Ancient Goat Genomics

Structure, Selection, and Admixture

Kevin Daly

A thesis submitted to The University of Dublin
for the degree of Doctor of Philosophy.

Trinity College Dublin

The University of Dublin

Smurfit Institute of Genetics

2019

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed

_____

Date

_____

Summary

This thesis explores the history of domestic goat (*Capra hircus*) using ancient genomic data. The wild bezoar ibex (*Capra aegagrus*) is thought to have come under human control in southwest Asia *c.* 8,000 cal BC, beginning a ten millennium-long partnership. By selectively targeting the petrosal part of the temporal bone of the skull, high quality genomes were generated from goat populations before and throughout this period, allowing the demographic and selective forces that shaped this livestock species be characterized.

Chapter 1 gives an overview of our current understanding of goat domestication, based on zooarchaeological and genetic research. Additionally, context will be given to the fields of population genetics, ancient DNA, and the use of both to investigate the process of domestication in other species.

Chapter 2 details the screening of 254 ancient bone elements, sampled primarily from southwest Asia, establishing the species identity for 154 samples. Of these, 52 were established to be from the *Capra* genus, and supplemented with a total of 34 ancient *Capra* samples identified by collaborators. A combination of mitochondrial enrichment and high-throughput short read sequencing produced 82 mitochondrial sequences and nuclear genome data for 62 samples. Following the removal of likely identical individuals, 51 genomes of >0.01× coverage remained, median 1.05×. These data form the basis of Chapters 3, 4, and 5.

Chapter 3 explores the distribution of mtDNA haplotypes and patterns of diversity over time and space. Strong geographic partitioning of diversity is observed in the Neolithic era, with distinct mitochondrial profiles in eastern, western, and Levantine parts of southwest Asia. Divergent haplogroups are observed in Neolithic ‘Ain Ghazal, Jordan, not represented in the modern mtDNA gene pool. This genetic structure declines dramatically in periods following the Neolithic; haplogroup A, previously restricted to western Turkey and southeast Europe, appears at a high frequency in other regions. In addition, mitochondrial sequences related to the Caucasian Tur were found in the Epipaleolithic site of Direkli Cave, southern Turkey.

Chapter 4 investigates the nuclear genome diversity of 62 *Capra* specimens. Goat from domestic contexts all fell within a single clade sister to modern wild goat suggesting a single primary domestication, a conclusion which is supported with demographic modelling. Strong regional differentiation defined Neolithic goat populations with a major divide observed

between eastern and Levantine/western regions, mirroring mitochondrial results. These localized differences were partially a consequence of gene flow from wild populations, which is explicitly detected in western Neolithic populations and inferred for eastern and Levantine goat. There is also nuclear genome evidence of admixture between regions, particularly in the Levant were goat populations show increased eastern ancestry.

Chapter 5 attempts to identify genomic signals of early selection in Neolithic goat, by means of a F_{ST} outlier approach conditioned on reduced genetic diversity compared to modern wild bezoar. A total of 21 outlier regions were detected, of which 2 were shared between western and eastern Neolithic populations and are associated with pigmentation genes: one overlapping the *KIT* locus and a non-genic region 350kb upstream of *KITLG*. Other outlier regions were found to overlap genes involved in detoxification, fecundity, other production-related traits.

Chapter 6 details additional sequencing data from 14 historic *Capra* samples from the Muséum National d'Histoire Naturelle, Paris, and include the first nuclear genome-wide data for several *Capra* species. A phylogeny constructed using patterns of allele sharing suggests a primary divide between domestic/wild goat and ibex species. The position of the markhor *Capra falconeri* was not resolved in this tree. There is also evidence of admixture between domestic goat and *Capra* species, but the directionality of this gene flow proved difficult to resolve. An additional sample from Direkli Cave and dating to *c* 14,000 cal BC is identified as sharing nuclear genome ancestry with the Caucasian Tur, suggesting that the genetics of wild *Capra* populations are poorly characterized and currently understudied.

Chapter 7 summarizes the results of this thesis, and details outstanding methodological avenues and questions regarding goat domestication. To address these, preliminary results of additional screening of 52 specimen are presented, giving a total of 26 new *Capra* samples after contributions from collaborators. These data give further indications on unresolved complexities of the goat populations involved in the early stages of goat herding, and will act as a starting point for future research.

Acknowledgements

First, I have to express my gratitude to Daniel Bradley for his supervision and mentorship over the last four years. He has been forgiving of my mistakes, has always been willing to hear and discuss ideas, and has encouraged me to take every opportunity to learn and grow as a researcher. Thank you.

I would like to thank everyone involved in CodeX: Amelie Scheu for sample contributions and advice; Marta Verdugo, Victoria Mullin, and Andrew Hare for goat samples they identified; Matthew Teasdale and Valeria Mattiangeli for mentorship and lab assistance; Pierpaolo Maisano Delser for his advice and the extensive modeling he performed; Matthew Collins for the enthusiasm and interest he showed in my work.

I would also like to thank each of the many archaeologists and institutions who contributed what are irreplaceable specimen for sampling purposes. In particular, I would like to acknowledge Marjan Mashkour, Liora Horwitz, Benjamin Arbuckle, and David Orton for going above and beyond what was expected of them. They each made themselves available to me at any time, and I learned a great deal from my exchanges with them, which undoubtedly helped in the writing of this thesis and interpretation of results. They each showed a generosity of time, patience, and learning that is the hallmark of an ideal collaborator, which should be the bedrock of the inherently-collaborative field of paleogenomics.

Some personal thanks - and it should go without saying that many aspects of this thesis would have been impossible were it not for the contributions, time, and patience of those listed. Matthew, a model of patience and insight. I learned a great deal from Matthew since joining the lab in 2013 (!) and would be lucky to continue doing so in the future. Valeria, I'll never be able to fully express my appreciation for all you do both the lab (allowing it to function, your wet-lab knowledge that puts us all to shame, organizing social events, being the ever-reliable "Lab Mom" and being so much more than that) and myself (countless petrous bones drilled, extractions performed, and libraries amplified) - a simple "thanks" is not enough, any of the lab's successes or achievements were massively dependent on your contributions and hard work. Amelie, who was always more than generous with her expertise and the samples she worked, you deserve and will no doubt earn success. To various members of the Bradley Lab (Eppie, Ian, Andrew, Emily, Bruno), it was a pleasure to work with and get to know you. Russell and his growing lab (Ross, Mark, and Jenny), for being

good company with great futures ahead of them all. To the technical staff of the Smurfit Institute, thank you for everything you do.

Rui, as misery loves company (and for making us all look bad). To Victoria, who definitely put up with more rubbish from me than she had to, but who was always happy to discuss pipelines and hash out results - whatever you apply yourself to, you will absolutely crush it. To Lara, as you were one of the few people who *taitneamh a bhaint as gabhar*, I would be lucky to half as smart or as driven as you. To Pierpaolo, who undoubtedly put up with more of my blunders, Slack messages, and figure redesigns than anyone else, yet was always ready for a chat, a coffee, or a brainstorming session - it was a pleasure getting to work so much with you. And thanks for your help in the printing of this thesis, I will genuinely never forget it. And to Marta, my first and only bench neighbour. You're okay, I suppose. Definitely in the competition for my favourite Portuguese person. Still waiting on that "Kevin's Top Ten Lab Disasters" ranking, featuring such hits as "Let's try to sequence a bone directly", and "Deleting bash in one easy step".

To my closest friends, in order of our meeting - Colin, Baris, Thomas, Charles, Rhys, James, Cat, John, and Fogey (the Boys ♂ Next ♂ Door); I can never express just how fortunate I am for having met and befriended you all. To my parents, Mary and Gerard (they'll hate that) - for raising me, for never stopping supporting me, and for waiting with me at the bus stop on Sunday evenings. I would not be the person I am now (for better or worse!) were it not for you both. And yes, I wish you had remembered to give me a middle name too.

And to Niall, for being the best thing in my life over these past four years. Woof.

And to the countless goats that have been born, lived, and died with humanity over the last 10,000 years. This thesis is a memorial to you. Gone, buck not forgotten.

"'Oh, mercy, sir', the goat replied
'Pray let me tell my story O
I am no Rogue, no Ribbon man
No Croppy, Whig, or Tory O
I'm guilty not of any crime
Of petty or high treason O
I'm sadly wanted at this time
This is the milking season O'"

The Peeler and the Goat
—Darby Ryan

“In a dead white field an untethered goat gave them sardonic greeting.”

— Anthony Burgess

Table of Contents

Chapter 1: General Introduction

Goat and its domestication

The domestic goat

The domestic goat, *Capra hircus*, is one of the oldest of our animal partners. The bezoar ibex or wild goat, *Capra aegagrus*, is generally accepted as its wild progenitor (Figure 1.1), and is currently found in southwest Asia (Takada et al. 1997; Mannen et al. 2001). Current zooarchaeological evidence places the earliest evidence of goat herd management in southwestern Syria in the late 9th millennium BC (Ducos 1993; Helmer 2008; Ibáñez et al. 2010), the Zagros Mountains of present-day Iran *c.* 8,000 BC (Zeder & Hesse 2000), and also at the same time in southern Levant (Arbuckle & Atici 2013).

Ancient use

Historically, goat have been exploited for their meat, hair, skin, milk, dung, and muscle-power (Hatziminaoglou & Boyazoglu 2004). Goat breeding is recorded in Early Bronze Age Sumerian cuneiform tablets (Stepien 1996), with certain animals selected for grazing and others for fattening. Goat were relied upon in Bronze Age Egypt (Zeuner 1963), likely due to their robust nature and capacity to survive where other animals fail to thrive (Mason 1986). Goat are depicted as draught animals pulling a chariot on a Cretan seal from *c.* 1,600 BC (Zeuner 1963). Goat appear recurrently in myth as a symbol of fertility, procreation and abundance (Bonney 1993). In the Bible, they appear to have been sacrificed or cast out into the desert to remove the sins of the larger community - a practise leading to the term *scapegoat*, and which may have been performed as early as *c.* 2,300 BC in Ebla, Syria (Zatelli 1998). In Homer's Odyssey, goats are milked by the cyclops Polyphemus to produce cheese, evidencing their use as a source of secondary products (Homer 1997).

Modern use

As of 2017 there are approximately 1 billion domestic goats, having risen from approx. 460 million in 1980, 710 million in 1999, and 920 million in 2011 (FAOSTAT 2018). Their reputation as the "poor man's cow" (MacHugh & Bradley 2001) is not inaccurate given their current importance in developing countries. Of the ten nations with the largest goat

Figure 1.1 - Wild and domestic goat. Wild goat (the bezoar ibex, *Capra aegagrus*) are displayed in the top left corner. Domestic goat (*Capra hircus*) are shown in the remaining panels, demonstrating the phenotypic variation of this livestock species. Image files were obtained from wikimedia.org via Creative Commons; attributions are presented in Appendix Table 1.1.

populations, five are in sub-Saharan Africa and the remaining five in central or southeast Asia. Their relative success in regions inhospitable to other livestock animals has been attributed to their diminutive body size, broad and neophilic feeding behaviour, and efficient feed utilization and cellulose digestion (Devendra 2010), in addition to their mobile upper lip and bipedilic ability (Malechek & Provenza 1983). Together, these allow goat to browse in regions as diverse as semi-arid African scrubland, rainy tropical countries such as Bangladesh, and the harsh environment of the Himalayas (Maijala 1991). Their adaptability is apparent in repeated cases of goat feralization (Porter et al. 2016). In developed countries, goat are maintained for the production of luxury goods such as cashmere fibre and their meat; other domesticates are more efficient in terms of land use.

Agriculture and domestication

The Fertile Crescent of southwest Asia, often referred to as the Near East, was the home to a series of remarkable technological innovations *c*. 10,000 BC. The management and eventual domestication of plants (including rye, barley, and lentils) and later, animals (sheep, goat, cattle and pig), permitted a predictable source of nutrition at the expense of sedentarism and risks associated with population and density growth (Cauvin, 2000). This new subsistence strategy gave to rise to the term "Neolithic" or "New Stone Age", which has been split into several distinct periods (Table 1.1). The agricultural way of life spread from southwest Asia over the course of several millennia, reaching continental Europe *c*. 7,000-6,500 BC and Africa by *c*. 5,000 BC, likely by demic diffusion but also with incipient or cross-cultural development in Iran (Lazaridis et al. 2016; Broushaki et al. 2016). This process resulted in managed, domestic populations of sheep, goat, cattle, and pig being transported from their original habitat and into novel environments, often outside of their progenitor species' range.

Management of the wild progenitors of sheep, goat, cattle, and pigs is thought to have begun in southwest Asia *c*. 8,500-8,000 BC (Early PPNB). It is only by the late 7th millennium BC (Final PPNB/PPNC) that all four were incorporated into integrated economies and were distinctly domestic in morphology (Buitenhuis & Caneva 1998; Arbuckle 2012), although there was a large degree in local variability in these economies (Arbuckle et al. 2014). The distinction between a wild or domestic population is usually made by reduced size, but horncore shape, changes to facial structure and tooth size, and the appearance outside their natural range are also taken as a marker of a domestic population. There is a growing

appreciation that these markers are problematic when applied to the earliest stages of the domestication process (Zeder 2006b), and this will be explored more fully below.

The spread of these various domesticates has often been interpreted as being part of a homogenous "Neolithic package" (Çilingiroğlu 2005; Arbuckle 2014), often associated with the "core area" of the Upper Euphrates-Tigris basin (Lev-Yadun et al. 2000; Helmer, Gourichon, et al. 2005b). More recent evidence suggests that this model of the origin and spread of agriculture does not entirely fit the available data. Despite their co-occurrence in the core area by 8,000 cal BC, the full assemblage of sheep, goat, cattle and pigs did not spread through southwest Asia in a concerted fashion; instead, there is evidence of much regional variation of farming practices and livestock keeping (Arbuckle 2014; Arbuckle et al. 2014). Early domestic crops were likely managed in regions south and east to the core area prior to their use there, with numerous failed attempts at domestication elsewhere in the Near East (Fuller et al. 2012). Further east at Chogha Golan in the Zagros Mountains of Iran, wild barley, lentil and wheat were cultivated contemporaneously with PPNA/PPNB (Pre-Pottery Neolithic A/B) sites of the supposed core area (Riehl et al. 2013). The simple model of diffusion of the complete farming set of practices from the Upper Euphrates-Tigris Basin must be updated in light of evidence of a decentralized set of innovations, which were likely influenced by local hunter-gather traditions.

Table 1.1 - Cultural periods in the Near East. Dates for Neolithic time periods are drawn from (Helmer, Gourichon, et al. 2005b), and dates for post-Neolithic time periods are from (Renfrew & Bahn 2008).

Period	Time Range (BC/AD)	Period	Time Range (BC/AD)
Pre-Pottery Neolithic A (PPNA)	8,000 - 7,500	Chalcolithic (Copper Age)	4,500 - 3,800
Early PPNB	7,500 - 7,200	Bronze Age	3,800 - 1,200
Middle PPNB	7,200 - 6,500	Iron Age	1,200 - 700
Late PPNB	6,500 - 6,000	Classical Antiquity	700 BC - 400 AD
Final PPNB (PPNC) and Pottery Neolithic	6,500 - 5,500	Medieval Period	400 AD - 1500 AD

Defining domestication

Defining domestication is a topic of complexity and debate. As summarized by Zeder, definitions vary in the degree of emphasis placed on the human or animal (or plant) component (Zeder 2012; Zeder 2015). Her definition stresses a sustained, multigenerational asymmetric relationship that benefits both partner organisms, and does not require genetic or phenotypic changes in the domesticate. Others have required specific genetic and phenotypic changes to be predominate in a population before it can be considered domestic as opposed to pre-domestic (Fuller et al. 2010). Vigne (2015) proposed a looser definition of a domesticate being "organisms, populations, lineages or species that show biological change", while expanding the concept of domestication itself as to include man-made ecosystems with which both humans and animals interact.

There is a general consensus that a dichotomous, "hunted or herded" view of domestication is overly simplistic, due to the many examples of intermediate cases (Helmer, Joris, et al. 2005). Consequently, domestication can be viewed as a process with a beginning but no end (Larson & Burger 2013). Zeuner proposed a model beginning with some contact between the two species and free breeding, intensifying to confinement, captive breeding, and eventual selective breeding and breed creation (Zeuner 1963). In the case of goat, the emergence of crop-keeping would result in increased interaction between humans and the "crop robber" wild goat, creating the potential for the beginning of management by humans. Others have argued that the domestication relationship is merely an intensification of local hunter-gatherer traditions established in the Pleistocene (Higgs & Jarman 1969). A degree of wild herd management (*i.e*. intentional selective hunting) is also considered as a phase of this process (Harris 1996).

The concept of a domestication process has been recently formalized in two distinct models. The first decouples the domestic animal from domestication itself, presenting domestication as a continuum from anthropophily and commensalism to reproductive control and eventually the use of the animal as a household pet (Vigne 2011). A given population of animals/plants would fall somewhere on this spectrum and are considered domestic once biological modifications are observed. These changes could be chromatic (*e.g.* coat colour), morphological (*e.g*. changes in brain and skull structure/ volume; degree of sexual dimorphism), or behavioural (*e.g.* aggressiveness) in nature. The second proposed model separates domestication into three broad pathways - commensal, prey, and directed - which

vary by the initial relationship between domestication partners and the intention, if any, of the human actor (Zeder 2012). These pathways proceed by intensifying of the relationship between both species, as farmers exerted greater control of the reproductive behaviour of the animal partner.

As the earliest evidence of goat management occurs in regions which showed substantial wild goat exploitation in the earliest stages of the Neolithic and the preceding Epipaleolithic, goat fall into the prey pathway of domestication. Wild goat hunting was an important and in some cases predominant component of Epipaleolithic and early Neolithic economies in the Taurus Mountains of southern Turkey, northern Levant (Lebanon and Syria), southern Levant (Israel and Palestine), the Zagros Mountains of western Iran, and some regions in the centre of the Fertile Crescent (such as Cafer Hökük in eastern Turkey) (Arbuckle 2014). Some have suggested that domestication began as a means of maximizing the supply of an increasingly-limited resource (animal meat), a result of overhunting by sedentary communities (Redding 2005). There is evidence that this occurred at Sha'ar Hagolan, Israel, where pigs and cattle were over-hunted prior to their appearance as a domesticate (Marom & Bar-Oz 2013). It is possible that such changes in behaviour and hunting practices led to "unintentional entanglement" of the community to the new strategy, and the intensification of the human-animal relationship (Fuller et al. 2010; Larson & Fuller 2014). Selection resulting from changing hunting/management strategies resulted in domestic phenotypes that were in some sense more reliant on the early farmer, requiring additional time and energy investment. In the case of crops, this may have included soil fertilization and additional processing steps such as threshing; for animals on the prey pathway it may have been the time and manpower required to protect and manage an increasingly tame, low-aggression population (Trut et al. 2009).

Despite the increasing reproductive control of domestic species by humans through time, genetic isolation between domesticates and their wild progenitors likely did not occur or occurred relatively recently. Zeuner hypothesized that different phases of domestication would be characterized by different degrees of admixture with wild populations. Early managed herds may have interbred relatively freely with wild individuals, while later captive animals were prevented from breeding with their wild counterparts (Zeuner 1963). A desire to breed larger animals in the Bronze Age may have led to managed out- breeding, while the development of specific breeds by 3,000 BC would have discouraged such practices and led to the extermination and/or assimilation of wild populations.

Consistent with this hypothesis, archaeological data suggests that there was interaction between wild and managed goat, sheep, and cattle populations in central and western Anatolia, for example (Arbuckle et al. 2014). A small degree of gene flow between these divergent populations is expected to alter the genetic makeup of the invading domestic herd. Simulations indicate that even low levels of admixture success (1%) can result in massive influx from local populations to invaders, regardless of whether they are in competition with each other (Currat et al. 2008). There is substantial genetic evidence that in species such as pig (Larson et al. 2007), apples (Cornille et al. 2012), and cattle (Park et al. 2015), introduction to a new region led to gene flow from local wild populations into the domestic invader's gene pool. These have been argued as being cases of introgressive capture, rather than domestication of local wild populations (Larson & Burger 2013; Larson & Fuller 2014).

Defining domestication in archaeological terms

Given the complexities of defining what should be considered a domestic individual or population, particularly at the earliest stages of the domestication process, it is unsurprising that zoo- archaeological field has struggled with the empirical criteria for domestic status. An additional complication is the attempt to fit discrete categories onto what is considered a continuum. General changes have long been noted between wild species and their domestic counterparts and have been suggested as markers of domestication. These include - but are not limited to - an overall reduction in size, colouration, morphological changes to the skull, reduced horn size, altered hairiness, changes to soft tissue, and behavioural changes (Zeuner 1963). Many of these do not persist in the zoo- archaeological record or are the product of more recent artificial selection, and therefore are of little use when asking questions concerning the earliest stages of domestication.

Some of these changes have been used as morphological markers of domestication (or "how domestic" an animal population was) by zooarchaeologists. Body size reduction was historically the most employed criteria (Zeder 2006b). For example, Uerpmann documented a reduction in sheep and goat body size *c.* 9,500 uncal BP, attributed to selection by humans for smaller body size or the result of a change in their environment *i.e.* management by humans (Uerpmann 1978; Uerpmann 1979). Horn size reduction or morphological change is also used as an indicator of domestication in sheep, goat, and cattle. This has been explained by selection by humans for smaller horn size, a relaxation of selection for large horn size, or nutritional impoverishment of domestic animals causing reducing horn growth (Zeder

2006a). Other morphological changes, such as smaller tooth size (particularly shorter third molars) have also been employed (Helmer, Gourichon, et al. 2005b). There are substantial flaws with these approaches, which will be explored later in the context of useful markers for early goat management.

Zooarchaeologists have also employed a range of non-morphological criteria to assess the degree of domestication (Zeder 2006b). Changes in taxonomic representation (i.e. a sudden increase in the abundance of a species) may indicate altered hunting practises or specialization. Changes in exploitation patterns, for example sex-profiles and kill-off patterns, also suggests a change in the relationship between human and animal. The appearance of a species outside of its natural range, such as wild boar (Vigne 2015) and cats (Vigne et al. 2004) in Cyprus, has been taken as evidence of management by humans. The zoogeographical argument is dependent on knowledge of the wild distribution, which may have varied through time (Meadow 1989). Recent innovative approaches in distinguishing wild from domestic specimen include directly measuring changes in sexual dimorphism between wild and domestic populations (Helmer, Gourichon, et al. 2005a), and the use of stable isotope ratios in collagen and bioapatite to infer changes in diet as a consequence of increased interactions with humans, foddering, and early weaning (Makarewicz & Tuross 2012; Vigne 2015). In some cases, cultural evidence can be strongly suggestive of domestic status, such as the discovery of two graves both containing a buried human with the articulated remains of five goat kids, arranged in a semicircle at the human's feet, in the 7th millennium BC site of Mehrgarh, Pakistan (Lechevallier et al. 1982). These markers all have unique issues, and usually a mixture is used when assessing "how domestic" an animal population was (Zeder 2006b).

Young Male Kill-Off

The most robust indicator of early management of wild herds is considered to be selective killing of young males and the delayed slaughter of females, known as Young Male Kill-Off (YMKO). This behaviour is consistent with a given herd having a meat production function: males are kept alive until weight gain is no longer optimal, at which point they are killed; females are retained for breeding and secondary products such as milk (Payne 1973). Modern pastoralist herds may have a female:male sex ratio between 10-30:1, indicating the preference for breeding females in flock composition (Davis 1987). Such behaviour is distinct from that

of hunters, who preferentially target adults (particularly males), or neonates if the hunted herd is a nursery herd.

In a landmark paper, Zeder demonstrated that the caprid remains of Ganj Dareh, a site in the Zagros Mountains of Iran occupied briefly *c.* 8,000-7,500 cal BC (Figure 1.2), were likely subject to YMKO (Zeder & Hesse 2000; Zeder 2005). Zeder utilized a modern reference collection of wild and domestic goat from the Zagros Mountains for comparison with the archaeological collection (Zeder 2001). This modern dataset indicated that sex and geography/climate were strongly correlated with size, while domestic status was not. Young adult males were consistent larger than adult females, leading to the conclusion that a reduction in the number of adult males would shift size distribution profiles to appear as if animals were morphologically smaller.

Climate and geography may have also confounded previous attempts to attribute domestic status based on size; cooler, wetter regions were inhabited by wild goats that were larger on average than those from warmer, more arid climates. This phenomenon, known as the Bergmann effect, is a consequence of how body volume and surface area scales with size and how this affects that processes of heat generation and loss. Additionally, climate fluctuations and trends may result in warmer environments, resulting in environmental selection for smaller body size rather than human-directed selection. A later analysis using a time series of gazelle suggested that non- domesticated species also underwent size reduction during the same period that it is thought to occur in sheep and goat, although factors such as over-hunting, habitat degradation, and migration of herds may also explain this observation (Zeder 2005). As such, Zeder has argued that body size as a leading marker for goat domestication or management is highly flawed, due to the range of possible causal factors that may be unrelated to the process of domestication (Zeder 2006a).

