EXPLORING THE EVOLUTIONARY HISTORY OF ANATOLIAN NEOLITHIC SHEEP USING MODERN AND ANCIENT GENOMICS

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

BY

ERİNÇ YURTMAN

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

SEPTEMBER 2019

Approval of the thesis:

EXPLORING THE EVOLUTIONARY HISTORY OF ANATOLIAN NEOLITHIC SHEEP USING MODERN AND ANCIENT GENOMICS

submitted by **ERİNÇ YURTMAN** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Ayşe Gül Gözen
Head of Department, **Biology** _____

Assoc. Prof. Dr. Mehmet Somel
Supervisor, **Biology, METU** _____

Dr. Füsun Özer
Co-Supervisor, **Anthropology Dept., Hacettepe University** _____

Examining Committee Members:

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

Assoc. Prof. Dr. Mehmet Somel
Biology, METU _____

Prof. Dr. Ergi Deniz Özsoy
Biology Dept., Hacettepe University _____

Assoc. Prof. Dr. Emel Özkan Ünal
Animal Science Dept., Tekirdağ Namık Kemal University _____

Assist. Prof. Dr. Alexey Yanchukov
Biology Dept., Zonguldak Bülent Ecevit University _____

Date: 06.09.2019

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, Surname: Erinç Yurtman

Signature:

ABSTRACT

EXPLORING THE EVOLUTIONARY HISTORY OF ANATOLIAN NEOLITHIC SHEEP USING MODERN AND ANCIENT GENOMICS

Yurtman, Erinç
Master of Science, Biology
Supervisor: Assoc. Prof. Dr. Mehmet Somel
Co-Supervisor: Dr. Füsun Özer

September 2019, 58 pages

The transition from hunting-gathering to sedentism happened in West Asia in the early Holocene, eventually giving way to the establishment of agriculture and livestock breeding. In this process, domestication of wild animals played crucial role for human settlements. The domestication center of sheep, among the main four livestock species, is thought to have been within Anatolia. Previous archaeozoological studies also suggested that after domestication this species migrated with human populations to other parts of the world. There has yet been no ancient whole-genome study investigating the evolutionary history of Anatolian sheep from the Neolithic period. In this project, we analyzed ancient genome data belonging to Neolithic sheep from Tepecik Höyük (Central Anatolia) and Ulucak Höyük (West Anatolia), comparing these with modern sheep genomic data. Three different statistical methods (PCA, *D-statistics* and *f3-statistics*) were used to understand the relationship between modern European and Asian sheep breeds and Anatolian Neolithic domestic sheep. Possible migration routes (the Land Route and the Maritime Route) that led from Anatolia to neighboring regions were investigated. We report that Anatolian Neolithic sheep samples are genetically closer to modern European sheep breeds (showing particularly high affinity to Central and Northern European breeds) than to non-European

populations. This implies that Anatolian Neolithic sheep might have contributed to modern domesticated European breeds gene pool by having migrated through the Land Route. This work thus provides a first genomic insight into domestic sheep history using ancient DNA.

Keywords: Ancient DNA, Sheep, Neolithic, Domestication, Migration

ÖZ

ANADOLU NEOLİTİK KOYUNUNUN EVRİMSEL GEÇMİŞİNİN ANTİK VE MODERN GENOM KULLANILARAK İNCELENMESİ

Yurtman, Erinç
Yüksek Lisans, Biyoloji
Tez Danışmanı: Doç. Dr. Mehmet Somel
Ortak Tez Danışmanı: Dr. Füsun Özer

Eylül 2019, 58 sayfa

İnsanlığın ilk olarak avcı-toplayıcı düzenden yerleşik hayata Doğu Akdeniz ve Mezopotamya coğrafyasında geçmiş olması bölgede geçmişte varlığını sürdürmüş insan popülasyonlarının diğer canlılarla ve diğer insan popülasyonlarıyla etkileşimlerinin sık olmasını sağlamıştır. Neolitik Dönüşüm ve yerleşik yaşama geçiş beraberinde tarım ve hayvancılığın başlamasını sağlamıştır. Bu noktada yabani hayvanların (koyun, keçi, vs.) evcilleştirilmesi insanlık için kritik bir rol oynamıştır. Dört ana çiftlik hayvanından koyunun evcilleştirme merkezinin Anadolu'nun içinde olduğu düşünülmektedir. Önceki arkeozoolojik çalışmalar da evcilleştirme sonrası bu türün insanlarla beraber dünyanın çeşitli bölgelerine göç ettiğini önermektedir. Bu projede Tepecik Höyük (Orta Anadolu) ve Ulucak Höyük (Batı Anadolu) kazı yerlerlerinden alınan Neolitik döneme ait koyun örneklerinin antik genomu, modern koyun ırklarının genomlarıyla karşılaştırarak analiz edildi. Modern Avrupa ve Asya ırklarının ve evcil Anadolu Neolitik koyunların arasındaki ilişkiyi anlamak için üç farklı istatistiksel yöntem (TBA, *D-istatistiği* ve *f3-istatistiği*) kullanıldı. Anadolu'dan yakın bölgelere göç edilirken kullanıldığı düşünülen olası göç yolları (Kara Yolu ve Deniz Yolu) incelendi. Sonuç olarak Anadolu Neolitik koyunlarının modern Avrupa ırklarına (Orta ve Kuzey Avrupa ırklarına daha fazla) Avrupa'dan olmayan ırklara

göre genetik olarak daha yakın olduğunu gözlemledik. Bu sonuç bize Anadolu Neolitik koyunlarının modern Avrupa ırklarının gen havuzuna Kara Yolu'ndan göç ederek katkı yapmış olabileceğini gösteriyor. Böylelikle bu çalışma evcil koyunların tarihine antik DNA kullanılarak yapılmış ilk genomik incelemeyi sunuyor.

Anahtar Kelimeler: Antik DNA, Koyun, Neolitik, Evcilleştirme, Göç

To My Family

ACKNOWLEDGEMENTS

Firstly, I would like to thank to my advisor Assoc. Prof. Dr. Mehmet Somel, my co-advisor Dr. Füsun Özer and my unofficial co-advisor Prof. Dr. İnci Togan for their scientific guidance during my master's years. They inspired me with their endless enthusiasm about ancient DNA research. They always lead me to think with other aspects.

I would like to thank to Dilek Koptekin and Mustafa Özkan who helped me by showing how to analyze ancient DNA data. I always benefited from their scientific guidance in every problem that I encountered during analysis. I also wanted to express my deep gratitude to Dr. Füsun Özer, Nihan Dilşad Dağtaş Kılıç and Dr. Eren Yüncü who gathered and prepared ancient samples in METU aDNA laboratory. I would like to thank to all CompEvo Laboratory members for their friendship and support.

I am very grateful to Reyhan Yaka, Evrim Fer, Elif Bozlak and Zeliha Gözde Turan for their endless moral support in every crisis moments that I faced during my master years. They really gave me strength for further challenges. They also provided scientific support by answering my endless questions about my analysis and giving me so many different perspectives about science.

Finally, I want to express my deepest gratitude to my family. They always supported my ideas and choices about life. They gave me the biggest support by showing their love to me in every situations. Their wise advices about life would always shed light on my darkest moments. I am very lucky to have them.

This study was supported by the Scientific and Technical Research Council of Turkey (TÜBİTAK) as a part of the project "Güneydoğu Anadolu'da evcilleştirildiği bilinen koyunun, evcilleştirme merkezinden Batı Anadolu'ya doğru gidiş yolunda geçirdiği evcilleştirme sürecinin antik mtDNA kullanılarak araştırılması" under the grant number 114Z356.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

ANS: Anatolian Neolithic sheep

DNA: deoxyribonucleic acid

aDNA: ancient DNA

PCA: principal component analysis

SNP: single nucleotide polymorphism

PCR: polymerase chain reaction

BP: Before Present

BCE: Before Common Era

mtDNA: mitochondrial DNA

PMD: post mortem damage

HPG: haplogroup

CHAPTER 1

INTRODUCTION

1.1. Brief introduction to history of sheep domestication

The history of humankind has been changed by the Neolithic Transition. This process is thought as one of the most important milestones for our journey of life. In fact, this transition process had a major impact on human life style by including farming and herding into our lives (Childe, 1925; Larsen, 1995). Herding management systems and domestication were important since obtaining products of livestock animals, such as meat, milk and wool, which were crucial or became crucial for human populations, who had embarked on sedentism under environmental conditions that kept changing over time (Zeder, Emshwiller, Smith, & Bradley, 2006).

Domestication is a process that was carried out by human populations targeting animal and plant populations (Zeder et al., 2006). The first main aim of animal domestication probably was to benefit from desired features of targeted populations by selecting and breeding individuals from wild populations that have properties that human populations required, such as fullfilling the need of meat (Sherratt, 1981). Even though meat acquirement was the first aim of domestication in early periods of Neolitic Time, especially during 5,000 BP in South-West Asia and 4,000 BP in Europe wool production of domesticated animals as secondary product gained importance. In Chessa et al. (2009) study, it was claimed that a special sheep breed was selected for high wool production in Near East and this breed spread through other continents. Therefore it can be said that purposes of domestication process have been changed with strategies that have been used for domestication through time. Therefore it is very important to explore early domestication strategies of human populations to

understand both Neolithic transition process and evolutionary history of domesticated animals.

In early periods of the Neolithic Transition in west Eurasia, human populations started herd management of four livestock species: sheep (*Ovis aries*), goat (*Capra hircus*), cattle (*Bos taurus*) and pig (*Sus scrofa*) (Zeder, 2008). In this region, the center of domestication appears to have located between Central Anatolia and Southwestern Anatolia and the Northern Zagros Mountains, which are parts of the Fertile Crescent region where first farming and domestication of crop plants activities thought to be established (Hillman, 1996). It is suggested that sheep and goat species were first domesticated around the same time period, around 11,000 BP, and then pig and cattle species were subjected to the herd management system around 10,500 and 10,000 BP (Demirci et al., 2013; Zeder, 2008) (Figure 1.1).

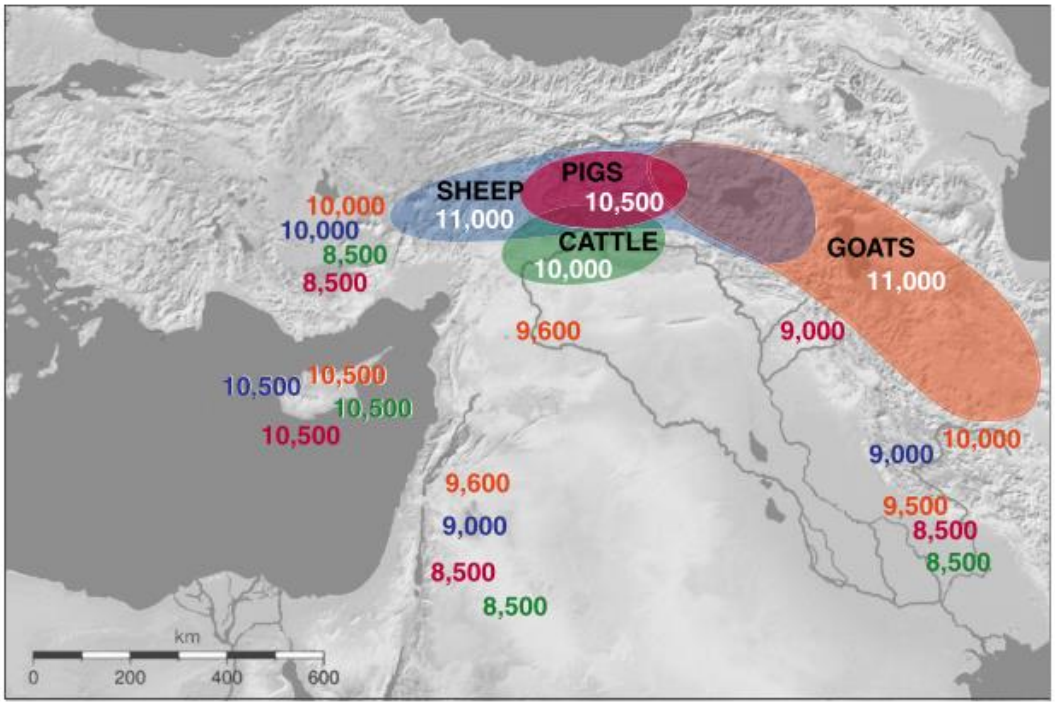

Figure 1.1. Domestication center of four main livestock species. Colored regions show the areas of initial domestication processes for each species and numbers show the approximate times (Before Present, BP). Figure was taken from Zeder (2008).

In the last century, archaeozoological studies carried out at excavation sites in this region revealed signs of domestication since there was more than one domestication strategy that early farmers followed, zoorarchaeologists also need to trace signs of domestications using more than one marker (Arbuckle et al., 2014). One of the most commonly employed marker to detect domestication is the sharp decrease of overall body size of prey individuals found at excavation sites. It has been shown that reduction of body size and morphological change of bones that was observed after domestication compared to wild relatives is inevitable for the prey population (Zeder & Hesse, 2000). Arbuckle et al. (2014) claimed that especially in later Neolithic periods, the degree of sexual dimorphism, which is observable at high levels in wild populations, was reduced as a result of the domestication process. However, since the reduction of overall body size that occurred by human selection is probably not a quick process, in fact it might have taken multiple generations to observe such overall body size reduction in general population. Therefore overall reduction of body size that was observed in early Neolithic times would not be interpreted as result of artificial selection. It has been indicated by Arbuckle et al. (2014) that one of the herding strategies of Neolithic human populations was 'young male culling' and this procedure might have led archaeozoologists to observe overall body size reduction. This strategy is based on slaughtering excess male individuals at young age in a population, which are not fruitful for herd reproduction. Therefore, unlike hunter gatherers who targeted large animals in a hunt, populations which adopted herding butchered female animals beyond reproductive ages and young males not useful for their herd (Zeder, 2008). Arbuckle et al. (2014) also argued that the observed overall body size reduction (smaller bones of female and young male individuals) in Neolithic sites might be because of different herd management applications like culling strategies, but not the direct effect of selection pressures to achieve smaller size, especially in early Neolithic times. Another strong sign of domestication that archaeozoologists use is the arrival of foreign prey taxa in a region (Meadow, 1989).

