Karayaka breed. Furthermore for each of the breeds the given conditions have been re-optimized several times. In other words, breed specific modifications had to be done and this made the study very demanding. PCRs for each primer of each enJSRV have been repeated at least as three times to obtain stable and consistent results. Breed specific PCR conditions for each of the integration sites used in the present study are provided in Appendix B.

Presence of these obstacles during experimental process can be explained, at least partly, by the use of primers which have been previously designed by Chessa *et al*. (2009). First question that came to mind is that: were these difficulties encountered due to the DNA quality of the samples used? The same samples have been employed in Uzun's study (2012)

for prion protein polymorphism examination, recently. They were able to successfully amplify the protein region from all of the tested individuals. These observations suggest that DNA quality may not be the (only) reason for the amplification of enJSRV-6 integration site.

Similar problems about primers have been encountered in Demirci's (2012) study in which mitochondrial DNA polymorphism was examined. Primers that have been designed in a previous study (Bruford and Townsend, 2006) were not successfully amplifying the DNA of the samples from Turkish sheep breeds. Therefore new set of primers were searched and used (Demirci, 2012).

In the study of Kijas *et al*. (2012) it was observed that Asian and African sheep populations had an excess of low minor allele frequency (MAF) SNP when compared to European derived populations. The observation was attributed to ascertainment bias in SNP discovery by Kijas *et al.* (2012). In the present study, parallel to the Kijas *et al.*'s (2012) interpretation we are suggesting the presence of ascertainment bias in the retrovirus insertion polymorphisms and we are and providing strong evidence in this line of suggestion.

All of the insertionally polymorphic enJSRVs in Chessa *et al.*'s (2009) study were cloned from genomic DNA of Texel ram based BAC library. In that study it was argued that since Texel is highly admixed and it must be representing all the derived products of sheep evolution. On the other hand, Texel is a well-defined breed (well isolated). Turkish breeds are more admixed (Açan, 2012) and were never subjected to selection hence they must be representing wider spectrum of genetic diversity than that of Texel. Thus the primers may not be suitable for some of the genetic variations that is present among the Turkish breed and, moreover, if some insertional polymorphism(s) went undetected, then the data are not representing the genetic diversity.

In relation to the former argument¸ when we tried to use the same primers designed by Chessa *et al* (2009), for Turkish breeds, there might be problems due to mutational differences in primer sites of 5' and 3' flanking regions of the enJSRV integration sites for some individuals of our breeds. May be the reasons of the most of the problems that have been encountered during experimental part of our study were partly due to these mutations.

In relation to observed total genetic diversity in sheep on the basis of retrotype insertion polymorphisms, according to Chessa *et al.*'s (2009) study, the highest genetic variation in European sheep breeds is observed in Northern European breeds. Even those Northern sheep breeds may be the subset of Turkish breeds because Middle East is predicted as the domestication center for sheep (Zeder, 2008). Since our breeds are very close to domestication center geographically and they may be closer to the start point of secondary population expansion, which was stated in Chessa *et al* (2009), when compared to European breeds, it can be suggested that primers that have designed by Chessa *et al* (2009) may not

fully represent our breeds. Furthermore, as was stated before samples from Turkey were few (n=42) and not from Central and Western Turkey (Chessa *et al.*,2009).

Also, there was an interesting observation in relation to the distribution of retrotypes (Figure 1.5): neither R0 nor R1 retrotypes were observed for Asian breeds in Chessa *et al*'s. (2009) study. Retrotype profile of Asian breeds may not be exactly determined due to same bias that has been stated in Kijas *et al.*'s (2012) study. Only 9 groups (n=121) examined from Asia where there were total of 42 groups (n=897) examined from Europe. This huge difference in the total number of individuals tested for the presence of insertionally polymorphic enJSRVs between Asian and European sheep breeds might have led to incomplete conclusions about Asian sheep breeds.

4.2 Absence of enJSRV-8 enJSRV-15 and enJSRV-16 integration sites and R3 and R5 retrotypes in Turkish sheep breeds

In the original study (Chessa *et al.,* 2009) 6 insertionally polymorphic enJSRVs were analyzed. Three of them have been used in this study: enJSRV-7, enJSRV-18 and enJS5F-16. Other three (enJSRV-8, enJSRV-15 and enJSRV-16) of them were intended to be used but they were not being able to. The latter 3 enJSRVs were not included in this study because they could not be optimized by using samples from Turkish breeds. Optimization experiments have been conducted on randomly selected 4 different DNA samples from Karayaka, Kıvırcık, İvesi, Hemşin and Norduz breeds (as a total of 20 samples) for each of those 3 enJSRVs but there were no single positive results for any of them.

One of the possible reasons that we were unable to optimize PCR conditions for these 3 specific enJSRVs is that they are very rare in general and also in Turkey (Chessa *et al.*, 2009). In the latter study enJSRV-8, enJSRV-15 and enJSRV-16 were only observed in only 3%-5% of the samples. On the other hand enJSRV-18 was the most abundant one with 85% and enJSRV-7 and enJS5F-16 were present in 27% and 30% of the samples respectively. Frequencies of each enJSRV locus in 133 breeds divided into 65 groups of which 42 are from Europe which were used in Chessa *et al.*'s (2009) study is given in Figure 4.1.

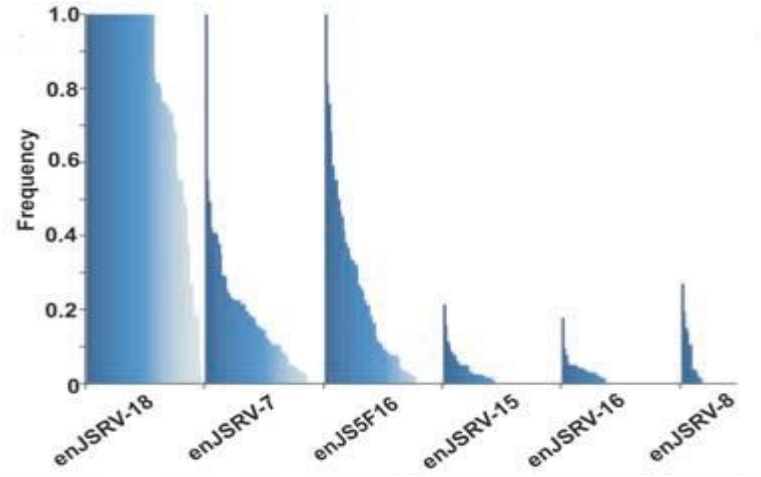

Figure 4.1 Frequencies of each enJSRV loci for 65 groups which were analyzed in Chessa *et al.*'s (2009) study. Each bar in the cluster corresponds to a group. The figure was taken from Chessa *et al.*'s (2009) study.

In addition to being rare, distributions of these unemployed enJSRVs (enJSRV-8, enJSRV-15 and enJSRV-16) in the present study are observed at sites which are far beyond the borders of Turkey. Especially enJSRV-8 is present in only Northern European breeds and enJSRV-16 is mainly present in Northern European and Central Asian breeds. These two enJSRVs are predicted to be absent in Turkey. Similarly enJSRV-15 is highly unlikely to be present in Turkey. Interpolation maps displaying spatial distribution of estimated frequencies of these three enJSRVs were given in Figure 4.2, 4.3 and 4.4.

Figure 4.2 Spatial distribution of estimated enJSRV-8 frequencies for 65 groups analyzed in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al.*'s (2009) study.