A final bias accounted for by Zeder was the practice of discarding unfused skeletal elements from young animals from analysis, an issue noticed first by Hesse (Hesse 1982). Different postcranial skeletal elements become fused at various times in a caprid's life; some fuse early, such as the astragalus, whereas others fuse much later in life, like the distal metacarpal and distal radius. When males are slaughtered at a younger age, they will contribute more unfused elements which were subsequently ignored in metric analyses. Due to strong sexual dimorphism in goat, this resulted in size profiles with reduced maxima, which were taken as evidence for the diminutization associated with domestication (Darwin 1868). By combining

the age information inherent to the fused/unfused status of different postcranial skeletal elements, and the marked sexual dimorphism of such elements, Zeder constructed sex-specific age profiles indicative of the age of death/slaughter of caprine at several archaeological sites. By the age of 24 months there was a substantial reduction in male survival at Ganj Dareh, a pattern not observed at the earlier (*c.* 8,500 cal BC) nearby site of Asiab at which hunting was likely practised. Delayed female slaughter was also observed at Ali Kosh, a later (*c.* 7,500 cal BC) lowland site outside the habitat of wild goat, where there is also evidence of horn morphology changes (Frank Hole, Kent V. Flannery, James A. Neely 1969). Taken together, this indicates that the goat of Ganj Dareh were part of a managed herd, despite not showing size decreases once the bias towards fused elements was accounted for.

As it is a useful marker of the earliest phase of goat management, 78 PPNA and PPNB assemblages throughout southwestern Asia have been assessed for signatures of YMKO (Arbuckle & Atici 2013). Among earlier Neolithic sites there is some evidence (35% of sites examined dating to before 7,500 BC) of the practice: at Ganj Dareh; at Beidha in Jordan (which may be confounded by the presence of Nubian ibex (Uerpmann 1987)); further north at Middle PPNB Tell Aswad, Syria; and possibly southeast Turkey at Çayönü, late ninth/early eighth millennium BC. These sites are distributed from opposite extremes of the Fertile Crescent, and neighbour sites with no evidence of the practice. YMKO becomes much more frequent following 7,500 BC, occurring at 84% of examined sites. However, different management strategies persisted at many settlements. These observations indicate a diffuse, locally-driven process of early goat management and herding, with a regional diversity of strategies persisting into the late Neolithic.

Other lines of evidence hint at the complexity and diffuse nature of the early stages of goat domestication. At Abu Ghosh in the Jordan Valley, there is isotopic evidence of foddering and movement between pastures of morphologically wild goat *c*. 8,000 cal BC (Makarewicz & Tuross 2012). In Cyprus, morphologically wild goats were transported to the island as early as mid-ninth millennium BC, likely for hunting (Vigne et al. 2011). Together, these analyses indicate the earliest indicators of a change in relationship between hunters and caprine - increased control of the wild herd by humans - are not morphological and require other lines of investigation.

The spread of goat in the Fertile Crescent

Beyond 7,500 BC, there is a clear shift in the animal economies of southwestern Asia and the domestic status of caprine assemblages becomes less contentious. At Yiftah'el in the Southern Levant, there is a dramatic change in survival patterns at the PPNC to one consistent with meat production (Horwitz 2003). Of note is the suggestion by the authors of autochthonous domestication due to an increase in goat frequency from early stages of the site, and a gradual decrease in the maximum (but not minimum) caprid size. In eastern Jordan, goat are clearly introduced to the region at the PPNC (Horwitz et al. 1999; Garrard et al. 1996). At 'Ain Ghazal in the Jordan Valley, there is evidence of YMKO *c*. 7,500 BC (PPNB), and the sudden appearance of morphologically domestic goats at the PPNC, based on size and horncore morphology (Kohler-Rollefson 1989; Wasse 2002; Arbuckle & Atici 2013). Coinciding with the PPNC is the appearance of domestic sheep for the first time in southern Levant, which were present at extremely low numbers or absent entirely prior.

In the upper Euphrates-Tigris basin it has been argued the sites of Göbleki Tepe, Nevalı Çori and Gürcütepe, which span from the PPNA to the PPNC, chronicle the early progression of domestication (Helmer, Gourichon, et al. 2005b; Arbuckle & Atici 2013). This is based on gradual size decreases, a shift in slaughtering towards more juveniles from the mid-PPNB, and likely YMKO at the PPNC. In Central Anatolia the assemblages at Suberde and Aşikli Höyük (both occupied in the Mid PPNB, with Aşikli also occupied earlier) have been

Figure 1.2 - Archaeological sites associated with early goat management. Shaded regions indicate current distribution of the wild goat *Capra aegagrus*, which likely differed during the early Holocene.

interpreted as management of morphologically wild goat, possibly as experiments in animal control (Arbuckle 2008b; Buitenhuis 1998). Later sites in the same region, such as Çatal Hökük (Hongo et al. 2005), Bademağacı, (De Cupere et al. 2008) and Erbaba (Arbuckle 2008a) show robust evidence of morphologically domestic goat and YMKO.

The spread of goat outside of the Fertile Crescent

In Europe, farming spread via two distinct routes: a maritime route along the Mediterranean, and a slightly later Danubian route into Northern Europe. Goat remains have been identified in Portiragnes, southern France *c.* 5,600 BC, likely introduced directly from Italy (Tresset & Vigne 2007). Goat were present in the earliest lake-dwellings in Switzerland, and in *Linearbandkeramik* Neolithic settlements in southwest Germany (Zeuner 1963). The earliest evidence for goat in Britain is at Windmill Hill, Wiltshire, dating to the 4th millennium cal BC (Grigson & Nobis 1984). In eastern Asia, goat are thought to have been introduced via two routes: to Mongolia and northern China likely via the Steppe corridor, and to India and southeast Asia via the Khyber pass (Porter et al. 2016). Goat are present and are likely representatives of a domestic herd in late eighth / early seventh millennium BC Mehrgarh, eastern Pakistan (Meadow 1996). Dispersal in southeast Asia may have occurred partially along a maritime route, and was not rapid: goat likely did not arrive until the 1st millennium AD, possibly not until the Arab invasion of Indonesia in the 14th century AD. In Africa, goat were likely introduced from southwest Asia 5,000-4,000 BC, appearing in the Ethiopian highlands and central Sahara 4,500-3,000 BC, and only reaching southern Africa by 3rd-7th centuries AD (Newman 1995; Porter et al. 2016).

The neolithisation of Iran is less well understood, but this situation is improving. The Central Zagros appears to be a key region, with evidence of crop use *c*. 9,500 BC (Matthews et al. 2013), and goat management in the early to mid-8th millennium BC (Zeder 2005). In southwest Iran, the site of Rahmatabad provides evidence of a Pre-Pottery Neolithic phase *c.* 7,000-6,700 cal BC (Kharanaghi et al. 2013). There is a dearth of sites associated with the Neolithic in the Central Plateau, particularly aceramic cultures (Nashli & Matthews 2013). The earliest evidence for ceramic Neolithic culture in central Iran is found in Tepe Sialk, *c.* 5,700-5,200 cal BC (Nashli et al. 2013). Further north in the Qazvin Plain, ceramic Neolithic sites date to 6,300-5,600 cal BC at the earliest; to the east on the Tehran Plain, Pre-Pottery Neolithic cultures have also not been discovered, with ceramics and farming appearing *c.* 5,600 cal BC at the earliest. In contrast, the site of Tappeh Sang-e Chakmaq in northeastern

Iran spans Pre-Pottery, Pottery Neolithic, and Transitional Chalcolithic periods, from the 8th to 6th millennium BC, across two mounds (Roustaei et al. 2015).The sparse data from southeast Iran indicate Pre-Pottery Neolithic settlements *c.* 5,500-5,000, but these dates are not absolute (Garazhian & Shakooie 2013). Faunal analyses in the regions beyond the Zagros are somewhat lacking; for example, the faunal assemblage of Tappeh Sang-e Chakmaq is currently unpublished (Roustaei et al. 2015).

Ancient DNA, Next Generation Sequencing, and Population Genetics

Introduction to ancient DNA

The process of domestication is amenable to study using population genetics. By humans initiating greater control on the lives of animals - restricting their movement, controlling breeding activity, and altering the environment to which they are exposed - the gene pool and allele frequencies of the domestic species change relative to their wild progenitor. Genetic drift is likely to occur as interbreeding with free-roaming populations is restricted, which intensifies as managed animals are brought outside their natural range and successively found new herds. Selection for specific traits by humans - either consciously or unconsciously - would have caused favoured alleles to rise in frequency in the domesticate population, and likely resulted in reduced genetic diversity surrounding the selected loci. Using DNA from ancient remains - aDNA - it is possible to trace these events chronologically, circumventing the issue of recent processes obscuring older ones. As such, it is necessary to review the technological innovations that form the bedrock of present-day population genetics, phylogenetics and aDNA studies. I will give an overview of the last 30 years of both, and briefly describe their application to domesticates.

A brief history of population genetics

The foundation of population genetics lies in the first seventy or so years of the 20th century. The discovery by Karl Landsteiner (Landsteiner 1901) that humans differed by the presence or absence of particular molecules - in this case blood antigens and antibodies - introduced the concept of molecular variation between individuals. Variation in the makeup of proteins was demonstrated by Linus Pauling by use of electrophoresis, allowing the separation of proteins with slightly different electrical charge which reflected variation of the constituent amino acids (Pauling & Itano 1949). This was eventually extended to quantifying the population level variation of several polypeptides - known to reflect genetic variation - simultaneously in both humans and *Drosophila pseudoobscura* (Lewontin & Hubby 1966; Harris 1966). Direct measurement of DNA variation was first performed using restriction fragment length polymorphisms - exploiting the digestion patterns produced by endonucleases either recognizing or not recognizing a cleavage site - both within (Robberson et al. 1974) and between species (Potter et al. 1975). The development of DNA sequencing technology in the mid-1970s (Sanger et al. 1977; Maxam & Gilbert 1977) allowed the

variation across molecular regions to be quantified directly, first in *Drosophila melanogaster* (Kreitman 1983). This represented a breakthrough in the analysis of populations as genetic variation previously unobservable could be measured, particularly in noncoding regions where mutations most likely to be neutral and not under selection are found.

Another technical hurdle was cleared with the development of the polymerase chain reaction (PCR) (Mullis & Faloona 1987). PCR allows for the amplification of small amounts of DNA, by separating of the duplex strands using heat, synthesizing the complementary strands using a polymerase, and annealing the newly formed duplexes by cooling. Repeated cycles of these steps result in the exponential amplification of the starting DNA molecules. The method was subsequently improved using the heat-stable *Thermus aquaticus*, or *Taq*, polymerase (Saiki et al. 1988). Combined with locus-specific primers determining the target region to be amplified, PCR allowed laborious cloning steps to be skipped, and consequently large-scale population sequencing to be performed (Arnheim et al. 1990).

PCR would prove to be the basis of the aDNA field, as it addressed the first issue inherent to the analysis of ancient DNA molecules: the small quantity of surviving molecules. The half-life of mitochondrial DNA from bone has been estimated to be approximately 521 years, and lower for nuclear DNA (Allentoft et al. 2012). In the ideal preservation medium of permafrost, sequenceable molecules may remain after one million years *post-mortem*. By the estimation of Allentoft et al, the half-life of 30bp mtDNA molecules is just 500 years at 25°C. However, this rate calculation was principally calculated using targeted qPCR of a fragment of the mitochondrial genome, of tibiotarsus specimen obtained from three nearby sites. Nuclear genome fragmentation is likely more rapid and complex, due to greater vulnerability of linear DNA to exonuclease compared to circular DNA, possible protective effects of histone proteins, and tissue-specific effects. A more recent analysis indicates that precipitation and thermal fluctuations are better predictors of fragment length than age, and that a more complex model of fragmentation is preferable (Kistler et al. 2017). The authors suggested that the rate estimated by Allentoft et al is consistent with bulk DNA loss to the surrounding environment, rather than hydrolytic depurination. This mechanism is consistent with the observation that certain tissue types - such as the petrosal part of the temporal bone, to be discussed later - are associated with higher DNA preservation rates.

Ancient DNA - the PCR stage

The first attempts at analysis of ancient DNA molecules were before the invention of PCR, and involved the cloning of molecules attributed to the quagga, an extinct horse relative (Higuchi et al. 1984), and from mummified human remains (Pääbo 1985). Once PCR was demonstrated to be applicable to aDNA (Pääbo 1989), it was used to amplify mtDNA from multitude of species remains including the extinct marsupial wolf (Thomas et al. 1989), European rabbits (Hardy et al. 1995), the Neanderthal (Krings et al. 1997), and Beringian Brown Bears (Barnes et al. 2002). These studies allowed modern genetic variation be placed in the context of ancient variation, and shed light on the phylogenetic placement of extinct species.

In the context of domestic animals, PCR amplification of mtDNA has been applied to several species. In cattle, British aurochs have been demonstrated to have a mitochondrial lineage divergent from but related to that of modern European, African and Near Eastern cattle (Bailey et al. 1996), consistent with southwest Asian domestication (Troy et al. 2001). The detection of haplotype T (the most common in modern European cattle) in Neolithic Syria, and its absence in pre-Neolithic auroch in Europe further supported domestication of *Bos taurus* in the Fertile Crescent (Edwards et al. 2007). Using similar methods, it has been shown that pigs were introduced to Europe from the Near East, at which point Near Eastern mitochondria were replaced by those of European wild boar, suggesting large scale local recruitment of native boar (Larson et al. 2007). The maternal gene pool of wild horses was demonstrated to be highly diverse, with a decline in haplogroup diversity of the past 4,500 years (Cieslak et al. 2010). Using aDNA, processes affecting the genetic diversity of domestic species have been revealed, which may have otherwise not been detected. PCR-based analyses of ancient goat mitochondria have also been performed and will be discussed in Chapter 3.

Ancient DNA and Next Generation Sequencing

The invention of massively parallel, short read sequencing - known as Next Generation Sequencing (NGS) - in the mid-2000s represented a sea change in the possibilities of genomic studies. Despite improvements to Sanger sequencing and other technological innovations, the cost to sequence a 3 billion base pair genome - approximately the size of the average mammalian genome - was between $10-25 million in 2004 (National Human

Genome Research Institute 2004). By sequencing many short fragments of DNA in parallel (commonly called shotgun sequencing), NGS allowed the sequencing of an entire genome over a few days, about 100 times faster than Sanger sequencing (Margulies et al. 2005; Rogers & Venter 2005). As of 2017 the cost of sequencing a 3 Gb genome had fallen to just over $1000 (Wetterstrand 2018), enabling the large-scale genome sequencing of populations.

Although the comparatively short read length (50-500 bp, versus 1-1.2 kbp of Sanger) produced by NGS limit its application (Zhang et al. 2011), it is coincidentally suited to sequencing the DNA fragments associated with ancient specimens. NGS was used to generate draft genomes from a range of ancient specimen including the wooly mammoth (Miller et al. 2008) and the Neanderthal (Green et al. 2010). These efforts have been eclipsed by more recent feats such as the 1× genome of a Middle Pleistocene (560-850 kya) horse (Orlando et al. 2013), a 30× genome of the extinct Denisovan hominin (Meyer et al. 2012), and a 57× genome of a Siberian Mesolithic hunter-gatherer (Günther et al. 2018). It is not infrequent for genomic data of dozens if not hundreds of ancient samples to be reported in single publication (Mathieson et al. 2018; de Barros Damgaard et al. 2018), and such volumes are not limited to human studies (Gaunitz et al. 2018).

This explosion in ancient sequencing occurred in parallel with technical and sampling advances specific to the area of aDNA research. A well-documented issue is post-mortem damage resulting in base error (Hofreiter et al. 2001; Stiller et al. 2006; Briggs et al. 2007). The hydrolytic deamination of cytosine bases results in its conversion to uracil, which occurs at a greater rate closer to the 5’ end of DNA fragments where single-strand overhangs are found. During amplification these uracil sites are effectively replaced with thymine residues due to polymerase activity, resulting in 5’ C>T misincorporations. When deamination occurs on the complementary strand, it causes elevated 3’ G>A error. Post-mortem damage can be used as a means of identifying reads likely endogenous to the sample (as opposed to modern exogenous contamination) (Skoglund et al. 2014), and can be exploited to recreate the epigenome of ancient individuals due to the fact that methylated cytosines deaminate directly to thymine, potentially preserving pre-mortem methylation signals at CpG sites. However, post-mortem deamination represents a potential bias to aDNA datasets, which are also often characterized by low mean sequencing depth. As deamination is correlated with both age and preservation temperature (Sawyer et al. 2012; Kistler et al. 2017), it can result in samples of

similar thermal ages appearing more closely related than they are in reality, and increase divergence estimates of more damaged samples.

A simple means by which to address deamination error is to limit analyses to transversions (A/G↔C/T changes), ignoring transitions (C↔T, G↔A changes). The beginnings and ends of reads can also be trimmed (*i.e.* by 4-10 bp), removing the bases most likely to contain misincorporations. The damage present can be directly modelled by use of Bayesian analysis, and quality scores of bases most likely to be affected rescaled appropriately (Jónsson et al. 2013). Alternatively, uracil bases can be removed prior to library construction using Uracil-DNA-Glycosylase (known as USER or UDG) to remove the uracil moiety and producing an abasic site, which Endonuclease VIII targets for cleavage (Briggs et al. 2010). Adding smaller quantities of USER enzyme results in partial removal of damage, reducing this source of error while allowing the discrimination of ancient endogenous molecules from modern contamination (Rohland et al. 2014).

A more serendipitous advance has been the realization that the petrous part of the temporal bone - the densest bone of the mammalian skeleton (Lam et al. 1999) - is a potential reservoir of ancient endogenous DNA. Typically, a small fraction (e.g. 1-10%) of reads obtained from ancient samples could be robustly associated with the sample itself (*i.e.* the endogenous DNA content), with the remainder derived from exogenous sources such as microbes or modern humans (Noonan et al. 2005; Skoglund et al. 2012; Allentoft et al. 2015). This represented an economic limit to the degree of sequencing that could be performed on a given sample, as lower endogenous samples require more sequencing to obtain equivalent genomic coverage. The discovery that the petrous can have an endogenous content over one hundred times higher than tooth or postcranial skeletal elements from the same individual (Gamba et al. 2014) allowed researchers to focus sampling efforts on bones most likely to have high endogenous content. In this manner the amount of researcher time spent searching for bone containing DNA is reduced, the potential for sequencing more samples to a higher depth is increased, and the unnecessary destruction of precious archaeological specimens which may not contain workable amounts of DNA is minimized.

There have been a host of other innovations that have been important to the field of aDNA study. In the laboratory these include an established standard for clean room protocol to minimizing contamination by researchers (Pääbo et al. 2004), and modifications to DNA extraction protocols maximizing endogenous DNA recovery such a pre-extraction digest or a

bleach wash (Korlević et al. 2015; Boessenkool et al. 2017). An additional advance is the use of DNA barcoding during library construction to permit simultaneous processing of multiple individuals on the same sequencing lane while minimizing sample cross-contamination (Meyer & Kircher 2010). RNA bait, in-solution enrichment strategies have allowed researchers to sequence specific genomic regions such as the mitochondria (Briggs et al. 2009; Cassidy et al. 2017), nuclear loci (Fu et al. 2013; Haak et al. 2015), or enrich for the whole genome (Carpenter et al. 2013), even in samples with poor endogenous content. Some effort has also been made to optimize probe hybridization conditions and design for ancient molecules (Cruz-Dávalos et al. 2017). At the bioinformatics stage of processing, alignment settings optimized for ancient reads have been suggested (Schubert et al. 2012), and imputation used to statistically circumvent missing data by leveraging modern reference populations (Gamba et al. 2014; Martiniano et al. 2017). Finally, there have been successful attempts at extracting aDNA from novel sources including parchment (Campana et al. 2010; Teasdale et al. 2017), lake sediments (Parducci et al. 2017), clothing (O'Sullivan et al. 2016), marine mollusc shells (Der Sarkissian et al. 2017), waterlogged wood (Wagner et al. 2018), and even chewing gum (Jensen et al. 2018; Kashuba et al. 2018).

Mitochondrial DNA

Domesticates and ancient mtDNA

Ancient DNA presents an opportunity to investigate the genetic context and consequences of domestication, as inferences of ancient events and patterns using modern genomes can lead to erroneous conclusions. For example, modern European human genomes do not, with the methods currently available, allow for the direct inference of Europe's history of population and genomic upheavals. Instead, these are best observed by use of time series of ancient genomes (Lazaridis et al. 2014; Gamba et al. 2014; Lipson et al. 2017). The study of domestication using present-day genomes is similarly problematic as humans subjected species to intense artificial selection and may have transported animal populations for a variety of reasons (*e.g.* trade, raiding, colonization). There is evidence that the later occurred early in or even before the development of farming, with wild boar transported to the island of Cyprus by the mid-10th millennium BC (Vigne et al. 2009). More recently, Chinese pigs have been introduced to Europe with the aim of creating new breeds (White 2011), although this event is supported by written records and is still detectable using modern genomes (Bosse et al. 2014). Additionally, it is likely that domestic and wild populations have experienced bidirectional admixture since their supposed separation. It is these "known unknowns" - the possible migratory and gene flow events that have gone unrecorded by human history - that aDNA can reveal.

There are several features which make mtDNA an attractive locus to amplify in aDNA studies. A major advantage of targeting this molecule is cellular high copy number, with each mitochondrion containing 1-10 copies of mtDNA, and a variable number of mitochondria occurring per cell (*e.g.* ~80 to ~2000 in mammalian somatic cell) (Cole 2016). PCR-based amplification of a subregion of the mitochondrial genome or enrichment of mitochondrial reads is more likely to produce useful results than shotgun sequencing of random reads in poorly preserved samples. Additionally, mammalian mitochondrial genomes are approximately 16.5 kb in length (Anderson et al. 1981; Bibb et al. 1981), are haploid and do not recombine, making them relatively inexpensive to sequence and simple to analyze, compared with the nuclear genome.

Given these advantages, mtDNA has been the targeted locus of many aDNA analyses addressing the dynamics of domestication. Mitochondrial analysis of ancient domestic pig

has revealed that the population which initially colonized European likely experienced introgression from wild boar in western Turkey prior to their introduction to Europe (Ottoni et al. 2013). The Anatolian mitochondrial haplogroup was then replaced by a European haplogroup derived from local wild boar, suggesting domestication or recruitment of indigenous animals (Larson et al. 2007). Gene flow back into Anatolia beginning in the Bronze Age resulted in the eventual replacement of the original domestic haplogroups by the European boar lineage. Mitochondrial typing of ancient cat remains suggests an Anatolia origin of *Felis catus*, with a lineage associated with Egyptian feline mummies becoming frequent across the Mediterranean during the Greco-Roman period (Ottoni et al. 2017). In cattle, Approximate Bayesian Computation (ABC) and serial coalescent simulation of modern and ancient mtDNA D-loop fragments from southwest Asia estimated a maternal founding population of ~80 individuals, suggesting that the auroch domestication process was not trivial (Bollongino et al. 2012). A more extensive dataset supported a female auroch founding population of <100, and colonization of Europe from Anatolia resulting in reduced genetic diversity consistent with a serial dilution model (Scheu et al. 2015).

In horses, it is well established that their mitochondrial diversity is high relative to their low Y chromosomal diversity (Lippold et al. 2011), suggesting mares were incorporated from a broad range of populations, while males were selectively maintained. Alternative explanations for this are a substantially increased mtDNA mutation rate, leading to the high observed diversity, or that the founding domestic mare population was from a metapopulation with little phylogeographic structure. Several analyses of ancient horse specimen have suggested that it is extensive and asynchronous capturing events that most likely explains the observed diversity (Vilà et al. 2001). A study by Cieslak utilizing data from over 200 ancient horses and 1000 modern horses revealed high pre-domestic diversity in horse mtDNA, some of which entered the early domestic gene pool (Cieslak et al. 2010). By the Iron Age there is evidence of novel domestic haplotypes, with 22 out of 47 newly observed in the ancient domestic gene pool, suggestive of wild mare capture (although this observation is susceptible to bias in success of amplifying DNA from more recent samples *i.e.* those from the Iron Age compared to Bronze Age). Many (70%) of the observed ancient domestic mtDNA lineages are no longer present in modern populations, implying that the high present-day diversity is in fact reduced compared to the past. In contrast, the available ancient dromedary camel mtDNA sequences imply that many of the early domestic maternal lineages are represented in modern camels, which also likely admixed with an unknown dromedary ghost population (Almathen et al. 2016).