Since their natural habitat is not the region where they are secondarily found, it can be thought that their appearance is result of herd management applications of human populations (Meadow, 1989).

1.2. Spread of domesticated livestock and possible migration routes

After domestication, domesticated animals spread through Europe continent along with human populations. Two main migrations routes were proposed as a result of both archaeozoological and genetic studies. They are called as the Danubian (Land) Route and the Maritime (Mediterranean) Route (Bökönyi, 1974; Fernandez et al., 2006; Ryder, 1983) (Figure 1.2). These two routes possibly had been used during the same periods by the first farmers of Southwest Asia and domesticated livestock animals have been introduced to Europe.

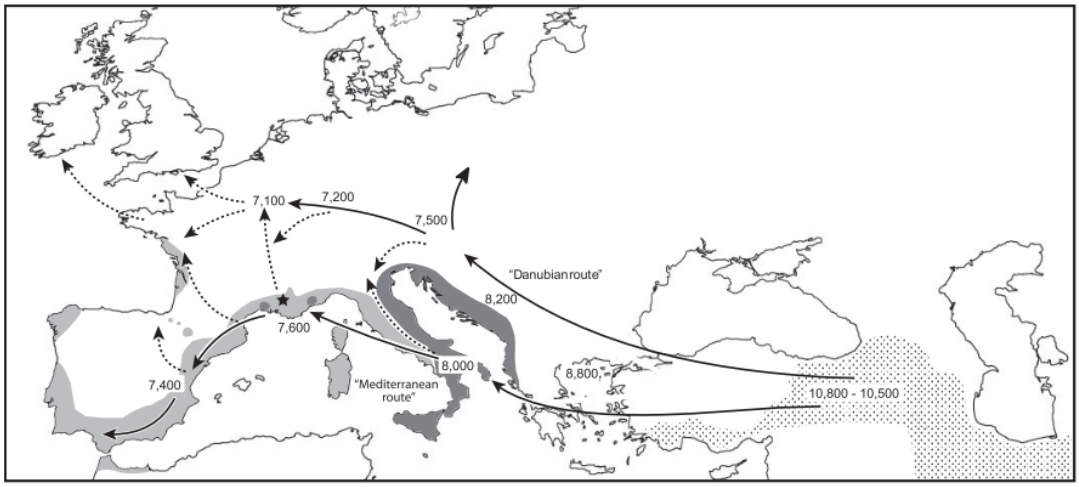

Figure 1.2. The two main routes of spread to west; The Danubian and The Mediterranean routes including other possible small movements. Dates are approximate time (Before Present) of migrations from domestication center. Figure was taken from Fernandez (2006).

It is suggested that the Neolithic colonization of Central Europe which then spread to Northern and Southern Europe has started by migrations followed through the Land route (Lazaridis et al., 2014; Mathieson et al., 2015; Skoglund et al., 2014). Previous studies showed that this route was beginning from upper part of domestication center and the Fertile Crescent then continuing across the Central and Northern-West Anatolia. Following along the Danube river this route reaches to Central Europe (Figure 1.2). The Mediterranean route is also thought to start from the Northern Fertile Crescent which followed Northern Mediterranean coastline including regions of Anatolia, Greece, Italy, France and Spain (Figure 1.2.). Improved seafaring activities facilitated human movement along the coastal areas (Ammerman, 2010) so that another major migration route was established.

Human populations that migrated through this route must have taken their domesticated animals with them. Many genetic studies were interested in to untangle the dynamics of these migration routes. Larson et al. (2007) has conducted a mtDNA study suggested that two main migration routes were used while spreading domesticated pigs into Europe in fact they claimed that very first domesticated pigs were introduced to Europe along the Land route. Consistent with these findings, another mtDNA study indicated that domesticated goats were carried to Europe using both of the routes. However, their results pointed out that the primary route which goats were carried to the Western Europe was the Mediterranean route (Fernandez et al., 2006). Genomic analysis of the modern domesticated breeds by Kijas et al. (2012) could also reveal signs of Mediterranean route but they emphasized that for detection of Danubian route further investigations are needed. Alberto et al. (2018) also detected shared selected genomic features between modern domesticated breeds of Iranian and Moroccon sheep which might indicate diffusion through Mediterranean route after initial domestication. Moreover, in their endogenous retrovirus integration sites study in modern domestic and wild sheep breeds, Chessa et al. (2009) claimed migration second wave of migration to Europe without clarifying the route. In their study, endogenous retrovirus families on both modern domesticated and wild species of

sheep were analyzed. According to Chessa et al.'s (2009) study, one sheep breed was selected for high production of quality wool and it migrated through different parts of world to substitute first domesticated sheep breed.

1.3. Study of ancient DNA

In the last few decades, in order to explore demographic and evolutionary history of human and animal populations, DNA obtained from archaeological samples that were excavated from ancient settlements have been used. Studying ancient DNA (aDNA) could provide direct information about migrations and interactions between past animal as well as human populations Additionally, genomic changes that was caused by evolutionary forces like selection pressure on populations could also be detected via aDNA studies (Gamba et al., 2014; Mathieson et al., 2015).

There are many difficulties in studying ancient DNA . First and most problematic one is the amount of DNA that is obtained from thousands, even if not tens of thousands of years old remains. After death, degradation of DNA starts by certain factors. After death, certain enzymes such as nucleases and postmortem reactions such as deamination and depurination could break down DNA molecule into small fragments and make it impossible to obtain a whole intact DNA molecule. The degree of postmortem damage (PMD) increases as time progresses and damages such as transitions from Cytosine (C) to Thymine (T) accumulates at the ends of the broken DNA fragments. The level of salinity, humidity and pH of soil that expose buried samples for a long time and the high environmental temperatures are known to affect preservation of endogenous DNA (Hofreiter, Serre, Poinar, Kuch, & P. äbo, 2001). In very low temperatures such as in permafrost, high amount of aDNA can be preserved for long time periods since enzymatic reactions and other reagents which would degrade DNA are inhibited (Poinar & Stankiewicz, 2002).

Secondly, contamination is another issue that may cause misinterpretations about the ancient sample under investigation. There are multiple methods that confirm aDNA content. Analysis of amino acid content in specimen would give information about authenticity of DNA fragments sequences. Nuclear DNA contains mtDNA sequences within itself so it can also lead false positive result. Therefore, control of mtDNA contamination is another precaution that help to determine aDNA (Hofreiter et al., 2001). Moreover, PMD is an important obstacle for population genetics analyses since it increases number of mismatches in DNA so it eventually could affect results. On the other hand, the authenticity of DNA can be validated by checking the presence of C to T transitions. In fact, this lead scientists to distinguish ancient and modern DNA and filter out modern contamination before downstream bioinformatics and population genetics analysis on them (Briggs et al., 2007; Krause et al., 2010).

1.4. Studies of sheep ancient DNA and genomics

After recent advances in aDNA analysis techniques, scientists have been able to answer more questions about the Neolithic Transition to better understand human evolutionary history. Considering that animal and plant domestication was one of the most important steps that caused a dramatic shift in human lifestyle by allowing food production, the importance of elucidating the evolutionary history of livestock animals is receiving high attention.

It is known that sheep was the one of the first domesticated livestock animals in and the proposed domestication center of sheep is nearly fully located inside Anatolia. To date, however, there are relatively few published ancient DNA studies on sheep. Further, these generally involve investigating mitochondrial haplogroups of archaeological samples from different regions of the world.

In 2006 and 2011, mtDNA haplogroup analysis were conducted by Cai's group in China. Samples taken from excavation site dated for Bronze age. They indicated that in Bronze age HPG A was dominant mtDNA haplogroup among sheep in East Asia.

They also found one HPG B individual but they claimed that Chinese domestic sheep could be originated from lineage of HPG A (D. Cai et al., 2011; D. W. Cai, Han, Zhang, Zhou, & Zhu, 2006).

First discovery of HPG B lineage in Africa was made in South Africa Western Cape (Ann Horsburgh & Rhines, 2010). Sheep samples were dated 2000 years before present (BP) and they all belong to lineage B. Since it is claimed that lineage B could be originated from Near East (Stefan Hiendleder, Kaupe, Wassmuth, & Janke, 2002), Horsburg and Rhines claimed that those sheep most probably were carried from Near East to Africa.

Niemi et al. (2013) determined haplogroup of 26 sheep samples from Iron, Medieval and post-Medieval Ages as HPG A (4 samples) and HPG B (22 samples) in Finland. They also used Y-chromosome polymorphism to detect Y chromosome diversity in sheep breeds (Niemi et al., 2013).

Five haplogroup (A, B, C, D and E) were determined among modern domesticated and wild sheep breeds in previous studies (S Hiendleder, Mainz, Plante, & Lewalski, 1998). However Demirci et al.'s (2013) study revelead first detection of HPG C and E among ancient samples. They found HPG C and E among ancient sheep samples which excavated from Oylum Höyük (Southern-East Anatolia) that were dated to 1880-330 BCE. It is an important findings because it showed high diversity of mtDNA that we see in modern breeds could be detected in ancient samples as well.

Besides known mtDNA haplogroups, Dymova et al. (2017) claimed that they found undiscribed new haplogroups in their samples. They investigated mtDNA haplogroups of 17 ancient sheep samples taken from various excavation sites in Altai dated to the Early Bronze Age. They found that HPG A and HPG B were dominant lineages among ancient samples but unidentified haplogroups also presented. They suggested that Altai region acted as a bridge for migrations of sheep from Near East to Far East in that times.

In very recent study, Sabatini et al. (2019) conducted both ancient DNA and mtDNA analyses of 8 Bronze Age sheep samples from Italy and Hungary. They could identify oldest sheep sample (1500-1400 BCE) that belongs to HPG A in Europe so far. Because extensive wool production was started in Europe during Bronze Age (Vretemark, 2010), they claimed that this might be indication of new sheep migration that introduce to Europe for its secondary product wool.

Studying ancient DNA of sheep populations can be more challenging than in humans, since domestic animals are usually not buried, unless the ancient societies had cultural ritual for animal burials. Even worse, animals could be cooked so that their skeletal elements may have been carbonized already. In those circumstances their DNA may not be preserved at all. Although mtDNA cannot provide as much genetic information about sample as information from the whole genome, scientists verged to obtain genetic information from mtDNA, because of its high abundance compared to nuclear DNA. Since mtDNA is inherited from only the mother, studies that investigate mtDNA haplogroups can obtain information about history of maternal lineages of ancient animals. Importantly, in domestic animal breeding, certain female individuals that belong to the selected population are kept as a stock population until they cannot reproduce anymore. Consequently, constant breeding of stock populations can create continuous mtDNA heritage through generations. Therefore, lineages of domesticated animals can be traced by mtDNA haplogroups (Bruford, Bradley, & Luikart, 2003).

Whole-genome studies on sheep have yet been conducted only using modern domesticated and wild sheep samples (Alberto et al., 2018; Kijas et al., 2012). However, there is no ancient DNA yet to study genome-wide variation in ancient domestic sheep, nor any genetic investigation of domesticated sheep samples from the Neolithic Period from west Asia.

1.5. Objectives of the present study

Throughout history, sheep is the one of the most valuable livestock animals for human populations in terms of its beneficial first and secondary products. It is known that

Anatolia hosted the initial domestication processes of sheep and also acted as a bridge for the spread of domestic sheep into Europe along with human populations, during the Neolithization of Europe. Therefore, exploring whole genomes of domesticated sheep samples from Anatolia can provide substantial information about both the evolutionary history and migration events of domesticated sheep. Since their migrations were mediated by humans, investigating sheep will also contribute to understanding of human movements during the Neolithic Period.

In order to understand the history of Anatolian domestic sheep, whole-genome sequencing data of four Anatolian Neolithic sheep individuals were analyzed, and compared to eighteen modern domesticated sheep breeds from Europe, Asia and Africa. The main objectives of this study are:

1. To explore the genetic relations between Anatolian Neolithic sheep sample and modern domesticated sheep breeds,
2. To determine the genetic relations among Anatolia Neolithic sheep individuals,
3. To investigate the spread of domesticated sheep through possible migrations routes after initial domestication events,
4. To gain a better understanding into the evolutionary history of domesticated sheep.

CHAPTER 2

MATERIALS AND METHODS

2.1. Modern and ancient genome data

In the present study two different DNA sequence datasets were used: 1) A modern sheep dataset and 2) an ancient sheep dataset. Populations that belong to the modern dataset were chosen from the Kijas et al. (2012) study (see below)

2.1.1. Modern dataset

The modern dataset consisted of two major parts: Domestic breed populations and outgroup individuals. Whole genome sequences of modern domestic sheep breeds and wild sheep individuals were chosen from Kijas et al.'s (2012) study. In this study the authors present genotype information of 74 different sheep breeds from across the world. In addition to modern domestic breeds, genotype information of wild sheep (Argali sheep (*Ovis ammon*)) was used in this study as outgroup population, produced by several contributors of the International Sheep Genomic Consortium using the ovine SNP50 Beadchip array ("International Sheep Genomics Consortium," n.d.) (see Appendix A).