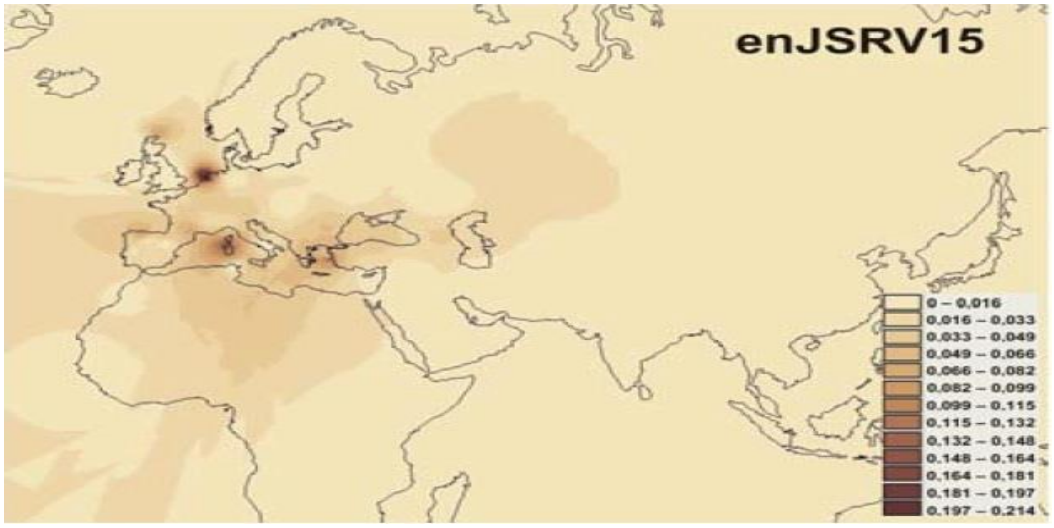

Figure 4.3 Spatial distribution of estimated enJSRV-15 frequencies for 65 groups analyzed in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al*'s. (2009) study.

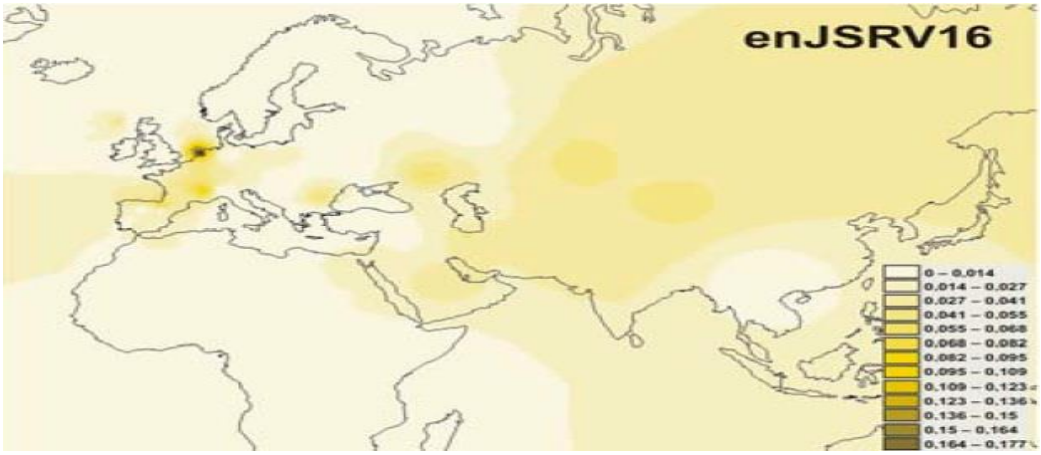

Figure 4.4 Spatial distribution of estimated enJSRV-16 frequencies for 65 groups analyzed in Chessa *et al.*'s (2009) study. The figure was taken from Chessa *et al.*'s (2009) study.

As it can be inferred from interpolation maps, estimated frequencies of these three enJSRVs are very low in Turkey (as estimated frequencies were between 0 and 0.018 for enJSRV-8, between 0.033 and 0.049 for enJSRV-15 and between 0 and 0.014 for enJSRV-16). In addition to being rare in the world they seemed to be absent in Turkey (Figures 4.2, 4.3 and 4.4). Therefore they may not have been detected in the optimization studies of the sites within the Turkish sheep breeds. However, enJSRV-16 being the marker of Asian sheep may have higher chance to be detected especially among the fat tailed sheep of Turkey because fat tail sheep of Turkey harbor a Y chromosome marker (H4) and mtDNA marker (HPG C) as was explained by Öner *et al.*, (2010) and Demirci *et al.* (manuscript in preparation) respectively.

With the enJSRVs tested in this study, we could have observed R3 and R5 retrotypes in our samples but frequencies of these both retrotypes were zero for all of the breeds analyzed (Table 3.3). In the study of Chessa *et al.* (2009) R3 retrotype (enJS5F-16 only) is called as "Nordic" retrotype because it was observed in only Northern European breeds and these breeds are characterized by the low frequency of enJSRV-18 and high frequency of enJS5F-16. Similarly R5 retrotype (enJSRV-7 + enJS5F-16) was observed in only Northern European breeds with only exception of *Ovis musimon* (European mouflon) whose samples collected from Sassari, Sardinia (Italy). R5 is characterized by the presence of enJSRV-7 and enJS5F-16 in the absence of enJSRV-18. Thus, it can be suggested that enJS5F-16 integrated into domestic sheep genome after the domesticated sheep migrated out of Anatolia.

Furthermore, in some sheep from Turkey, in combination with some other integration sites (for instance enJSRV-18) enJS5F-16 was detected. It can be claimed that absence of R3 is

not due to a technical problem. Similar argument applies for the absence of R5. Perhaps their absence also supports the undetected integration sites of enJSRV-8, enJSRV-15 and enJSRV-16 in parallel to their absence/very low frequencies in Turkish sheep breeds which was the expectation drawn from the previous study.

4.3 Observed integration sites and retrotypes in Turkish sheep breeds

For all of the 220 samples in this study, enJSRV-6 integration is present with only one exception from Karagül breed (Table 3.2). This is expected because enJRV-6 is regarded as fixed in all domesticated sheep breeds (Chessa *et al.*, 2009) and presence of enJSRV-6 is used to assess DNA quality of the samples.

For enJSRV-6 case 7 of 220 individuals have amplified product for only one of the PCRs either 5'- or 3'- PCR. Since enJSRV-6 was used to assess DNA quality of the samples they should not be tested for presence of insertionally polymorphic enJSRVs. But in the present study they have been tested and if they have been found positive for any other polymorphic site, they have been taken into consideration. This exception has been made because even if a sample is not positive for enJSRV-6, it can be positive for a polymorphic enJSRV. Thus these results are still informative and can lead to novel observations.

On the other hand, as a total of 34 samples from 7 different breeds (Karayaka, Dağlıç, Kıvırcık, İvesi, Norduz, Akkaraman and Sakız) had been eliminated from the data before checking for insertionally polymorphic enJSRVs because there were no amplifications in PCRs which were conducted for enJSRV-6 integration for those samples. Thus they were regarded as samples with low DNA quality.

Discarding 34 samples especially from 7 different breeds due to absence of enJSRV-6 has raised the question: Can enJSRV-6 be only fixed in almost all of the European domestic breeds but it is polymorphic for some Middle Eastern and Asiatic sheep breeds? In order to answer that question, observations on Turkish domestic sheep breeds in terms of the presence of enJSRV-6 gained importance. Efforts made to answer this specific question and it will be discussed in section 4.5.

Among the 3 insertionally polymorphic enJSRVs (enJSRV-7, enJSRV-18 and enJS5F-16) used in this study, enJSRV-18 is the most abundant one in Turkish breeds. This result is also consistent with literature because for most of the sheep breeds in the world, fixation or high frequency of enJSRV-18 is regarded as a characteristic feature and related with the second migration of sheep (Chessa *et al.*, 2009). enJSRV-7 and enJS5F-16 have lower frequencies than enJSRV-18 as it was observed in Chessa *et al.*'s (2009) study (Figure 4.1)

R2 retrotype (enJSRV-18 only) is the predominant retrotype in Turkey. However, frequency of R2 retrotype is decreasing at the west part of the Turkey and this decrease is not corresponding to increase in a single retrotype frequency. Towards the west part of Turkey and R1 and R4 retrotypes are increasing with varying frequencies in different breeds (Figure 3.5). This distribution of retrotypes in Turkish sheep breeds changes the characterization of Turkish sheep breeds in Chessa *et al.*'s (2009) study where Turkish sheep breeds were defined as mostly uniform in retrotype frequencies and highly dominated by R2 retrotype. Domination of R2 is still present in Turkish breeds but it is not as high as previously stated, especially for the breeds present in west part of Turkey.

The high frequencies of R0 and R1 are characteristics of primitive breeds. Very high frequency of R0 is especially characteristics of native primitive Scandinavian breeds and R1 is mainly observed in isolated populations like Asiatic (Cyprus) and European mouflon (Sardinia) and some of the Nordic breeds (Chessa *et al.* 2009). Thus, R1 might be at least partly the marker of primitive sheep traveled with sea-farers. Since Turkey is harboring center of the first sheep domestication (Zeder, 2008) it was expected to observe remnants of R0 and R1 in the Turkish breeds.