A more complex picture of domestication is observed in the dog using mtDNA. Modern data has been interpreted as evidence of extensive admixture between dogs and wolf (Vilà et al. 1997). Alternatively, distinct wolf populations may have been domesticated and subsequently intermixed. The relative likelihoods or contributions of these explanations are difficult to determine in the absence of ancient genetic data, as modern-day patterns of wolf and dog diversity are probably different from those of the distant past. For example, present-day American dogs derive the large majority of their mtDNA lineages from European canids, while the pre-Columbian population they succeeded possessed a haplogroup absent from most extant American populations (Leonard et al. 2002), suggesting that the two populations did not extensively admix during the European colonization of the Americas. The clade was subsequently detected in pre-Columbian Arctic dogs, and also in some modern-day arctic breeds (Brown et al. 2013). An extensive analysis of modern dog from indigenous American human populations indicates that the pre-Columbian clade also persists in these groups (van Asch et al. 2013). A later paper utilizing ancient pre-Columbian nuclear genomes found that this persistence was minimal, and population replacement was almost absolute (Ní Leathlobhair et al. 2018). In Europe, ancient dog mitochondria suggest a large population turnover: dogs prior to *c.* 1,000 BC having haplogroups C and D, and modern dogs predominantly A and B (Frantz et al. 2016). This has been interpreted as supporting a model of two distinct wolf domestications, one in Europe and one in East Asia, that was obscured by the aforementioned replacement; an explanation not wholly supported by nuclear data (Botigué et al. 2017).

The domestication history of fowl has also been examined using ancient mitochondrial DNA. mtDNA from turkey remains have been used to argue for domestication of distinct populations of wild fowl in North America (Speller et al. 2010). Additionally, turkey mtDNA fragments have been used to support a migratory event from the settlements of the central Mesa Verde region of Colorado to Northern Rio Grande, New Mexico in the 13th century AD (Kemp et al. 2017). And in the Pacific, the spread of chicken from their wild habitat in southeast Asia has been characterized using mtDNA, with studies in disagreement as to whether there was several waves of dispersal or if stasis followed their introduction (Storey et al. 2012; Thomson et al. 2014). In these examples, processes which may have gone unrecorded are recovered by the direct sequencing of domestic samples before, during, and after past demographic events.

The current understanding of ancient and modern goat mtDNA variation will be discussed in Chapter 4.

Weaknesses of mtDNA

Some consideration should be made on the weaknesses of mtDNA analyses, in particular in the context of domestication. Although there is information unique to mtDNA patterns due to its maternal inheritance, the mitogenome is a single locus and subject to genetic drift. Drift of mtDNA that creates a pattern inconsistent with the "true" species/population phylogeny will result in incomplete lineage sorting (Tajima 1983). When divergence times between populations/species are short there is a great probability of incomplete lineage sorting to occur (Nei 1987; Pamilo & Nei 1988; Rosenberg 2002). Inferences restricted to a single locus will be more likely to be erroneous (i.e. not reflective of the "true" tree), while those that take multiple independent loci into account (*e.g.* whole genome analysis) can average the stochasticity of drift across loci and is more likely to recover the "true" tree (Maddison & Knowles 2006). The most famous instance of this phenomenon is the mitochondrial DNA of the Neanderthal, which suggested a sister relationship between the hominin and modern human, to the exclusion of the Denisovan hominin (Krause et al. 2010). However, whole genome data from both species later indicated that Denisovans and Neanderthal were sister groups (Reich et al. 2010). Initial mitochondrial DNA evidence did not support a contribution from European auroch to European cattle (Troy et al. 2001; Edwards et al. 2007). Later data, including an ancient auroch nuclear genome, contradicted this (Achilli et al. 2008; Park et al. 2015).

Phylogenies are frequently used in concert with geographic data to describe the history of a group of species or populations in a spatial context, an approach referred to as phylogeography (Nielsen & Beaumont 2009). The underlying assumption is that demographic processes will have predictable effects on the genetic makeup of populations involved, reflected in gene phylogenies such mtDNA trees. For example, a dispersal event into an uninhabited environment is expected to create a shallow branching pattern, as the loss of lineages associated with a population bottleneck will increase the likelihood that a smaller number of individuals will contribute alleles to the next generation at the expense of other individuals. When certain key assumptions hold (e.g. migrations are associated with population bottlenecks; no admixture has occurred), trees from single loci can indeed accurately reflect the history of the species or group of populations. However, these assumptions do not always hold; recent admixture events can obscure older patterns of

diversity. For example the brown bears of the Alexander Archipelago in Alaska have polar bear-like mitochondria due to recent gene flow between the groups, concealing their deep divergence (Miller et al. 2012).

An additional consideration when making inferences from mtDNA is that uniparental markers behave differently to nuclear ones following admixture. As mitochondria are passed on only by females, no males will contribute to the transmission of mtDNA during admixture. This reduces the mitochondrial migration rate and in effect increasing the likelihood of a structured mitochondrial distribution, assuming equal sex ratios and reproductive opportunities (Birky et al. 1989). In simulations of migration, restricted gene flow between demes results in elevated introgression from the local population to the invaders, as there are less competing invader alleles that can rise in frequency at the expanding population wavefront (Currat et al. 2008). As uniparental markers show this reduced gene flow, they are more likely to rise in frequency following admixture during migration, overestimating genetic exchange. This phenomenon has been documented in flies (Bachtrog et al. 2006), newt (Zieliński et al. 2013), and gulls (Pons et al. 2014). On shorter timescales, incomplete lineage sorting can also be mistaken for admixture, complicating inferences (Ballard & Whitlock 2004).

Furthermore, the application of single-locus phylogeography to the study of domestication has frequently led to the equating of distinct mitochondrial haplogroups with evidence of independent domestications (or wild admixture). The underlying misinterpretation is that geographically-distinct wild populations always show genetic structure that is accurately reflected in the locus under investigation (*e.g.* the mitochondria). A particular region may contain individuals with diverse haplogroup lineages, assuming no major demographic events have recently occurred (Lorenzen et al. 2007). Domestication of wild individuals from such a population would create a herd with several segregating mitochondrial haplogroups. Such observations in modern domestic populations have been taken as evidence for several domestications in donkey (Beja-Pereira et al. 2004), and goat (Luikart et al. 2001; Han et al. 2010; Colli et al. 2015). A flaw of these studies is the assumption that divergent lineages within a domestic species must be the consequence of distinct domestications rather the large amount of time taken for the final two lineages to coalesce. When this observation is supported with strong phylogeographic structure in wild populations or additional evidence (*e.g.* archaeological), then the conclusion may be valid, as in the case of taurine/indicine cattle or Western/Asiatic pig. However historical admixture between wild and domestic

animals, and human activity, can alter the genetic makeup of wild populations such that phylogeographic arguments alone for multiple domestications are not robust.

Final considerations for analyses dependent on mtDNA include the possibility of selection and sweeps affecting the entire molecule (Meiklejohn et al. 2007); homoplasmy due to the higher mutation rate of mtDNA (or possibly parental mtDNA leakage and recombination); mutation rate heterogeneity across regions of the mitogenome and between different mammalian species and groups (Ballard & Whitlock 2004; Galtier et al. 2009); possible bias in molecular diversity captured when RNA enrichment probes are employed; and the presence of nuclear copies of mitochondrial DNA (Hazkani-Covo et al. 2010). These complicate estimates of divergence times and effective population sizes, both of which are of interest in population-level studies. Furthermore, these compound the wide confidence intervals associated with mtDNA-based estimates of such population parameters (Edwards & Beerli 2000), and the high variance of coalescence time estimates between loci (Hudson, Turelli, and Whitlock 2003). With perhaps a touch of hyperbole, Galtier et al described the molecule as "perhaps intrinsically the worst population genetic and phylogenetic molecular marker we can think of" (Galtier et al. 2009).

Advantages of mtDNA

Despite these concerns, mtDNA is a useful tool to analyse patterns of genetic diversity through time (in the case of aDNA studies) and space. A key aspect to this is the low effective population size of mtDNA compared to nuclear loci: as mtDNA is haploid and is maternally-inherited, it has a four-fold smaller effective population size. As lineage sorting occurs at rates proportional to the effective population size of the group, the mtDNA lineage sorts at a faster rate than the nuclear genome (Zink & Barrowclough 2008). mtDNA is expected to be a locus more reflective of the "true" population history, particularly important for rapid or recent processes. Discrepancies between nuclear and mtDNA phylogenies can also be informative, such as suggesting a sex-biased process (Funk & Omland 2003). In addition, the use of mtDNA in aDNA research has been popular due to the higher copy number relative to nuclear loci, its short length, and the extensive modern datasets available for comparison.

The availability of methods and toolkits to explicitly test different population models is another advantage of mtDNA studies of the domestication process. Rather than positing a

preferred domestication scenario as the explanation of the observed genetic diversity, it is better to define different models and test which, if any, fit the data best (Gerbault et al. 2014). As well as accounting for the bias in the narratives favoured by the researchers, explicit model testing addresses the fact that due to stochasticity, the same evolutionary process can produce different genetic patterns (Wilson & Balding 1998), and vice-versa. The development of sequence simulators that can generate temporal data (Laval & Excoffier 2004; Excoffier et al. 2013) also aids in the testing of different domestication models, by allowing ancient data points to be incorporated. Explicit model testing has been used to assess likely domestication scenarios of horses and goat (Gerbault et al. 2012; Warmuth et al. 2012), the population history of the wild auroch (Mona et al. 2010), and olive domestication (Diez et al. 2015).

Approximate Bayesian computation (ABC) is a popular and flexible method of model comparison that can be applied to mtDNA sequences (Beaumont et al. 2002). Briefly, data simulations are performed based on defined models, from which summary statistics are extracted. These simulations are defined by parameters (*e.g.* migration rate; time of separations; population size) which are given prior distributions. Summary statistics (*e.g.* pairwise differences between populations) are then used to compare each simulation to the real data, and are rejected if they are not similar enough. The parameter values of retained simulations are used to approximate the posterior parameter distributions. The appropriate summary statistics to use can vary between datasets, as different parameters will be more efficiently represented by different summary statistics (Nunes & Balding 2010; Aeschbacher et al. 2012). A major advantage of this approach is its speed when complex models are defined, compared to the computational intractability of likelihood-based approaches (Gerbault et al. 2014). With regards to model testing using mtDNA, simulations should be less time consuming and computationally expensive as only one locus must be simulated, but this comes at the cost of the wide parameter confidence intervals obtained with mtDNA (Edwards & Beerli 2000).

Nuclear DNA

The nuclear genome

Analyses which leverage the nuclear genome are considerably more powerful than those that rely on a single locus such as the mtDNA. As the autosomes contain hundreds of thousands of independent loci, phylogenetic inferences are more likely to recover the "true" phylogeny relating the groups, ameliorating the issue of variance in coalescence times and incomplete lineage sorting (Ballard & Whitlock 2004). The nuclear genome is also a more accurate record of admixture: as chromosomes are diploid and recombining, hybrid individuals will likely contribute recombinant chromosomes composed of both ancestries to the next generation. Subsequent admixture episodes will produce recombinant chromosomes of mixed ancestries. In this way, the genetic signals of admixture are preserved in the gene pool rather than lost due to genetic drift of a single locus. The nuclear genome preserves additional information in the pattern of linkage disequilibrium and regions of homozygosity, allowing for the estimation of the demographic history. Combined, a more complex account of a population can be computed using whole genome data.

Ancient DNA and WGS

As described above, technical advances (*e.g.* short-read shotgun sequencing; UDG treatment) and other discoveries (*e.g.* the petrous bone) have led to the possibility of whole genome sequencing of ancient samples. When applied to domestic animals, the powerful methods of analyses based on nuclear data can be applied to elucidate their domestication history. Explicit tests can be performed to test how domestic animals may have admixed with their wild progenitors or with other domestic populations; whether domestication was associated with a strong bottleneck, and the timing of that bottleneck; and the patterns and timing of selection acting on domestic animals. In this manner, genomic events that may have been obscured by the intensification of animal breeding and artificial selection during breed creation can be revealed.

Horses, cattle and dogs have been the primary focus of ancient whole genome studies addressing their domestication processes. In horses, ancient genomes initially indicated that the Przewalski horse represents the last extant wild horse population (Orlando et al. 2013). However genomes from the earliest horse-herding culture, the Botai of the Kazakh steppe,

indicate the Przewalski horse is actually a feralized remnant of the Botai horses, and that modern domestic horses derive almost all of their ancestry from a different source (Gaunitz et al. 2018). There is also substantial evidence that horse domestication has been associated with positive selection at many loci, a general relaxation of selection as expressed by a recent increase in deleterious mutations (known as the cost of domestication), and a very recent population bottleneck particularly affecting stallions (Schubert et al. 2014; Librado et al. 2017). In cattle, there is evidence that the European auroch admixed with North European domestic herds prior to their extinction (Park et al. 2015), and also possibly with Iberian cattle (Upadhyay et al. 2017). Late Neolithic cattle genomes from North-Central China have also indicated gene flow from a European-like source into the region over the last *c.* 4,000 years (Chen et al. 2018).

Finally in dogs, an ancient Siberian wolf genome dated to *c.* 33,000 BC suggests that certain high-latitude dog breeds have ancestry derived from a related extinct (or unsampled) wolf population, and that the divergence between the ancestors of present-day wolves and dogs is older than previous estimates (~25,000–38,000 BC) (Skoglund et al. 2015). A high coverage Neolithic dog genome from Newgrange, Ireland, showed high degree of similarity with modern European breeds (Frantz et al. 2016). Additionally, a second domestication for dogs was inferred by a change in the mtDNA pool over the last 3,000 years, the distinction of the Newgrange dog genome from modern European dogs by PCA analysis, by simulations suggesting partial divergent ancestry of the Newgrange dog, as well as the relatively late arrival of dog remains in Central Asia. However, two Neolithic German dog genomes suggested a continuity through the Neolithic and into the present, with models supporting a single origin of dogs (Botigué et al. 2017). The authors found direct evidence of Indian dog-like ancestry in Neolithic dog genomes, absent in present-day European dogs, and a dog-wolf divergence time consistent with Skoglund et al. Evidently, broad temporal sampling of domesticate genomes, including those from pre-domestication wild populations, are required to disentangle thousands of years of gene flow and selection.

Our current understanding of the nuclear variation of domestic and wild goat populations will be explored in Chapter 4.

Selection in Domesticates

The domestic phenotype

The process of selection has substantially changed the phenotypic and genetic characteristics of our partner species, including the goat. Domestic animals generally show a suite of changes relative to their wild counterparts, a fact noted and explored by Darwin (Darwin 1868). These include, but are not limited to, a reduced degree of aggression, coat colour changes, altered estrous cycles, reduced brain volume and the size of specific brain regions, and adrenocorticotropic hormone production changes. These have been characterized as features of a "domestication syndrome" common to domestic animals, which has been suggested to be a consequence of changes in neural crest cell development brought upon by human selection for reduced reaction to external stimuli (Wilkins et al. 2014). This hypothesis requires selection to act on only on a few keys genes which have pleiotropic effects, particularly those genes expressed during development (Zeder 2015).

In addition to the domestication phenotype, the process of domestication resulted in changes in the environment of the animal, with consequences to the selection acting on the animal population. Increased population density due to the corralling of large herds may have affected stress levels, the frequency of parasites, and the rate at which infections spread throughout the group. Humans may have exposed the domesticate to new food sources such as fodder, or plant types outside of their natural distribution. The transportation of animals into new continental or eco-regions would have exposed managed animal herds to precipitation rates, temperatures, and altitudes they may have not have been adapted to. These stressors would have consequences in the form of selection for traits better suited to the novel, anthropogenic herd environment. Additionally, visible traits such as coat colour, horn size, and body size may have been favoured by early farmers and selected for, possibly without the active intent of the human actor, although such changes are more frequently associated with later phases in the domestication process (Larson & Fuller 2014).

Positive selection vs relaxation of selection

The relative contributions of the relaxation of selection versus positive selection by humans to the genetic changes leading to domestic phenotypes have been debated (Larson & Fuller 2014). Relaxation of selection suggests that a phenotype which was under intense pressure in

the wild (*e.g.* large horn size) was suddenly not favoured in the domestic environment, leading to greater variation (*e.g.* smaller and more variable horn sizes). Alternatively, the anthropogenic environment may have actively favoured smaller horns, as such animals may have been easier for farmers to manage. Large horns may be energetically-costly to grow and may have been actively selected against within the herd environment and also by herders themselves (Zeder 2006b). The same applies to coat colour, where a relaxation of selection or positive selection could have produced the increased variety of phenotypes observed in domestic populations (Larson & Fuller 2014).

These two hypotheses have different predictions as to the nature of genetic change that would have occurred in the domesticate genome. In dog (Cruz et al. 2008), horse (Schubert et al. 2014), and cattle (MacEachern et al. 2009), there is evidence of an increased burden of likely-deleterious nonsynonymous mutations in modern domestic genomes, suggesting a relaxation of selection. Alternatively, positive selection at loci can cause reduced selection against nearby slightly-deleterious variants by Hill-Robertson interference (Hill & Robertson 1966). There are many examples of positive selection in the genomes of domesticates, including coat colour in pig (Fang et al. 2009) and in horses (Ludwig et al. 2009), body size and other traits in pig (Frantz et al. 2015), neural crest cells and steroid hormone binding in horses (Librado et al. 2017), altered thyroid-stimulating hormone receptor activity in chicken thought to influence growth, reproductive seasonality and metabolism (Rubin et al. 2010), and a variety of domestication-related traits such as coat colour and neurobehavioral functioning in cattle (Qanbari et al. 2014). There was likely interplay between positive selection by humans and a relaxation of selection, with the latter producing the raw material for the former (Larson & Fuller 2014).

Leveraging ancient whole genomes in selection studies

Ancient genomes have been leveraged to detect regions under positive selection in domesticates. Neolithic dogs appear to have fewer *AMY2B* gene copies than most modern dog breeds, suggesting selection for greater starch digestibility due to dietary change in humans and consequently dogs also (Botigué et al. 2017). Using Iron Age horse genomes from Kazakhstan, Librado used the Population Branch Statistic (PBS) to detect outlier regions of differentiation in the ancient population relative to modern domesticates, with over-representation of genes involved in bone morphology and expressed in breast and mammary glands (Librado et al. 2017). Additionally, the authors developed a multi-

population framework to detect regions of excessive Levels of exclusively Shared Differences (LSD) within specific clades and also within internal branches. In this manner they detected outlier regions in the internal branch associated with early domestication, which contained genes involved in neural crest cell morphology, androgen and steroid hormone binding, and associated learning, as well as genes expressed in neural crest derivatives.

Selection studies in goat will be reviewed in Chapter 5, in addition to a scan for possible selection signatures in ancient goat genomes.

The *Capra* Genus

Capra

True goat are members of the *Capra* genus, falling into the *Caprini* tribe in the cloven-hoofed ruminant family *Bovidae*. The *Caprini* tribe also includes true sheep (*Ovis*) and other goat-antelope species, including the Himalayan tahr, Arabian tahr, Himalayan blue sheep, and the American Mountain goat; the latter is a species closer to sheep than to *Capra hircus*/*aegagrus* (Macdonald 1990). In general, members of this tribe show adaptation to extreme environments, are gregarious, and show a high degree of sexual dimorphism.

There are around nine recognized species within *Capra*, including both the domestic goat and its likely progenitor, the bezoar ibex (Pidancier et al. 2006; Groves & Grubb 2011). These species designations are contentious and fail the Biological Species Concept, as they are known to interbreed in captivity (Mason 1986). The *Capra* species consist of the domestic goat and the bezoar (usually considered a single species), seven other species of ibex (Alpine, Iberian, Nubian, Walia, Siberian, East and West Caucasian), and the markhor (Table 1.2). *Capra* species are adapted to cliffs and high-altitude environments, and in some cases inhabitant alpine or arid regions (Macdonald 1990). The genus is currently distributed in a fragmented fashion (Figure 1.3): from the Himalayas in central Asia, to the Alps and Pyrenees Mountains in western and central Europe, and also extending south across the Arabian peninsula to the Simien Mountains of Ethiopia (Shackleton 1997). The distribution of each species overlaps in a limited number of regions, where they typically assort to different altitudes or areas (Schaller 1977).

The evolution of the *Capra* genus

The *Capra* genus is thought to have evolved via rapid adaptive radiation during the late Pliocene/early Pleistocene time period *c.* 1-6 million years ago, based on early biochemical and genetic analyses (Hartl et al. 1990; Manceau et al. 1999). Fossil evidence places the first appearance of the genus in Central Asia (Pilgrim 1947), but may represent an ancestor shared between the tahr and true goat. The phylogeny of *Capra* and their relatives has proven difficult to resolve, whether by morphological features (Groves & Grubb 2011), mitochondrial (Manceau et al. 1999), Y chromosomes (Pidancier et al. 2006), or nuclear loci (Ropiquet & Hassanin 2006). Within *Capra*, uniparental markers have produced incongruent

phylogenies, leading to the suggestion of gene flow between species or distinct ancestral taxa (Pidancier et al. 2006). The occurrence of hybridization between *Capra* species in captivity (Mason 1986), and between domestic goat and other *Capra* in the wild (Zalikhanov 1967), lends credibility to these suggestions.

Nuclear markers indicate *Capra* appear to be a sister of the Himalayan blue sheep or bharal *Pseudois nayaur*, the Himalayan tahr *Hemitragus jemlahicus*, the Arabian tahr *Arabitragus jayakari*, and the Barbary sheep *Ammotragus lervia* (Ropiquet & Hassanin 2006). Like the *Capra* phylogeny, the *Caprini* nuclear tree is not consistent with the mitochondrial genome tree. These incongruent genetic histories have been attributed to ancient interspecific hybridization between the ancestors of *Capra* and *Hemitragus jemlahicus*, although incomplete lineage sorting is another possible explanation.

The evolutionary history of the *Capra* genus itself is not well understood, in part due to the poor fossil record characteristic of mountainous regions species of the genus tend to inhabit (Schaller 1977). A two-population model for the genus has been suggested based on incongruous mtDNA and Y chromosome phylogenies (Pidancier et al. 2006), with a central

Table 1.2 - Capra species, binomial names and countries in which they are currently found in the wild. Table is replicated in Chapter 6 (Table 6.1) with precise geographic information. Species definition are those proposed by Heptner et al. 1961. Species distributions are drawn from Shackleton 1997.

Common Name	Binomial Name	Geographic Distribution
Bezoar Ibex	*Capra aegagrus*	Southwest and Central Asia
Markhor	*Capra falconeri*	Central Asia
East Caucasian Tur	*Capra cylindricornis*	Eastern Caucasus Mountains
West Caucasian Tur	*Capra caucasica*	Western Caucasus Mountains
Alpine Ibex	*Capra ibex*	The European Alps
Iberian Ibex	*Capra pyrenaica*	Iberian Peninsula
Nubian Ibex	*Capra nubiana*	Northeast Africa, Arabian Peninsula, Mediterranean Levant
Walia Ibex	*Capra walie*	Semien Mountains of Ethiopia
Siberian Ibex	*Capra sibirica*	Central Asia and Siberia

Asian ancestral ibex population admixing with a southeast Asian ancestral bezoar group, producing the Caucasian, Nubian, and European ibex. In Europe, Pleistocene (*c.* 80,000 BP) remains in the French Pyrenees attributed to *Capra caucasica* (Crégut-Bonnoure 1991; Rivals 2006) have led to a two migration model of wild goat colonization. Based principally on paleontological evidence, this model suggests that the ancestors of the alpine ibex first migrated to Europe *c*. 300 kya; they were followed by a population related to *Capra caucasica*, which migrated to Iberia with little interaction with alpine ibex and produced the Iberian ibex species (Ureña et al. 2018). Molecular studies have demonstrated the monophyly of the European ibex species and thus support a single colonization model (Pidancier et al. 2006; Hassanin et al. 2012). Finally, a model for the evolution of the Caucasian Tur has suggested the Western species *Capra caucasica* is a stable hybrid population of bezoar ibex and the Eastern Caucasian Tur *Capra cylindricornis*, a product of successive glacial/interglacial vicariance events (Groves & Grubb 2011). The reliance of these models on morphological features indicates our uncertainty and the deficit of *Capra* molecular data.

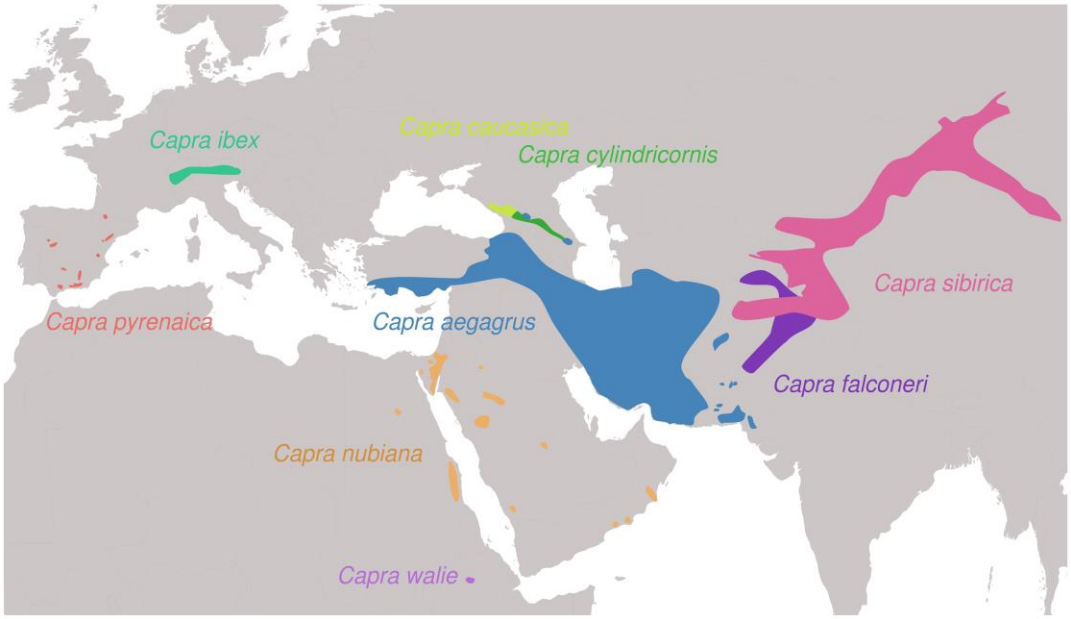


Figure 1.3: Distribution of wild caprid species. Approximate distributions taken from (Shackleton 1997), and inspired by (Pidancier et al. 2006). Reproduced in Figure 6.1.