The provenances of the sheep samples were variable, and there was more than one breed for each region. Out of all domestic breeds 18 populations were selected for population genetics analysis (Table 2.1).

Table 2.1. Chosen modern domestic sheep breeds from the Kijas et al. (2012) study. "N" stands for sample size of each population

Breed Name	**Abbreviation**	***N***	**Origin**
Afshari	AFS	37	Middle East
Australian Merino	MER	50	South West Europe
Bundner Oberlander Sheep	BOS	24	Central Europe
Changthangi	CHA	29	Asia
Chios	CHI	23	East Aegean
Churra	CHU	120	South West Europe
Comisana	COM	24	South West Europe
Cyprus Fat Tail	CFT	30	South West Asia
Deccani	IDC	24	Asia
Engadine Red Sheep	ERS	24	Central Europe
Ethiopian Menz	EMZ	34	Africa
Finn sheep	FIN	99	Northern Europe
Leccese	LEC	24	South West Europe
Norduz	NDZ	20	Middle East
Old Norwegian Spaelsau	NSP	15	Northern Europe
Sakiz	SKZ	22	East Aegean
Sardinian Ancestral Black	SAB	20	South West Europe
Valais Blacknose sheep	VBS	24	Central Europe

There were multiple reasons why these 18 were selected for comparative analyses in this study. Firstly, as can be seen the map in Figure 3.1 describing the locations of from modern and ancient sheep samples, we chose modern populations close to the different routes of possible migrations of the first domesticated livestock animals described in previous sections (section 1.2). Secondly, for each geographic region we chose breeds that were less variable, or less admixed, as described in Kijas et al.'s

(2012) study. We preferred to have less admixed modern breeds for simplicity in interpreting the demographic history.

2.1.2. Ancient dataset

In this study we used ancient DNA information of four Anatolian Neolithic sheep individuals, with IDs TEP03, TEP62, TEP83, and ULU31. The data was produced as part of a previous sheep ancient genomics project (TÜBİTAK 114Z356). In that project, ancient DNA laboratory work was performed on 34 individuals' samples from different excavation sites within Anatolia, shown in Table 2.2.

Table 2.2. Information about whole genome sequenced samples

Excavation Site	**Number of Sample**	**Approximate C-14 ages of samples (BCE)**
Tepecik Çiftlik Höyük	11	7500-6650
Ulucak Höyük	13	7000-6000
Barçın	4	6300-6100
Yeşilova	6	6250-5800

2.1.2.1. Summary of laboratory work for ancient samples

Here I provide brief information about the background of the ancient sheep genome data, which is yet unpublished. Sheep bone and tooth samples studied in this project were mainly acquired by Dr. Füsun Özer, Nihan Dilşad Dağtaş and Dr. Eren Yüncü

from excavation sites, and provided by archaeologists or archaeozoologists from different excavations.

Laboratory procedures of samples were carried out in the METU Ancient DNA Laboratory clean room (dedicated for aDNA work); UV light treatment before experiments and constant bleach and DNA AWAY™ (surface decontaminant Thermo Scientific™) was applied during experiments, and researchers used specialized lab clothes to prevent any human DNA contamination. The DNA extraction method of Dabney et al. (2013) was used with number of modifications. The main steps could be summarized by following:

- Sample surface removing
- Cutting a small piece from sample with a drill and grinding
- Overnight incubation of sample powder with treatment of extraction buffer (0.45 M EDTA, 0.25 mg/ml Proteinase K, pH: 8.0)
- Separation of extracted DNA with multiple steps by centrifugation and applying the elution buffer (Qiagen PE buffer)
- Double stranded DNA library preparation of each sample for whole genome sequencing on the Illumina Hiseq platform using blunt-end ligation as described in Meyer & Kircher (2010)
- Purification of amplified libraries using AMPure XP beads (Agencourt)
- Profiling of the libraries on a 2100 Bioanalyzer using High Sensitivity Kit (Agilent Technologies) for quantification and quality control, and then pooling at equimolar concentrations for initial sequencing (screening).

After initial sequencing, obtained reads were mapped in sheep referance genome by following methodology that was described in (section 2.1.2.2). Then proportion of mapped reads was calculated for each sample as sheep DNA proportion each sample. TEP62 sample had highest sheep DNA proportion was 0.11 % percentage of all reads and others were lower than 0.03 % percentage all reads so sheep DNA proportions of samples were very low. Among all libraries four individuals that had relatively higher

DNA proportions and low clonality levels were chosen for further analysis. Clonality of samples means that proportion of reads which have duplicate so low clonality and a high endogenous proportion together indicate the possibility of obtaining sufficient genetic information from capture sequencing. Because of budget limitations, instead of shotgun deep sequencing, the laboratory team chose to use the in-solution hybridization capture targeting single nucleotide polymorphisms (SNPs) approach to increase coverage per individual. For this, from the OvineSNP50 Bead Chip (Illumina, San Diego CA, USA) 20,000 SNPs were chosen. An approach very similar to what is described in Haak et al. (2015) was used for designing capture probes. Three main criteria were taken into consideration while choosing target SNPs:

1. Include SNPs related with phenotypic traits such as high production of milk (Cecchinato, Bittante, & Vacca, 2016; Kijas et al., 2012), wool (Kijas et al., 2012) and deposition of fat (Moradi, Nejati-javaremi, Moradi-shahrbabak, & Dodds, 2012), possibly subject to positive selection in sheep history.
2. Include SNPs linked to X and Y chromosomes (Heaton et al., 2014).
3. Include SNPs which are mainly far from sides that are vulnerable to post mortem damage.

Eventually DNA information for each of the four individuals was acquired from both whole genome shotgun sequencing (screening data) and the hybridization capture sequencing processes.

2.1.2.2. FASTQ preprocessing and alignment

For population genetics analyses, the initial step is to extract genotype information from FASTQ files, which contain sequence information. For this, the FASTQ files need to be preprocessed in different ways, and aligned to the genome.

First, the residual adapter sequences in FASTQ files were removed, and paired-end sequencing reads were merged into a single sequence. These two steps were performed using a published script for processing ancient genome data "MergeReadsFastQ_cc.py" (Kircher, 2012).

The merged reads then were mapped to the reference sheep genome (Oar_v3.1) using "aln" program of BWA (v. 0.7.12) software (Li & Durbin, 2009). The parameters of mapping, -n, -o and -1 indicate maximum edit distance, maximum number of gap openings, and the seed length, respectively (Li & Durbin, 2009). Mapping was performed with -n 0.01, -o 2 and -1 16500 (no seed) parameters, resulting in BAM files that contain information about base quality, mapping quality and mapping location of read sequences to the reference genome. Libraries of each samples were merged and PCR dublicates were excluded by using FilterUniqueSAMCons.py (Kirscher et al., 2012). We thus created eight BAM files, one from the shotgun screening and one from capture FASTQ files. These two different BAM files for each individual were then merged using Samtools software (Li et al., 2009). Here the "-merge", "-view", and "-index" commands were used for the merging processes of the two files, filtering out dublicates, and indexing, respectively. Importantly, after merging we filter out PCR duplicates, since otherwise they might be creating false high numbers of reads and leading to potential bias on SNPs calling process (Kircher, 2012).

Postmortem damage (PMD) is an important issue in ancient DNA studies, which was mentioned in the Introduction section. Mostly post-mortem damage appear close regions of read ends. There is no exact base number but to avoid PMD-induced mismatches, ten base pairs from each end of each DNA read was trimmed to prevent unwanted postmortem damaged sites. The trimming of BAM files was performed by assigning the letter "N" to the first ten base pair of two end of DNA, so that those parts would be excluded from the downstream genotyping step. Here, the merged BAM file of each individual was trimmed ten base pairs from the two ends using the bamUtil software (Jun, Wing, Abecasis, & Kang, 2015). After those processes ancient dataset that belong to four individual were ready for SNP calling (Table 2.3).

Table 2.3. Information about four Anatolian Neolithic sheep individuals after preprocesses for population genetic analysis

Sample name	Shotgun seq. genome coverage	Shotgun +Capture merged genome coverage	Shotgun number of SNP	Shotgun+ Capture merged number of SNP	Molecular sex	C14 age (BC)
TEP03	0.006191	0.103	147	8142	F	7059-6756
TEP62	0.056317	0.273	1262	10406	F	7731-6687
TEP83	0.004179	0.022	97	3113	M	6469-6361
ULU31	0.006048	0.022	149	4319	F	6227-6071

2.2. SNP calling

SNP calling process of all ancient individuals was carried out separately. The SNP reference dataset file that was generated for genotyping process of modern dataset before and sheep reference genome (Oar_v3.1) was used for SNP calling . Except for both ends of DNA, transition and transversion sites were included to maximize caught SNP number. Additionally random haplodizing process was applied all individuals by choosing random allele from heterozygous sites. Since ancient data has low coverage and mostly only one allele from one locus is represented, to avoid technical bias that can be arose from coverage differences, one allele is randomly chosen (Barlow, Hartmann, Gonzalez, Hofreiter, & Johanna, 2018; Günther & Nettelblad, 2019). In

order to SNP calling was conducted by “mpileup” program of Samtools software (v. 1.3) (Li et al., 2009) with “-q 30” and “-Q 30” parameters of minimum mapping quality and base quality scores, respectively. Eventually by using PLINK software (Purcell et al., 2007) PED and MAP files of each individuals were ready for next step.

2.3. Principal Component Analysis (PCA)

Genomic data in general carries multiple dimensional and information which may be hard to interpret without summarizing. Principle component analysis (PCA) is a mathematical approach that can reduce high dimensional data into two or three dimensional spaces described by eigen vectors (Patterson, Price, & Reich, 2006). Those vectors reflect major components of variation within the dataset. Therefore, this method can reveal both common and unique variances which populations have. Since we can visualize the result of this analysis, it allows easily assessing the quality of the whole data and see differentiation among samples included in the dataset. PCA has been used in population genomic studies since similarity of genetic background makes populations cluster together (Patterson et al., 2006); it is also commonly used in ancient genomic studies.

In this study two different PCA graphs were generated. For each analysis modern and ancient dataset were merged using the “-mergelist” command of PLINK (v. 1.9) software (Purcell et al., 2007), so that eventually the final set had eighteen modern populations and four ancient samples. Eigenvalues of each PCA were calculated using the “smartpca” program of EIGENSOFT (v. 7.2) software (Patterson et al., 2006), and using all eighteen modern domesticated breeds' genotypes. The four ancient samples' genotypes were then projected onto the space described by these eigenvalues using the “lsqproject:YES” parameter of the program. All graphs were drawn by the “scatterplot3d” package (Ligges & Maechler, 2003) and “plot” function of Rstudio (v. 3.5) software (RStudio Team, 2015).

2.4. D-statistics

D-statistics is a very powerful detection method of possible gene flow between two populations, performed using genome-wide polymorphism data. In previous studies this method was used to investigate ancestral past of modern human and find out admixture events of populations (Patterson et al., 2012). It is demonstrated as *D*(C,D;E,F). *D-statistics* can be explained by simple one common rooted tree figure. In this figure C is the outgroup and other letters D,E and F represent three different populations (Figure 2.1).

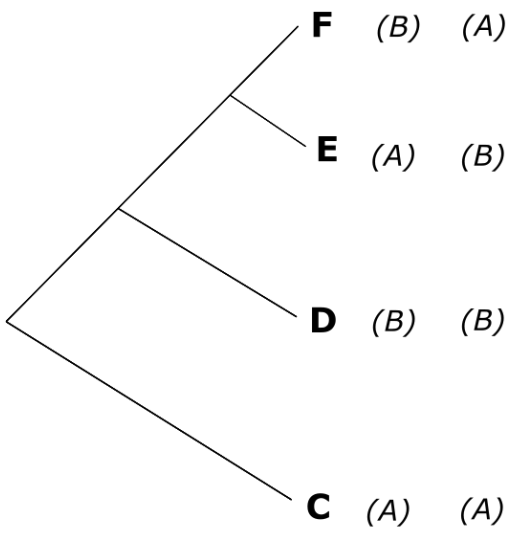

Figure 2.1. Example one rooted tree of populations C (outgroup), D, E, and F.

Patterson et al. (2012) explained that, using biallelic genotype data from 4 populations as shown in Figure 2.1, the *D-statistic* can be calculated by this formula:

$$D=\Sigma(P(\text{BABA-ABBA})/P(\text{BABA+ABBA})),$$

where BABA represents the situation that when allele that is randomly chosen from four populations C and E carry one type, and D and F carry the other type. The ABBA

situation happens when an allele is randomly chosen from the four populations and C and F carry one type, and D and E carry the other. If the phylogeny is true, then D will be on average zero. If there has been gene flow between populations D and E, the result of *D*(C,D;E,F) will be less than zero, because they have more shared alleles compared to population F. However, if gene flow has happened between D and F, the result is going to be higher than zero this time because they have more shared alleles compared to population E.

In this study, one Argali sheep (*Ovis ammon*) individual was designated as outgroup. Argali sheep is an unadmixed subspecies (S Hiendleder et al., 1998) belonging to the Ovis genus and distantly related to modern domesticated sheep populations (*Ovis aries*), which may therefore be used in *D-statistics* tests. In order to the confirm the suitability of Argali sheep as outgroup, the genetic distance of the Argali sample to other sheep was checked by conducting multiple D-tests. In these tests a more distant group to all Ovis species, the Angora Goat (*Capra aegagrus hircus*), was chosen as outgroup, and Argali sheep as the test population. Other populations were from modern domestic sheep populations. P-values were calculated from Z scores and multiple testing correction was done using the "p.adjust" function of R (v.3.5) with the Benjamini-Hochberg method (Benjamini & Hochberg, 1995), in order to control for the false discovery rate.