When R0 and R1 are considered together, either R0 or R1 is present in 6 Turkish breeds and both are present in one breed (Sakız). In total 7 out of 11 Turkish breeds contains animals which have retrotypes related to primitive breeds even at low frequencies. This observation points out that current Turkish breeds still carrying the genetic signs of first domestication event but in a much diluted way.

When the results of this study are compared to Chessa *et al.*'s (2009) study, in which only 42 samples from 5 different Turkish breeds (Karayaka, Akkaraman, Morkaraman, Hemşin and Tuj) were analyzed, there are significant differences. In Chessa *et al.*'s (2009) study only R2, R4 and R6 were observed for Turkish breeds whose frequencies are consistent with the results of this study, In addition to those retrotypes we also observed R0, R1 and R7 retrotypes in Turkish breeds. Even at low frequencies, presence of these 3 retrotypes, especially R0 and R1, propose a new view point in understanding of the migration events. R0 was observed in 5 breeds and R1 was observed in 3 breeds out of 11 Turkish breeds at low frequencies. R1 was observed only in western part of Turkey which is noteworthy because it was observed in the coastal periphery of Southern and Eastern Europe suggesting that Anatolian breeds; Sakız, Gökçeada and Kıvırcık must be related to Mediterranean route of the first sheep migration.

4.4 Spatial patterns related with the distribution of retrotypes in Turkey

In the present study spatial analyses of retrotype distributions were used in the concept of landscape genetics which is an approach to understand how geographical and environmental

features shape the genetic variation at both of the individual and population levels. Genetic differences between the populations were expressed by F_{ST} values which are a widely used measure to describe the genetic differences between the populations. Since retrotypes are based on neutral loci the genetic divergence between breeds must be under the effect of genetic drift and/or migration. In order to understand the directions of gene flow between breeds in the course of evolutionary history spatial patterns were analyzed based on retrotypes.

Different from Chessa *et al.*'s (2009) study where interpolation maps were used for spatial analysis, in this study Kernel density estimation (KDE) maps were used for the same purpose because we have only 11 sampling sites and they are unevenly distributed. Another reason to use KDE is because in our study we are interested in much smaller geographical region (Turkey) than Chessa *et al.*'s (2009) study where samples from Europe, Africa and Asia were analyzed. Also there are geographical gaps between breeds where no population is presenting these regions. Most of these spaces between breeds are large when the total geographical area is considered. In order to fill these gaps KDE analysis was the suitable method for spatial analysis.

The significant distinction of west part of Turkey from other parts was observed in spatial analysis of retrotype integrations within Turkey. For both 2 class (Figure 3.6) and 3 class (Figure 3.7) KDE maps west of Turkey was different from the rest. Since the difference involves the occurrence of R0 and R1 on the west of Turkey and if R0 and R1 were not observed among the sheep of second migration then it can be suggested that except the west of Turkey second expansion of sheep heavily invaded the east of Turkey.

The sheep accompanying the pastoral nomadic Turks and entering from Eastern/ Southeastern Anatolia may have contributed to the pattern of sheep retrotype frequency distribution in Turkey. These results might be useful in determining (for conservation purposes) the ancient native domestic sheep of Anatolia. The breeds which were determined as enJSRV-6 negative would probably be the best candidates.

In order to test if there is a significant relation between genetic distance and geographical distance, Mantel's test was used. Two different Mantel's tests were conducted for relationship between pairwise F_{ST} values and geographical distance (Figure 3.9) and for relationship between pairwise F_{ST} values and logarithm of geographical distance (Figure 3.10). Significance values were found as 0.615 and 0.755 respectively which indicate that there is no significant relation between genetic distance and geographical distance. Use of log-transformed distances in Mantel test is especially useful to observe "isolation by distance" type of migration in both equilibrium and non-equilibrium populations.

In relation to the observed KDE patterns and Mantel's tests, it might be said that there are at least two different relatively homogenous groups of domestic sheep in Turkey. That can explain the observation the observation of west part of Turkey from KDE analysis results.

Then, it can be suggested that different groups have different origins. May be the breeds on the west are related with the earliest Anatolian sheep and also with the Southeastern European sheep (Peter *et.al.*, 2007). Whereas Central and Eastern Anatolian sheep are related with Middle Eastern / Southwestern Asian sheep as was also observed by microsatellites (Peter *et al.*, 2007) and SNP (Kijas *et al.*, 2012). In the context of animal husbandry practices in Turkey rams are largely exchanged between the breeds. These exchanges between breeds may have been carried out for a long time. However, the breeds in Turkey are not genetically homogenous and genetic differences between the breeds are not related with their geographic differences.

In Demirci's (2012) study based on mtDNA polymorphisms, two groups were separated (northern and western) from the rest of the Turkish domestic sheep breeds. Since that study was conducted on maternal markers, it can be said that different sexes can have different dynamics. Thus, using a sex dependent marker can lead to different results when compared to the present study where nuclear markers were used.

In addition to geographical distance, pairwise F_{ST} values of retrotype frequencies have been tested by Mantel's test for their relationship to pairwise F_{ST} values of mtDNA haplogroup frequencies (Demirci, 2012) and pairwise F_{ST} values of Y-linked haplogroup frequencies (Parmaksız, 2013). These two comparisons were made to determine if retrotype distribution was related with the polymorphism distribution pattern of one of the sexes. The results of these Mantel's tests were given in Figure 4.5 and 4.6.

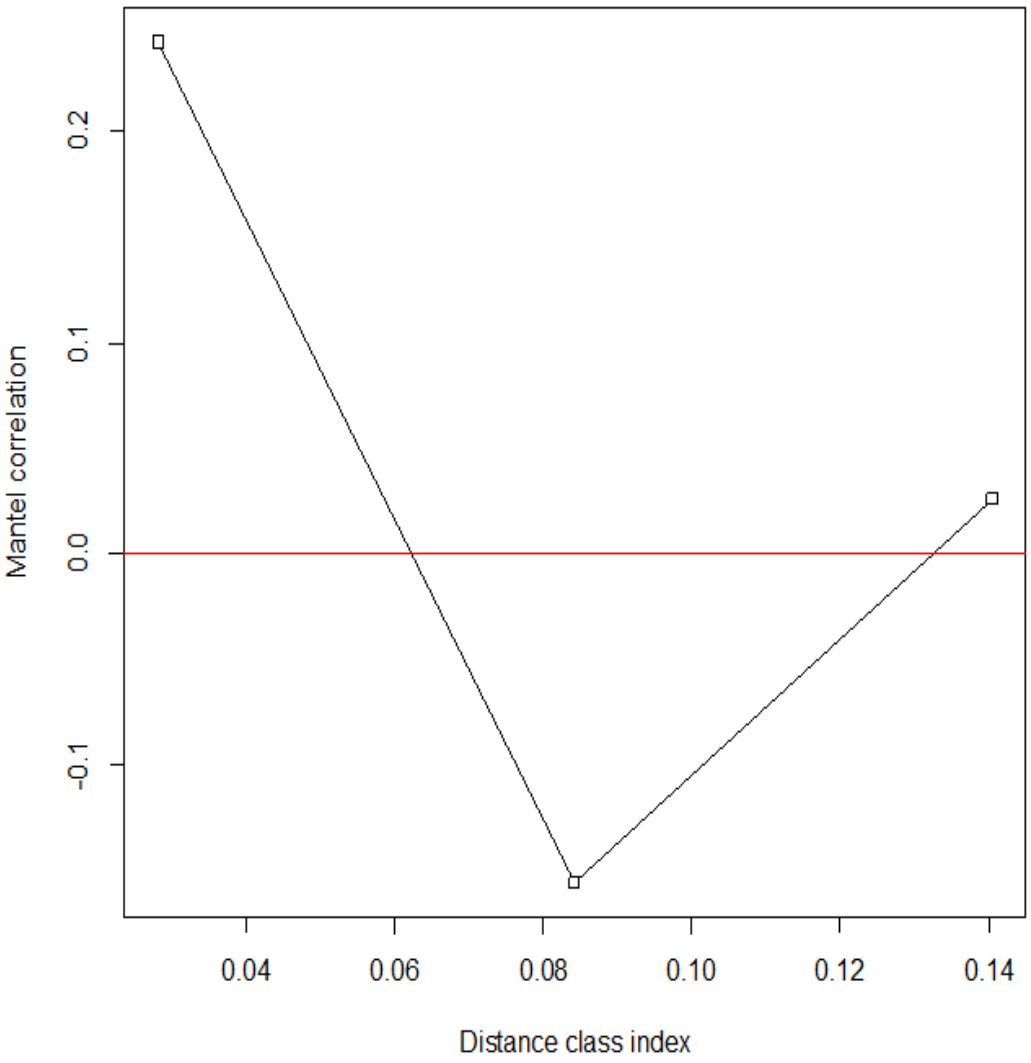

Figure 4.5 Mantel's test for the relationship between pairwise F_{ST} values of retrotype frequencies and pairwise F_{ST} values of mtDNA haplotype frequencies.