Capra prisca

The existence of a wild European goat and hypothetical ancestor to domestic goat showing homonymous horn twisting (as opposed to the heteronymous twisting of markhor horns), was suggested in the early 20th century (Maijala 1991). While most bezoar ibex and early domestic goat have scimitar-shaped horns, spiral-horned domestic goat appear early in the zooarchaeological records, and have been attributed to a genetic input from a species distinct from the bezoar ibex (Zeuner 1963). This existence of this species, referred to as *Capra prisca*, was posited following the discovery of cranial and horncore remains from Ukraine, attributed to a Pleistocene context (2.58 mya - 11.2 kya) (Adametz 1914). A second purported *Capra prisca* specimen was reported being found in loess in Lower Austria (Sickenberg 1930). A remarkably complex set of explanations were constructed around these specimen, including suggestions that *Capra prisca* was domesticated in Europe and spread the homonymous horn type to Africa and Asia by the seventh millennium BC (Epstein 1971). The *Capra prisca* hypothesis has been largely discredited by independent assessments of the contexts the samples were found. In both cases, they likely derived from post-Pleistocene (*i.e.* post-Neolithic) contexts (Thenius et al. 1961), and the ibex being the only *Capra* species with strong evidence of existing in Pleistocene Europe (Epstein 1971).

Bezoar ibex as matrilineal ancestor of domestic goat

The identification of the bezoar ibex as the progenitor of goat is robust, despite suggestions that the markhor may be the progenitor of some breeds (Bökönyi 1974). The current distribution of bezoar ibex across the Near East overlapping with the earliest sites featuring evidence of domestic goat, their presence in the archaeological record in this region, and the general absence of other *Capra* are evidence in favour of their progenitor status (Uerpmann 1987). There is also substantial molecular evidence. mtDNA sequences demonstrated a lower number of differences between goat and bezoar than between either and markhor, with phylogenetic trees of both cytochrome *b* and the D-loop supporting monophyly of goat and bezoar ibex (Takada et al. 1997; Luikart et al. 2001). Y chromosome phylogenies also support goat-bezoar monophyly (Pidancier et al. 2006).

This does not preclude some contribution of other *Capra* species to the gene pool of domestic goat. As mentioned above, interspecific hybridization between domestic goat and other *Capra* members occurs in the wild (Zalikhanov 1967). Gene flow from domestic goat to

Alpine ibex has been reported and is the source of variation at a Major Histocompatibility immune locus in this recently-diminished wild species (Grossen et al. 2014). However, as of 2018, no major survey of the nuclear genome contribution of *Capra* species to domestic goat has been performed.

The role played by other *Capra* members in the genetic diversity of ancient and modern domestic goat will be explored in Chapter 6, using historical samples from the Muséum National d'Histoire Naturelle (MNHN) of Paris, France.

Project Overview

This thesis will attempt to address the major topics discussed in this chapter, specifically in relation to the domestic goat *Capra hircus*. A large screen of 254 ancient bone remains from southwest Asia and other areas, with a sampling focus on the petrous bone, is described in Chapter 2, and an overview of the *Capra* species identified and sequencing presented. Chapters 3, 4, and 5 discuss in detail data previously published in "Ancient goat genomes reveal mosaic domestication in the Fertile Crescent" (Daly et al. 2018). Using a dataset of 82 ancient mitochondrial sequences, Chapter 3 describes the mitochondrial DNA distributions of ancient goat, and how they change throughout the course of their domestication history. Chapter 4 dissects the variation present in 62 nuclear genomes, examining the relationship between pre-domestic wild goat and goat derived from some of the earliest farming communities, and their relationships with later goat population. Chapter 5 describes an attempt to identify possible selection signatures in these early domestic genomes, using an F_{ST} outlier and genetic diversity approach. Chapter 6 presents novel genomic data generated from 14 historical *Capra* specimen, to assess if other *Capra* species contributed to the gene pool of ancient and/or modern *Capra hircus*. Finally, Chapter 7 reviews the major findings of this project and identifies outstanding questions relating to the process of goat domestication. To address these, preliminary screening and analyses of an additional 58 unpublished samples is presented.

Chapter 1 References

Achilli, A. et al., 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current biology: CB*, 18(4), pp.R157–8.

Adametz, L., 1914. Untersuchungen über Capra prisca, einer ausgestorbenen neuen Stammform unserer Hausziegen. *Mitteilungen der landw. Lehrkanzeln dkk Hochschule f. Bodenkultur in Wien. Wien*, 3(1), pp.1–21.

Aeschbacher, S., Beaumont, M.A. & Futschik, A., 2012. A novel approach for choosing summary statistics in approximate Bayesian computation. *Genetics*, 192(3), pp.1027–1047.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Allentoft, M.E. et al., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological sciences / The Royal Society*, 279(1748), pp.4724–4733.

Almathen, F. et al., 2016. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), pp.6707–6712.

Anderson, S. et al., 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806), pp.457–465.

Arbuckle, B.S., 2012. Animals in the Ancient World. In D. T. Potts, ed. *A Companion to the Archaeology of the Ancient Near East*. Malden: Wiley-Blackwell, pp. 201–219.

Arbuckle, B.S., 2008a. Caprine exploitation at Erbaba Hoyuk: A pottery Neolithic village in central Anatolia. In L. Gourichon & E. Vila, eds. *Archaeozoology of Southwestern Asia and Adjacent Areas VIII*. Paris: Travaux de la Maison de l'Orient, pp. 345–365.

Arbuckle, B.S. et al., 2014. Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PloS one*, 9(6), p.e99845.

Arbuckle, B.S., 2014. Pace and process in the origins of animal husbandry in Neolithic Southwest Asia. *Bioarchaeology of the Near East*, 8, pp.53–81.

Arbuckle, B.S., 2008b. Revisiting Neolithic Caprine Exploitation at Suberde, Turkey. *Journal of Field Archaeology*, 33(2), pp.219–236.

Arbuckle, B.S. & Atici, L., 2013. Initial diversity in sheep and goat management in Neolithic south-western Asia. *Levantina*, 45(2), pp.219–235.

Arnheim, N., White, T. & Rainey, W.E., 1990. Application of PCR: Organismal and Population Biology. *Bioscience*, 40(3), pp.174–182.

van Asch, B. et al., 2013. Pre-Columbian origins of Native American dog breeds, with

only limited replacement by European dogs, confirmed by mtDNA analysis. *Proceedings. Biological sciences / The Royal Society*, 280(1766), p.20131142.

Bachtrog, D. et al., 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. *Evolution; international journal of organic evolution*, 60(2), pp.292–302.

Bailey, J.F. et al., 1996. Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proceedings. Biological sciences / The Royal Society*, 263(1376), pp.1467–1473.

Ballard, J.W.O. & Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Molecular ecology*, 13(4), pp.729–744.

Barnes, I. et al., 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science*, 295(5563), pp.2267–2270.

de Barros Damgaard, P. et al., 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*, p.eaar7711.

Beaumont, M.A., Zhang, W. & Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics*, 162(4), pp.2025–2035.

Beja-Pereira, A. et al., 2004. African origins of the domestic donkey. *Science*, 304(5678), p.1781.

Bibb, M.J. et al., 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell*, 26(2 Pt 2), pp.167–180.

Birky, C.W., Jr, Fuerst, P. & Maruyama, T., 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, 121(3), pp.613–627.

Boessenkool, S. et al., 2017. Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular ecology resources*, 17(4), pp.742–751.

Bökönyi, S., 1974. *History of domestic mammals in Central and Eastern Europe*, Budapest: Akademiai Kiado.

Bollongino, R. et al., 2012. Modern taurine cattle descended from small number of near-eastern founders. *Molecular biology and evolution*, 29(9), pp.2101–2104.

Bonney, M., 1993. *World of Sheep and Goats*, M. & A. Bonney.

Bosse, M. et al., 2014. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nature communications*, 5, p.4392.

Botigué, L.R. et al., 2017. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature Communications*, 8, p.16082.

Briggs, A.W. et al., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, 104(37), pp.14616–14621.

Briggs, A.W. et al., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic acids research*, 38(6), p.e87.

Briggs, A.W. et al., 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science*, 325(5938), pp.318–321.

Broushaki, F. et al., 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science*. Available at: http://dx.doi.org/10.1126/science.aaf7943.

Brown, S.K., Darwent, C.M. & Sacks, B.N., 2013. Ancient DNA evidence for genetic continuity in arctic dogs. *Journal of archaeological science*, 40(2), pp.1279–1288.

Buitenhuis, H., 1998. Asikli Höyük: a "protodomestication" site. *Anthropozoologica*, 25-26, pp.655–662.

Buitenhuis, H. & Caneva, I., 1998. Early animal breeding in south-eastern Anatolia: Mersin-Yumuktepe. In P. Anreiter, ed. *Man And The Animal World*. Budapest: Archaeolingua, pp. 121–130.

Campana, M.G. et al., 2010. A flock of sheep, goats and cattle: Ancient DNA analysis reveals complexities of historical parchment manufacture. *Journal of archaeological science*, 37, pp.1317–1325.

Carpenter, M.L. et al., 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American journal of human genetics*, 93(5), pp.852–864.

Cassidy, L.M. et al., 2017. Capturing goats: documenting two hundred years of mitochondrial DNA diversity among goat populations from Britain and Ireland. *Biology letters*, 13(3).

Chen, N. et al., 2018. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature communications*, 9(1), p.2337.

Cieslak, M. et al., 2010. Origin and history of mitochondrial DNA lineages in domestic horses. *PloS one*, 5(12), p.e15311.

Çilingiroğlu, Ç., 2005. The concept of "Neolithic package": considering its meaning and applicability. *Documenta Praehistorica*, 32, pp.1–13.

Cole, L.W., 2016. The Evolution of Per-cell Organelle Number. *Frontiers in cell and developmental biology*, 4, p.85.

Colli, L. et al., 2015. Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC genomics*, 16(1), p.1115.

Cornille, A. et al., 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS genetics*, 8(5), p.e1002703.

Crégut-Bonnoure, E., 1991. Intérêt biostratigraphique de la morphologie dentaire de Capra (Mammalia, Bovidae). *Annales zoologici Fennici*, 28(3/4), pp.273–290.

Cruz-Dávalos, D.I. et al., 2017. Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular ecology resources*, 17(3), pp.508–522.

Cruz, F., Vilà, C. & Webster, M.T., 2008. The legacy of domestication: Accumulation of deleterious mutations in the dog genome. *Molecular biology and evolution*, 25, pp.2331–2336.

Currat, M. et al., 2008. The hidden side of invasions: massive introgression by local genes. *Evolution; international journal of organic evolution*, 62(8), pp.1908–1920.

Daly, K.G. et al., 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, 361(6397), pp.85–88.

Darwin, C., 1868. *The Variation of Animals and Plants Under Domestication*, O. Judd.

Davis, S.J.M., 1987. *The archaeology of animals*, London: Batsford.

De Cupere, B., Duru, R. & Umurtak, G., 2008. Animal husbandry at the Early Neolithic to Early Bronze Age site of Bademağacı (Antalya province, SW Turkey): evidence from the faunal remains. In E. Vila et al., eds. *Archaeozoology of the Near East VIII*. Lyon: Maison de l'Orient et de la Méditerranée, pp. 367–405.

Der Sarkissian, C. et al., 2017. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. *Molecular ecology resources*, 17(5), pp.835–853.

Devendra, C., 2010. Exploiting goats for maximum productivity. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*, 91(1-4), pp.246–255.

Diez, C.M. et al., 2015. Olive domestication and diversification in the Mediterranean Basin. *The New phytologist*, 206(1), pp.436–447.

Ducos, P., 1993. Some remarks about Ovis, Capra and Gazella remains from two PPNB sites from Damascene, Syria, Tell-Aswad and Ghoraifa. *Archaeozoology of the Near East I, Universal Book Services, Leidan*, p.37e45.

Edwards, C.J. et al., 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings. Biological sciences / The Royal Society*, 274(1616), pp.1377–1385.

Edwards, S.V. & Beerli, P., 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution; international journal of organic evolution*, 54(6), pp.1839–1854.

Epstein, H., 1971. *The origin of the domestic animals of Africa*, New York: Africana Publishing Corporation.

Excoffier, L. et al., 2013. Robust demographic inference from genomic and SNP data. *PLoS genetics*, 9(10), p.e1003905.

Fang, M. et al., 2009. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS genetics*, 5(1), p.e1000341.

FAOSTAT, 2018. FAOSTAT. Available at: http://www.fao.org/faostat/en/.

Frank Hole, Kent V. Flannery, James A. Neely, 1969. *Prehistory and Human Ecology of the Deh Luran Plain: An Early Village Sequence from Khuzistan, Iran*, Ann Arbor: University of Michigan Press.

Frantz, L.A.F. et al., 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature genetics*, 47(10), pp.1141–1148.

Frantz, L.A.F. et al., 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science*, 352(6290), pp.1228–1231.

Fuller, D.Q., Allaby, R.G. & Stevens, C., 2010. Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. *World archaeology*, 42(1), pp.13–28.

Fuller, D.Q., Willcox, G. & Allaby, R.G., 2012. Early agricultural pathways: moving outside the "core area" hypothesis in Southwest Asia. *Journal of experimental botany*, 63(2), pp.617–633.

Funk, D.J. & Omland, K.E., 2003. Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annual review of ecology, evolution, and systematics*, 34(1), pp.397–423.

Fu, Q. et al., 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America*, 110(6), pp.2223–2227.

Galtier, N. et al., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular ecology*, 18(22), pp.4541–4550.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Garazhian, O. & Shakooie, M., 2013. Tell-e Atashi (Bam, Southeastern Iran) and the Neolithic of the Eastern Near East. In R. Matthews & H. F. Nashli, eds. *The Neolithisation of Iran: The Formation of New Societies*. Oxford: Oxbow Books, pp. 285–296.

Garrard, A.N., Colledge, S.M. & Martin, L., 1996. The emergence of crop cultivation and caprine herding in the "Marginal Zone" of the southern Levant. In D. R. Harris, ed. *The origins and spread of agriculture and pastoralism in Eurasia.* The origins and spread of agriculture and pastoralism in Eurasia. London: UCL Press, pp. 204–226.

Gaunitz, C. et al., 2018. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science*, 360(6384), pp.111–114.

Gerbault, P. et al., 2012. Domestication and migrations: Using mitochondrial DNA to infer domestication processes of goats and horses. In E. Kaiser, J. Burger, & S. Wolfram, eds. *Population Dynamics in Prehistory and Early History. New Approaches Using Stable Isotopes and Genetics*. Berlin: De Gruyter, pp. 17–30.

Gerbault, P. et al., 2014. Storytelling and story testing in domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), pp.6159–6164.

Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *Science*, 328(5979), pp.710–722.

Grigson, C. & Nobis, G., 1984. The domestic animals of the earlier Neolithic in Britain. *Die Anfänge des Neolithikums vom Orient bis Nordeuropa. Teil*, 10, pp.205–220.

Grossen, C. et al., 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. *PLoS genetics*, 10. Available at: http://dx.doi.org/10.1371/journal.pgen.1004438.

Groves, C. & Grubb, P., 2011. *Ungulate Taxonomy*, Baltimore: JHU Press.

Günther, T. et al., 2018. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS biology*, 16(1), p.e2003703.

Haak, W. et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555), pp.207–211.

Han, L. et al., 2010. Mitochondrial DNA analysis provides new insights into the origin of the Chinese domestic goat. *Small ruminant research: the journal of the International Goat Association*, 90(1–3), pp.41–46.

Hardy, C. et al., 1995. Rabbit mitochondrial DNA diversity from prehistoric to modern times. *Journal of molecular evolution*, 40(3), pp.227–237.

Harris, D.R., 1996. Introduction: themes and concepts in the study of early agriculture. In D. R. Harris, ed. *The Origins and Spread of Agriculture and Pastoralism in Eurasia*. London: UCL Press, pp. 1–11.

Harris, H., 1966. Enzyme polymorphisms in man. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society* , 164(995), pp.298–310.

Hartl, G.B. et al., 1990. On the Biochemical Systematics of the Caprini and the Rupicaprini. *Biochemical systematics and ecology*, 18(2-3), pp.175–182.

Hassanin, A. et al., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes rendus biologies*, 335(1), pp.32–50.

Hatziminaoglou, Y. & Boyazoglu, J., 2004. The goat in ancient civilisations: from the Fertile Crescent to the Aegean Sea. *Small ruminant research: the journal of the International Goat Association*, 51(2), pp.123–129.

Hazkani-Covo, E., Zeller, R.M. & Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS genetics*, 6(2), p.e1000834.

Helmer, D., Gourichon, L., et al., 2005a. Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 86–95.

Helmer, D., 2008. Révision de la faune de Cafer Höyük (Malatya, Turquie): apports des méthodes de l'analyse des mélanges et de l'analyse de Kernel à la mise en évidence de la domestication. *Travaux de la Maison de l'Orient et de la Méditerranée*, 49(1), pp.169–195.

Helmer, D., Gourichon, L., et al., 2005b. The upper Euphrates-Tigris basin: Cradle of agro-pastoralism? In J. Peters, J.-D. Vigne, & D. Helmer, eds. *First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 86–124.

Helmer, D., Joris, P. & Vigne, J.-D., 2005. New archaeozoological approaches to trace the first steps of animal domestication. In D. Helmer, P. Joris, & J.-D. Vigne, eds. *First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 1–16.

Heptner, V.G., Nasimovich, A.A. & Bannikov, A.G., 1961. *Mammals of the Soviet Union* V. G. Heptner & N. P. Naumov, eds., Moscow: Vysshaya Shkola.

Hesse, B., 1982. Slaughter Patterns and Domestication: The Beginnings of Pastoralism in Western Iran. *Man*, 17(3), pp.403–417.

Higgs, E.S. & Jarman, M.R., 1969. The Origins of Agriculture: a Reconsideration. *Antiquity*, 43(169), pp.31–41.

Higuchi, R. et al., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 312(5991), pp.282–284.

Hill, W.G. & Robertson, A., 1966. The effect of linkage on limits to artificial selection. *Genetical research*, 8(3), pp.269–294.

Hofreiter, M. et al., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research*, 29(23), pp.4793–4799.

Homer, 1997. *The Odyssey*, Penguin.

Hongo, H. et al., 2005. Sheep and goat remains from Çayönü Tepesi, southeastern Anatolia. In H. Buitenhuis et al., eds. *Archaeozoology of the Near East VI: Proceedings of the sixth international symposium on the archaeozoology of southwestern Asia and adjacent areas*. ARC Publications 123. Groningen: ARC Publications, pp. 112–123.

Horwitz, L.K. et al., 1999. Animal Domestication in the southern Levant. *Paléorient*, 25(2), pp.63–80.

Horwitz, L.K., 2003. Temporal and Spatial Variation in Neolithic Caprine Exploitation Strategies : a Case Study of Fauna from the Site of Yiftah'el (Israel). *Paléorient*, 29(1), pp.19–58.

Ibáñez, J.J. et al., 2010. The early PPNB levels of Tell Qarassa North (Sweida, southern Syria). *Antiquity*, 84, pp.1–5.

Jensen, T.Z.T. et al., 2018. Stone Age "chewing gum" yields 5,700 year-old human genome and oral microbiome. *bioRxiv*, p.493882. Available at: https://www.biorxiv.org/content/10.1101/493882v1.

Jónsson, H. et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , 29(13), pp.1682–1684.

Kashuba, N. et al., 2018. Ancient DNA from chewing gums connects material culture and genetics of Mesolithic hunter-gatherers in Scandinavia. *bioRxiv*, p.485045. Available at: https://www.biorxiv.org/content/10.1101/485045v1.

Kemp, B.M. et al., 2017. Prehistoric mitochondrial DNA of domesticate animals supports a 13th century exodus from the northern US southwest. *PloS one*, 12(7), p.e0178882.

Kharanaghi, H.A., Nashli, H.F. & Nishiaki, Y., 2013. Tepe Rahmatabad: A Pre-Pottery and Pottery Neolithic Site in Fars Province. In R. Matthews & H. F. Nashli, eds. *The Neolithisation of Iran*. Oxford: Oxbow Books, pp. 108–123.

Kistler, L. et al., 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic acids research*, 45(11), pp.6310–6320.

Kohler-Rollefson, I., 1989. Changes in Goat Exploitation at 'Ain Ghazal between the Early and Late Neolithic : A Metrical Analysis. *Paléorient*, 15(1), pp.141–146.

Korlević, P. et al., 2015. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques*, 59(2), pp.87–93.

Krause, J. et al., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, 464(7290), pp.894–897.

Kreitman, M., 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. *Nature*, 304(5925), pp.412–417.

Krings, M. et al., 1997. Neandertal DNA sequences and the origin of modern humans. *Cell*, 90(1), pp.19–30.

Lam, Y.M., Chen, X. & Pearson, O.M., 1999. Intertaxonomic Variability in Patterns of Bone Density and the Differential Representation of Bovid, Cervid, and Equid Elements in the Archaeological Record. *American antiquity*, 64(2), pp.343–362.

Landsteiner, K., 1901. Agglutination phenomena in normal human blood. *Wiener klinische Wochenschrift*, 14, pp.1132–1134.

Larson, G. et al., 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 104(39), pp.15276–15281.

Larson, G. & Burger, J., 2013. A population genetics view of animal domestication. *Trends in genetics: TIG*, 29(4), pp.197–205.

Larson, G. & Fuller, D.Q., 2014. The Evolution of Animal Domestication. *Annual review of ecology, evolution, and systematics*, 45(1), pp.115–136.

Laval, G. & Excoffier, L., 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* , 20(15), pp.2485–2487.

Lazaridis, I. et al., 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), pp.409–413.

Lazaridis, I. et al., 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617), pp.419–424.

Lechevallier, M., Meadow, R.H. & Quivron, G., 1982. Dépôts d'animaux dans les sépultures néolithiques de Mehrgarh, Pakistan. *Paléorient*, 8(1), pp.99–106.

Leonard, J.A. et al., 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science*, 298(5598), pp.1613–1616.

Lev-Yadun, S., Gopher, A. & Abbo, S., 2000. Archaeology. The cradle of agriculture. *Science*, 288(5471), pp.1602–1603.

Lewontin, R.C. & Hubby, J.L., 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*, 54(2), pp.595–609.

Librado, P. et al., 2017. Ancient genomic changes associated with domestication of the horse. *Science*, 356(6336), pp.442–445.

Lippold, S. et al., 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature communications*, 2, p.450.

Lipson, M. et al., 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, 551(7680), pp.368–372.

Lorenzen, E.D. et al., 2007. Phylogeography, hybridization and Pleistocene refugia of the kob antelope (Kobus kob). *Molecular ecology*, 16(15), pp.3241–3252.

Ludwig, A. et al., 2009. Coat color variation at the beginning of horse domestication. *Science*, 324(5926), p.485.

Luikart, G. et al., 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America*, 98, pp.5927–5932.

Macdonald, D. ed., 1990. *Hoofed Mammals (All the World's Animals)*, New York: Torstar Books.

MacEachern, S. et al., 2009. Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? *BMC genomics*, 10, p.179.

MacHugh, D.E. & Bradley, D.G., 2001. Livestock genetic origins: goats buck the trend. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), pp.5382–5384.

Maddison, W.P. & Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology*, 55(1), pp.21–30.

Maijala, K., 1991. *Genetic resources of pig, sheep, and goat*, Elsevier Science

Publishers.

Makarewicz, C. & Tuross, N., 2012. Finding Fodder and Tracking Transhumance: Isotopic Detection of Goat Domestication Processes in the Near East. *Current anthropology*, 53(4), pp.495–505.

Malechek, J.C. & Provenza, F.D., 1983. Feeding behaviour and nutrition of goats on rangelands. *World Animal Review (FAO)*. Available at: http://agris.fao.org/agris-search/search.do?recordID=XF8335667.

Manceau, V. et al., 1999. Systematics of the genus Capra inferred from mitochondrial DNA sequence data. *Molecular phylogenetics and evolution*, 13, pp.504–510.

Mannen, H., Nagata, Y. & Tsuji, S., 2001. Mitochondrial DNA reveal that domestic goat (Capra hircus) are genetically affected by two subspecies of bezoar (Capra aegagurus). *Biochemical genetics*, 39, pp.145–154.

Margulies, M. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), pp.376–380.

Marom, N. & Bar-Oz, G., 2013. The prey pathway: a regional history of cattle (Bos taurus) and pig (Sus scrofa) domestication in the northern Jordan Valley, Israel. *PloS one*, 8(2), p.e55958.