We further compared multiple combinations of modern and ancient populations (using Argali sheep as outgroup) to address questions about the relationships among populations and thus infer demographic history. In total, 2390 D-statistics were conducted using the qpDstat program of AdmixTools (v. 5.1) software (Patterson et al., 2012). Modern, ancient and outgroup datasets were merged using the "mergelist" command and extra chromosomes of sheep samples that could not recognize by PLINK program were excluded by the "exclude" command of PLINK. PED, PEDIND and MAP file formats were converted into EIGENSOFT file formats (Patterson et al., 2012), namely GENO, IND and SNP file formats, before D-statistics tests were performed.

2.5. Outgroup f3-statistics

This method is used to measure shared genetic drift along the whole genome among three populations. The statistics is shown as *f3*(A; B,C) (Peter, 2016). A is the outgroup population that is assumed to be highly diverged from both population B and C. B is the unknown test population of interest, and C is another panel population. In order to find out the most related population to population B, this method explores shared derived alleles of the test population among panel populations. For the comparative analysis widely diverged outgroup is used (Peter, 2016).

In this study for calculating *f3-statistics* modern and ancient dataset were merged and converted into EIGENSOFT file format using the same procedure applied in *D-statistics*. One randomly chosen Argali sheep individual, which was already in the modern dataset, was used as outgroup. Suitability of this individual as an outgroup was tested as mentioned above. The test populations were Anatolian Neolithic sheep individuals, and they were compared with modern sheep breeds populations one by one, so that we performed the following test *f3*(Argali; Anatolian Neolithic, modern).

CHAPTER 3

RESULTS

3.1. Modern and ancient sheep datasets

All the genotype and individual information of modern domesticated sheep and outgroup populations were retrieved from Kijas et al.'s (2012) study, which was conducted by multiple working groups of the International Sheep Genomics Consortium. There were 71 modern domesticated sheep populations which have been genotyped in that dataset, but a subset of 18 populations (Table 2.1) and one randomly choosen Argali sheep were chosen for further genetic analysis. The final modern dataset included 644 individuals and 49,034 SNPs in total.

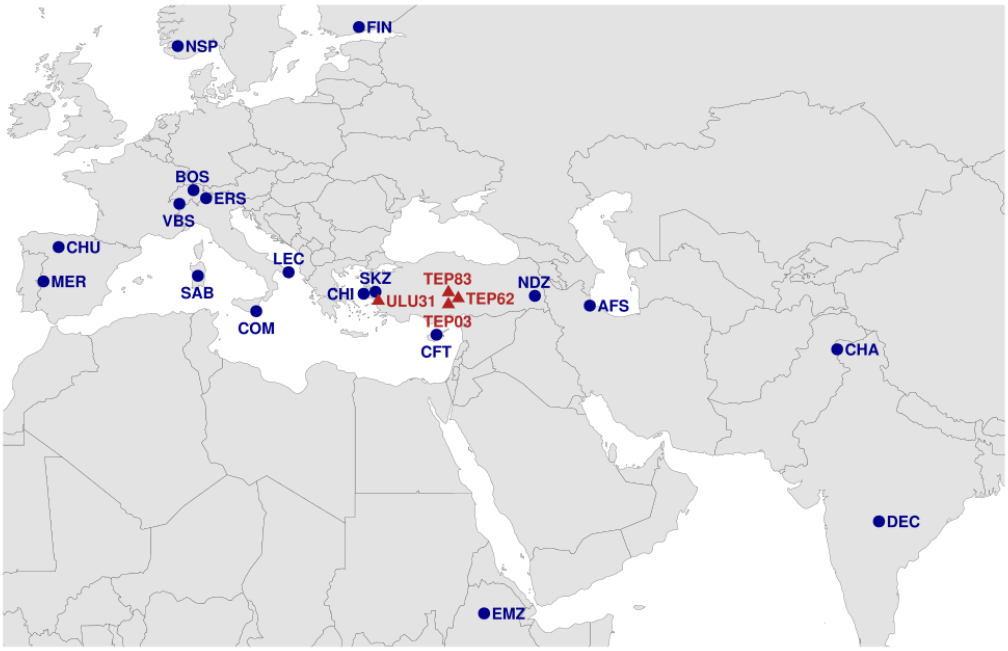

Figure 3.1. The locations of chosen modern domestic breeds (blue dots) and Anatolian Neolithic sheep samples (red triangles) used in this study

The Anatolian Neolithic sheep (ANS) data was produced by a combination of shotgun sequencing and SNP capture approaches (section 2.1.2.1). After screening material from 29 individuals, we could obtain useable data for 4 individuals, three from the Pottery Neolithic site of Tepecik-Ciftlik Hoyuk in Central Anatolia (ages 7059-6361 (BC)), and one from Pottery Neolithic site of Ulucak Hoyuk in West Anatolia (age 6227-6071 (BC)). In order to genotype these individuals we used the same SNP panel containing 40,225 SNPs as the one used for genotyping the modern populations (Kijas et al., 2012). After merging shotgun and capture sequenced libraries, we obtained 10,406-3113 SNPs for these four Anatolian Neolithic sheep individuals (see Table 2.3.). Hence the ancient individuals contained large amounts (90-75%) of missing data compared to modern domesticated sheep individuals.

3.2. Principal Component Analysis (PCA) suggests closer affinity between ANS and modern European breeds

We first used principal component analysis (PCA), a highly used method in ancient DNA-based demography studies, to summarise the genetic affinities of the ancient individuals with modern populations (Patterson et al., 2006). Eigenvectors were calculated using modern samples only. The first three eigenvectors explained 3.6%, 3.1% and 1.7% of variation in the modern dataset. Hence in two dimensional PCA explained 6.7% and the three dimensional PCA 8.4% of total variation. We then visualized PCA results in two ways. Firstly, two principal components of modern sheep populations were drawn on a two dimensional surface and ANS individuals were projected onto these two dimensions (Figure 3.2). Secondly, in order to increase information about genomic variation, we used the first 3 eigenvectors to draw a three dimensional plot and projected the ANS individuals again on this space (Figure 3.3).

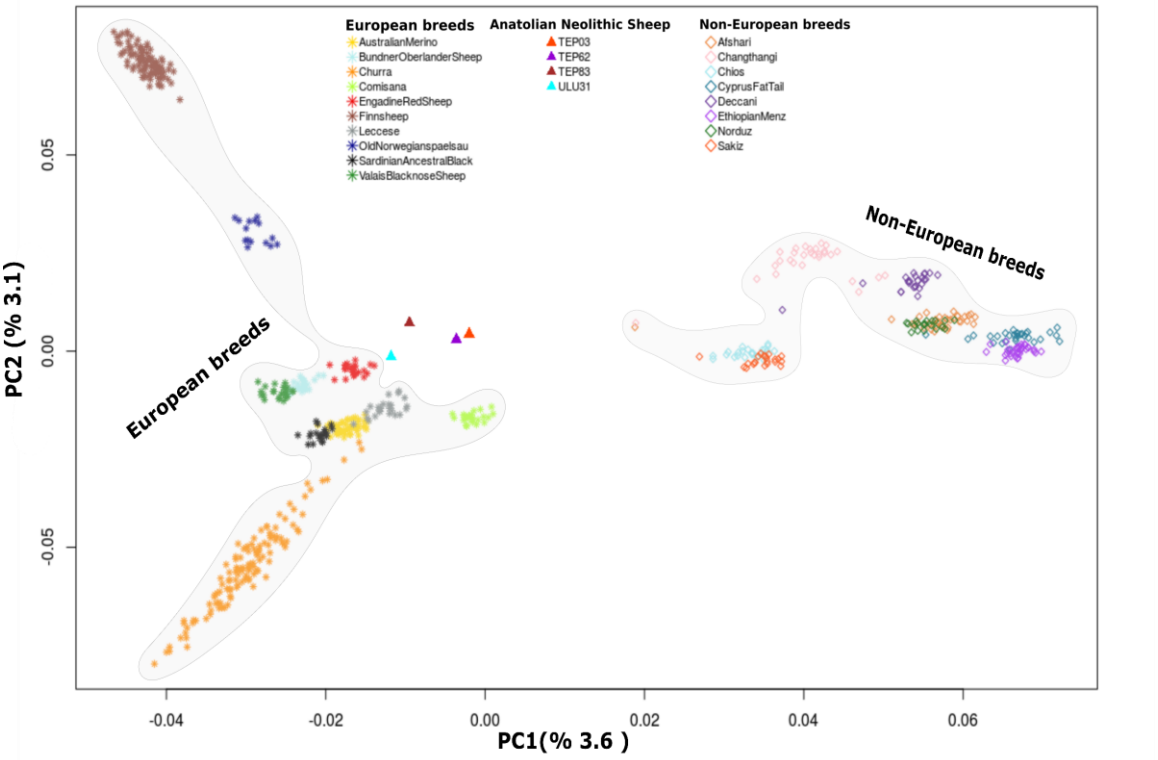

Figure 3.2. The two-dimensional PCA graph of the chosen eighteen modern domesticated breed populations with four ANS individuals projected onto the PC space. In PCA graphs each point represented one individual of one population. Clusters that include same colored points represented one population. Each of them were labeled categorized according to populations. Anatolian Neolithic sheep individuals were represented with bigger rectangles to make them easy to distinguish.

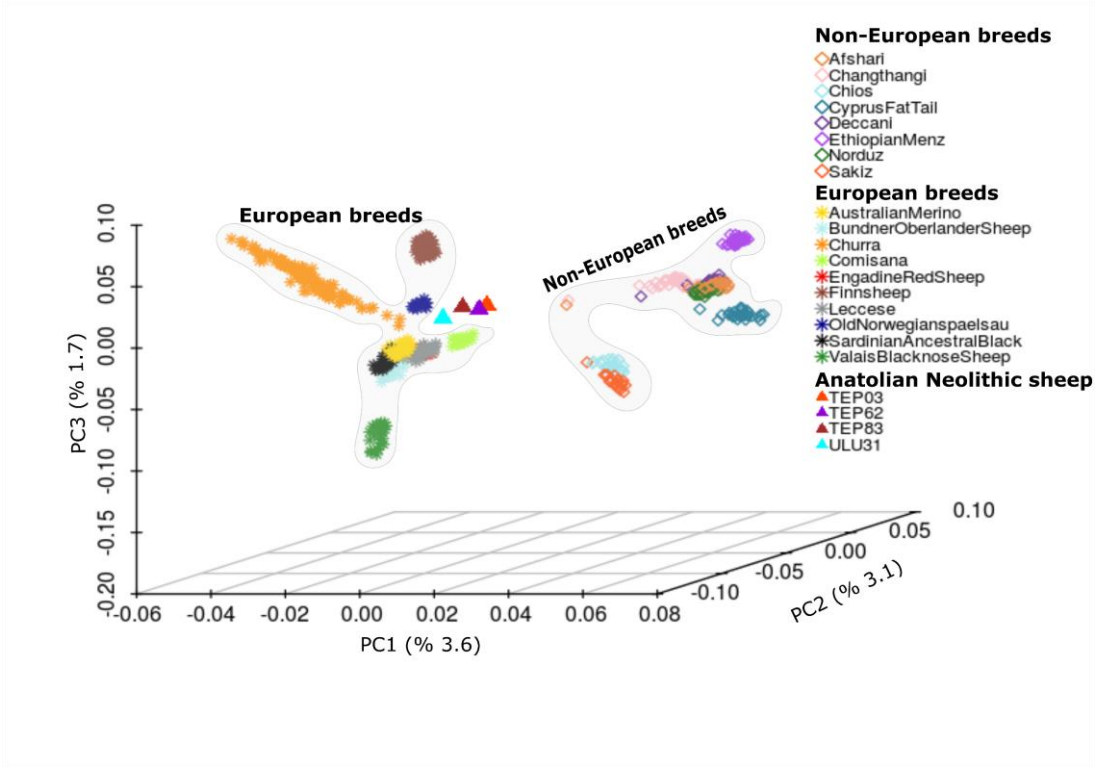

Figure 3.3. The three-dimensional PCA graph of the chosen eighteen modern breed populations with four ANS individuals which were projected onto the PC space. In PCA graphs each point represented one individual of one population. Clusters that include same colored points represented one population. Each of them were labeled categorized according to populations. Anatolian Neolithic sheep individuals were represented with bigger rectangles to make them easy to distinguish.

As seen in the PCA graphs European and non-European breeds were clustered separately, and formed two major poles. This difference might be explained by model of isolation by distance (Wright, 1943). European and non-European breeds geographically located seperately (see Figure 3.1). Therefore, individual of European populations cannot mate with non-European populations so gene pool of two seperated region cannot be mixed. Eventually, their genetic background become different by evolutionary forces such as genetic drift and founder effect. On the other hand, populations that can be mixed would share more genetic background. Since clusters represent genetically similar individuals this scenario could be expected. Additionally, it can be seen that there are breed clusters among both European and

non-European breeds. We observe three main groups among of European breeds, which fit Kijas et al.'s (2012) categorization of modern samples according to their geographic regions: Middle European breeds (BOS, ERS and VBS) are clustered together and are located intermediate between North European breeds (FIN and NSP) and South European breeds (LEC, SAB, MER, CHU and COM) in the PC space. A similar pattern can be observed among the group of non-European breeds in the PCAs. Populations from geographically nearby locations appear clustered together similar to what is seen on the European side of PCA. Notably, East Aegean breeds (CHI and SKZ) and Middle Eastern breeds (AFS and NDZ) overlapped on the PCA graph.