The Mantel statistics r, which is based on Pearson's product-moment correlation, is calculated as: "0.134". Significance value is 0.215 and this test is based on 999 permutations.

Empirical upper confidence limits of r are given as 0.248 for 90% confidence level, 0.350 for 95% confidence level, 0.401 for 97.5% confidence level and 0.465 for 99% confidence level.

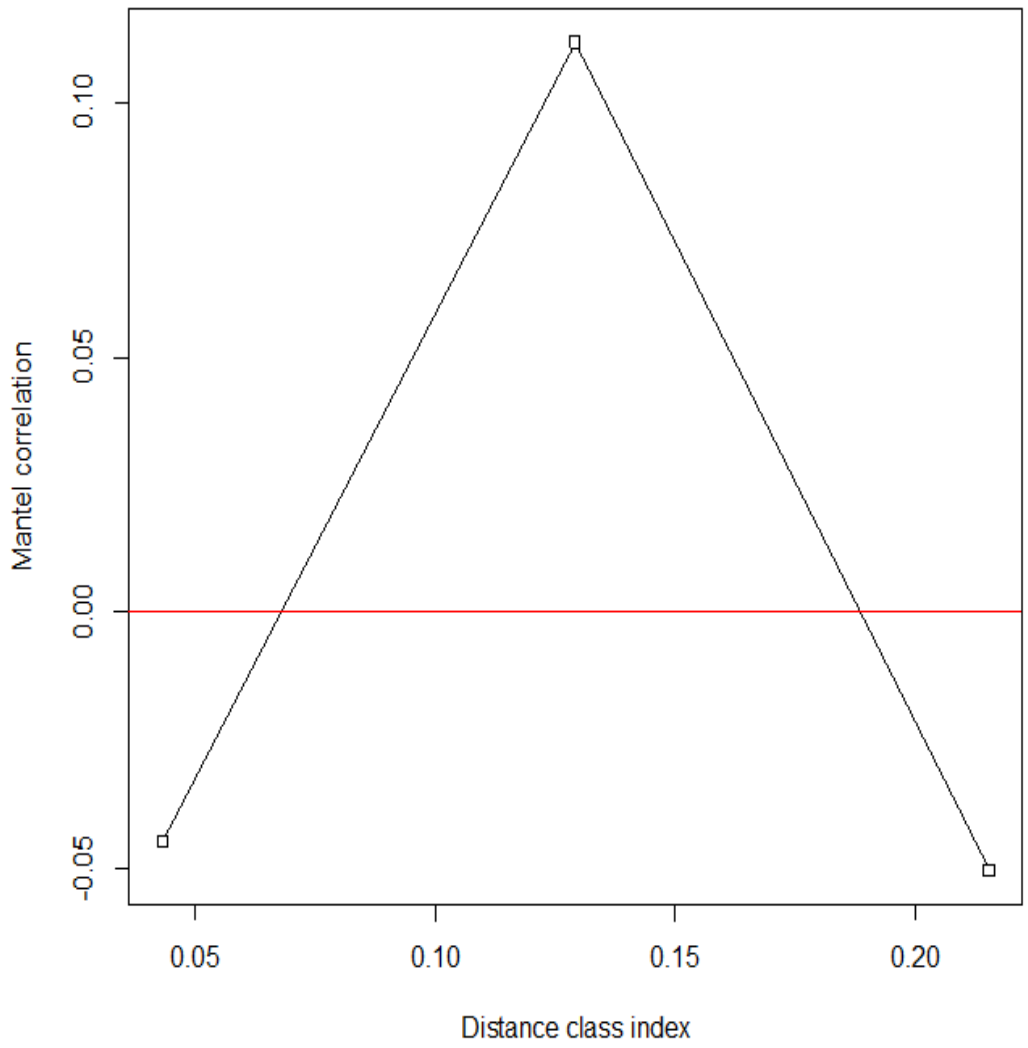

Figure 4.6 Mantel's test for the relationship between pairwise F_{ST} values of retrotype frequencies and pairwise F_{ST} values of Y-linked haplotype frequencies.

The Mantel statistics r, which is based on Pearson's product-moment correlation, is calculated as: "0.02308". Significance value is 0.099 and this test is based on 999 permutations.

Empirical upper confidence limits of r are given as 0.228 for 90% confidence level, 0.313 for 95% confidence level, 0.435 for 97.5% confidence level and 0.535 for 99% confidence level.

According these two Mantel's tests, it can be said that relationship between retrotype frequencies and mtDNA haplogroup frequencies are not significant and relationship between retrotype frequencies and Y-linked haplotype frequencies are significant with 90% confidence level.

Pairwise genetic distances between domesticated Turkish breeds in terms of retrotype frequencies are nearly significantly (p=0.099) related with the pairwise genetic distances between domesticated Turkish breeds in terms of Y-linked haplotype frequencies. In other words, findings of the nuclear markers (enJSRVs) are consistent with the results of paternal (Y-linked haplotypes). It may be said that these two markers are following a similar pattern. On the other hand, similar relationship between the nuclear markers and the maternal markers (mtDNA haplotypes) was not observed. All of these observations suggest that the second "major expansion of sheep" marked with the distribution of retrotypes (R0-R1-R2-R4-R6-R7) with secondary products as wool might be mainly mediated by males.

4.5 Absence of enJSRV-6 integration sites in some Turkish breeds

Retrotypes for Turkish native sheep breeds determined based on the criterion that enJSRV-6 integration must be amplified successfully to ensure the DNA quality of samples. As stated before during this procedure 34 samples (enJSRV-6 negative) from 7 different breeds were eliminated from the data because of the reason discussed in section 4.3.

Eliminating 34 samples from total of 254 samples (with a percentage of 9.6%) were not very high in proportion and it was thought that this could happen because DNA samples were not in good quality.

When Anatolian wild sheep (*O.g. anatolica*) samples were tried to be amplified for enJSRV-6 locus only for 7 out of 30 samples (with a percentage of 23.3) were able to be amplified. In other words, 76.7% of the samples supposedly have low DNA quality.

After this observation and knowing that part of the same DNA samples were used in previous studies without much problem has raised the argument of if enJSRV-6 integrations could be polymorphic for some of our samples and how much of the data are lost because of the assumption of enJSRV-6 integration site is fixed in sheep?

In order to answer these questions we amplified 5 *O.g.anatolica* and 7 native samples (all were determined as enJSRV-6 negative) were amplified with empty enJSRV-6 locus primers. All of the 12 samples were amplified successfully by those primers. Furthermore, all of the amplified products were sequenced to prove that these products are empty enJSRV-6 locus.

After proving the point that being enJSRV-6 negative is not always due to low DNA quality, in some individuals it is just not integrated, we wanted to reconsider part of the data where enJSRV-6 integration is considered as a polymorphic integration site. For that purpose 20 samples (10 from Karayaka and 10 from İvesi breed), which were identified as enJSRV-6 negative before, were selected. Samples were selected from Karayaka and İvesi breed because they have the highest number of discarded samples due enJSRV-6 criterion, n=10 and n=12 respectively.

Those 20 samples were amplified for empty enJSRV-6 locus. 4 out of 10 (40%) of Karayaka samples and 6 out of 10 (60%) İvesi samples showed positive results. These results mean that being negative for enJSRV-6 integration can be due to low DNA quality represented by 60% of the examined Karayaka samples and 40% of the examined İvesi samples. On the other hand, remaining samples have been identified as enJSRV-6 negative just because there is really no enJSRV-6 integration.