Martiniano, R. et al., 2017. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS genetics*, 13(7), p.e1006852.

Mason, I.L., 1986. *Evolution of Domesticated Animals*, John Wiley & Sons, Incorporated.

Mathieson, I. et al., 2018. The genomic history of southeastern Europe. *Nature*, 555(7695), pp.197–203.

Matthews, R. et al., 2013. Investigating the neolithisation of society in the central Zagros of western Iran. In R. Matthews & H. Nashli Fazeli, eds. *The Neolithisation of Iran: The Formation Of New Societies.* Oxford: Oxbow Books, pp. 14–34.

Maxam, A.M. & Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), pp.560–564.

Meadow, R.H., 1989. Osteological evidence for the process of animal domestication. In J. Clutton-Brock, ed. *The Walking Larder: Patterns of Domestication, Pastoralism, and Predation*. London: Unwin-Hyman, pp. 80–90.

Meadow, R.H., 1996. The origins and spread of agriculture and pastoralism in northwestern South Asia. In D. R. Harris, ed. *The origins and spread of agriculture and pastoralism in Eurasia*. London: UCL Press, pp. 390–412.

Meiklejohn, C.D., Montooth, K.L. & Rand, D.M., 2007. Positive and negative selection on the mitochondrial genome. *Trends in genetics: TIG*, 23(6), pp.259–263.

Meyer, M. et al., 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), pp.222–226.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, 2010(6), p.db.prot5448.

Miller, W. et al., 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), pp.E2382–90.

Miller, W. et al., 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, 456(7220), pp.387–390.

Mona, S. et al., 2010. Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC evolutionary biology*, 10, p.83.

Mullis, K.B. & Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology*, 155, pp.335–350.

Nashli, H.F. et al., 2013. Mapping the Neolithic occupation of the Kashan, Tehran and Qazvin plains. In M. Rogers & H. F. Nashli, eds. *The Neolithisation of Iran*. Oxford: Oxbow Books, pp. 124–146.

Nashli, H.F. & Matthews, R., 2013. The Neolithisation of Iran: patterns of change and continuity. In H. F. Nashli & R. Matthews, eds. *The Neolithisation of Iran: The Formation of New Societies*. Oxford: Oxbow Books, pp. 1–13.

National Human Genome Research Institute (NHGRI), 2004. Press Release: NHGRI Seeks Next Generation of Sequencing Technologies. *National Human Genome Research Institute (NHGRI)*. Available at: https://www.genome.gov/12513210/.

Nei, M., 1987. *Molecular Evolutionary Genetics*, Columbia University Press.

Newman, J.L., 1995. *The Peopling of Africa: A Geographic Interpretation*, Yale University Press.

Nielsen, R. & Beaumont, M.A., 2009. Statistical inferences in phylogeography. *Molecular ecology*, 18(6), pp.1034–1047.

Ní Leathlobhair, M. et al., 2018. The evolutionary history of dogs in the Americas. *Science*, 361(6397), pp.81–85.

Noonan, J.P. et al., 2005. Genomic sequencing of Pleistocene cave bears. *Science*, 309(5734), pp.597–599.

Nunes, M.A. & Balding, D.J., 2010. On optimal selection of summary statistics for approximate Bayesian computation. *Statistical applications in genetics and molecular biology*, 9, p.Article34.

Orlando, L. et al., 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 499(7456), pp.74–78.

O'Sullivan, N.J. et al., 2016. A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing. *Scientific reports*, 6, p.31279.

Ottoni, C. et al., 2013. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Molecular biology and evolution*, 30(4), pp.824–832.

Ottoni, C. et al., 2017. The palaeogenetics of cat dispersal in the ancient world. *Nature Ecology & Evolution*, 1(7), p.0139.

Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, 86(6), pp.1939–1943.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Pääbo, S., 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, 314(6012), pp.644–645.

Pamilo, P. & Nei, M., 1988. Relationships between gene trees and species trees. *Molecular biology and evolution*, 5(5), pp.568–583.

Parducci, L. et al., 2017. Ancient plant DNA in lake sediments. *The New phytologist*, 214(3), pp.924–942.

Park, S.D.E. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome biology*, 16, p.234.

Pauling, L. & Itano, H.A., 1949. Sickle cell anemia a molecular disease. *Science*, 110(2865), pp.543–548.

Payne, S., 1973. Kill-off Patterns in Sheep and Goats: the Mandibles from Aşvan Kale. *Anatolian Studies*, 23, pp.281–303.

Pidancier, N. et al., 2006. Evolutionary history of the genus Capra (Mammalia, Artiodactyla): discordance between mitochondrial DNA and Y-chromosome phylogenies. *Molecular phylogenetics and evolution*, 40(3), pp.739–749.

Pilgrim, G.E., 1947. The Evolution of the Buffaloes, Oxen, Sheep and Goats. *Zoological journal of the Linnean Society*, 41(279), pp.272–286.

Pons, J.-M. et al., 2014. Extensive mitochondrial introgression in North American Great Black-backed Gulls (Larus marinus) from the American Herring Gull (Larus smithsonianus) with little nuclear DNA impact. *Heredity*, 112(3), pp.226–239.

Porter, V. et al., 2016. *Mason's World Encyclopedia of Livestock Breeds and Breeding*, Boston: CABI.

Porter, V. & Tebbit, J., 1996. *Goats of the World*, London: Farming Press Ltd.

Potter, S.S. et al., 1975. Specific cleavage analysis of mammalian mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 72(11), pp.4496–4500.

Qanbari, S. et al., 2014. Classic selective sweeps revealed by massive sequencing in

cattle. *PLoS genetics*, 10(2), p.e1004148.

Redding, R.W., 2005. Breaking the mold: A consideration of variation in the evolution of animal domestication. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow, pp. 41–48.

Reich, D. et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468(7327), pp.1053–1060.

Renfrew, C. & Bahn, P., 2008. *Archaeology: Theories, Methods and Practice* 5 edition., Thames and Hudson Ltd.

Riehl, S., Zeidi, M. & Conard, N.J., 2013. Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science*, 341(6141), pp.65–67.

Rivals, F., 2006. Discovery of Capra caucasica and Hemitragus cedrensis (Mammalia, Bovidae) in the Late Pleistocene levels of the Caune de l'Arago (Tautavel, France): Biochronological implication in the Mediterranean Basin context. *Geobios . Memoire special*, 39, pp.85–102.

Robberson, D.L., Clayton, D.A. & Morrow, J.F., 1974. Cleavage of replicating forms of mitochondrial DNA by EcoRI endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 71(11), pp.4447–4451.

Rogers, Y.-H. & Venter, J.C., 2005. Genomics: massively parallel sequencing. *Nature*, 437(7057), pp.326–327.

Rohland, N. et al., 2014. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660).

Ropiquet, A. & Hassanin, A., 2006. Hybrid origin of the Pliocene ancestor of wild goats. *Molecular phylogenetics and evolution*, 41(2), pp.395–404.

Rosenberg, N.A., 2002. The probability of topological concordance of gene trees and species trees. *Theoretical population biology*, 61(2), pp.225–247.

Roustaei, K., Mashkour, M. & Tengberg, M., 2015. Tappeh Sang-e Chakmaq and the beginning of the Neolithic in north-east Iran. *Antiquity*, 89(345), pp.573–595.

Rubin, C.-J. et al., 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464(7288), pp.587–591.

Saiki, R.K. et al., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839), pp.487–491.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp.5463–5467.

Sawyer, S. et al., 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one*, 7(3), p.e34131.

Schaller, G.B., 1977. *Mountain Monarchs: Wild Sheep and Goats of the Himalaya*, London: University of Chicago Press.

Scheu, A. et al., 2015. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC genetics*, 16, p.54.

Schubert, M. et al., 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics*, 13, p.178.

Schubert, M. et al., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences*, 111 (52), pp.E5661–E5669.

Shackleton, D.M., 1997. *Wild sheep and goats and their relatives: status survey and conservation action plan for Caprinae*, IUCN. Available at: https://portals.iucn.org/library/sites/library/files/documents/1997-006.pdf.

Sickenberg, O., 1930. Eine Wildziege der Capra prisca-Gruppe aus dem Plistozän Niederösterreichs. *Palaeobiologica*, 3, pp.92–102.

Skoglund, P. et al., 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current biology: CB*, 25(11), pp.1515–1519.

Skoglund, P. et al., 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, 336(6080), pp.466–469.

Skoglund, P. et al., 2014. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), pp.2229–2234.

Speller, C.F. et al., 2010. Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.2807–2812.

Stepien, M., 1996. *Animal Husbandry in the Ancient Near East: A Prosopographic Study of Third-Millennium Umma*, Capital Decisions Ltd.

Stiller, M. et al., 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37), pp.13578–13584.

Storey, A.A. et al., 2012. Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures. *PloS one*, 7(7), p.e39171.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105(2), pp.437–460.

Takada, T. et al., 1997. Bezoar (Capra aegagrus) is a matriarchal candidate for ancestor of domestic goat (Capra hircus): evidence from the mitochondrial DNA diversity. *Biochemical genetics*, 35(9-10), pp.315–326.

Teasdale, M.D. et al., 2017. The York Gospels: a 1000-year biological palimpsest. *Royal*

Society open science, 4(10), p.170988.

Thenius, E., Hofer, F. & Preisinger, A., 1961. Capra" prisca „Sickenberg und ihre Bedeutung für die Abstammung der Hausziegen: Ein Beitrag zur Verwendbarkeit des Fluortestes auf röntgenographischer Basis für Herkunft und Alter subfossiler Knochen. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*, 76(1-4), pp.321–325.

Thomas, R.H. et al., 1989. DNA phylogeny of the extinct marsupial wolf. *Nature*, 340(6233), pp.465–467.

Thomson, V.A. et al., 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proceedings of the National Academy of Sciences of the United States of America*, 111(13), pp.4826–4831.

Tresset, A. & Vigne, J.-D., 2007. Substitution of species, techniques and symbols at the Mesolithic/Neolithic transition in Western Europe. *Proceedings of the British Academy*, 144, pp.189–210.

Troy, C.S. et al., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410(6832), pp.1088–1091.

Trut, L., Oskina, I. & Kharlamova, A., 2009. Animal evolution during domestication: the domesticated fox as a model. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 31(3), pp.349–360.

Uerpmann, H.-P., 1978. Metrical analysis of faunal remains from the Middle East. *Approaches to faunal analysis in the Middle East*, 2.

Uerpmann, H.-P., 1979. *Probleme der Neolithisierung des Mittelmeerraums*, Weisbaden: Reichert.

Uerpmann, H.-P., 1987. *The ancient distribution of ungulate mammals in the Middle East: fauna and archaeological sites in Southwest Asia and Northeast Africa*, Wiesbaden: L. Reichert Verlag.

Upadhyay, M.R. et al., 2017. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. *Heredity*, 118(2), pp.169–176.

Ureña, I. et al., 2018. Unraveling the genetic history of the European wild goats. *Quaternary science reviews*, 185, pp.189–198.

Vigne, J.-D., 2015. Early domestication and farming: what should we know or do for a better understanding? *Anthropozoologica*, 50(2), pp.123–150.

Vigne, J.-D. et al., 2004. Early taming of the cat in Cyprus. *Science*, 304(5668), p.259.

Vigne, J.-D. et al., 2009. Pre-Neolithic wild boar management and introduction to Cyprus more than 11,400 years ago. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), pp.16135–16138.

Vigne, J.-D. et al., 2011. The Early Process of Mammal Domestication in the Near East: New Evidence from the Pre-Neolithic and Pre-Pottery Neolithic in Cyprus. *Current anthropology*, 52(S4), pp.S255–S271.

Vigne, J.-D., 2011. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies*, 334(3), pp.171–181.

Vilà, C. et al., 1997. Multiple and ancient origins of the domestic dog. *Science*, 276(5319), pp.1687–1689.

Vilà, C. et al., 2001. Widespread origins of domestic horse lineages. *Science*, 291(5503), pp.474–477.

Wagner, S. et al., 2018. High-Throughput DNA sequencing of ancient wood. *Molecular ecology*, 27(5), pp.1138–1154.

Warmuth, V. et al., 2012. Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), pp.8202–8206.

Wasse, A., 2002. Final Results of an Analysis of the Sheep and Goat Bones from Ain Ghazal, Jordan. *Levantina*, 34(1), pp.59–82.

Wetterstrand, K.A., 2018. DNA Sequencing Costs: Data. *National Human Genome Research Institute (NHGRI)*. Available at: https://www.genome.gov/27541954/dna-sequencing-costs-data/.

White, S., 2011. From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History. *Environmental history*, 16(1), pp.94–120.

Wilkins, A.S., Wrangham, R.W. & Fitch, W.T., 2014. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197(3), pp.795–808.

Wilson, I.J. & Balding, D.J., 1998. Genealogical inference from microsatellite data. *Genetics*, 150(1), pp.499–510.

Zalikhanov, M.C., 1967. *Tur in Kabardin-Balkaria*, Nal'chik: Kabardin-Balkarian Publishers.

Zatelli, I., 1998. The Origin of the Biblical Scapegoat Ritual: The Evidence of Two Eblaite Texts. *Vetus Testamentum*, 48(2), pp.254–263.

Zeder, M.A., 2006a. A critical assessment of markers of initial domestication in goats (Capra hircus). In M. Zeder et al., eds. *Documenting Domestication: New Genetic and Archaeological Paradigms*. Berkeley: University of California Press, pp. 181–208.

Zeder, M.A., 2001. A Metrical Analysis of a Collection of Modern Goats (Capra hircus aegargus and C. h. hircus) from Iran and Iraq: Implications for the Study of Caprine Domestication. *Journal of archaeological science*, 28(1), pp.61–79.

Zeder, M.A., 2006b. Archaeological Approaches to Documenting Animal Domestication. In M. A. Zeder et al., eds. *Documenting Domestication: New Genetic and Archaeological Paradigms*. Berkeley: University of California Press, pp. 171–180.

Zeder, M.A., 2005. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. In J. Peters, A. V. Den Driesch, & D. Helmer, eds.

The First Steps of Animal Domestication: New Archaeozoological Approaches. Oxford: Oxbow Books, pp. 125–146.

Zeder, M.A., 2015. Core questions in domestication research. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), pp.3191–3198.

Zeder, M.A., 2012. The Domestication of Animals. *Journal of anthropological research*, 68(2), pp.161–190.

Zeder, M.A. & Hesse, B., 2000. The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* , 287, pp.2254–2257.

Zeuner, F.E., 1963. *A History of Domesticated Animals*, London: Hutchinson & Co.(Publishers) Ltd.

Zhang, J. et al., 2011. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, 38(3), pp.95–109.

Zieliński, P. et al., 2013. No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (Lissotriton montandoni). *Molecular ecology*, 22(7), pp.1884–1903.

Zink, R.M. & Barrowclough, G.F., 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology*, 17(9), pp.2107–2121.

Chapter 2: Screening and Sequencing

Introduction

Screening

As described in Chapter 1, various discoveries and innovations over the past decade have led to ancient whole genome sequencing becoming a reality. Optimized DNA extraction techniques, sampling of the petrous bone, and damage-mitigating procedures such as USER-treatment have allowed populations of ancient individuals to be sequenced and more robust inferences made of past demographic events. However, there is an element of survivorship bias in this positive description of the state of aDNA. Very often, dozens of bones with little-to-no endogenous DNA will have been screened prior to genome sequencing. Even if a bone contains sufficient endogenous DNA for species identification, it is often too low (*e.g.* <5%) to be useful for whole genome analysis. For such samples it may be more appropriate to perform targeted enrichment prior to sequencing.

An additional complexity is encountered when attempting to screen bones identified as being likely goat. Goat and sheep are difficult to distinguish based on postcranial skeletal elements (Boessneck 1971; Zeder & Lapham 2010). To quote the archaeologist Charles A. Reed, "The greatest obstacle to an analysis of the origin and spread of prehistoric sheep and goats is my complete disbelief in the validity of most of the published identifications" (Reed 1960). It is not uncommon for zooarchaeological studies to simply combine likely goat and sheep into a "ovicaprid" classifier (*e.g.* demographic profiling of Helmer et al. 2005). Any attempt to target goat samples that have been morphometrically-identified will result in unintentional sampling of sheep, and vice-versa.

Aims

The aims of this chapter are:

1. To extract DNA from ovicaprid bone elements.
2. To identify the species when DNA is present and estimate the endogenous DNA content.
3. To perform shotgun sequencing on mtDNA-enriched or non-enriched goat samples.
4. To determine the molecular sex of goat samples, and test for possible sex ratio biases.

Materials and Methods

Materials

A total of 254 bone samples were initially processed, with the vast majority (242/254) derived from the petrous. The remaining samples derived from postcranial elements such as long bones, phalanges and carpals. These samples were supplemented by 13 bones screened by others (Victoria Mullin and Andrew Hare) and identified to be *Capra hircus/aegagrus*. Finally, 21 samples of non-petrous bone powder from which control region fragments were previously analyzed by Amelie Scheu (Scheu 2012) were added to the total sample set of 288.

A summary of all samples described above, including the collaborator from which the sample was obtained, the material type (petrous, tooth root etc), assigned species, and endogenous DNA percentage is presented in Appendix Table 2.1.

Methods

Drilling

All bones/teeth were prepared in a dedicated ancient DNA facility at Trinity College Dublin, Ireland, following standard protocols (Pääbo et al. 2004). Briefly, these protocols amount to:

- Forbidding access to any item that had previously been in work spaces containing amplified DNA.
- Requiring full body suits, face masks, hair nets and shoe covers be worn by researchers while in the aDNA facilities.
- Requiring two layers of gloves be worn by researchers at all time, with the outermost glove replaced when moving between samples or stages in the workflow.
- Cleaning all surfaces and equipment with a 5% sodium hypochlorite solution or DNA-ExitusPlus™ (AppliChem GmbH, Esser et al. 2006) and ethanol.

All samples (bone and teeth) were first subject to UV decontamination (30 minutes per side) prior to handling. Samples were prepared within a fume hood in order to minimize external contamination of the sample and also contamination of the work environment by the sample.

For bone samples, the upper surface layer was removed and cleaned using a dentist drill and drill bit. A dremel cutter/saw was then used to remove a section of the bone. A Mixer Mill (MM 400, Retsch) was then used to reduce the removed bone piece to a fine powder. For tooth samples, the tooth was cut in half along a vertical line through the centre, when possible, and one half reduced to a fine powder using the Mixer Mill. Approximately 150mg of bone/tooth powder was transferred to an Eppendorf tube for extraction. Any remaining powder was stored in a separate Eppendorf tube.

Environment controls were included to detect work environment contamination; an air control by placing open Eppendorf tubes in the work fume hood for 30 minutes, and the water control by adding 500ml H_2O to the Mixer Mill shaker and subjecting it to the same procedures as described above. The aims of these controls were to account for environmental contamination in the air and contamination in the shakers respectively. Controls were subsequently included in all extraction and library preparation steps.

DNA extraction

Powdered samples and environmental controls were subject to DNA extraction as described by (Yang et al. 1998) and later modified by MacHugh et al. 2000 and Gamba et al. 2014. One further modification was introduced to the protocol: a total of three 24-hour proteinase K incubation digests were performed. Control (empty) tubes were introduced to estimate contamination introduced during the extraction process and treated identically to sample tubes.

Briefly, lysis buffer (1M Tris-HCl; 2% SDS; 0.5M EDTA; 100 µg/ml Proteinase K) was prepared and 1ml added to each tube and briefly vortexed to mix. Tubes were incubated at 37°C for 24 hours using an Eppendorf ThermoMixer™, mixing at 700 rpm. Following this, tubes were spun down using a centrifuge at 13,300 rpm, resulting in an undissolved pellet and supernatant. The supernatant was removed by use of a pipette, with care taken not to disturb the pellet. An additional 1ml of freshly prepared extraction buffer was added to each tube, vortexed, and incubated under the same conditions (37°C, 24 hours, 700 rpm). This was repeated for a total of three extractions.

The final supernatant was obtained by centrifugation of tubes for 10 minutes at 13,300 rpm and transferred to an Amicon filter (Amicon Ultra-4 Centrifugal Filter Unit 30 kDA)

containing 3ml 1× Tri-EDTA. Amicon tubes were spun down for approximately 20 minutes, until the 250µl mark was reached by the solution. The effluent was discarded and an additional 3m 1× Tri-EDTA added to the filter. The tubes were spun down until the 100µl mark was reached. The remaining 100µl was purified using a MinElute Silica column according to the manufacturer's instructions (MinElute PCR Purification Kit, Qiagen, Hilden, Germany). Purified DNA was eluted in 40µl EBT (Elution Buffer containing 0.05% Tween).

NGS library construction and amplification

DNA libraries were constructed according to Meyer & Kircher 2010 with modifications (Gamba et al. 2014). 16.25µl of purified DNA was used as the starting material. Control tubes (16.25µl H2O × 2) were included.

Blunt End Repair was performed using NEB Next End Repair Module (New England BioLab Inc.), with a final reaction volume of 70 µl (all volumes were scaled by 0.7). The reaction mix was incubated at 25℃ for 15 minutes, followed by 5℃ for 5 minutes. Qiagen MinElute PCR purification was performed according to manufacturer's instructions and eluted in 20 µl EBT. Adaptor ligation using T4 DNA ligase was performed on the elute, followed by MinElute purification as previously described. Adaptor fill-in was then performed, with an additional heat inactivation step (20 minutes at 80℃) in place of a further MinElute purification.

3µl library was amplified using 1µl of a unique index oligo (5 µM) and 21 µl of amplification master mix (20.5 µl AcuPrime Pfx Polymerase (Invitrogen), 0.5 µl primer IS4 (10 µM)). Blank PCR controls (3µl H_2O) were included. 12 cycles of amplification were performed for all samples and controls (95℃ for 5 min; 12 × 95℃ for 15 sec, 60℃ for 30 sec, 68℃ for 30 sec; 68℃ for 5 min). Amplified product was purified using MinElute columns as previously described, and then eluted in 10 µl EB.

DNA quantification was performed using an Agilent 2200 Tapestation (Agilent Technologies) and Qubit® Fluorometric quantitation (dsDNA HS assay, Invitrogen).

MiSeq sequencing

10ng of each library was pooled and then sequenced on an Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland), using 70bp single-end sequencing, and a PhiX control at 1X.

Quality assessment and read trimming

Sequenced reads were obtained in the.fastq.gz format. The quality of reads was assessed using FastQC (Andrews 2010), to determine if systematic biases may have occurred during sequencing. When no issue was detected, reads were trimmed using cutadapt v1.9.1 (Martin 2011), to remove reads shorter than 30bp in length and if more than 1bp of overlap between the read and adaptor sequence occurred:

```
cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30
```

Species identification

Species identity was determined using Fastq Screen (Wingett & Andrews 2018). Briefly, several reference genomes (Appendix Table 2.2) are combined into a dataset and sample reads aligned using bowtie (Langmead et al. 2009) to each genome in that dataset. Each read is scored as aligning to one unique location in one genome, aligning multiple times in one genome, aligning once to several genomes, or aligning multiple times to several genomes. For each sample, the genome producing the highest observed “one hit, one genome” statistic was selected as the likely identity for that sample. When the highest “one hit, one genome” percentage was below 0.2%, the sample was considered too poorly preserved for species identification and removed from subsequent analysis. An example output of Fastq Screen is presented below (Figure 2.1).

Read alignment and endogenous DNA

Following species identification, trimmed reads were aligned to the appropriate reference genome, using the assembly specified in Appendix Table 2.2. Alignment was performed using bwa aln (Li & Durbin 2009) with seeding disabled (bwa aln -l 1024) (Schubert et al. 2012). Sai (suffix array index) files were then used to produce Sam (Sequence/Alignment

Map) files which were converted to Bam (Binary Sam) files with Samtools 0.1.19. Bam files were sorted and filtered for reads mapping quality less than 30, where mapping quality is a Phred-based probability of the alignment being incorrect, and duplicates were removed using Samtools. Endogenous DNA was calculated as number of unique reads aligned following mapping quality filtering, divided by total reads following trimming step.