All four ANS samples clustered close to each other in PCA space. Notably, one could observe slight trends, such as TEP03 and TEP62 individuals being genetically even closer to each other compared to TEP83 and ULU31, and the ULU31 individual being closer to both Central and Southern European breeds compared to Tepecik-Çiftlik individuals. Still, overall the four individuals did not appear genetically distinct. Moreover, ANS individuals formed a cluster clearly close to European breeds than to non-European breeds.

However, because PCA does not represent the uncertainty in the results, a formal statistical test is needed to confirm these patterns.

3.3. Outgroup f3-statistics suggest higher affinity of ANS to modern European breeds

The f3 method (Peter, 2016) is a method to measure genetic similarity between two populations in terms of shared derived allele frequency, relative to an outgroup (Peter, 2016). The f3 value can vary between 0-1 and a higher value indicates that the target population is closer to the second population (Peter, 2016). Here, to further investigate the genetic similarity patterns observed in the PCA in terms of the relationship between Anatolian Neolithic sheep individuals and modern domestic breeds, we used the *f3-statistic* of the form *f3*(Outgroup; ANSX, MDSX), where the outgroup was Argali sheep, ANSX one Anatolian Neolithic sheep individual, and MDSX one

modern breed population. All *f3-statistics* tests were significant. The results are displayed in (Figure 3.4).

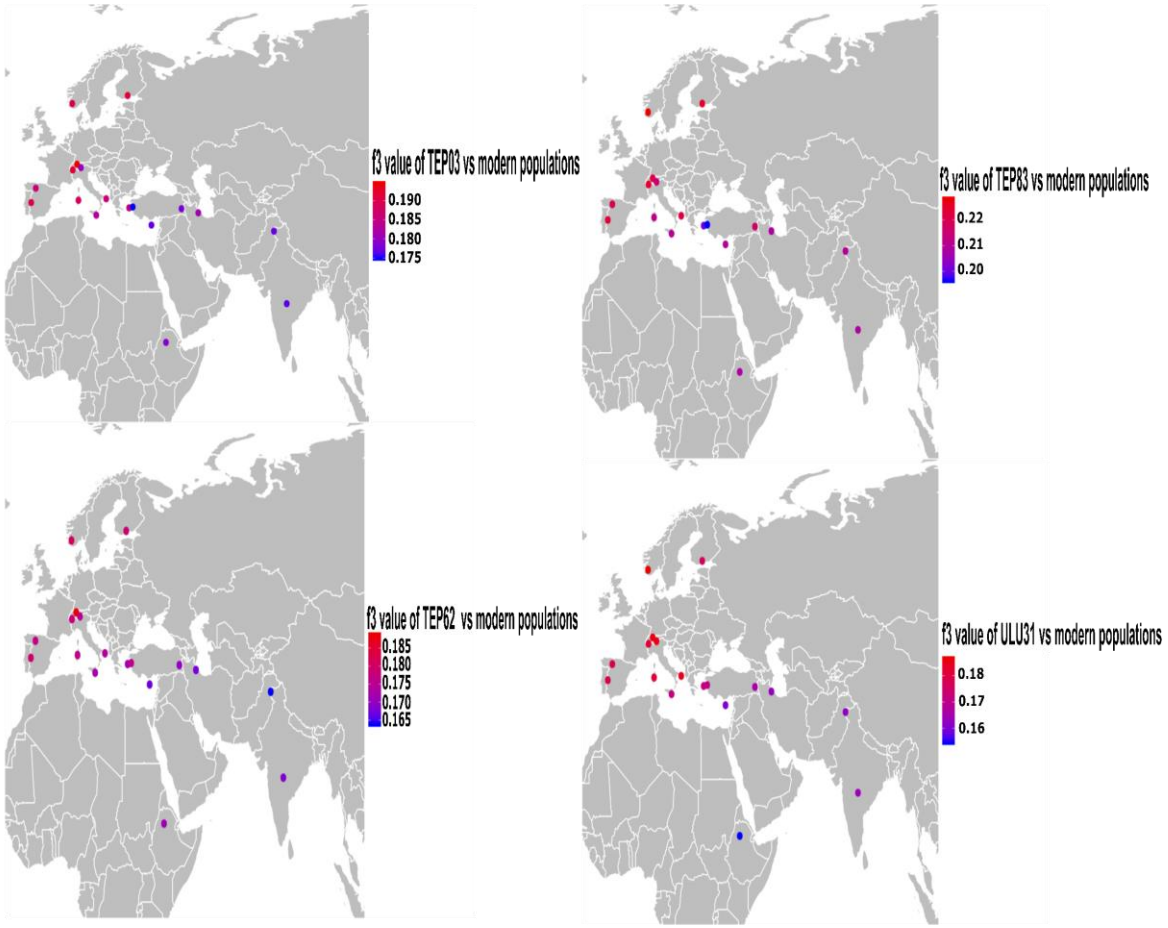

Figure 3.4. The map of f3-statistics describing affinity between ANS and modern populations. Every points represent a modern domesticated breed population and colors from blue to red indicate increasing f3 values, i.e. increasing genetic affinity between an ANS individual and the target modern population.

As it is seen in f3 maps both TEP03, TEP62 and ULU31 individuals showed higher similarity to European breeds (red color), and lower affinity to non-European breeds (blue color). The trend was slightly weaker for TEP083. Also, ANS affinities appeared higher for Central and Northern European breeds than South European breeds.

3.4. ANS individuals have significant affinity to each other comparing to modern breeds

D-statistics (see section 2.4) is a powerful and relatively simple method to measure genetic affinities and to infer gene flow and population admixture. In order to confirm some of the possible genetic similarities seen in the PCA graphs, D-statistics were calculated between different combinations of populations to study relationships between modern and ancient populations, and tested for significance using the jackknife procedure. We performed in total 2390 D-tests; so that to control for false positive results, we calculated statistical significance using Benjamini-Hochberg (BH) multiple testing correction, as mentioned in section 2.4. We obtained both significant and non-significant results, and in this section the results will be explained according to their significance. Accordingly, I only represented combinations that had significant results in the plots. The non-significant results are not trustworthy to make conclusions about test results, although they might give some clues about trends.

Here 216 individual and population combinations were tested, in the form of *D*(Outgroup,ANSX; ANSY,MDSX), where ANSX and ANSY are a pair of ANS individuals, and MDSX is a modern domestic sheep population. All comparisons had significant results at BH corrected jackknife $p < 0.05$ (Figure 3.5.). ANS individuals chose ANS individuals against all 18 modern domesticated sheep breeds for every comparison significantly. Therefore, it may be said that ANS individuals genetically closer to each other comparing the modern breeds. Additionally, we might see a technical effect on these result. Since all ANS individuals dataset were prepared by same protocols and different than modern sheep dataset, it might lead technical artefact. This might contribute affinity that was observed from tests.

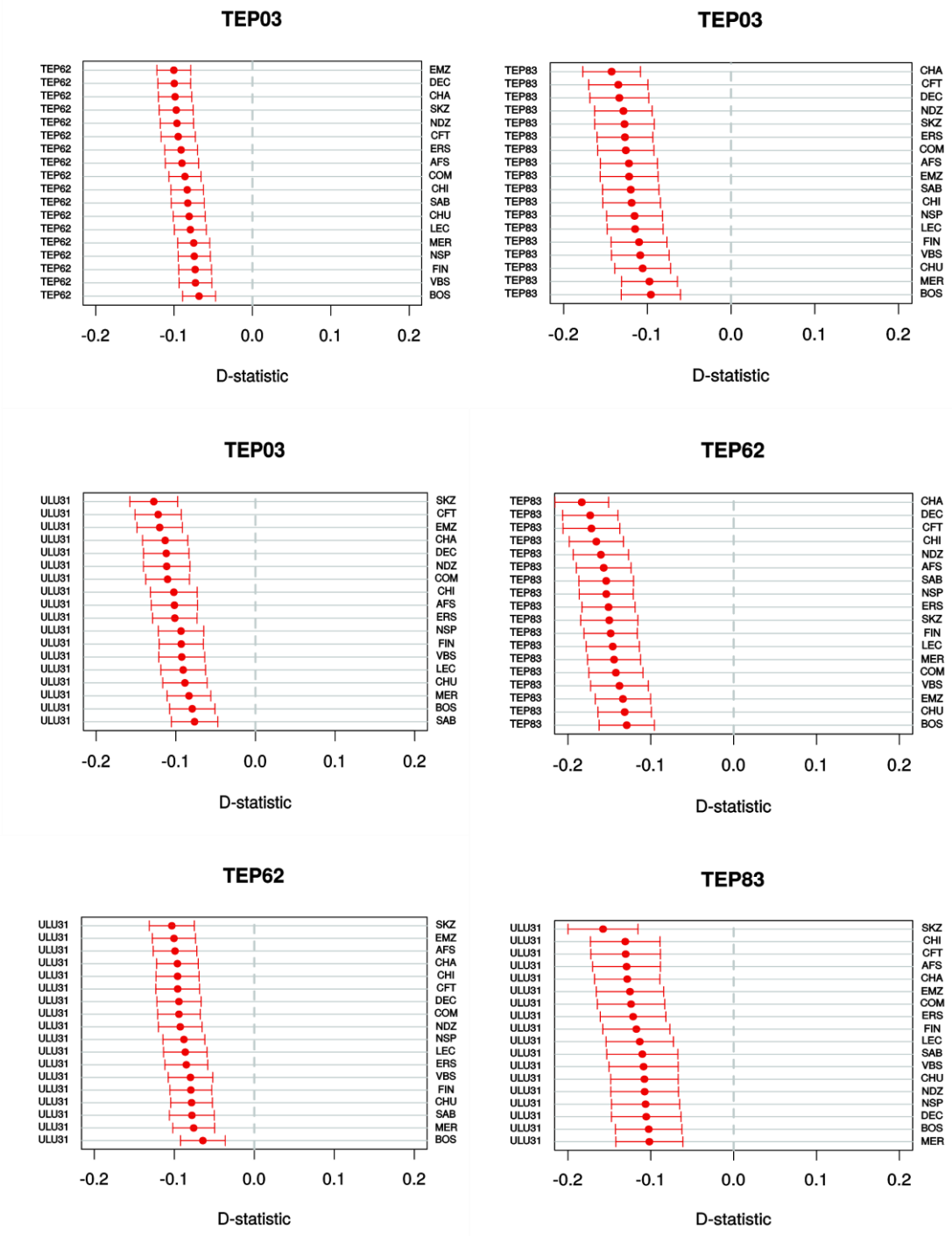

Figure 3.5. The significant results of tests D(Outgroup,ANSX; ANSY,MDSX), where ANSX and ANSY are a pair of ANS individuals, and MDSX is a modern domestic sheep population. Each row in a figure represents one test, the red points are the D-statistic values, while the red horizontal lines around each point show standard error. All tests were significant (BH corrected jackknife $p<0.05$).

3.4.1. No clustering among ANS individuals

In order to test if any ANS individual pairs were significantly closer to each other than other ANS individuals, we performed 24 D-tests of the form *D*(Outgroup,ANSX; ANSY,ANSZ), where ANSX, ANSY and ANSZ are ANS individuals. There were no significant results among these combinations. It can thus be said that none of the ANS individuals appears closer to another ANS individual in this sample . We may note that the TEP83 individual tended to be the most chosen, and the ULU31 individual was the least chosen among all non-significant comparisons, but we caution that these trends may represent technical biases - e.g. TEP83 is the lowest covered individual and may be most subject to reference allele bias (Günther & Nettelblad, 2019) .

3.4.2. ANS individuals show higher affinity to modern European breeds compared to non-European breeds

Here we tested which modern populations ANS are most strongly associated with, using D-tests of the form *D*(Outgroup,ANSX; MDSX,MDSY), where ANSX is an ANS individual, MDSX and MDSY are a pair of modern domestic sheep population, with 1224 such combinations. There are both significant and non-significant results among these comparisons. It is easy to interpret the results if combinations would be split into two major line up.

3.4.2.1. ANS individuals make preference between modern European and non-European breeds

Here we tested whether ANS individual have higher affinity to European breeds compare to non-European breeds. We performed 320 such tests of the form *D*(Out, ANSX; EURX, non-EURX), where EURX and non-EURX are pair of modern European and non-European populations, and ANSX is one of ANS individuals. Overall, except for the TEP83 individual all ANS samples seemed to prefer European breeds to non-European breeds, from 320 comparisons 7.5 % of cases significantly .

Specifically, the TEP03 individual was closer to all European breeds (except for CHI) against non-European populations, though always non-significantly. In one combination, the Central European BOS population was chosen significantly by TEP03 compared to CHA population (Figure 3.6.).