Total of 10 samples which are determined to have empty locus for enJSRV-6 integration were amplified for polymorphic enJSRVs. At least one polymorphic enJSRV is present in 8 of those samples while two of them have no polymorphic enJSRVs (Table 3.7).

From those 10 samples, 2 of them (KRY11 and IVE 11) were negative for all of the insertionally polymorphic enJSRVs examined (enJSRV-7, enJSRV-18 and enJS5F15). In the light of this information, they can be defined as R0 retrotype as defined by Chessa *et al*. (2009) but they are also negative for enJSRV-6 integration. This indicates that to the first approximation, 20% of the enJSRV-6 negative samples have retrotype characteristics of R0 (no polymorphic enJSRVs) whose frequency was found as 0.032 (3.2%) among 219 Turkish native sheep samples (enJSRV-6 positive) examined. Noticeable difference is present when the two cases (enJSRV-6 positives and negatives) are considered. The difference in this ratio is important because having no polymorphic enJSRVs is the characteristic of the primitive sheep breeds. In other words, important information will be lost about primitive breed by using enJSRV-6 as the DNA quality marker at least for the Turkish sheep breeds.

In the concept of primitive breeds, as stated before 7 Anatolian wild sheep individuals which were identified as enJSRV-6 positive and examined for the presence of insertionally polymorphic enJSRVs exhibits R0 (no insertionally polymorphic enJSRVs) retrotype. R0 is one of the characteristic retrotypes of primitive breeds so observation of this retrotype in a wild sheep population was expected. On the other hand, there were 23 enJSRV-6 negative individuals where, 5 of them were shown to contain empty enJSRV-6 locus in their genome. This observation leads to definition of a new retrotype "R00" in which there were no enJSRV integrations, including enJSRV-6. Similar to R0 and R1, this new retrotype can be the characteristic of primitive breeds too because it was mostly observed in a wild sheep population. As it was discussed before Anatolian wild sheep population has passed through bottleneck as low as to 30 individuals and after this event they were highly isolated from the

rest of the sheep (now population size is approximately 500). The effect of this event might have led to increase of the proportion of enJSRV-6 negative individuals. However, the absence of enJSRV-6 insertion being observed in İvesi and Karayaka confirms the existence of the polymorphism before/around the time of domestication. It is not an observation only made in *O.g. anatolica*..

As it was stated before there was no observation of remnants of the first sheep migration (R0 and R1) in Asian sheep breeds. In the light of the observations made for Anatolian wild sheep population, it can be suggested that Asian breeds may contain individuals which exhibits this new retrotype "R00". This observation might reshape the characterization of Asian breeds especially in the concept of primitive breeds and results in the better understanding of the evolutionary history of domestic sheep migration.

The retrotype R3 which is defined by the presence of enJS5F-16 only was not observed in 219 samples from Turkish native sheep breeds. This retrotype is called as "Nordic" retrotype because it was mainly observed in Northern European breeds (Chessa et al., 2009) so absence of R3 retrotype in Turkish native sheep breeds was consistent with the definition. The interesting observation among enJSRV-6 negative samples is one sample (IVE46) have only enJS5F-16 integration (on the basis of enJSRVs considered in the present study). This observation is interesting because a retrotype which was not observed in 219 samples was observed in one of the 10 enJSRV-6 negative sample. In the light of this information, it can be said that presence of enJS5F-16 integration by itself (even if it is present in a few individuals) was missed for Turkish native sheep breeds because of the use of presence of enJSRV-6 integration as a DNA quality control.

When the frequencies of polymorphic enJSRV integrations were compared according to use of enJSRV-6 integration as DNA quality control (Table 3.9), frequency changes for all of the three polymorphic enJSRVs was observed. For of the breeds (Karayaka and İvesi) enJSRV-7 frequency was increasing while enJSRV-18 integration frequency was decreasing when the enJSRV-6 negative samples were included in the data. enJS5F-16 frequency changes were present for both of the breeds in different ways. It was decreasing for Karayaka breed while it was increasing for İvesi breed. In the light of this observation it can be suggested that using enJSRV-6 integration as DNA quality control affects the results. enJSRV-6 can be fixed in European sheep breeds which account for the majority of samples in the study of Chessa *et al.* (2009) but it may not be fixed in all Turkish native sheep breeds. It can be suggested that assuming that enJSRV-6 integration is fixed may lead to loss of data before the screenings for the presence of polymorphic enJSRVs especially in Middle East region and Asia.

The changes of the enJSRV integration frequencies when enJSRV-6 considered as a polymorphic integration is noteworthy. From that point it may be proposed that there might be need of definitions for new retrotypes where enJSRV-6 is also considered polymorphic.

This new classification will lead to increase in the variation of retrotypes at least for Turkish breeds and probably for Asian breeds. With the increase of the retrotype variation direction and time of migration waves may be explained in a more detailed way than before.

CHAPTER 5

CONCLUSION

In the present study in addition to R2, R4 and R6 retrotypes which were the observed retrotypes for Turkish sheep breeds in Chessa *et al.*'s (2009) study, R0, R1 and R7 retrotypes were observed in 219 samples of Turkish native sheep breeds (Karayaka, Dağlıç, Kıvırcık, Sakız, Norduz, Karagül and Gökçeada).

Spatial analyses showed that the breeds in Western Turkey are different from the breeds present in Central and Eastern Turkey which indicates that sheep breeds in these two groups were probably originated from different ancestors.

Comparative studies of neutral dominant nuclear enJSRV markers with sex dependent markers (mtDNA and Y-linked) showed that possible second migration of the sheep may be mediated by males.

Results of enJSRV-6 positive samples but no polymorphic integrations in of Anatolian wild sheep (*Ovis* gmelinii anatolica) indicate that this population harbors only R0 retrotype which is the characteristic of one class of primitive breeds.

Integration of enJS5F-16 only itself (no other enJSRVs) has been observed in enJSRV-6 negative samples with a frequency of 0.1. This type of integration was not observed for 219 samples from 11 different breeds which are enJSRV-6 positive. These results are prompting the reinterpretation of the argument that integration of enJS5F-16 only itself is a characteristic of Nordic breeds.

Most importantly, it has been shown in 22 samples (5 from O.g. anatolica and 17 from native sheep samples) that enJSRV-6 integration was absent but it was not due to DNA quality. Also 10 of those enJSRV-6 negative samples (6 from Ivesi breed and 4 from Karayaka breed) have been investigated for the presence of polymorphic enJSRVs. In other words, on the basis of those 10 samples enJSRV-6 integration is considered as a polymorphic integration

Increasing the number of sample sites and sample sizes of Asian sheep breeds and while using enJSRV integrations as markers may lead to higher resolution in evolutionary studies of sheep. Also, using more enJSRV loci which were accepted as fixed in sheep breeds (for instance enJSRV-6) as polymorphic markers may contribute to the better understanding the evolutionary history.

enJSRV integrations can be used as highly informative molecular markers for understanding the evolutionary history of sheep but considering enJSRV-6 as a polymorphic integration and examining more enJSRV loci may lead to different and maybe more interesting results.

REFERENCES

Acar, H. (2010). Bioinformatic Analyses in Microsatellite-Based Genetic Diversity Of Turkish Sheep Breeds. MSc Thesis. The Graduate School Of Informatics, Middle East Technical University.

Açan, S.C. (2012). Genetic Diversity Of Turkish Sheep Breeds In The Focus Of Conservation Reasearch. PhD Thesis, Department of Biology, Middle East Technical University, Ankara, Turkey.

Allendorf, F. W., & Luikart G. (2007). *Conservation and the Genetics of Populations*. Malden, MA, USA: Wiley-Blackwell.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410.

Arıhan, O. (2000). Population Structure, Habitat Use, and Reintroduction of Anatolian Wild Sheep Ovis gmeliniii anatolica. MSc thesis. Biology Department, Middle East Technical University, Ankara, Turkey.

Arnaud, F., Caporale, M., Varela, M., Biek, R., Chessa, B., Alberti, A., . . . Palmarini, M. (2007). A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. PLoS Pathog, 3(11), e170.