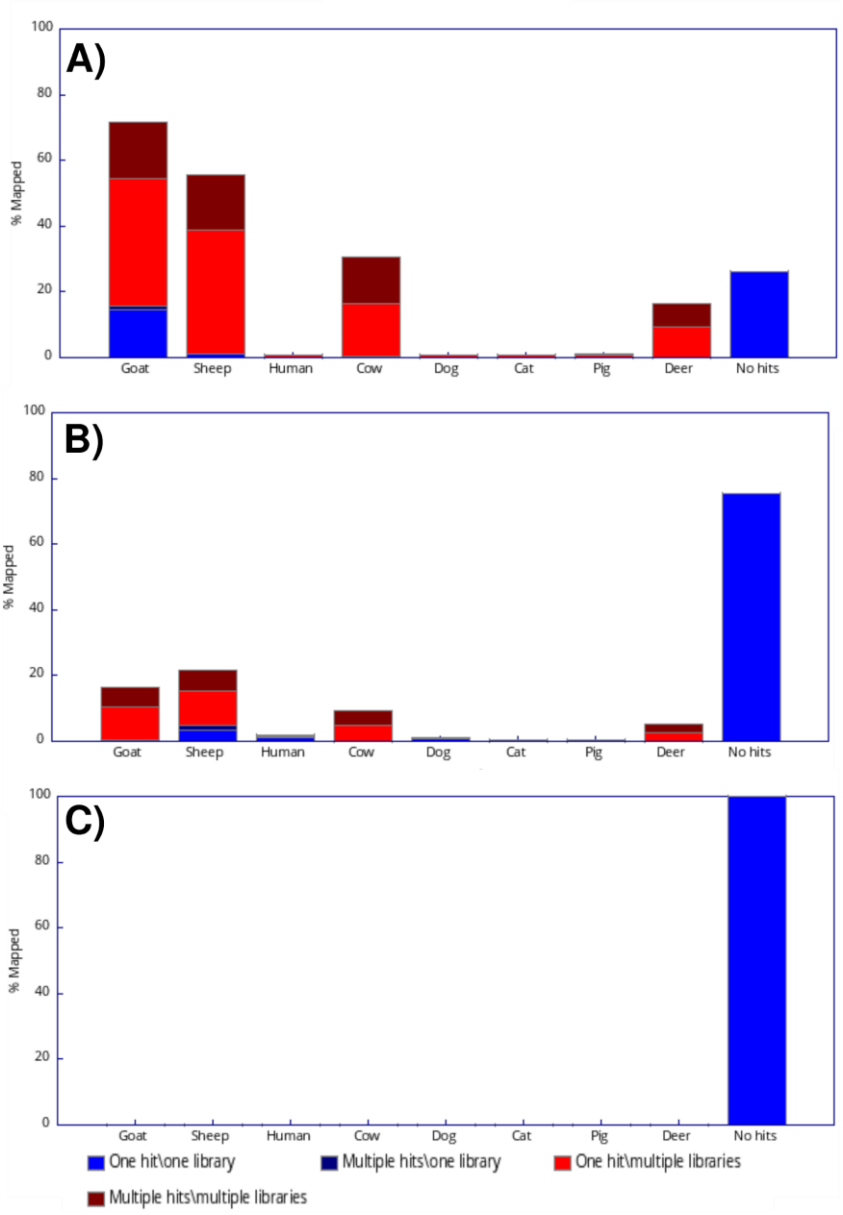


Figure 2.1 - Fastq Screen output for species identification. Examples of identified A) goat, B) sheep, and C) when species ID could not be determined.

Damage assessment

As described in Chapter 1, aDNA is characterized by short fragment length (~ 45bp) and 5' C>T, 3'G>A misincorporations. To assess the damage for those samples identified as *Capra*, bam files were analyzed using maDamage2.0 (Jónsson et al. 2013). Among other utilities, mapDamage can calculate the read length distributions and the rate of various substitution patterns at both ends of reads in an aligned bam file.

Molecular sex identification

Due to the absence of a complete Y chromosome in CHIR_1.0, molecular sex of all samples identified as *Capra* was determined using the relationship between the number of reads aligned to each chromosome versus the length of that chromosome (Park et al. 2015). The ratio of reads aligned to the X chromosome and the length of that chromosome were then used to estimate the sex of each sample. A linear regression model of the number of reads aligned to a chromosome versus the length of that chromosome is constructed using the rstudent function of R (R Core Team 2016), and studentized residuals extracted. The probability of the sample being female is then determined by determining the probability of observing the X chromosome residual from a Student's t distribution (24 degrees of freedom). This is then repeated with a linear model in which the individual is male *i.e.* twice as many reads should be aligned to the X chromosome to get a well-fitting linear regression. The ratio of these probabilities is used to compute a p value from a Chi-Squared distribution (one degree of freedom), first by calculating the likelihood ratio of the two probabilities. An example of the output plots for male and females is displayed in Figure 2.2.

To test for an excess of females to males, a 95% confidence interval for the female proportion was calculated using the following equation:

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}},$$

where p is the proportion of females and n is the total number of individuals. This was repeated excluding samples derived from a pre-Neolithic context.

Sequencing strategy

Samples identified as *Capra* species were selected for further sequencing. In general, samples with endogenous DNA percentages greater than 10% were chosen for whole genome on Illumina HiSeq platforms, while samples with less than 10% endogenous content were enriched for mitochondrial reads using RNA baits. In some cases, samples with less than 10% endogenous DNA but from an under-sampled time period or location were also sequenced on a HiSeq platform.

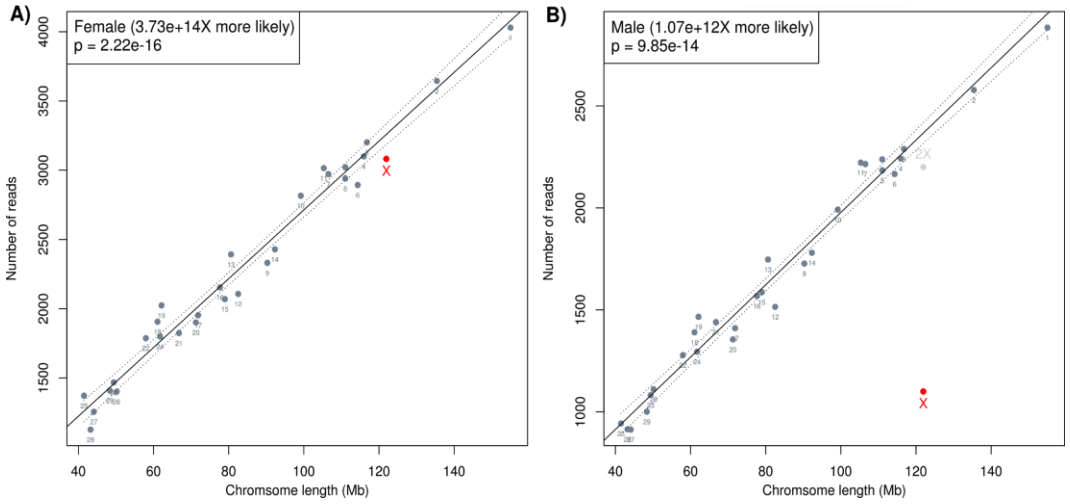


Figure 2.2 - Example plots for sexing based read alignment to the X chromosome. A) demonstrates a typical plot for a sample identified as female while B) is typical for a male sample.

USER treatment

Treatment of ancient DNA with Uracil-DNA glycosylase (UDG) and Endonuclease VIII has been demonstrated to remove misincorporation associated with ancient DNA (Briggs et al. 2010; Rohland et al. 2015). A proprietary-ratio mixture of the two enzymes is known as USER (New England Biolabs). UDG first acts to cleave the Uracil bases that are a result of hydrolytic deamination of cytosine, resulting in an abasic site (Lindahl et al. 1977). As abasic sites can interfere with the activity of certain polymerases (McDonald et al. 2006), endonuclease VIII is required to cleave DNA fragments at abasic sites before amplification steps.

UDG-treated libraries for all *Capra* samples were prepared as above, with an additional step prior to library construction: 5 µl USER (1,000U/ml; Uracil-Specific Excision Reagent, NEB) was added to 16.25 µl purified DNA and incubated in an Eppendorf ThermoMixer for 3 hours at 37°C prior to library construction. In the subsequent Blunt End Repair step, 5 µl less H_2O was used (total reaction volume 70 µl).

Whole genome sequencing

USER-treated libraries of samples selected for whole genome sequencing were amplified (as described in above) using unique index oligos for a total of 6 indexes per lane of sequencing. The number of amplification cycles for each sample was chosen in order to both minimize the number of cycles and obtain the minimum amount of DNA (15 ng) required for sequencing. Amplified product was purified and quantified as described above. The purified product was pooled such that each index was present in equimolar amounts for each lane of sequencing. Pools were then sequenced using a HiSeq 2000 or 2500 platform, single end, read length 1x100bp (Macrogen Inc., 1002, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 153-781, Republic of Korea).

Ancient whole genome data processing

Read quality and potential systematic bias was assessed using FastQC. Read trimming and length filtering was performed using cutadapt 1.1, removing reads less than 30 bp in length (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30).

For samples selected for whole genome sequencing, alignment to CHIR_1.0 (Dong et al. 2013) was performed using bwa aln (Li & Durbin 2009) with seeding disabled (bwa aln -l 1024) (Schubert et al. 2012). Bam files were produced with samtools 0.1.19 (Li & Durbin 2009), with read groups assigned to each unique PCR reaction. Clonal PCR products (PCR duplicates) were removed using samtools rmdup, following which reads with mapping quality less than 30 were removed. Reads from the same sample were merged using the MergeSamFiles option of Picard (The Broad Institute 2018), and duplicates removed again prior to indel realignment using GATK (McKenna et al. 2010).

Damage patterns were assessed using mapDamage2 (Jónsson et al. 2013), and were substantially reduced compared to libraries constructed without USER-treatment (see Figure

2.6 below). To ameliorate remaining damage, bam files were rescaled using mapDamage2, reducing the base qualities of sites likely to be affected by deamination. As a final precaution against damage, bam files were soft-clipped (reducing base quality to 0) by 4 bp at either end of each read. Analyses taking into account base quality scores would therefore ignore the soft-clipped bases.

Mean genome coverage for each individual was calculated using the GATK DepthOfCoverage tool (McKenna et al. 2010). Endogenous content was estimated by dividing the number of reads aligned following the removal of reads with less than 30 mapping quality (0.1% chance of misalignment), divided by the total reads following adaptor trimming and removal of reads <30 bp in length.

Mitochondrial capture

An in-solution bait-and-capture approach (Gnirke et al. 2009; Maricic, Whitten, and Pääbo 2010) was taken, using custom RNA baits designed to target domesticate species (MYcroarray, 5692 Plymouth Road, Ann Arbor, MI 48105, USA). MYbaits v2.3.1 (Mycroarray) capture system was used according to the manufacturer's protocol (O'Sullivan et al. 2016).

For each sample selected for mitochondrial enrichment, five aliquots from USER-treated libraries were PCR amplified, using unique indexes for each amplification and sample combination, according to the protocol described above. After MinElute purification and quantification, samples were pooled such that (i) each sample had an equal amount of endogenous DNA and (ii) there was a total of 2,000 ng of DNA present. This pool was desiccated for 8 hours and then re-suspended in 8.4 µl H2O. RNA baits and blocks were added to the pool as manufacturer's instructions, with a single modification: Block #1 was replaced with an additional 2.5 µl of pooled DNA (total 8.4 µl). Baits and DNA were incubated for 40 hours at 65°C. Captured DNA was recovered using Dynabeads® MyOne™ Strepavidin C1 magnetic beads (ThermoFisher Scientific) and resuspended in 30 µl H2O. 15 µl of the captured DNA was amplified for 14 cycles using KAPA HiFi DNA Polymerase (Kapa Biosystems), according to the Mybaits protocol. Single-end, 70 bp sequencing was performed on an Illumina MiSeq platform (Trinity Genome Sequencing Laboratory, Trinity College Dublin, Ireland).

Mitochondrial alignment and sequence generation

All samples, including those selected for whole genome sequencing and those subject to targeted mitochondrial capture, were aligned to a circularized version of the revised mitochondrial reference NC_005044.2 (Hassanin et al. 2010). Circulatization was performed as the mitochondrial genome itself is circular but is represented by a linear sequence. Reads overlapping the region of the mitochondria that are found at the beginning and end of the linear reference sequence will therefore not align when using bwa aln. The mitochondrial sequence was circularized by concatenating 15bp from either end to the opposite end of the mitogenome, repeating such that each end has been extended. Alignment was performed following identical steps described above.

Consensus fasta sequences were generated using ANGSD (Korneliussen et al. 2014) (angsd -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30). For samples with less than 3× mean coverage, -setMinDepth 1 was used. Sequences were then de-circularized by removing 15bp from each end. An initial maximum likelihood (ML) phylogeny was constructed using a dataset of reference goat and other *Capra* species (Appendix Table 2.3). Sheep was selected as an outgroup. A multi-sequence alignment was generated using MUSCLE (Edgar 2004), and a ML phylogeny constructed using phyML, using Bayesian information criteria (BIC) automated model selection (Guindon et al. 2010). Samples were then assigned to haplogroups according to position within this phylogeny.

Each sample was then realigned to a circularized representative sequence (Appendix Table 2.3) from the appropriate haplogroup, using the pipeline described above, to generate final mitochondrial sequences for each sample. Realignment was performed to maximize reads aligning to the control region, which evolves faster compared to the average mitochondrial rate (Duchêne et al. 2011). This increases the divergence between the chosen reference sequence and the sample of interest, decreasing the number of reads aligning to this particular part of the mitogenome. For poorly preserved or divergent samples this is an issue, and as such all samples were realigned to a closer reference mitochondrion.

Merging of identical samples

Due to the complexity of zooarchaeology assemblages, samples were screened for relatedness or if they were the same individual. Samples which were from petrous bones of

opposite orientations (left and right), and had the same mitochondrial sequence, archaeological identifier, and molecular sex were identified and merged. Shared identity of samples was confirmed later (Chapter 4) with lcMLkin (Lipatov et al. 2015).

C_{14} Radiocarbon dating

For samples selected for radiocarbon dating, ~1g of bone powder was provided to either ^{14}CHRONO Centre (Belfast) or Beta Analytic (Miami) for standard Accelerated Mass Spectronomy estimation of age. Uncalibrated dates were calibrated (2 sigma) using OxCal 4.3 (Bronk Ramsey 1994; Ramsey & Lee 2013) and IntCal 13 (Niu et al. 2013).

Results

Species identification

Of the 254 bone samples screened, 154 produced libraries with sufficient endogenous DNA to determine species identity; results are summarized in Figure 2.3. The assigned species for all samples is presented in Appendix Table 2.1. The majority of those identified were either goat (52) or sheep (81). At this stage, *Capra* species were not differentiated from each other as Fastq Screen was performed using only CHIR_1.0. When combined with goat already identified by other researchers, there was a total of 85 goat samples; after identical samples were combined (see below), 83 were included in subsequent sequencing and analyses (Table 2.1).

There was also a small number of cattle (11), pigs (5), dogs (2), and various ungulates (Appendix Table 2.1). The remaining 100 libraries produced inconclusive Fastq Screen results (Figure 2.1C). These non-*Capra* and poorly preserved libraries were not included in subsequent steps.

Endogenous DNA content

Percentage endogenous DNA content was calculated for each sample, including those screened by other researchers. Values for all samples are displayed in Appendix Table 2.1; a summary of the distribution of endogenous content for goat samples is present in Figure 2.4. The majority of goat samples had endogenous content values less than 1%. 26 samples had an endogenous content of at least 20%, and 15 had endogenous content values of at least 50%.

Archaeological sites which yielded ancient goat DNA are presented in Figure 2.5. A table summarizing the archaeological sites which yielded *Capra* DNA is presented at end of the chapter in Table 2.2.

Damage assessment

Damage rates and read length distributions for all samples identified as *Capra* were plotted and presented in Figures 2.6 and 2.7. Damage patterns were in line with those reported from previous aDNA publications (Librado et al. 2017), with USER-treated libraries showing a ~10-fold reduction in C->T misincorporations.

For samples Direkli4, Direkli5 and Direkli6, no UDG-treated library was prepared due to scarcity of material. For samples previously reported by Amelie Scheu (Scheu 2012), only UDG-treated libraries were prepared as mitochondrial haplogroups assigned here were in concordance with previous work.

Molecular sex identification

Molecular sex could be assigned for 79 *Capra* samples (Table 2.1). In cases where a non-significant p value was obtained, samples were marked as Cannot Determine. 51 of the sexed samples were female, and the remaining 28 males (1.8:1 female:male ratio). The 95% confidence interval for the

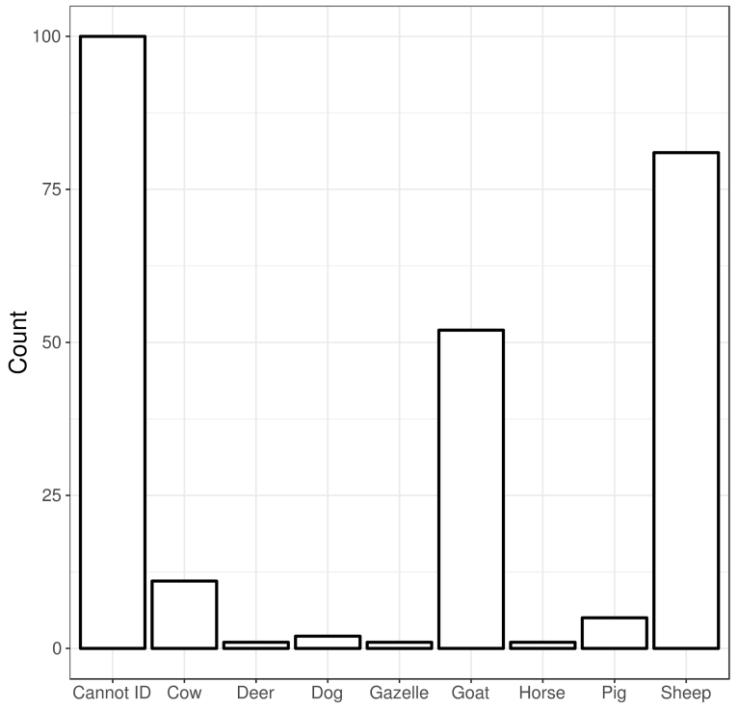


Figure 2.3 - Summary of identified species frequencies. Species identification was performed using Fastq Screen.

proportion of females was calculated as $0.65 \pm 1.96\sqrt{\frac{(1-0.65)}{79}}$ = 0.65 ± 0.13, suggesting a statistically significant bias towards females. When pre-Neolithic individuals were excluded (leaving 49 females and 25 males, a female proportion of 0.66), the female proportion confidence interval was similarly calculated to be 0.66 ± 0.13, supporting a bias towards females in managed goat populations.

Whole genome shotgun sequencing

62 samples were selected for HiSeq shotgun sequencing (before identical samples were merged). As some samples were identified as being from the same individual (see below), they were combined into single individual samples in the following tables. Reads generated, alignment metrics, and coverage statistics are displayed in Appendix Table 2.4. ×-fold autosomal coverage and mitochondrial results are displayed in Table 2.1. 52 individuals produced mean genome-wide coverage of >0.01× (median coverage 1.05×); 27 genomes

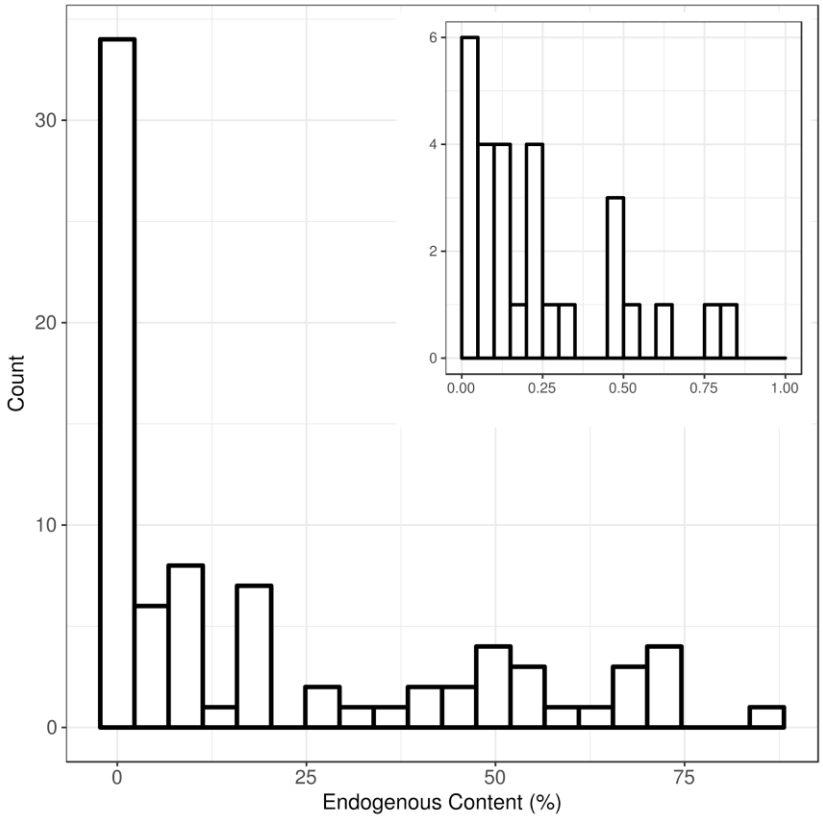


Figure 2.4 - Distribution of endogenous content of samples identified as goat. Inset shows distribution between 0 and 1% endogenous content.

were at least 1× mean coverage. Mitochondrial genomes generated from shotgun sequencing were generally high coverage (median 128.4×) and produced nearly complete mitochondrial sequences (median percentage of called bases = 99.9%). In a small number of cases, mitochondrial data quantity was low (*e.g.* <5× coverage). These samples were included in the mitochondrial enrichment and additionally sequenced. mtDNA alignment results for HiSeq sequenced samples are presented in Appendix Table 2.4 and are differentiated from mtDNA-enriched samples (below) by the absence of raw read count (Appendix Table 2.5).

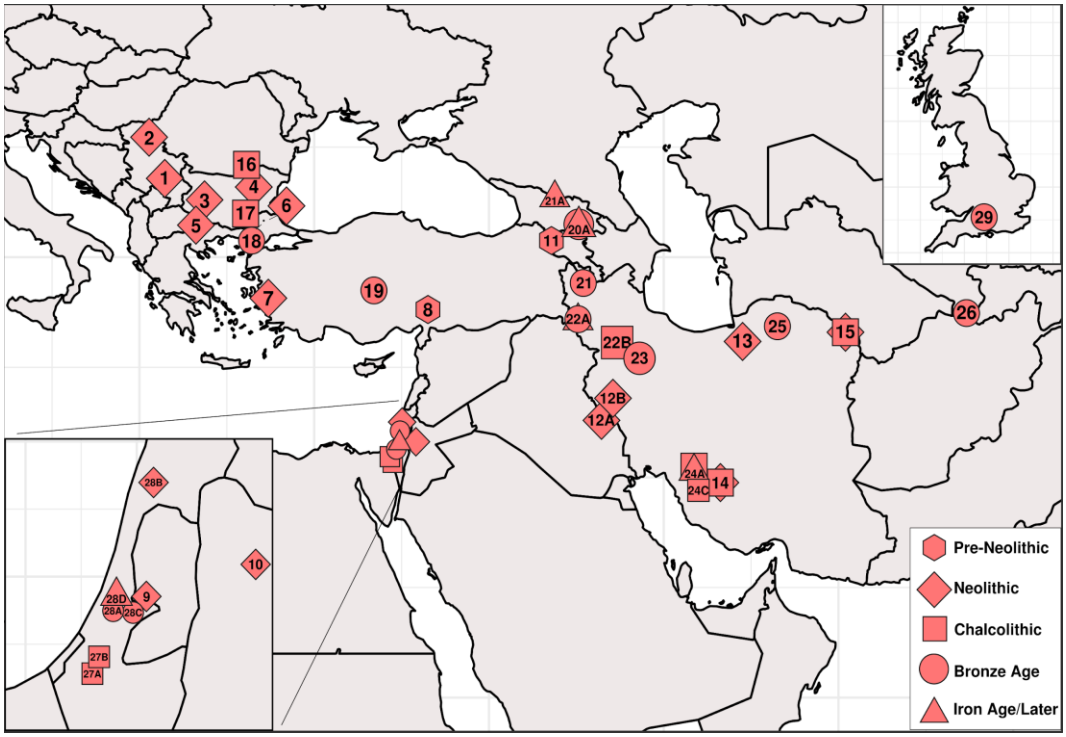


Figure 2.5 - Archaeological sites which yielded ancient goat DNA: 1. Blagotin-Poljna, 2. Uivar, 3. Čavdar, 4. Ovčarovo-gorata, 5. Kovačevo, 6. Aşağı Pınar, 7. Ulucak Höyük, 8. Direkli Cave, 9. Abu Ghosh, 10. 'Ain Ghazal, 11. Hovk-1 Cave, 12A. Kelek Asad Morad, 12B. Tepe Abdul Hosein, 13. Sang-e Chakmaq, 14. and 24B. Rahmat Abad, 15. Monjukli Depe, 16. Pietrele, 17. Drama-Merdžumekja, 18. Kanlıgeçit, 19. Acemhöyük, 20A. Tachti Perda, 20B. Dariali Tamara Fort (Kazbegi), 21. Kohneh Tepesi, 22A. Tepe Hasanlu, 22B. Soha Chay Tepe, 23. Tepe Chizar, 24A. Darre-ye Bolāghi, 24C. Mianroud, 25. Chalow, 26. Tilla Bulak, 27A. Shiqmim, 27B. Gilat, 28A. Tel Yarmuth, 28B. Tel Yoqne'am, 28C. Tel es-Safi/Gath, 28D. Tel Miqne-Ekron, 29. Potterne.

Mitochondrial enrichment

47 samples were selected for mitochondrial enrichment using RNA baits. Sequencing and alignment results are displayed in Appendix Table 2.5; summary mitochondrial results are displayed at end of the chapter in Table 2.1.