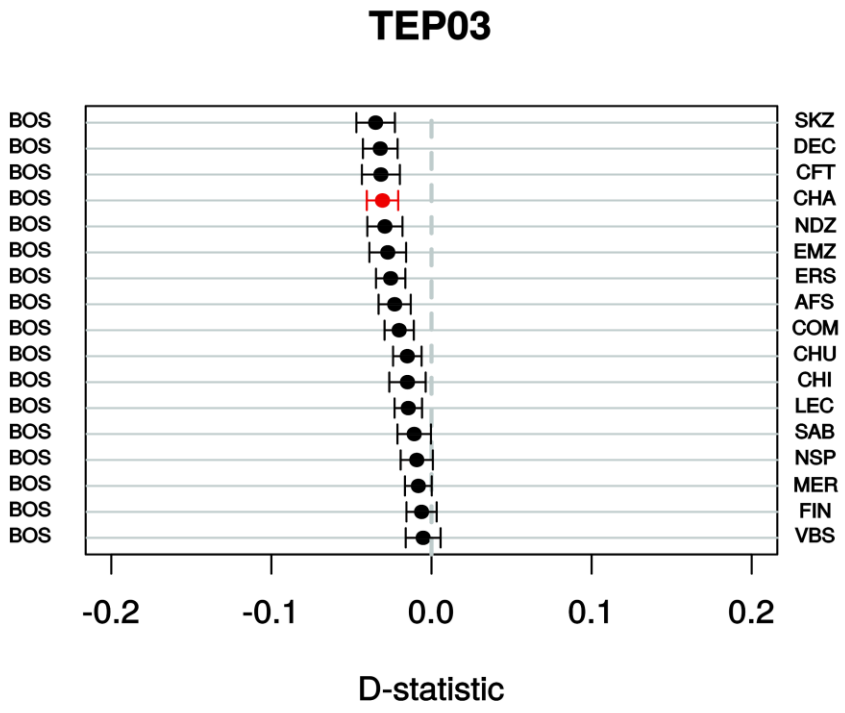

Figure 3.6. Graph shows result of choices TEP03 individual between BOS breed and other modern breeds. Red colour indicates significant result.

The TEP62 individual tended to choose European breeds always against non-European populations, significantly in 11.25 % of comparisons. It can be noted that significant choices of TEP62 individual were increased for both BOS, NSP, FIN and MER populations, which are Central, North and South European populations, respectively (Figure 3.7.)

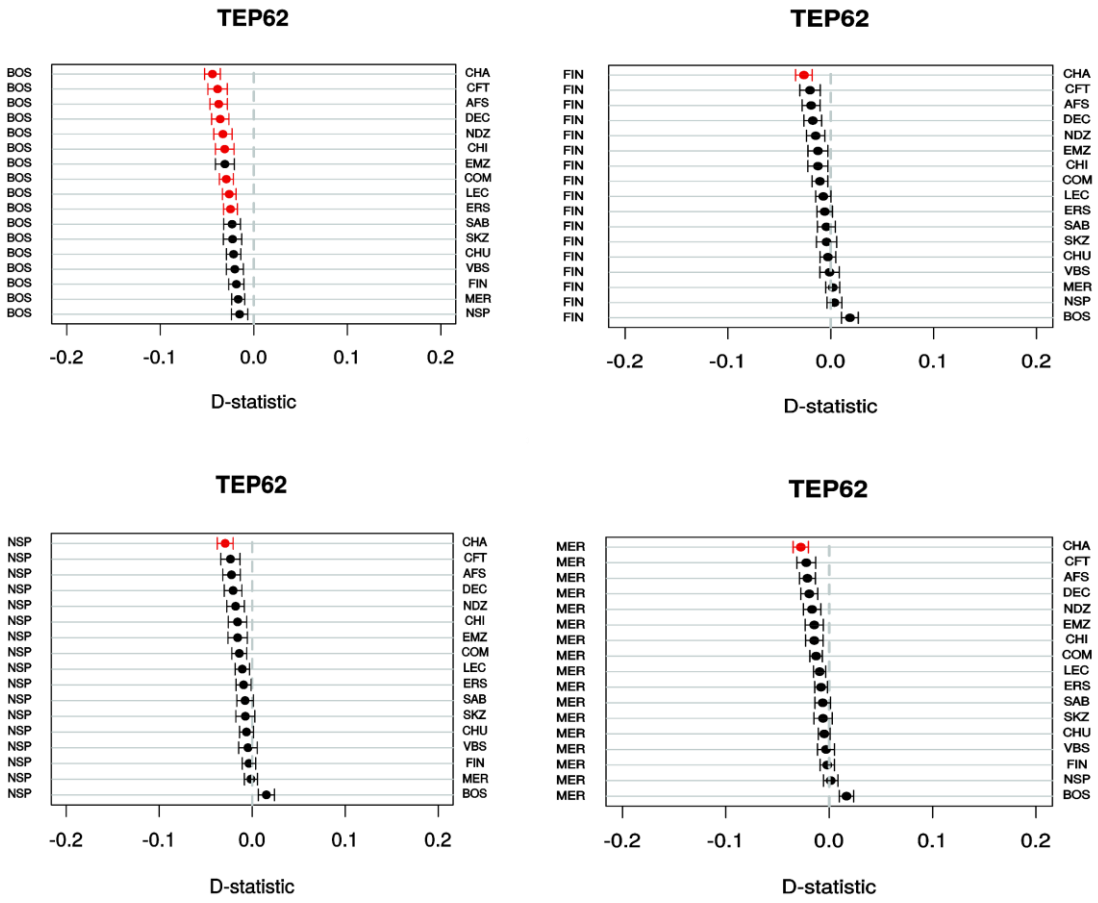

Figure 3.7. Graph shows result of choices TEP62 individual between BOS, NSP, FIN and MER breeds and other modern breeds. Red colour indicates significant result.

The TEP83 individual nearly always tended to prefer European breeds (except for NDZ breed) against non-European populations, although none of the comparisons involved TEP83 were significant.

The ULU31 individual from Ulucak Höyük showed slightly different choices than Tepecik-Çiftlik individuals against non-European breeds. Among these combinations ULU31 seemed to have significant choices of LEC and CHU breed which is from South European regions, compared to (AFS, CHA, CFT, EMZ) (which was not

observed for the Tepecik-Ciftlik individuals). This individual also prefers two of the Central European breeds (ERS and BOS) and one of the Northern European breeds (NSP) significantly in 37.5 % of (ERS, BOS and NSP vs. non-European) comparisons, and it also always had higher affinity to European breeds (except for CHI) against non-European breeds, although non-significantly (Figure 3.8.).

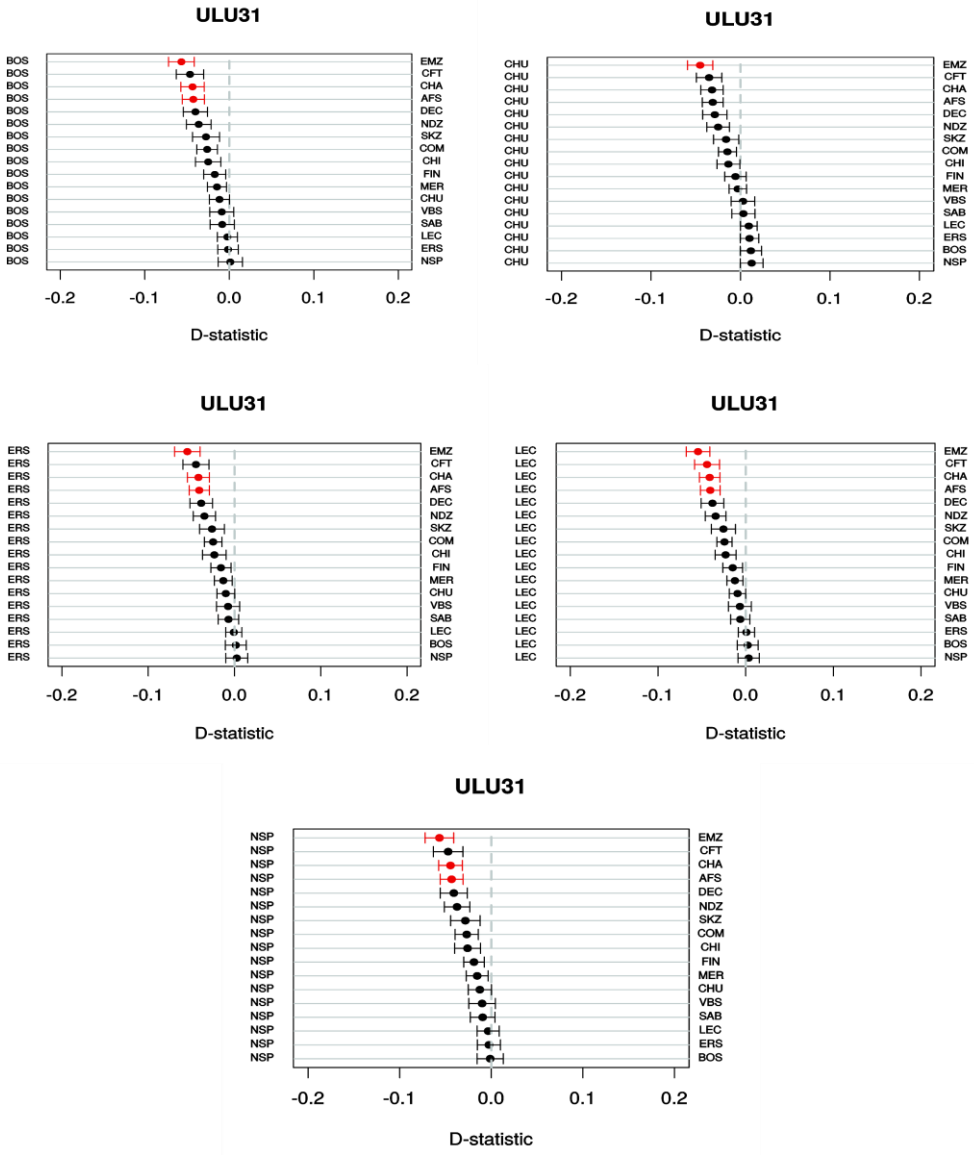

Figure 3.8. Graph shows significant result of ULU31 affinity to BOS, CHU, ERS, LEC and NSP breeds vs. other modern breeds. Red lines indicate significant results.

3.4.2.2. ANS individuals make preference between modern European breeds

Among 360 comparisons of the form *D*(Outgroup, ANSX; EUR1, EUR2), where EUR1 and EUR2 are pair of modern European populations, and ANSX is one of ANS individuals. ANS individuals tended to choose Central and North European breeds relative to South European breeds, although in most comparisons (98%) non-significantly. Among all ANS individuals, the strongest trend was observed for TEP62, which preferred one of the Central European breeds (BOS) to other European populations, significantly in 33% of comparisons, respectively (Figure 3.9).

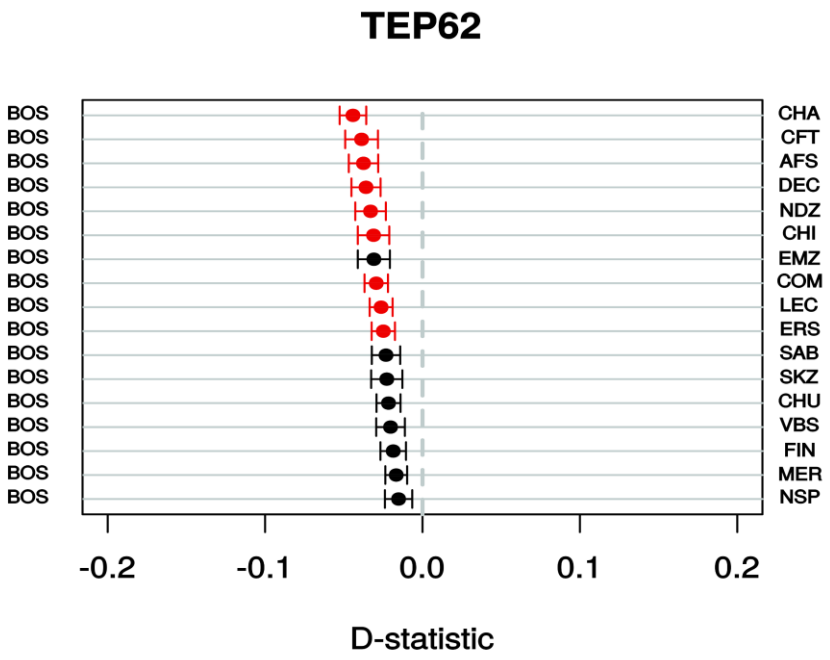

Figure 3.9. Graph shows result of choices TEP62 individual between the BOS breed and other modern breeds. Red colour indicates significant results ($p<0.05$).

3.4.3. ANS individuals make preference between modern non-European breeds

Among 224 comparisons of the form *D*(Outgroup, ANSX; non-EUR1, non-EUR2), where non-EUR1 and non-EUR2 are pair of modern non-European populations, and ANSX is one of ANS individuals. ANS individuals tended to choose SKZ, CHI, AFS, NDZ and CFT breeds relative to CHA, EMZ and DEC breeds, although all comparisons non-significantly. Among all ANS individuals, there was no strong trend observed.

In summary, ANS individuals have higher affinity to each other compare to modern popoulations but no clustering could be detected among them. Additionally, compare to non-European populations ANS individuals higher affinity trend to European populations especially for Central and Northern European breeds. This trend might be giving us a signal for first migrations routes. Moreover it is also important to detect signal of other migration that Chessa et al. (2009) pointed out. If there is a second wave migration from Near East, it might be detectable in comparisons that only modern populations included.