Arnaud, F., Varela, M., Spencer, T. E., & Palmarini, M. (2008). Coevolution of endogenous betaretroviruses of sheep and their host. *Cell Mol Life Sci*, 65(21), 3422-3432.

Barker, G. (2002). *Examining the Farming/Language Dispersal Hypothesis*. P. Bellwood & C. Renfrew (Eds). Cambridge: McDonald Institute for Archaeological Research.

Bar-Yosef O, Meadow RH (1995). The origins of agriculture in the Near East. *Last Hunters, First Farmers* (pp. 39-94). In Price TD, Gebauer, A-B (Eds.). Santa Fe, NM: School of American Research Press.

Bökönyi S. (1969). *The Domestication and Exploitation of Plants and Animals*. Ucko, P. J., & Dimbleby, G.W. (Eds.). London: Duckworth.

Bruford M.W., Townsend S.J. (2006) Mitochondrial DNA diversity in modern sheep: Implications for domestication. In Zeder M.A., Bradley D.G., Emshwiller E., Smith B.D. (Eds.) *Documenting Domestication: New Genetic and Archaeological Paradigms* (pp 307-317). California, CA,USA: University of California Press.

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. Nat Rev Genet, 4(11), 900-910.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., . . . Palmarini, M. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science, 324*(5926), 532-536.

Clutton-Brock, J. (1994). *Animals and human society: Changing perspectives.* A. Manning & J. A. Serpell (Eds.). London: Routledge.

Danford, C. G. & Alston, E.R. (1877). On the mammals of Asia Minor. *Proceedings of the Zoological Society of London* 1 270-282

Demirci, S. (2012). mtdna Based Genetic Diversity of Native Sheep Breeds and Anatolian Mouflon (*Ovis gmelinii anatolica*) in Turkey. MSc thesis. The Graduate School Of Informatics, Middle East Technical University, Ankara, Turkey.

Demirci, S., Yüncü, E., Koban Baştanlar E., Togan, I. (Unpublished) Comparison of three simple methods in relation to mtDNA haplogroup identification of Domestic Sheep. Submitted Manuscript.

Dobney K. and Larson G. (2006). Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* 269 (2): 261–271.

Doğan A.Ş. (2009). Reassessment of Genetic Diversity in Native Turkish Sheep Breeds with Large Numbers of Microsatellite Markers and Mitochondrial DNA (mtDNA). MSc Thesis. Biology Department, Middle East Technical University, Ankara, Turkey.

Excoffier L., Laval G. and Schneider S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1:47-50.

Gatrell, A.C., Bailey, T.C., Diggle P.J. and Rowlingson B. S. (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Trans Inst Br Geogr NS*, 21: 256-274.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* 53: 325–328.

Guilaine J. (2003). *The Neolithic of Cyprus*. LeBrun A. (Ed). Athens: École Français d'Athèns.

Hernández Fernández, M., & Vrba, E. S. (2005). A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biol Rev Camb Philos Soc*, 80(2), 269-302.

Hiendleder S., Kaupe B.,Wassmuth R. and Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proceedings: *Biological Sciences*, 269:893-904.

Hiendleder, S., Mainz, K., Plante, Y., & Lewalski, H. (1998). Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. J Hered, 89(2), 113-120.

J. D. Boeke, J. P. Stoye (1997). *Retroviruses.* J. M. Coffin, S. H. Hughes, H. E. Varmus (Eds.). Plainview, NY: Cold Spring Harbor Laboratory Press.

Kaya M. A., Bunch T. D. and Konuk M. (2004). On Konya wild sheep, Ovis orientalis anatolica, in the Bozdag protected area. *Mammalia* 68 (2-3): 229-232.

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., . . . International Sheep Genomics Consortium, Members. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol, 10(2), e1001258.

Koban, E. (2004). Genetic Diversity of Native and Crossbreed Sheep Breeds in Anatolia. PhD Thesis. Biology Department, Middle East Technical University, Ankara, Turkey

LeBrun A, Cluzan S, Davis SJM, Hansen J, Renault-Mislkovsky (1987). The pre-ceramic Neolithic of Cyprus. *L'Anthropologie* 91, 283–316.

Manly, B.F.J. (1991). *Randomization and Monte Carlo methods in biology.* London: Chapman and Hall.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res*, 27(2), 209-220.

Meadow RH (1989). Osteological evidence for the process of animal domestication. *The Walking Larder* (pp. 80-90). In Clutton-Brock J (Ed.). London: Unwin Hyman.

Meadows, J. R., Cemal, I., Karaca, O., Gootwine, E., & Kijas, J. W. (2007). Five ovine mitochondrial lineages identified from sheep breeds of the near East. Genetics, 175(3), 1371-1379.

Meadows, J. R., Hanotte, O., Drogemuller, C., Calvo, J., Godfrey, R., Coltman, D., . . . Kijas, J. W. (2006). Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. Anim Genet, 37(5), 444-453.

Meadows, J. R., Hawken, R. J., & Kijas, J. W. (2004). Nucleotide diversity on the ovine Y chromosome. Anim Genet, 35(5), 379-385.

Meadows, J. R., Hiendleder, S., & Kijas, J. W. (2011). Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity (Edinb), 106(4), 700-706.

Palmarini, M., Sharp, J. M., de las Heras, M., & Fan, H. (1999). Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. J Virol, 73(8), 6964-6972.

Pamaksız, A. (2004). Contrubitons to understanding of sheep domestication process with investigation on polymorphism of chromosome Y of Anatolian sheep Breeds. PhD. Thesis. Biology Department, Harran University, Şanlıurfa, Turkey.

Pedrosa, S., Uzun, M., Arranz, J. J., Gutierrez-Gil, B., San Primitivo, F., & Bayon, Y. (2005). Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. Proc Biol Sci, 272(1577), 2211-2217.

Peltenburg E (2004). Introduction. *Neolithic Revolution* (pp. xi–xx). Peltenburg E, Wasse A (Eds.). Chippenham, UK: Anthony Row.

Pereira, F., Davis, S. J., Pereira, L., McEvoy, B., Bradley, D. G., & Amorim, A. (2006). Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. Mol Biol Evol, 23(7), 1420-1426.

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G., Erhardt, G., & Consortium, Econogene. (2007). Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. Anim Genet, 38(1), 37-44.

Peters J, von den Driesch A, Helmer D (2005). Theupper Euphrates-Tigris Basin, cradle of agro-pastoralism? *The First Steps of Animal Domestication* (pp. 96-124). In Vigne J-D, Peters J, Helmer D (Eds.). Oxford: Oxbow Books.

Poplin, F. (1979). Ann Genet Sel Anim, 11(2), 133.

Price, T.D. (2000). *Europe's First farmers*. Cambridge: Cambridge University Press.

Sambrook, J., Foritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Newyork, NY: Cold Spring Harbor Laboratory.

Sanger, F., Donelson, J. E., Coulson, A. R., Kossel, H., & Fischer, D. (1973). Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. *Proc Natl Acad Sci U S A*, 70(4), 1209-1213.

Sezen (Yüksel), Z. (2000). Population viability analysis for reintroduction and harvesting of Turkish mouflon *Ovis gmelinii anatolica valenciennes* 1856. MSc thesis. Department of Biology, Middle East Technical University METU, Ankara, 2000.

TAGEM (2009). *Türkiye Evcil Hayvan Genetik Kaynakları*. Ankara: TAGEM

Turan, N. (1984). Memeliler. *Türkiye'nin Av ve Yaban Hayvanları* (pp 1-177). Ankara: Olgun Kardeşler Matbaacılık Sanayi.

Uerpmann H-P. (1979). *Problems of the Neolithization of the Mediterranean Regio*n. Weisbaden: Ludwig Reichert.

Uzun, B. (2012). Surveillance of prion protein (PrP9) gene polymorphisms in Turkish native sheep breeds. MSc Thesis. Biology Department, Middle East Technical University, Ankara, Turkey.

Uzun, M., Gutierrez-Gil, B., Arranz, J. J., San Primitivo, F., Saatci, M., Kaya, M., & Bayon, Y. (2006). Genetic relationships among Turkish sheep. Genet Sel Evol, 38(5), 513-524.

Wright, S. (1951). The genetical structure of populations. *Ann. Eugenics*, 15:323-354.

Wright, S. (1965). The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution*, 19(3), 395-420.