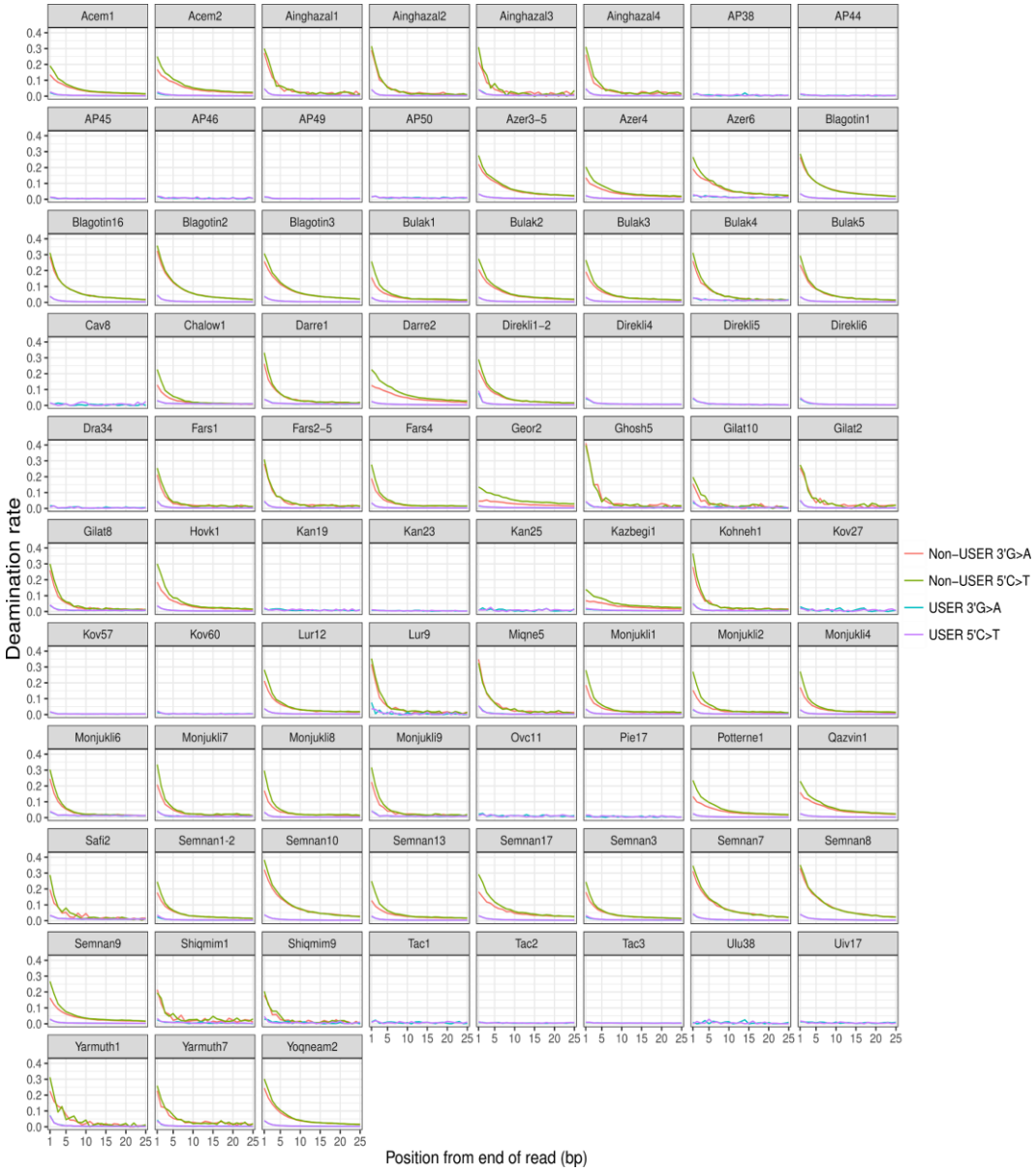


Figure 2.6 - Damage patterns of goat samples, relative to distance to end of read. USER-treated and non-USER-treated libraries are plotted together when available.

Despite the generally low endogenous content of these samples (median 0.66%), mitochondrial genomes were successfully recovered (median 11.1× coverage). Of the mitochondrial sequences generated, 33 had >90% of bases called. 6 mitogenomes had less than 3× mean coverage, while 4 had less than 1× mean coverage.

Haplogroup assignments for each sample is displayed in Table 2.1. The distribution of haplogroups through space and time will be discussed in detail in Chapter 3.

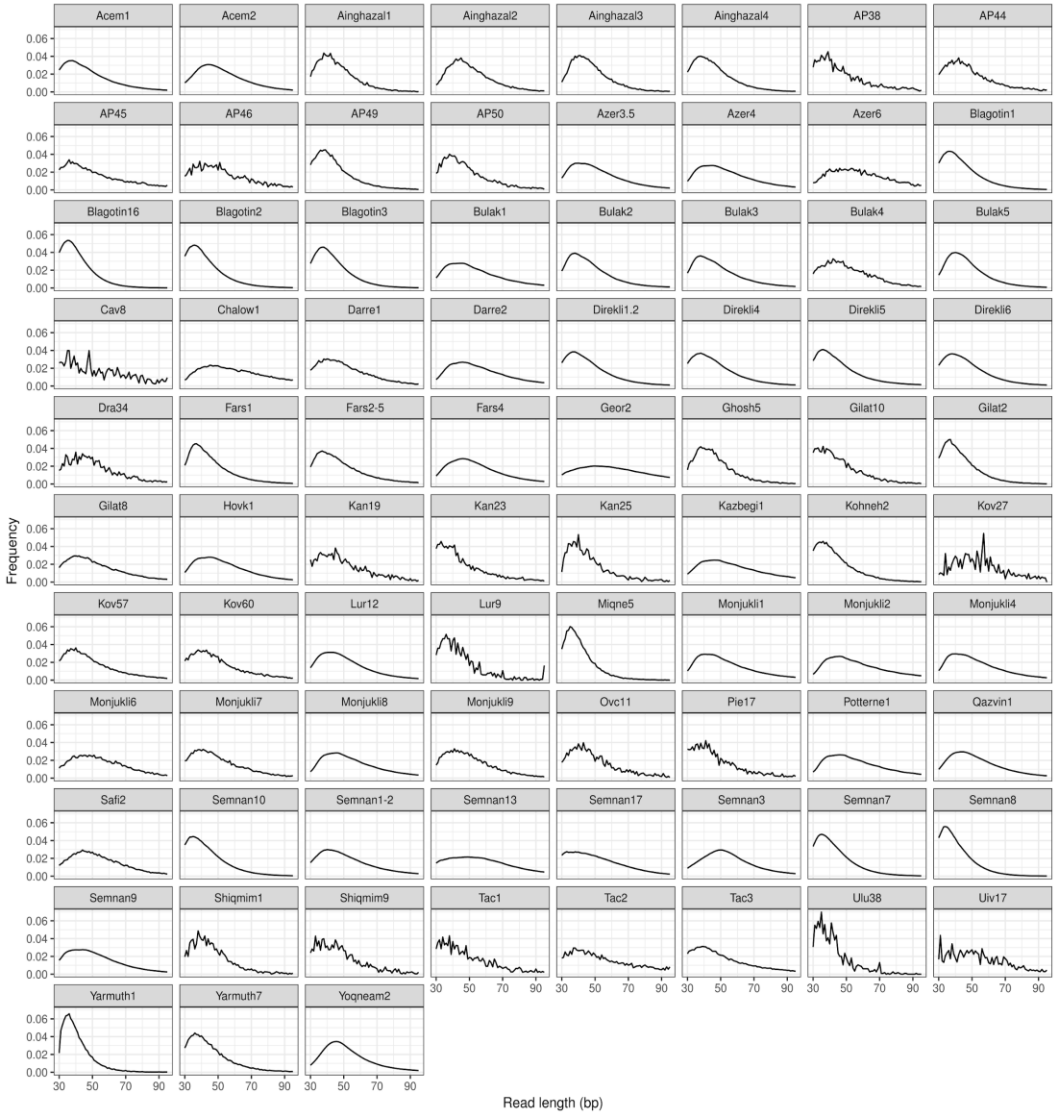


Figure 2.7 - Fragment length distribution of goat samples. USER-treated libraries are plotted when available.

Identical individuals

Three pairs of samples met the defined criteria: Direkli1 and Direkli2, Azer3 and Azer5, Semnan1 and Semnan2; additionally, the pairs had pi-HAT values >0.9 (Chapter 4). These individuals were combined and considered a single individual (Direkli1-2, Azer3-5, Semnan1-2). A fourth pair of samples, Bulak1 and Bulak3, met these criteria but were from petrous bones of the same orientation. One of these samples (Bulak3) was removed from subsequent analyses. A final pair of samples were identified as having met the criteria above (Fars2 and Fars5), but due to low endogenous DNA did not have sufficient coverage to estimate pi-HAT (Chapter 4). These samples were also combined into a single individual, Fars2-5.

C_{14} radiocarbon dating

A total of 13 samples were selected for radiocarbon dating. These were in addition to 4 dated samples previously published by Amelie Scheu (Scheu 2012). Uncalibrated conventional dates, calibrated dates, and C14 Codes are presented in Appendix Table 2.6. When applicable, calibrated ages are displayed in Table 2.1.

In the majority of cases, calibrated age was in line with the age expected from the archaeological context. One sample, Darre2, was assigned a calibrated age (1473-1641 cal AD) inconsistent with the contextually-assigned age (5th millennium BC). As such, Darre2 was considered a medieval sample for all analyses. An additional sample, Hovk1, could not be calibrated (unbounded age distribution of at least >47074 BP). This was in line with the Middle Paleolithic context from which the sample (Pinhasi et al. 2011).

Discussion

Screening success rate

61% of the 254 screen bones produced sufficient DNA for species identification. This is despite the high number of Neolithic and older samples screened (120 and 15, respectively), and the arid environments from which many of the samples derive. A recent published model of DNA fragmentation found that precipitation and thermal fluctuation correlated with fragmentation rates rather than age, while deamination rates were correlated with thermal age (Kistler et al. 2017). Assuming that fragmentation rather than deamination affects endogenous DNA preservation and hence species identification, these screening results are consistent with the proposed model. Bulk loss of DNA to the environment was suggested as an explanation to loss of endogenous DNA. The "semi-closed" system of the non-vascular petrous bone, which was preferentially targeted during screening, likely limits this bulk diffusion to the environment, better preserving endogenous DNA molecules.

Capra samples identified

After accounting for samples that likely derived from the same animal, a total of 83 *Capra* individuals were identified, of which 52 were obtained via screening reported here. This number does not differentiate between *Capra hircus* and *Capra aegagrus*, or indeed any *Capra* species. In contrast, 81 sheep were identified during screening. This bias towards sheep is unsurprising, as the animal assemblages of southwest Asia became dominated by sheep at the PPNC and Pottery Neolithic (Horwitz et al. 1999; Zeder 2008). Sheep and goat are difficult to distinguish morphometrically (Boessneck 1971), and are commonly combined into a "ovicaprid" taxonomic unit during analysis. aDNA analysis presents a means by which the uncertainty in the species composition of zooarchaeological assemblages can be resolved, as relatively little sequencing data is required to differentiate between the two species.

Sex bias in ancient herds

The transition to management of goat populations is expected to coincide with a greater proportion of females, compared with that of a hunted population (Zeder & Hesse 2000). Modern pastoralist herds can have extremely skewed sex ratios, ranging 10-30:1 (Payne 1973), although we should not assume that modern and ancient animal herders practised entirely analogous management strategies. Given that, the sex ratios presented here do show a

bias towards females when restricted to Neolithic-and-later specimen. The maximum of the 95% confidence interval, 0.79, would be consistent with a sex ratio of ~4:1, whereas if the 0.66 proportion was accurate, the ratio would be 2:1. Even if the true proportion is the minimum of the confidence interval, 0.53, this result is consistent with the expectation of a higher proportion of females to males in managed populations.

Prior to management and herding, prime-age male bezoar are thought to be the primary target of Paleolithic hunters (Zeder 2008; Payne 1973). The small sample size of pre-Neolithic goat (5) precludes a statistically-useful female proportion being calculated. Addressing the possibility of sex bias in hunted goat assemblages should be assessed in the future if sufficient samples can be obtained.

Sequencing: mtDNA and nuclear genomes

The high endogenous content of many screened *Capra* samples permitted nuclear data to be generated for 62 individuals. 52 individuals had at least 0.01× nuclear genome average, with a median ×-fold coverage of 1.05× (with or without the removed individual Bulak3). This dataset was used as the basis of the nuclear genome analysis presented in Chapters 4 and 5.

Though 38 of the 82 unique ancient goats identified had endogenous content values less than 5%, mtDNA enrichment using RNA baits successfully generated mtDNA sequences for the large majority of these samples. Although limited by being a single locus, mtDNA are still useful in phylogeographic analysis. This is particularly true in the case of species in which only a small amount of nuclear genome data is available, such as goat. These 82 ancient goats mitogenomes are used as the basis of Chapter 3, which examines the patterns of goat mtDNA diversity across geographical space and time.

Conclusion

Sufficient DNA was recovered from 154 screened bones to identify the likely genus; 52 of these were determined to be *Capra*. When supplemented with *Capra* identified by others (Victoria Mullin, Amelie Scheu, and Andrew Hare) and after identical samples were combined, a total of 83 *Capra* specimen were available for further sequencing. mtDNA (82 following removal of likely identical individuals) and nuclear genomes (63 total; 51 unique genomes with > 0.01×, median coverage of 1.05×) were generated by a combination of shotgun sequencing of non-enriched and mtDNA-enriched libraries. These genomes are analyzed in the succeeding chapters to investigate the patterns of mtDNA (Chapter 3) and nuclear diversity (Chapter 4) before and following the onset of the domestication process. The genetic consequences of domestication as expressed by selection on nuclear variation is examined in Chapter 5.

Table 2.1 - Goat Sample Summary. Radiocarbon-dated ages are indicated with *; ages are otherwise a contextual estimate. Bulak3 was excluded from analyses due to high similarity with Bulak1. Mean ×-fold coverage is presented for nuclear and mitochondrial genomes.

Sample	Site	Region	Context	Age	Endog (%)	Sex	Nuclear cov	mtDNA cov	mtDNA hap
Acem1	Acemhöyük	Aksaray Plain, Turkey	Bronze Age	*2,346-2,040 cal BC	56.2	F	4.76	411.03	A
Acem2	Acemhöyük	Aksaray Plain, Turkey	Middle Bronze Age	~1,700 BC	73.76	M	8.67	849.8	A
Ainghazal1	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	2.22	M	0.03	5.46	F
Ainghazal2	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	2.58	F	0.06	12.78	F
Ainghazal3	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	0.21	F	<0.01	4.25	F
Ainghazal4	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	0.65	M	0.01	65.44	F
AP38	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	*5,468-5,316 cal BC	0.5	F	-	7.61	C
AP44	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,500-5,000 BC	0.8	F	-	3.57	A
AP45	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,500-5,000 BC	1.48	F	0.02	6.47	A
AP46	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	*5,293-5,057 cal BC	0.28	F	-	9.77	C
AP49	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,500-5,200 BC	2.02	F	0.02	12.69	A

AP50	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,300-5,000 BC	0.47	F	-	14.31	A
Azer3-5	Tepe Hasanlu	Azerbaijan, Iran	Early Bronze Age	3,000-2,100 BC	57.81	F	4.66	487.2	A
Azer4	Tepe Hasanlu	Azerbaijan, Iran	Iron Age	550-330 BC	63.64	M	2.57	309.59	A
Azer6	Soha Chay Tepe	Azerbaijan, Iran	Chalcolithic	~4,200 BC	19.13	F	0.28	44	A
Blagotin1	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,398-6,098 cal BC	50.79	M	6.99	544.63	A
Blagotin2	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,379-6,078 cal BC	38.81	F	4.02	253.67	A
Blagotin3	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,096-5,892 cal BC	66.45	M	11.47	885.45	A
Blagotin16	Blagotin-Poljna	Trstenik, Serbia	Neolithic	~6,100 cal BC	46.04	M	3.51	296.4	A
Bulak1	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	51.16	F	0.87	220.47	A
Bulak2	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	49.47	M	2.67	284.97	A
Bulak3	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	39.83	F	0.61	180.39	A
Bulak4	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	5.60	M	-	60.58	B
Bulak5	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	18.13	F	0.27	113.99	D
Cav8	Čavdar	Sofia District, Bulgaria	Neolithic	6,000-5,500 BC	0.03	F	-	0.58	A
Chalow1	Chalow	Khorasan, Iran	Bronze Age	2,300-2,000 BC	6.66	M	0.05	341.05	D

Darre1	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	5th mil. BC	7.1	F	0.04	95.66	A
Darre2	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	*1,473-1,641 cal AD	70.31	F	3.93	365.56	A
Direkli1-2	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	*11,351-11,166 cal BC	49.44	F	11.55	998.17	T
Direkli4	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	*12,191-11,882 cal BC	25.89	M	-	142.69	F
Direkli5	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	~9,500 cal BC	7.08	M	0.27	120.94	T
Direkli6	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	~9,500 cal BC	16.05	M	1.93	523.8	T
Dra34	Merdžumekja	Drama, Bulgaria	Chalcolithic	*4,580-4,354 cal BC	0.11	F	-	5.44	G
Fars1	Rahmat Abad	Fars, Iran	Early Chalcolithic	~4,600 BC	1.13	M	0.02	40.38	A
Fars2-5	Rahmat Abad	Fars, Iran	Pottery Neolithic	*7,047-6,772 cal BC	1.8	M	0.03	284.47	B
Fars4	Mianrud	Fars, Iran	Early / Mid Chalcolithic	*5,460-5,211 cal BC	16.19	F	1.05	109.24	A
Geor2	Tamara Fort	Kazbegi, Georgia	Medieval	11th-15th Century AD	68.31	M	1.5	157.36	A
Ghosh5	Abu Ghosh	Judean Hills, Israel	Middle Pre-Pottery Neolithic B	Late 9th-Early 8th mil. BC	0.04	F	<0.01	6.25	F
Gilat2	Gilat	Northern Negev, Israel	Chalcolithic	4,500-4,200 BC	0.34	M	<0.01	16.58	A
Gilat8	Gilat	Northern Negev, Israel	Chalcolithic	4,500-4,200 BC	0.83	M	0.02	133.77	D

Gilat10	Gilat	Northern Negev, Israel	Chalcolithic	4,500-4,200 BC	0.15	F	<0.01	2.05	A
Hovk1	Hovk-1 Cave	Tavush, Armenia	Paleolithic	*>47,074 BP	29.72	F	3.08	519.39	F
Kan19	Kanlıgeçit	Kırklareli, Turkey	Bronze Age	2,700-2,200 BC	0.06	C.D.	-	8.55	A
Kan23	Kanlıgeçit	Kırklareli, Turkey	Bronze Age	*2,833-2,465 cal BC	0.5	M	-	10.56	G
Kan25	Kanlıgeçit	Kırklareli, Turkey	Bronze Age	2,700-2,200 BC	0.04	C.D.	-	8.97	A
Kazbeg1	Tamara Fort	Kazbegi, Georgia	Medieval	10th Century AD	73.94	F	3.84	256.38	A
Kohneh2	Kohneh Tepesi	Azerbaijan, Iran	Early Bronze Age	3,300-3,000 BC	6.81	F	0.04	29.56	A
Kov27	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	6,200-5,600 BC	0.02	C.D.	-	3.12	A
Kov57	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	6,200-5,600 BC	5.74	F	0.07	5.86	A
Kov60	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	6,200-5,600 BC	0.52	F	-	6.97	A
Lur9	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	8,500-8,200 cal BC	0.03	F	<0.01	1.7	B
Lur12	Tepe Abdul Hosein	Luristan, Iran	Pre-Pottery Neolithic	*8,171-7,745 cal BC	8.42	F	1.05	480	G
Miqne5	Tel Miqne- Ekron	Shephelah, Israel	Iron Age	End of 8th Century BC	0.15	M	<0.01	12.22	A
Monjukli1	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	7.87	F	0.24	42.51	A
Monjukli2	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	8.18	F	0.21	96.78	D

Monjukli4	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	20.08	F	0.6	90.55	A
Monjukli6	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	5.84	M	0.03	123.04	D
Monjukli7	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic	~6,400-5,900 cal BC	3.18	F	-	31.87	D
Monjukli8	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic	~6,400-5,900 cal BC	17.43	M	2.57	411.74	D
Monjukli9	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic	~6,400-5,900 cal BC	1.99	M	-	33.02	G
Ovc11	Ovčarovo-gorata	Tărgovište, Bulgaria	Early Neolithic	5,700-5,500 BC	0.09	F	-	8.63	A
Pie17	Pietrele	Giurgiu, Romania	Chalcolithic	4,450-4,250 BC	0.18	F	-	6.98	A
Potterne1	Potterne	Wiltshire, UK	Bronze Age	2,040-990 BC	71.12	F	3.67	217.21	A
Qazvin1	Tepe Chizar	Qazvin, Iran	Middle Bronze Age	2,400-1,900 BC	54.71	F	3.16	448.01	A
Safi2	Tel es-Safi	Ashkelon, Israel	Early Bronze Age	2,570-2,900 BC	3.28	F	0.04	134.14	A
Semnan1-2	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic	*7,454-6,850 cal BC	27.31	F	6.85	533.55	B
Semnan3	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	*6,214-6,004 cal BC	66.23	F	14.89	887.12	D
Semnan7	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	85.89	M	3.28	204.62	D
Semnan8	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	9.91	F	0.21	76.45	D

Semnan9	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	53.27	M	3.05	394.12	G
Semnan10	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	19.17	M	1.43	369.08	G
Semnan13	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	34.70	F	2.54	109.05	D
Semnan17	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	7.37	F	0.12	149.93	D
Shiqmim1	Shiqmim	Northern Negev, Israel	Chalcolithic	4,300-3,700 BC	0.17	F	<0.01	7	D
Shiqmim9	Shiqmim	Northern Negev, Israel	Chalcolithic	4,300-3,700 BC	0.10	M	<0.01	0.85	D
Tac1	Tachti Perda	Kakheti, Georgia	Late Bronze Age	1,400-1,000 BC	0.09	F	-	11.45	A
Tac2	Tachti Perda	Kakheti, Georgia	Iron Age	1,000-700 BC	0.24	F	-	12.39	A
Tac3	Tachti Perda	Kakheti, Georgia	Late Bronze Age	1,400-1,000 BC	13.58	F	0.13	25.27	A
Uiv17	Uivar	Timişoara, Romania	Neolithic	5,250-5,050 BC	0.1	F	-	0.95	A
Ulu38	Ulucak Höyük	Turkey	Early Neolithic	6,400-6,100 BC	0.03	C.D.	-	0.46	A
Yarmut1	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	~2,700-2,500 BC	0.23	M	<0.01	40.41	A
Yarmut7	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	2,650-2,200 BC	0.14	F	<0.01	11.1	A
Yoqneam2	Tel Yoqne'am	Haifa, Israel	Middle Bronze Age	~1650-1550/ 1540 BC	45.66	F	2.2	472.98	A

Table 2.2 - Site Information for those which yielded *Capra* DNA. Latitude and Longitude coordinates are presented with five decimal points. Map label refers to Figure 2.5.