3.4.4. Modern Middle Eastern breeds are genetically closer to modern South European breeds compared to other European breeds

Because we observed possible signal of first migration in the earlier D-test comparisons, as well as f3 results, we hypothesized whether signals of second wave migration that mentioned in Chessa et al. (2009) study, could be detected by comparing Middle Eastern and European modern populations. To address this we performed 70 comparisons of the form *D*(Out, MEDSX; SEURX, CNEURX), where MEDSX is one of Middle Eastern populations,and SEURX are CNEURX pair of modern Southern European and Central or Northern European populations. In all comparisons Middle Eastern breeds (AFS and NDZ) showed higher affinity to Southern European breeds (MER, CHU, SAB, COM and LEC) against central and northern European breeds (BOS, VBS, ERS, NSP and FIN), with significant results

(38 % of comparisons). These results are consistent with a scenario where ancestors of contemporary Middle East breeds contributed more to Southern Europe than to Central and North Europe populations, although alternative scenarios would also be possible.

CHAPTER 4

DISCUSSION

The sheep was one of the first domesticated livestock animals, managed since the early stages of the Neolithic period in west Eurasia (Abell et al., 2019). Domestic sheep soon became a valuable food source for human populations in Neolithic times and were widely distributed by humans across Eurasia and to Africa (Bellwood & Renfrew, 2002; Harris, 1996; Price, 2000). Their importance grew even further with the so-called Secondary Products Revolution (Sherratt, 1981). It can be said that examining the evolutionary history of domesticated sheep will give scientists clues not only about the history and biology of livestock animals but also the Neolithic Transition and later demographic and sociocultural changes in human populations (Vigne, 2011).

After the first domestication events in west Eurasia during the 11,000 BP, archaeozoological studies suggest that (section 1.2) sheep from the domestication center, namely Anatolia, were carried during human migrations to other continents. Europe presumably did not host wild sheep populations that could admix with the incoming domestic sheep (Davis, 1987) (in contrast to pig (Larson et al., 2007)), although in west and south Asia such populations may have existed. It can thus be hypothesized that, at least in Europe, modern-day domesticated sheep populations were largely descendants of the first domesticated populations that were brought there during the early Neolithic migrations (Tapio et al., 2006). However, this idea has yet not been directly tested, at least using whole genome ancient DNA data. This study was conducted to fill this gap and contribute to the understanding of the history of

domesticated sheep, and understanding the genetic relationships between Anatolian Neolithic domesticated sheep and modern-day breeds.

4.1. Affinity among ANS individuals relative to modern domesticated sheep breeds

Both in the principal component analysis and in D-statistics it can be easily observed that the four studied Anatolian Neolithic sheep individuals were genetically closer to each other compared to modern sheep populations. These two analyses reflect genetic similarity based on number of shared alleles (Peter, 2016b). It should be noted that higher similarity among ancient genomes (produced in a distinct way) compared to modern genomes may be, at least partly, caused technical biases, such reference bias. However, such affinity among ANS individuals may also be biologically expected, given that domesticated animals from a single region (the domestication center and its perifery) and from the same approximate period may have similar genetic backgrounds. In other words, there was not enough time for evolutionary forces, such as genetic drift, founder effect, migrations, admixtures and selection pressures on specific traits (Hamilton, 2009), to act on the Ulucak and Tepecik-Ciftlik gene pools to differentiate their genetic backgrounds. ANS, as represented by these 4 individuals, thus appears to consitute a relatively homogeneous population.

In later periods, however, the evolutionary forces in question likely acted to differentiate sheep lineages from each other, as has been reported (Tapio et al., 2006). For instance, during migrations to new regions, only small part of the sheep populations was likely carried, and the migrant population would thus undergo founder effects and differentiate. Also, there were likely distinct multiple migrations waves towards different regions of Eurasia, and some of these migrant populations may have admixed with local wild sheep, also causing differentiation among populations (Barbato et al., 2017). Besides, sheep breeds likely underwent artificial selection in later periods, which can also contribute to differentiation at least around the loci in question. Indeed, Kijas et al.'s (2012) study pinpointed regions on the sheep

genome that may have undergone directional selection. These events may lead to genetic differences between ANS individuals and modern sheep breeds, as detected in this study.

4.2. Affinity of ANS individuals to different modern domesticated sheep breeds

Both PCA analysis and also *f3-statistics*, which summarized shared genetic drift between ANS and modern breeds, suggested that ANS were genetically closer to modern European than to non-European breeds, including modern Aegean breeds very close in location to Ulucak Hoyuk.

To formally verify and assign statistical significance to these patterns, we conducted 2390 *D-statistics* testing the genetic relationship between ANS individuals and pairs of the 18 chosen modern breed populations. However, we should note in advance that only a small number of these tests were statistically significant after multiple testing correction. This is not surprising, as our ANS individuals carries low number of SNPs (between 10,406 and 3113), and when the number of SNPs is low, the power of comparisons will become weaker in *F-statistics* (Patterson et al., 2012; Zheng & Janke, 2018). Indeed, the ANS individual with the lowest number of SNPs, TEP83, had the fewest significant *D-statistics* results. Given our limited power, we also report certain trends in non-significant *D-statistics*; while admitting that these results, if repeated in the future with more SNPs for ANS individuals, may not hold.

We obtained relatively large numbers of significant results when comparisons were conducted between European and non-European breeds, since genetic differences between those breeds are relatively huge and ANS individuals had clear affinity to European breeds. We found that, overall, ANS are closer to European groups. The TEP03 individual had significant affinity to the BOS population and the TEP62 individual showed significant preference to FIN and MER populations. In general, the most significant results were found when comparing ANS to Central / North European breeds vs. non-Europeans. One exception to this was the ULU31 individual, which chose one of the southern European population (LEC) significantly against non-

European breeds (a pattern not observed for Tepecik-Çiftlik individuals), in addition to its affinity to Central and Northern European populations. The TEP83 individual showed similar trends but not significant results; again this situation is most probably the result of the low number of SNPs in TEP83.

When ANS individuals were compared to pairs of European breeds with the D-test, we obtained only few significant results, which may be expected given the similarity between the tested European populations. The TEP62 individual showed significant affinity to the Central European BOS against other European populations. This significant signal was seen most probably because of relatively high number of SNP belong to TEP62 individual. In addition, all ANS individuals tended to show higher affinity to Central and Northern European breeds against other European populations.

4.3. Implications for the history of domestic sheep: Two gene pools

We find higher affinity of ANS to European breeds compared to non-European breeds, in principal component analysis, as well as analyses using *f3-statistics* and *D-statistics*. Further, the ANS in general show highest similarity to Central and Northern European breeds. These results raise a number of interesting points regarding the demographic history of domestic sheep:

First, ANS individuals appear to have contributed to more to the gene pool of the ancestors of modern European breeds than to those of non-European breeds. This finding seems to contradict a scenario where modern-day domestic sheep populations derived from a single homogeneous gene pool from Neolithic Anatolia. Instead, it supports a scenario where ANS individuals might be members of a genetic pool that colonized Europe, whereas other sheep gene pools, in Asia and Africa, had non-ANS ancestry from another source. Asian sheep could have differentiated from ANS already during the Neolithic Period; this could indicate another ancestral gene pool for domesticated sheep breeds, perhaps in the east of the Fertile Crescent, as reported for the goat (Daly et al., 2018). Alternatively, introgressions from local wild sheep populations to domestic sheep breeds, as a result of sheep breeding strategy, could

have differentiated Asian domestic sheep populations in a later period separately. Introgression from local Asian wild sheep populations to domesticated Asian sheep breeds have been reported in various previous genetic studies (D. W. Cai et al., 2006; Hu et al., 2019).

4.4. Implications for the putative second migration wave

Chessa et al. (2009) claimed that there occurred a second wave of sheep migration after the Neolithic Period, most likely originating from Southwest Asia and spreading into Europe, Africa and the rest of the Asia. Their suggestion was that domesticated sheep that spread by this second migration wave was very successful because of wool demand at that time, estimated to be 3000-2000 BCE. According to Chessa and colleagues, this new sheep have been selected for its high wool production and subrogated the first migrated sheep in early Neolithic time. This idea has not yet been tested using ancient DNA data.

Our results would be consistent with a second migration wave scenario that had an Asian origin. We find that ANS sheep are not only closer to European populations, but particularly to Central and North European populations. Further, we observe that Southwest Asian sheep, Afshari (AFS) and Norduz (NDZ), are significantly closer to South European breeds compared to Central and Northern European breeds. A second migration wave from Southwest Asia that heavily influenced the Mediterranean region of Europe but less affected Central and North Europe could explain both our results. After domestication, Maritime route played crucial role for dispersion of goats and mainly shaped the domesticated goat profile of North Africa (Pereira et al., 2009). Additionally, during Bronze Age migrations, as reported in a previous pigs and cattle study (Meiri et al., 2017), Mediterranean coastline was highly used by human populations. Those examples might indicate that migrations have been happening throughout history. Therefore, perhaps with more recent migrations across

Mediterranean coastlines, all movements could lead to introgression from Middle Eastern breeds to Southern Europe breeds.

We note that the general affinity of ANS individuals to Central European and Northern European breeds might also be consistent with two routes linking Anatolia to Europe. This would assume that another early Neolithic sheep population dispersed along the Mediterranean, different from the ANS individuals we studied, while ANS descendants spread to Europea through the Danube Land Route, as mentioned in Bökönyi (1974). Here it is interesting that the only Aegean Neolithic individual we have, ULU31, showed the highest affinity to a South European population (LEC) in comparisons between European and non-European populations. This may also be taken as a hint for the Maritime Route, although in none of the comparisons among European breeds did any ULU31 show higher affinity to South Europe. It is thus safer to conclude that with the current dataset we cannot yet discern alternative routes of the early Neolithic migrations. This question would be best resolved comparing ANS with ancient sheep genomes from Europe.

4.5. Future directions

In future studies, we are planning to expand our SNP panel to have higher number of SNPs for ancient samples. It is crucial to have higher number of SNPs, because we could have more information about our samples and it would increase the power of statistical analyses. Therefore, we could make stronger and deeper interpretations about demographic history.

In the present study, because of the inadequate number of SNPs and small sample size, we could not perform phenotypic investigations of the samples. Higher number of SNPs and samples could provide an opportunity to study the phenotype. After 3000-2000 BCE, secondary products of sheep are thought to have increased in value for human populations (Chessa et al., 2009), and artificial selection was likely carried out to breed productive sheep in terms of wool and milk. Therefore, phenotype studies

could observe see specific changes on the genome under selection, and would provide precious information about the evolutionary history of domesticated sheep.

Finally, we are planning to increase both ancient and modern sample size. Higher sample size will increase reliability of the statistical tests. Additionally, ancient samples from different excavations sites and periods found in Anatolia and along possible migration routes will provide us an opportunity to make genomic comparisons between sheep samples coming from different ages and places. Therefore, further explorations can be done about the evolutionary history of domesticated sheep in detail.

REFERENCES

Abell, J. T., Quade, J., Duru, G., Mentzer, S. M., Stiner, M. C., Uzdurum, M., & Özbaşaran, M. (2019). Urine salts elucidate Early Neolithic animal management at Aşikli Höyük, Turkey. *Science Advances*, *5*(4). https://doi.org/10.1126/sciadv.aaw0038

Alberto, F. J., Boyer, F., Orozco-Terwengel, P., Streeter, I., Servin, B., De Villemereuil, P., … Pompanon, F. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03206-y

Ammerman, A. . (2010). The first Argonauts: towards the study of the earliest seafaring in the Mediterranean. In *The Global Origins and Development of Seafaring* (pp. 81–92).

Ann Horsburgh, K., & Rhines, A. (2010). Genetic characterization of an archaeological sheep assemblage from South Africa's Western Cape. *Journal of Archaeological Science*, *37*(11), 2906–2910. https://doi.org/10.1016/j.jas.2010.06.035

Arbuckle, B. S., Kansa, S. W., Kansa, E., Orton, D., Çakirlar, C., Gourichon, L., … Würtenberger, D. (2014). Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0099845

Barbato, M., Hailer, F., Orozco-Terwengel, P., Kijas, J., Mereu, P., Cabras, P., … Bruford, M. W. (2017). Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-07382-7

Barlow, A., Hartmann, S., Gonzalez, J., Hofreiter, M., & Johanna, L. A. (2018). Consensify : a method for generating pseudohaploid genome sequences from palaeogenomic datasets with reduced error rates. https://doi.org/10.1101/498915

Bellwood, P., & Renfrew, C. (2002). *Examining the Farming/Language Dispersal Hypothesis*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bökönyi, S. (1974). *History of domestic mammals in Central and Eastern Europe*.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prufer, K., … Paabo, S. (2007). Patterns of damage in genomic DNA sequences from a

Neandertal. *Proceedings of the National Academy of Sciences*, *104*(37), 14616–14621. https://doi.org/10.1073/pnas.0704665104

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, *4*, 900–910.

Cai, D., Tang, Z., Yu, H., Han, L., Ren, X., Zhao, X., … Zhou, H. (2011). Early history of Chinese domestic sheep indicated by ancient DNA analysis of Bronze Age individuals. *Journal of Archaeological Science*, *38*(4), 896–902. https://doi.org/10.1016/j.jas.2010.11.019

Cai, D. W., Han, L., Zhang, X. L., Zhou, H., & Zhu, H. (2006). DNA analysis of archaeological sheep remains from China. *Journal of Archaeological Science*, *34*(9), 1347–1355. https://doi.org/10.1016/j.jas.2006.10.020

Cecchinato, A., Bittante, G., & Vacca, G. M. (2016). Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep, 717–726. https://doi.org/10.1111/age.12474

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., … Arranz, J. J. (2009). Revealing the History of Sheep Domestication Using Retrovirus Integrations, *324*(April), 532–537.

Childe, G. (1925). *The Dawn of European civilization*.

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, *5*(7), 1–7. https://doi.org/10.1101/cshperspect.a012567

Daly, K. G., Delser, P. M., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., … Erek, C. M. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science , 85–88*, *361*, 24–27.