Yüncü, E. (2009). Mitochondrial DNA (mtDNA) Haplogroup Composition in Turkish Sheep Breeds. MSc Thesis. Biology Department, Middle East Technical University, Ankara, Turkey.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin:Origins, diffusion, and impact. *Proc Natl Acad Sci U S A, 105*(33), 11597-11604.

Zeder, M. A., Emshwiller, E., Smith, B. D., & Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends Genet, 22*(3), 139-155.

Zilhao, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci U S A*, 98(24), 14180-14185. doi: 10.1073/pnas.241522898.

Zohary, D., Tchernov, E., & Horwitz, L. (1998). The role of unconscious selection in the domestication of sheep and goats. *J. Zool.* 245, 129-135.

APPENDIX A

RETROTYPES RESULTS FOR EACH SAMPLE ACCORDING TO ENJSRV INTEGRATIONS AND THE COLLECTION SITES OF THE EACH SAMPLE.

Table A.1 Retrotypes and Locations of the Samples

Sample ID	Breed Name	Retrotype	Latitude	Longitude
KRY4	Karayaka	R2	36.82	40.6391
KRY5	Karayaka	R2	36.82	40.6391
KRY8	Karayaka	R4	36.82	40.6391
KRY10	Karayaka	R2	37.2129	40.5738
KRY13	Karayaka	R2	37.2129	40.5738
KRY14	Karayaka	R2	36.6067	40.541
KRY20	Karayaka	R2	36.4705	40.3219
KRY22	Karayaka	R4	36.7333	40.5359
KRY23	Karayaka	R2	36.249	40.1685
KRY28	Karayaka	R4	36.4705	40.3219
KRY32	Karayaka	R2	36.5955	40.7404
KRY34	Karayaka	R2	35.7256	40.167
KRY36	Karayaka	R2	36.5955	40.7404
KRY37	Karayaka	R2	36.5955	40.7404
KRY39	Karayaka	R2	36.5523	40.3351
KRY42	Karayaka	R2	37.4493	40.7503
KRY45	Karayaka	R0	37.4493	40.7503
KRY46	Karayaka	R2	37.4493	40.7503
KRY47	Karayaka	R6	37.4493	40.7503
KRY48	Karayaka	R6	37.2936	40.5417
DAG1	Dağlıç	R2	38.47	31.05
DAG2	Dağlıç	R0	38.47	31.05
DAG3	Dağlıç	R2	38.47	31.05
DAG4	Dağlıç	R2	38.46	31.06
DAG6	Dağlıç	R2	38.42	31.05
DAG7	Dağlıç	R2	38.42	31.05
DAG13	Dağlıç	R2	38.08	30.15
DAG15	Dağlıç	R2	38.59	30.3

Table A.1 (continued)

DAG16	Dağlıç	R2	38.59	30.3
DAG17	Dağlıç	R2	38.59	30.3
DAG19	Dağlıç	R2	38.51	30.31
DAG21	Dağlıç	R7	38.37	29.55
DAG24	Dağlıç	R4	38.47	31.05
DAG32	Dağlıç	R2	38.37	29.55
DAG34	Dağlıç	R2	38.47	31.05
DAG42	Dağlıç	R2	38.47	31.05
DAG45	Dağlıç	R2	38.47	31.05
DAG46	Dağlıç	R2	38.47	31.05
DAG47	Dağlıç	R2	38.08	30.15
DAG49	Dağlıç	R2	38.08	30.15
KIV1	Kıvırcık	R2	41.8377	27.4681
KIV5	Kıvırcık	R2	41.902	27.5458
KIV6	Kıvırcık	R2	41.8377	27.4681
KIV12	Kıvırcık	R2	41.8377	27.4681
KIV16	Kıvırcık	R2	41.902	27.5458
KIV17	Kıvırcık	R2	41.902	27.5458
KIV19	Kıvırcık	R2	41.902	27.5458
KIV21	Kıvırcık	R2	41.8377	27.4681
KIV24	Kıvırcık	R4	41.8377	27.4681
KIV27	Kıvırcık	R1	41.8377	27.4681
KIV31	Kıvırcık	R2	41.8377	27.4681
KIV34	Kıvırcık	R6	41.8377	27.4681
KIV36	Kıvırcık	R7	41.8377	27.4681
KIV37	Kıvırcık	R2	41.8377	27.4681
KIV38	Kıvırcık	R6	41.8377	27.4681
KIV40	Kıvırcık	R2	41.8377	27.4681
KIV41	Kıvırcık	R2	41.8377	27.4681
KIV42	Kıvırcık	R4	41.8377	27.4681
KIV43	Kıvırcık	R4	41.8377	27.4681
KIV48	Kıvırcık	R2	41.8377	27.4681
SAK2	Sakız	R0	38.3408	26.2908
SAK12	Sakız	R2	38.3	26.45
SAK13	Sakız	R2	38.3	26.45
SAK15	Sakız	R2	38.3	26.45
SAK17	Sakız	R2	38.3	26.45
SAK18	Sakız	R4	38.3	26.45
SAK19	Sakız	R2	38.25	26.3
SAK21	Sakız	R2	38.2833	26.2833
SAK22	Sakız	R2	38.2833	26.2833
SAK23	Sakız	R2	38.2833	26.2833
SAK25	Sakız	R2	38.2833	26.2833
SAK26	Sakız	R4	38.2833	26.2833
SAK28	Sakız	R1	38.2833	26.2833
SAK30	Sakız	R4	38.2833	26.2833
SAK33	Sakız	R4	38.2833	26.2833

Table A.1 (continued)

SAK34	Sakız	R6	38.2833	26.2833
SAK36	Sakız	R0	38.3	26.45
SAK38	Sakız	R2	38.3	26.45
SAK39	Sakız	R2	38.3	26.45
SAK40	Sakız	R2	38.3	26.45
AKK1	Akkaraman	R2	37.9	32.883
AKK2	Akkaraman	R6	37.9	32.883
AKK4	Akkaraman	R2	37.9	32.883
AKK5	Akkaraman	R6	37.9	32.883
AKK6	Akkaraman	R2	37.9	32.883
AKK9	Akkaraman	R2	37.9	32.883
AKK10	Akkaraman	R2	37.9	32.883
AKK15	Akkaraman	R2	37.9	32.883
AKK18	Akkaraman	R2	37.9	32.883
AKK23	Akkaraman	R2	37.833	32.783
AKK25	Akkaraman	R2	37.833	32.783
AKK27	Akkaraman	R2	37.966	32.9
AKK28	Akkaraman	R2	37.966	32.9
AKK29	Akkaraman	R4	37.966	32.9
AKK32	Akkaraman	R2	37.9	32.883
AKK37	Akkaraman	R2	37.9	32.883
AKK39	Akkaraman	R2	37.9	32.883
AKK40	Akkaraman	R2	37.9	32.883
AKK45	Akkaraman	R4	37.833	32.783
AKK48	Akkaraman	R2	37.833	32.783
IVE2	İvesi	R2	39.26	37.06
IVE3	İvesi	R4	39.26	37.06
IVE6	İvesi	R2	39.26	37.06
IVE10	İvesi	R2	39.26	37.06
IVE12	İvesi	R2	38.9	37.12
IVE13	İvesi	R2	39.14	36.99
IVE16	İvesi	R2	39.26	37.06
IVE17	İvesi	R2	38.93	37.1
IVE20	İvesi	R2	38.79	37.14
IVE22	İvesi	R2	38.79	37.14
IVE23	İvesi	R4	38.79	37.14
IVE24	İvesi	R2	38.79	37.14
IVE26	İvesi	R2	39.14	36.99
IVE28	İvesi	R2	39.15	37.01
IVE29	İvesi	R2	38.93	37.1
IVE30	İvesi	R2	39.26	37.06
IVE32	İvesi	R2	38.79	37.14
IVE39	İvesi	R2	38.79	37.14
IVE41	İvesi	R2	39.14	36.99
IVE42	İvesi	R6	38.93	37.1
NOR2	Norduz	R2	43.37	38.03
NOR3	Norduz	R2	43.44	37.87

Table A.1 (continued)

NOR6	Norduz	R0	43.44	37.87
NOR7	Norduz	R4	43.44	37.87
NOR9	Norduz	R6	43.44	37.87
NOR10	Norduz	R2	43.44	37.87
NOR12	Norduz	R0	43.37	38.03
NOR22	Norduz	R2	44.01	38.04
NOR23	Norduz	R2	44.01	38.04
NOR26	Norduz	R2	43.41	38.32
NOR30	Norduz	R2	43.41	38.32
NOR31	Norduz	R2	43.41	38.32
NOR32	Norduz	R2	43.41	38.32
NOR33	Norduz	R2	43.41	38.32
NOR34	Norduz	R2	43.41	38.32
NOR35	Norduz	R4	43.41	38.32
NOR37	Norduz	R2	43.41	38.32
NOR43	Norduz	R2	43.37	38.03
NOR45	Norduz	R2	43.41	38.32
NOR46	Norduz	R2	43.41	38.32
KRG2	Karagül	EnJSRV-6 (-)	36.15	40.6
KRG3	Karagül	R2	36.15	40.6
KRG4	Karagül	R2	36.15	40.6
KRG5	Karagül	R2	36.39	40.29
KRG9	Karagül	R2	36.39	40.29
KRG11	Karagül	R2	36.39	40.29
KRG12	Karagül	R2	36.39	40.29
KRG14	Karagül	R2	36.43	40.3
KRG15	Karagül	R2	36.43	40.3
KRG16	Karagül	R2	36.43	40.3
KRG19	Karagül	R2	36.39	40.29
KRG20	Karagül	R2	36.39	40.29
KRG21	Karagül	R2	36.39	40.29
KRG22	Karagül	R2	36.39	40.29
KRG23	Karagül	R2	36.39	40.29
KRG33	Karagül	R2	36.39	40.29
KRG41	Karagül	R2	36.39	40.46
KRG47	Karagül	R0	36.43	40.33
KRG49	Karagül	R4	36.39	40.29
KRG52	Karagül	R2	36.43	40.33
HEM1	Hemşin	R2	41.11	42.07
HEM2	Hemşin	R2	42.21	41.04
HEM3	Hemşin	R2	41.11	42.07
HEM5	Hemşin	R4	41.98	41.07
HEM15	Hemşin	R4	42.21	41.03
HEM16	Hemşin	R4	42.06	41.16
HEM17	Hemşin	R2	42.21	41.04
HEM20	Hemşin	R4	42.31	41.29
HEM22	Hemşin	R2	42.06	41.16

Table A.1 (continued)

HEM24	Hemşin	R2	42.21	41.04
HEM26	Hemşin	R2	42.31	41.29
HEM32	Hemşin	R2	41.98	41.07
HEM33	Hemşin	R2	42.06	41.16
HEM34	Hemşin	R2	42.16	41.04
HEM35	Hemşin	R6	42.06	41.16
HEM36	Hemşin	R2	42.21	41.04
HEM40	Hemşin	R2	42.21	41.04
HEM41	Hemşin	R2	41.11	42.15
HEM42	Hemşin	R2	42.06	41.16
HEM44	Hemşin	R2	42.06	41.14
GOK1	Gökçeada	R2	40.2321	25.9446
GOK2	Gökçeada	R2	40.2152	25.9128
GOK5	Gökçeada	R2	40.2152	25.9128
GOK8	Gökçeada	R2	40.1297	25.9493
GOK10	Gökçeada	R2	40.2148	25.9373
GOK15	Gökçeada	R2	40.2148	25.9373
GOK18	Gökçeada	R2	40.2148	25.9373
GOK20	Gökçeada	R2	40.2148	25.9373
GOK22	Gökçeada	R2	40.2148	25.9373
GOK25	Gökçeada	R2	40.2148	25.9373
GOK28	Gökçeada	R2	40.2305	28.9388
GOK30	Gökçeada	R2	40.2305	28.9388
GOK32	Gökçeada	R2	40.2152	25.9128
GOK35	Gökçeada	R2	40.2152	25.9128
GOK36	Gökçeada	R2	40.1297	25.9493
GOK40	Gökçeada	R2	40.1297	25.9493
GOK42	Gökçeada	R2	40.1297	25.9493
GOK43	Gökçeada	R2	40.1297	25.9493
GOK47	Gökçeada	R2	40.1297	25.9493
GOK49	Gökçeada	R2	40.1297	25.9493
MRK2	Morkaraman	R2	41.76	39.77
MRK4	Morkaraman	R2	41.8	40.07
MRK6	Morkaraman	R2	41.76	39.77
MRK7	Morkaraman	R2	41.82	40.01
MRK8	Morkaraman	R2	41.5	39.92
MRK10	Morkaraman	R2	41.49	39.99
MRK11	Morkaraman	R2	42.39	39.7
MRK14	Morkaraman	R2	41.67	39.89
MRK17	Morkaraman	R2	41.09	40.04
MRK18	Morkaraman	R2	41.67	39.89
MRK22	Morkaraman	R2	41.68	40.03
MRK23	Morkaraman	R2	41.72	39.94
MRK24	Morkaraman	R2	41.53	39.92
MRK26	Morkaraman	R2	41.8	40.07
MRK29	Morkaraman	R2	41.68	40.03
MRK30	Morkaraman	R6	41.82	40.01

Table A.1 (continued)

MRK32	Morkaraman	R6	41.68	40.03
MRK33	Morkaraman	R4	41.49	39.92
MRK37	Morkaraman	R4	41.68	40.03
MRK39	Morkaraman	R2	41.76	39.77

APPENDIX B

BREED SPECIFIC MODIFICATIONS OF PCR CONDITIONS

Table B.1 Breed specific modifications of PCR conditions given in Section 2.2.2.2.

Breed	**enJSRV type**	**PCR type**	**Modifications**
Dağlıç	enJSRV-6	5’	Annealing: 63^0C
Akkaraman	enJSRV-6	3’	Annealing: 62^0C and 45 secs
İvesi	enJSRV-6	5’	Annealing: 64^0C, 1 X NH_4 buffer, 2.5 mM $MgCl_2$
Karagül	enJSRV-6	3’	Annealing: 45 secs, 35 cycles
Gökçeada	enJSRV-6	3’	Annealing: 63^0C, Extension:50 secs, 35 cycles
Dağlıç	enJSRV-7	5’	Annealing: 59^0C, 30 cycles
Kıvırcık	enJSRV-7	Both	Annealing 58^0C, 1.5 X KCl buffer
Akkaraman	enJSRV-7	5’	Annealing: 60^0C and 1 min
İvesi	enJSRV-7	Both	Annealing 60^0C, 1 X NH_4 buffer, 2.5mM $MgCl_2$
Norduz	enJSRV-7	3’	Annealing:61^0C
Karagül	enJSRV-7	Both	Annealing 45 secs, 40 cycles
Gökçeada	enJSRV-7	3’	1 X NH_4 buffer, 2.5 mM $MgCl_2$
Morkaraman	enJSRV-7	5’	Annealing: 40 secs, 40 cycles
Dağlıç	enJSRV-18	Both	Annealing 63^0C, 1.5 X NH_4 buffer, 2mM $MgCl_2$
Kıvırcık	enJSRV-18	5’	Annealing: 64^0C and 50 secs
Akkaraman	enJSRV-18	5’	Annealing: 50 secs, 40 cycles
İvesi	enJSRV-18	5’	Annealing: 63^0C
Norduz	enJSRV-18	3’	Annealing: 61^0C, 1 X NH_4 buffer, 2.5 mM $MgCl_2$
Karagül	enJSRV-18	5’	1 X KCl buffer, 2.5 mM $MgCl_2$, 30 cycles
Hemşin	enJSRV-18	5’	Annealing: 60^0C
Gökçeada	enJSRV-18	5’	Annealing: 64^0C
Morkaraman	enJSRV-18	5’	Annealing: 63^0C, 30 cyles
Dağlıç	enJS5F-16	3’	Annealing: 58^0C, 35 cycles
Kıvırcık	enJS5F-16	3’	Annealing: 59^0C, 35 cycles
Sakız	enJS5F-16	5’	Annealing 58^0C, 1 X KCl buffer,
Akkaraman	enJS5F-16	3’	1 X KCl buffer, 2.5mM $MgCl_2$, 2.5mM $MgCl_2$
İvesi	enJS5F-16	Both	Annealing: 59^0C, Final extension:15 mins, 35 cycles