Davis, S. (1987). *The Archaeology of Animals*.

Demirci, S., Koban Baştanlar, E., Dağtaş, N. D., Pişkin, E., Engin, A., Özer, F., … Togan, I. (2013). Mitochondrial DNA diversity of modern, ancient and wild sheep (Ovis gmelinii anatolica) from Turkey: New insights on the evolutionary history of sheep. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0081952

Dymova, M. A., Zadorozhny, A. V., Mishukova, O. V., Khrapov, E. A., Druzhkova, A. S., Trifonov, V. A., … Filipenko, M. L. (2017). Mitochondrial DNA analysis of ancient sheep from Altai. *Animal Genetics*, *48*(5), 615–618. https://doi.org/10.1111/age.12569

Fernandez, H., Hughes, S., Vigne, J., Helmer, D., Hodgins, G., Miquel, C., … Taberlet, P. (2006). Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *PNAS*, *103*(42), 15375–15379.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., … Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, *5*, 1–9. https://doi.org/10.1038/ncomms6257

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLOS Genetics*, *15*(7), e1008302. https://doi.org/10.1371/journal.pgen.1008302

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., … Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, *522*(7555), 207–211. https://doi.org/10.1038/nature14317

Hamilton, M. B. (2009). *Population Genetics*. Wiley-Blackwell.

Harris, D. (1996). *The Origins and Spread in Agriculture and Pastoralism in Eurasia*.

Heaton, M. P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., Mcewan, J., … Genomics, S. (2014). SNPs for Parentage Testing and Traceability in Globally Diverse Breeds of Sheep, *9*(4), 1–10. https://doi.org/10.1371/journal.pone.0094851

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. *Journal of Heredity*, 113–120.

Hiendleder, Stefan, Kaupe, B., Wassmuth, R., & Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1494), 893–904. https://doi.org/10.1098/rspb.2002.1975

Hillman, G. C. (1996). Late Pleistocene changes in wild plant-foods available to hunter-gatherers of the northern Fertile Crescent: possible preludes to cereal cultivation. In *The origins and spread of agriculture and pastoralism in Eurasia*. UCL Press.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics*, *2*(May), 353–359.

Hu, X. J., Yang, J., Xie, X. L., Lv, F. H., Cao, Y. H., Li, W. R., … Li, M. H. (2019). The genome landscape of tibetan sheep reveals adaptive introgression from Argali and the history of early human settlements on the Qinghai-Tibetan Plateau. *Molecular Biology and Evolution*, *36*(2), 283–303. https://doi.org/10.1093/molbev/msy208

International Sheep Genomics Consortium. (n.d.). Retrieved from

https://www.sheephapmap.org/

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. *Genome Research*. https://doi.org/10.1101/gr.176552.114

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Neto, L. R., Cristobal, M. S., … Dalrymple, B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, *10*(2). https://doi.org/10.1371/journal.pbio.1001258

Kircher, M. (2012). Analysis of High-Throughput Ancient DNA Sequencing Data. In *Ancient DNA. Methods in Molecular Biology (Methods and Protocols)* (pp. 197–228). Humana Press. https://doi.org/https://doi.org/10.1007/978-1-61779-516-9_23

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., & Pääbo, S. (2010). A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Current Biology*, *20*(3), 231–236. https://doi.org/10.1016/j.cub.2009.11.068

Larsen, C. (1995). Biological Changes in Human Populations with Agriculture. In *Annual Review of Anthropology* (pp. 185–213).

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., … Dinu, A. (2007). Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *PNAS*, *104*(39), 15276–15281. https://doi.org/10.1037/h0059413

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows – Wheeler transform, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Sam, T. (2009). The Sequence Alignment / Map format and SAMtools, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Ligges, U., & Maechler, M. (2003). scatterplot3d - an R package for Visualizing Multivariate Data. *Journal of Statistical Software*, 1–20. https://doi.org/http://dx.doi.org/10.18637/jss.v008.i11

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient

Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Meadow, R. (1989). Osteological evidence for the process of animal domestication. In *The walking larder: Patterns of domestication, pastoralism, and predation.* (pp. 80–96).

Meiri, M., Stockhammer, P. W., Marom, N., Bar-Oz, G., Sapir-Hen, L., Morgenstern, P., … Finkelstein, I. (2017). Eastern Mediterranean Mobility in the Bronze and Early Iron Ages: Inferences from Ancient DNA of Pigs and Cattle. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-00701-y

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. https://doi.org/10.1101/pdb.prot5448

Moradi, M. H., Nejati-javaremi, A., Moradi-shahrbabak, M., & Dodds, K. G. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition.

Niemi, M., Bläuer, A., Iso-Touru, T., Nyström, V., Harjula, J., Taavitsainen, J. P., … Kantanen, J. (2013). Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (Ovis aries) in Finland: comparison with contemporary sheep breeds. *Genetics, Selection, Evolution : GSE*, *45*, 2. https://doi.org/10.1186/1297-9686-45-2

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pereira, F., Queirós, S., Gusmão, L., Nijman, I. J., Cuppen, E., Lenstra, J. A., … Amorim, A. (2009). Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. *Molecular Biology and Evolution*, *26*(12), 2765–2773. https://doi.org/10.1093/molbev/msp200

Peter, B. M. (2016). Admixture, Population Structure, and F -Statistics. *Genetics*, *202*(4), 1485–1501. https://doi.org/10.1534/genetics.115.183913

Poinar, H. N., & Stankiewicz, B. A. (2002). Protein preservation and DNA retrieval from ancient tissues. *Proceedings of the National Academy of Sciences*, *96*(15), 8426–8431. https://doi.org/10.1073/pnas.96.15.8426

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

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*,

81(September), 559–575. https://doi.org/10.1086/519795

RStudio Team. (2015). RStudio: Integrated Development for R. Retrieved from http://www.rstudio.com/

Ryder, M. (1983). *Sheep & man*.

Sabatini, S., Bergerbrant, S., Brandt, L. Ø., Margaryan, A., & Allentoft, M. E. (2019). Approaching sheep herds origins and the emergence of the wool economy in continental Europe during the Bronze Age. *Archaeological and Anthropological Sciences*. https://doi.org/10.1007/s12520-019-00856-x

Sherratt, A. (1981). Plough and pastoralism: aspects of the secondary products revolution. In *Pattern of the past* (pp. 261–305).

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., … Jakobsson, M. (2014). Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. *Science*, *344*(6185), 747–750. https://doi.org/10.1126/science.1253448

Tapio, M., Marzanov, N., Ozerov, M., Ćinkulov, M., Gonzarenko, G., Kiselyova, T., … Kantanen, J. (2006). Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. *Molecular Biology and Evolution*, *23*(9), 1776–1783. https://doi.org/10.1093/molbev/msl043

Vigne, J. (2011). The origins of animal domestication and husbandry : A major change in the history of humanity and the biosphere. *Comptes Rendus Biologies*, *334*(3), 171–181. https://doi.org/10.1016/j.crvi.2010.12.009

Vretemark, M. (2010). Subsistence strategies. In *Organizing Bronze Age societies; the Mediterranean, Central Europe, and Scandinavia compared.* (pp. 155–184). Cambridge University Press.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(114), 1514–1520.

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences*, *105*(NA), 11597–11604. https://doi.org/10.1073/pnas.0801317105

Zeder, M. A., Emshwiller, E., Smith, B. D., & Bradley, D. G. (2006). Documenting domestication: The intersection of genetics and archaeology. *Trends in Genetics*, *22*(3), 139–155. https://doi.org/10.1016/j.tig.2006.01.007

Zeder, M. A., & Hesse, B. (2000). The Initial Domestication of Goats (Capra hircus) in the Zagros Mountains 10 , 000 Years Ago, *287*(March), 2254–2257.

Zheng, Y., & Janke, A. (2018). Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC Bioinformatics*, *19*(1), 1–19. https://doi.org/10.1186/s12859-017-2002-4

APPENDICES

A. Appendix A

Table A.1: List of genotyped domesticated sheep breeds from Kijas et al. (2012) study

Breed Name	Abbrevation	N	Region	Contributor
African Dorper	ADP	21	Africa	Mikka Tapio
African White Dorper	AWD	6	Africa	James Kijas
Afshari	AFS	37	SouthWest Asia	Henner Simianer
Altamurana	ALT	24	SouthWest Europe	Elena Ciani
Australian Coopworth	CPW	19	Northern Europe	James Kijas
Australian Industry Merino	AIM/MER	88	SouthWest Europe	James Kijas
Australian Merino	MER	50	SouthWest Europe	James Kijas
Australian Poll Dorset	APD	108	Northern Europe	James Kijas
Australian Poll Merino	APM/MER	98	SouthWest Europe	James Kijas
Australian Suffolk	ASU/SUF	109	Northern Europe	James Kijas
Bangladeshi BGE	BGE	24	Asia	Faruque Mdomar
Bangladeshi Garole	BGA	24	Asia	Faruque Mdomar
Barbados Black Belly	BBB	24	Americas	Cyril Roberts
Black-Headed Mutton	BHM	24	Central Europe	Ottmar Distl
Border Leicester	BRL	48	Northern Europe	James Kijas
Boreray	BOR	17	Northern Europe	Josephine Pemberton
Brazilian Creole	BCS	23	Americas	Samuel Paiva
Bundner Oberlander Sheep	BOS	24	Central Europe	Cord Drogemuller

Castellana	CAS	23	SouthWest Europe	JJ Arranz
Changthangi	CHA	29	Asia	Jorn Benenwitz
Chinese Merino	CME	23	SouthWest Europe	Runlin Ma
Chios	CHI	23	SouthWest Europe	Georgios Banos
Churra	CHU	120	SouthWest Europe	JJ Arranz
Comisana	COM	24	SouthWest Europe	Fabio Pila
Cyprus Fat Tail	CFT	30	SouthWest Asia	Despoina Miltiadou
Deccani	IDC	24	Asia	Vidya Gupta
Dorset Horn	DSH	21	Northern Europe	John McEwan
East-Friesian Brown	EFB	39	Central Europe	Ottmar Distl
East-Friesian White	EFW	9	Central Europe	Ottmar Distl
Engadine Red Sheep	ERS	24	Central Europe	Cord Drogemuller
Ethiopian Menz	EMZ	34	Africa	Mikka Tapio
Finn sheep	FIN	99	Northern Europe	Juha Kantanen
Galway	GAL	49	Northern Europe	David Machugh
Garut	GUR	22	Asia	Herman Raadsma
German Texel	GTX/TEX	46	Northern Europe	Cord Drogemuller
Gulf Coast Native	GCN	94	Americas	Noelle Cockett
Indian Garole	GAR	26	Asia	Vidya Gupta
Irish Suffolk	ISF/SUF	55	Northern Europe	David Machugh
Karakas	KRS	18	SouthWest Asia	Ibrahim Cemal
Leccese	LEC	24	SouthWest Europe	Elena Ciani
MacArthur Merino	MCM	10	SouthWest Europe	James Kijas
Meat Lacaune	LAM/LAC	78	SouthWest Europe	Carole Moreno

Merino Landschaf	MLA	24	SouthWest Europe	Georg Erhardt
Milk Lacaune	LAC	103	SouthWest Europe	Carole Moreno
Moghani	MOG	34	SouthWest Asia	Henner Simianer
Morada Nova	BMN	22	Americas	Samuel Paiva
Namaqua Afrikaner	NQA	12	Africa	James Kijas
New Zealand Romney	ROM	24	Northern Europe	John McEwan
New Zealand Texel	NTX/TEX	24	Northern Europe	John McEwan
Norduz	NDZ	20	SouthWest Asia	Ibrahim Cemal
Ojalada	OJA	24	SouthWest Europe	JJ Arranz
Old Norwegian Spaelsau	NSO/NSP	15	Northern Europe	Matthew Kent
Qezel	QEZ	35	SouthWest Asia	Henner Simianer
Rambouillet	RMB	102	SouthWest Europe	Noelle Cockett
Rasa Aragonesa	RAA	22	SouthWest Europe	Jorge Calvo
Red Maasai	RMA	45	Africa	Mikka Tapio
Ronderib Afrikaner	RDA	17	Africa	James Kijas
Sakiz	SKZ	22	SouthWest Asia	Ibrahim Cemal
Santa Ines	BSI	47	Americas	Samuel Paiva
Sardinian Ancestral Black	SAB	20	SouthWest Europe	Antonello Carta
Scottish Blackface	SBF	56	Northern Europe	Lutz Bunger
Scottish Texel	STX/TEX	80	Northern Europe	Lutz Bunger
Soay	SOA	110	Northern Europe	Josephine Pemberton
Spael-coloured	NSC/NSP	3	Northern Europe	Matthew Kent
Spael-white	NSP	32	Northern Europe	Matthew Kent
St. Elizabeth	STE	10	Americas	Cyril Roberts

Sumatra	SUM	24	Asia	Herman Raadsma
Swiss Black-Brow nMountain Sheep	SBS	24	Central Europe	Cord Drogemuller
Swiss Mirror Sheep	SMS	24	Central Europe	Cord Drogemuller
Swiss White Alpine Sheep	SWA	24	Central Europe	Cord Drogemuller
Tibetan	TIB	37	Asia	Han Jianlin
Valais Blacknose Sheep	VBS	24	Central Europe	Cord Drogemuller
Valais Red Sheep	VRS	24	Central Europe	Cord Drogemuller
Wiltshire	WIL	23	Northern Europe	John McEwan
Outgroup Population				
Argali sheep	OAM	9	-	Stefan Hiendleder