Site Name	Location	Latitude	Longitude	Period	Map Label
Blagotin-Poljna	Trstenik, Serbia	43.72329	21.09945	Neolithic	1
Uivar	Timişoara, Romania	45.66180	20.90320	Neolithic	2
Čavdar	Sofia District, Bulgaria	42.66020	24.05580	Neolithic	3
Ovčarovo-gorata	Tărgovište, Bulgaria	43.11410	26.38240	Neolithic	4
Kovačevo	Blagoevgrad, Bulgaria	41.50000	23.48330	Neolithic	5
Aşağı Pınar	Kırklareli, Turkey	41.72167	27.22528	Neolithic	6
Ulucak Höyük	Turkey	38.46675	27.35212	Neolithic	7
Direkli Cave	Taurus Mountains, Turkey	37.55597	36.63767	Late Epipaleolithic	8
Abu Ghosh	Judean Hills, Israel	31.80480	35.11240	Mid PPNB	9
'Ain Ghazal	Amman, Jordan	31.99351	35.98115	Middle PPNB	10
Hovk-1 Cave	Tavush, Armenia	40.82510	45.05850	Late Pleistocene	11
Sang-e Chakmaq	Semnan, Iran	36.50416	55.00079	Neolithic	13
Monjukli Depe	Meana-Čaača, Turkmenistan	36.84850	60.41810	Neolithic - Chalcolithic	15
Pietrele	Giurgiu, Romania	44.05947	26.11947	Chalcolithic	16
Merdžumekja	Drama, Bulgaria	42.23330	26.43330	Chalcolithic	17
Kanlıgeçit	Kırklareli, Turkey	41.71981	27.22049	Bronze Age	18
Acemhöyük	Aksaray Plain, Turkey	38.54098	33.77123	Bronze Age	19
Kohneh Tepesi	Azerbaijan, Iran	38.89956	46.61242	Bronze Age	21
Tepe Chizar	Qazvin, Iran	35.78300	49.41600	Bronze Age	23
Chalow	Khorasan, Iran	37.10355	56.88528	Bronze Age	25
Tilla Bulak	Surkhandarja, Uzbekistan	37.70000	66.80000	Bronze Age	26
Potterne	Wiltshire, UK	51.32439	-2.00891	Bronze Age	29
Kelek Asad Morad	Luristan, Iran	33.15099	47.50848	Neolithic	12A

Tepe Abdul Hosein	Luristan, Iran	34.05000	48.13333	Neolithic	12B
Rahmat Abad	Fars, Iran	30.11208	53.05775	Neolithic - Chalcolithic	14, 24B
Tel Miqne-Ekron	Shephelah, Israel	31.77889	34.84992	Iron Age	18D
Tachti Perda	Kakheti, Georgia	41.47181	46.01196	Bronze Age - Iron Age	20A
Tamara Fort	Kazbegi, Georgia	42.74611	44.62349	Medieval	20B
Tepe Hasanlu	Azerbaijan, Iran	37.00483	45.45889	Bronze Age - Iron Age	22A
Soha Chay Tepe	Azerbaijan, Iran	36.31795	48.37936	Chalcolithic	22B
Darre-ye Bolāghi	Fars, Iran	30.20330	53.17900	Chalcolithic, Medieval	24A
Mianroud	Fars, Iran	29.88282	52.80988	Chalcolithic	24C
Shiqmim	Northern Negev, Israel	31.19464	34.62707	Chalcolithic	27A
Gilat	Northern Negev, Israel	31.32465	34.65259	Chalcolithic	27B
Tel Yarmuth	Bet Shemesh, Israel	31.70799	34.97528	Bronze Age	28A
Tel Yoqne’am	Haifa, Israel	32.66410	35.10830	Bronze Age	28B
Tel es-Safi	Ashkelon, Israel	31.69972	34.84694	Bronze Age	28C

Chapter 2 References

Andrews, S., 2010. *FastQC*, Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Boessneck, J., 1971. *Osteological differences between sheep (Ovis aries Linné) and goat (Capra hircus Linné)*, Thames and Hudson.

Briggs, A.W. et al., 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic acids research*, 38(6), p.e87.

Bronk Ramsey, C., 1994. Analysis of chronological information and radiocarbon calibration: the program OxCal. *Archaeological Computing Newsletter*, 41(11), p.e16.

Dong, Y. et al., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nature biotechnology*, 31(2), pp.135–141.

Duchêne, S. et al., 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PloS one*, 6(11), p.e27138.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–1797.

Esser, K.-H., Marx, W.H. & Lisowsky, T., 2006. DNA decontamination: DNA-ExitusPlus in comparison with conventional reagents. *Nature Methods | Application Notes*. Available at: http://dx.doi.org/10.1038/nmeth853.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–321.

Hassanin, A. et al., 2010. Comparisons between mitochondrial genomes of domestic goat (Capra hircus) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA*, 21, pp.68–76.

Helmer, D. et al., 2005. The upper Euphrates-Tigris basin: Cradle of agro-pastoralism? In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 86–124.

Horwitz, L.K. et al., 1999. Animal Domestication in the southern Levant. *Paléorient*, 25(2), pp.63–80.

Jónsson, H. et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , 29(13), pp.1682–1684.

Kistler, L. et al., 2017. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic acids research*, 45(11), pp.6310–6320.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next

Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), p.R25.

Librado, P. et al., 2017. Ancient genomic changes associated with domestication of the horse. *Science*, 356(6336), pp.442–445.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Lindahl, T. et al., 1977. DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. *The Journal of biological chemistry*, 252(10), pp.3286–3294.

Lipatov, M. et al., 2015. Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data. *bioRxiv*, p.023374. Available at: http://biorxiv.org/content/early/2015/07/29/023374.

MacHugh, D.E. et al., 2000. The extraction and analysis of ancient DNA from bone and teeth: a survey of current methodologies. *Ancient biomolecules*, 3(2), pp.81–103.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), pp.10–12.

McDonald, J.P. et al., 2006. Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic acids research*, 34(4), pp.1102–1111.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Meyer, M. & Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor protocols*, 2010(6), p.db.prot5448–pdb.prot5448.

Niu, M. et al., 2013. The Bayesian Approach to Radiocarbon Calibration Curve Estimation: The IntCal13, Marine13, and SHCal13 Methodologies. *Radiocarbon*, 55(4), pp.1905–1922.

Pääbo, S. et al., 2004. Genetic analyses from ancient DNA. *Annual review of genetics*, 38, pp.645–679.

Park, S.D.E. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome biology*, 16, p.234.

Payne, S., 1973. Kill-off Patterns in Sheep and Goats: the Mandibles from Aşvan Kale. *Anatolian Studies*, 23, pp.281–303.

Pinhasi, R. et al., 2011. Middle Palaeolithic human occupation of the high altitude region of Hovk-1, Armenia. *Quaternary science reviews*, 30(27), pp.3846–3857.

Ramsey, C.B. & Lee, S., 2013. Recent and Planned Developments of the Program

OxCal. *Radiocarbon*, 55(2–3), pp.720–730.

R Core Team, 2016. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Reed, C.A., 1960. A review of the archaeological evidence on animal domestication in the prehistoric Near East. In R. J. Braidwood & B. Howe, eds. *Prehistoric Investigations in Iraqi Kurdistan*. Studies in Ancient Oriental Civilization. Chicago: University of Chicago Press, pp. 119–146.

Rohland, N. et al., 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130624.

Scheu, A., 2012. *Palaeogenetische Studien zur Populationsgeschichte von Rind und Ziege mit einem Schwerpunkt auf dem Neolithikum in Südosteuropa*. Johannes Gutenberg-Universität Mainz. Available at: https://publications.ub.uni-mainz.de/theses/frontdoor.php?source_opus=18041&la=en.

Schubert, M. et al., 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics*, 13, p.178.

The Broad Institute, 2018. *Picard Tools*, The Broad Institute. Available at: https://broadinstitute.github.io/picard/.

Wingett, S.W. & Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, 7. Available at: https://f1000research.com/articles/7-1338/v1/pdf.

Yang, D.Y. et al., 1998. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American journal of physical anthropology*, 105(4), pp.539–543.

Zeder, M., 2008. Animal domestication in the Zagros: an update and directions for future research. *Travaux de la Maison de l'Orient et de la Méditerranée*, 49(1), pp.243–277.

Zeder, M.A. & Hesse, B., 2000. The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* , 287, pp.2254–2257.

Zeder, M.A. & Lapham, H.A., 2010. Assessing the reliability of criteria used to identify postcranial bones in sheep, Ovis, and goats, Capra. *Journal of archaeological science*, 37, pp.2887–2905.

Chapter 3: Ancient goat mtDNA

Introduction

Modern goat mtDNA phylogeny

The mitochondrial phylogeny of modern domestic goat is well established. Using D-loop sequences from 406 goat, Luikart et al demonstrated three mtDNA clades in domestic goat, to which the bezoar ibex (*Capra aegagrus*) was closest among wild caprids (Luikart et al. 2001). Haplogroup A was by far the most frequent of these domestic clades, with haplogroup B and C occurring at low frequencies in eastern Asia and Europe respectively. Despite this, phylogeographic structure was weak: 10% of variation was explained by intercontinental variation, and ~79% of variation within breeds. This distribution of haplogroups, together with the predomestic time of divergence of the haplogroups (~200,000-280,000 BP) and differences in estimated time of expansions of the haplogroups themselves were taken as evidence of distinct domestication histories. This was suggested to be a result of later admixture from the wild or distinct domestications. The alternative explanation of a large domestic ancestral population size containing multiple divergent lineages was also suggested, but ultimately rejected. Notably, goats were sampled from remote areas and not from centres of transportation/commerce; thus, the weak phylogeographic structure observed was considered to be indicative of a long history of movement by humans, more so than other livestock.

The mtDNA phylogeny of goat was later more fully resolved by a combined analysis of 2,430 control region sequences (Naderi et al. 2007). This expanded the canonical domestic haplogroups to include D (found in Asia and north Europe (Joshi et al. 2004)), F (reported only in three Sicilian goat (Sardina et al. 2006)), and G (found in north Africa and the Middle East). Again, the worldwide frequency of haplogroup A was >90%, and phylogeographic structure was weak (12% variation among regions, 77% within breed). A haplogroup-based estimate of population growth using pairwise mismatch distributions was again consistent with expansion since domestication, although with wide confidence intervals on the timing.

A vital element missing from these studies was wild bezoar ibex mtDNA; phylogeographic analyses were severely limited by the absence of wild haplotypes and their distributions. This was addressed by Naderi et al, which produced 473 bezoar mtDNA control region sequences

across their current habitat (Naderi et al. 2008). Domestic goat sequences were found to fall within the diversity of bezoar, which also showed many unique clades not present in goat. Phylogeographic structure was observed to be weak (~42% of variation within population, ~43% between populations within regions, and ~15% between regions), with the caveat that gene flow from domestic herds back into the wild may have reduced past genetic structure. Intriguingly, the most commonly observed haplogroup in sampled bezoar was C (~39%) rather than A (~6%). Domestic goat C sequences were more similar to bezoar sequences from eastern Turkey rather than Iran, where they were more common in bezoar. In contrast, Haplogroup A was observed in eastern Turkey and eastern Iran, but not in the Zagros Mountains of western Iran. This was taken as evidence that eastern Turkey was the source of both haplogroups, and the majority of the present-day domestic mtDNA gene pool. Minor domestic haplogroups (*i.e.* non-A) may have been incorporated in small-scale domestications or wild restocking.

It is important to note the selective use of human-induced migration as an explanation by Naderi et al; the discovery of bezoar haplogroup A in eastern Iran is argued as being due to admixture from domestics into the wild. The presence of haplogroup C in eastern Turkey, far from their highest density in the Central Zagros and Iranian Plateau, is explained as being due to movement of managed wild goat at the earliest stages of domestication. However, their data also fit an explanation that the bezoar haplogroup C in eastern Turkey is a result of recent wild-domestic gene flow, and that haplogroup A was transferred from eastern Iran to eastern Turkey. In addition, their detection of population expansion ~8,000 BC in the bezoar C haplogroup, but not in other haplogroups, was suggested to be a signal of early management of goat in the Zagros Mountains; they did not account for the possibility of domestic-to-wild gene flow, which would result in a similar signal. The arguments made by Naderi et al may indeed be correct, as they better fit the current understanding of early goat herding, but highlight the need for temporal datasets that circumvent recent population movement and admixture.

These surveys of wild and domestic mtDNA variability relied on the D-loop of the mitochondria, just a few hundred base pairs of the mitogenome. This fast-mutating region can be affected by homoplasy (mutations shared by chance rather than common origin) and mutation rate heterogeneity, which can confound analyses (McCracken & Sorenson 2005). This issue can be ameliorated by the use of whole mitochondrial sequences. Several early goat mitogenomes were in fact affected by sequencing errors and incorporation of numts -

nuclear sequences of mitochondrial origin (Hassanin et al. 2010). To address this, Licia Colli and colleagues generated a dataset of 76 domestic goat and 6 bezoar mitogenome sequences (Colli et al. 2015). There was perfect agreement between haplogroups assigned by control region sequences and mitogenome assignment. The bezoar representatives of each haplogroup fell as the outgroup of all domestic sequences, although none were sequenced for bezoar haplogroup A. All domestic clades independently coalesce *c*. 7,000-12,000 BC, using an estimated coding region synonymous mutation rate of 7.77×10^{-8} mutations per year, but older age estimates resulted when the entire molecule was used (3.95×10^{-8} mutations per year).

This work provided a substantial resource for goat mtDNA studies, albeit with some technical concerns. Mutation rates were calibrated using a fossil-based date of divergence between sheep and goat of 5-7 mya (Carroll 1988), which the authors acknowledged would likely be an underestimate of the divergence time, biasing mutation rate estimates downwards (Ho & Larson 2006). The time-dependent nature of the mutation rate can also introduce error when palaeontological calibration is employed (Ho et al. 2005; Ho et al. 2011). If the non-A domestic haplogroups did coalesce at a more recent non-overlapping time than haplogroup A, this would support a domestication of a population in which A was frequent, and with later assimilation (by independent domestication or herd restocking) of the other clades. However, calibration error and high variability of coalescence times severely limits confidence in such assertions. Additionally their use of a generation time of 4.5 years to scale a Bayesian Skyline Plot - showing population expansion 12-10 kya - is puzzling, as their citation supporting its use does not mention goat or 4.5 years (Thirstrup et al. 2009). The authors also suggest an equivalence between each domestic clade and a distinct population that underwent domestication; the older divergence time of haplogroup C from the ABGD clades was taken as supporting a "concomitant secondary domestication", similar to the argument of (Naderi et al. 2008).

As well as global analyses, there have also been many region-level assessments of goat mtDNA variability and phylogeny. These include studies of goat in populations in Portugal, Spain, Turkey, Pakistan, India, China, East Asia, Kenya, Nigeria, Ethiopia, Burkina Faso, Morocco, and Oman (Pereira et al. 2005; Azor et al. 2005; Akis, Oztabak, et al. 2014; Sultana et al. 2003; Joshi et al. 2004; Zhao et al. 2014; E et al. 2018; Lin et al. 2013; Kibegwa et al. 2016; Awotunde et al. 2015; Tarekegn et al. 2018; Royo et al. 2009; Pereira et al. 2009; Al-

Araimi et al. 2017). The consistent pattern emerging is high mtDNA diversity, a preponderance of haplogroup A, and low maternal geographic structure (~70-80% variability is explained within breeds or populations across studies). The general hypothesis as to the lack of observed structure is recent or historic movement by humans: with goat as subsistence during human migration, or as a commercial resource. In the absence of ancient data, it is difficult to discern whether the lack of structure is indeed a recent phenomenon, or the vestigial signal of extensive mixing during the early stages of domestication; an extremely large and diverse maternal domestic population would also be consistent with the data.

mtDNA of ancient goat

Several papers have indeed attempted to address dynamics of goat domestication using ancient mtDNA. The presence of domestic goat, bezoar ibex, and Nubian ibex was inferred throughout the occupations of Abu Ghosh, a Neolithic (PPNB and Pottery) site in Israel, based on the phylogeny of amplified mtDNA fragments (Bar-Gal et al. 2002). Similarly, PCR amplification was used to molecularly identify the presence of goat in the PPNA level of Hatoula, Israel (Bar-Gal et al. 2003).

The first haplogroup-level identification of ancient goat mtDNA was reported from Chalcolithic settlements of the Qazvin Plain, in north-central Iran, where three haplogroup A sequences were detected (Fernandez et al. 2005; Luikart et al. 2006). More significantly, haplogroup A and C mtDNA was identified in remains from Baume d'Oullen, an early Neolithic site in southern France (*c.* 5,300-4,900 BC) (Fernández et al. 2006). The early presence of both A and C haplogroups in domestic goat, representative of those introduced to southern Europe via the Mediterranean strongly suggests that the two haplogroups were present in goat herds prior to the introduction of farming from Turkey (Lazaridis et al. 2014; Hofmanová et al. 2016), or as part of successive waves of Anatolian migration. The presence of haplogroup A and C in Late Neolithic samples from western Turkey would support this interpretation (Scheu et al. 2012).

The absence of other haplogroups was taken by Fernández (2006) as evidence of a temporally- and geographically-distinct domestication history of A and C from those now in Europe and common elsewhere (*i.e.* B, D, G and F). One complexity to add to this was the detection of haplogroup G in Chalcolithic Bulgaria and Bronze Age Turkey (Scheu et al.

2012). Based on modern bezoar haplogroup distributions, G does not co-occur with A and C in eastern Turkey, but does so with C and D in the Zagros Mountains (Naderi et al. 2008).

In China, ancient mtDNA recovered from Bronze Age (*c.* 500 BC) goat in Inner Mongolia show the presence of haplogroups A (7/10), B (2/10), and D (1/10) (Han et al. 2010). The ancient B haplogroup sequences were identified as "founder" haplogroups, based on network analysis and sharing with a large number of modern Chinese goat, and as such China was suggested as being the origin of the B haplogroup. This ignored the fact that B haplogroup mtDNA was found in Iranian bezoar (Naderi et al. 2008), and *Capra aegagrus* is not found in China (de Smet et al. 2008). Nevertheless, the geographic partitioning of the C haplogroup in Europe, and both B and D in China is suggestive of distinct ancestries of goat in these continental regions. The discovery of the B haplogroup from Neolithic goatskin leggings found the Swiss Alps (Schlumbaum et al. 2010) does challenge this, or implies unresolved complexity in early goat mtDNA distribution.

Other attempts at recovering ancient or historic mtDNA from Europe have consistently indicated persistence of the high frequency of haplogroup A through time. Mitochondria recovered from the leggings and coat of the Tyrolean Iceman Ötzi, a Copper Age (*c.* 5,300 BP) natural mummy from the Italian Alps, derive from haplogroup A goat (O'Sullivan et al. 2016). In Corsica, 21 sequences obtained from 12th and 14th century AD medieval goat remains all were of the A haplogroup, and showed some affinity to modern sequences from the island (Hughes et al. 2012). More recently, mtDNA obtained from taxidermic samples from Britain and Ireland dating across the last 200 years were of the A haplogroup (Cassidy et al. 2017). Mitochondria from historical samples tended to cluster together, along with several modern unimproved feral goat from Mulranny, Ireland, possibly reflecting shared ancestry between the islands prior to improvement by Swiss breed introgression (Porter & Tebbit 1996).

These ancient samples from the peripheries of the goat distribution, although interesting, do not address the region of greatest importance to their domestication: southwest Asia. Tracing the appearance of haplogroups through early domestic sites in the Near East could give some indication of where the earliest goat populations derived from and their interactions with the wild.

Haplogroup A has been reported from the early Neolithic (middle 9th - early 8th millennium BC) site of Chia Sabz in the Central Zagros Mountains of Iran (Mazdarani et al. 2014). This is surprising, given the notable absence of haplogroup A in the wild goat of Iran (Naderi et al. 2008). This particular observation would suggest that the bezoar ibex of the Zagros Mountains, where the earliest robust evidence for management is found (Zeder & Hesse 2000), contributed the bulk of the modern domestic mtDNA gene pool. However, there are substantive technical concerns about the methodology of Mazdarani et al. (2014), most notably the length of the region targeted for amplification. The authors report the use of primers used for modern amplification (Naderi et al. 2008), and also primers used to amplify ancient sheep mtDNA, which in the original study generated fragments ~120-180 bp in length (Ann Horsburgh & Rhines 2010). In contrast, the Chia Sabz sequences ranged from 534-632 bp. Overlapping PCR primers can generate ancient sequence data from longer targeted regions, but as the average mtDNA fragment is ~50 bp in length (Hansen et al. 2017), this amplification strategy effectively enriched for modern contamination, inviting skepticism towards the study's findings.

There have been more credible aDNA studies performed on goat remains from the Near East. Intriguingly, haplogroup A has been reported from five goat remains from two Pottery Neolithic (*c.* 6,000-5,500 cal BC) sites in western Azerbaijan (Kadowaki et al. 2016). As these two sites represent some of the earliest agricultural sites in the southern Caucasus region, the early high frequency of A in domestic goat close to the region the domestication likely began in is evidence in favour of an eastern Turkish origin of haplogroup A (Naderi et al. 2008). Seven successfully-amplified control region sequences from Van-Yoncatepe, eastern Turkey, indicate the presence of haplogroup A in the Early Iron Age *c.* 1,000 BC (Akis, Onar, et al. 2014). Bronze and Iron Age samples from the western Turkey and southern Caucasus also possess the A haplogroup (Scheu et al. 2012). In contrast to the >500bp region amplified in Mazdarani et al. 2014, these studies amplified just ~100-200 bp of the control region.

The impression from the mtDNA landscape of ancient goat is that the patterns observed in the present were established early in the domestication process. The lack of phylogeographic structure within ancient haplogroup A (and the minor haplogroups) suggest that either the goat incorporated into domestic herds were from a very large population, or extensive gene flow occurred in the first few millennia following domestication. Accepting the usual caveats

that apply to mtDNA, these data present a useful starting pointing in generating hypotheses that can be addressed with additional ancient mtDNA sequences and whole genomes.

Aims

The aims of this chapter are:

1. To explore the geographic distribution of goat mitochondrial haplogroups during the earliest phase of domestication.
2. To investigate how this distribution may have changed through time.
3. To utilize radiocarbon-dated samples to calibrate the goat mtDNA phylogeny.

Materials and Methods

Materials

Ancient data

82 *Capra* specimens identified in Chapter 2 were selected for mitochondrial analysis. Sample identifiers, geographic and contextual information, and haplogroup assignments are displayed in Table 2.1. Mitochondrial sequences were first aligned to the goat mtDNA reference NC_005044.2, before realignment to a haplogroup-specific reference (Appendix Table 2.3).

Modern data

86 modern *Caprid* whole mitochondrial sequences were downloaded from the NCBI nucleotide database, and are displayed in Appendix Table 2.3. Nubian ibex was selected as an outgroup, based on a preliminary alignment presented in Appendix Figure 3.1.

Methods

Maximum likelihood phylogeny

The relationships between different mitochondria are commonly represented by a tree model, in which each branch tip is a sequence and each node is a common ancestor of two sets of sequences (Yang & Rannala 2012). Tree-building methods such as neighbour-joining, maximum likelihood or Bayesian inference compute on a distance matrix summarize how distant sequences are from each other. These approaches are explicitly model-based as they utilize substitution models describing the probability of different substitutions and the variations of these rates. Other methods such as maximum parsimony are not model-based, making them computationally efficient but susceptible to issues such as long-branch attraction (Felsenstein 1978). This systematic bias can also affect model-based algorithms if the model is overly simplistic (Yang 1996) and as such realistic models accounting for inter-site rate variation should be used in tree-building.

Maximum likelihood approaches to tree building attempt to maximize the likelihood function: the probability of the data given the parameters, where the parameters are the branch lengths and substitution parameters (Yang & Rannala 2012). Usually an iterative

optimization algorithm is used to find the maximum likelihood estimates of the parameters which maximize the tree likelihood. Compared to other tree-building approaches, maximum likelihood is more efficient (*i.e.* has a higher probability of returning the true tree) than other model-based approaches (Felsenstein 2004). However, the approach is computationally demanding, and misspecification of the substitution model can be problematic for statistical inference.

Mitochondrial sequences were aligned using MUSCLE (Edgar 2004). Nubian ibex was selected as an outgroup. A subsample of modern sequences was selected as representatives of their respective haplogroups (Appendix Table 2.3). The alignment was inspected visually using Seaview (Gouy et al. 2010). Modelgenerator v0.85 (Keane et al. 2006) was then used to determine the most appropriate substitution model for the multiple sequence alignment. A maximum likelihood tree was generated using PhyML 3.1 (Guindon et al. 2010) with 100 bootstrap replicates, using model parameters estimated by modelgenerator. The resulting phylogeny was visualized using Figtree v1.4.2 (Rambaut 2009).

To investigate if modern and ancient mtDNA from the same geographic region form clades *i.e.* they are more closely related than geographically-disparate samples the above steps were repeated with all modern and ancient samples, rather than a subset of moderns. Samples were labelled by their geographic origin.

Bayesian analysis

Bayesian approaches to tree inference assume that parameters in a model are not fixed values but are best described by statistical distributions (Yang & Rannala 2012). Prior distributions of these parameters are supplied before the analysis, which are used together with the data to produce posterior parameter distributions. Bayesian methods such as BEAST (Bayesian Evolutionary Analysis Sampling Trees) (Drummond & Rambaut 2007) use Markov Chain Monte Carlo (MCMC) algorithms to create an estimate of the posterior distributions. This is achieved by moving through tree and parameter space, sampling both in proportion to their posterior probabilities (Yang & Rannala 2012).

Bayesian analysis shares many of the same advantages of those based on likelihood maximization: model-based, robust to missing information, and consistent when the underlying model is correctly specified. Additionally, Bayesian approaches have been

extended to allow rates and substitution patterns to vary between lineages (Ho & Phillips 2009), and implemented in the BEAST program. BEAST can also incorporate uncertainty in sample ages during rate calibration (Molak et al. 2013), an approach that converts units of genetic change (branch lengths) into units of time, utilizing node (*e.g.* fossil) or branch tip (*e.g.* dated temporal samples) information (Rambaut 2000; Rieux & Balloux 2016). This implementation of branch tip calibration has been shown to improve the accuracy of rate estimation compared to node calibration (Rieux et al. 2014), and the uncertainty associated with dated samples can be easily accounted for (Drummond et al. 2006). Tip calibration can be confounded by tree imbalance - cases in which the tips are unevenly distributed among branches (*i.e.* many long branches, some comb-like branches), and can result in underestimates of timescales (D. Duchêne et al. 2015).

A BEAST analysis was performed using BEAST 2.4.2 (Drummond et al. 2012; Bouckaert et al. 2014). To estimate the mitochondrial mutation rate and split times for mitochondrial lineages, a multiple sequence alignment of modern goat/bezoar and radiocarbon dated ancient sequences which fell within domestic goat lineages (*i.e.* those not designated haplogroup T) was prepared using MUSCLE (Edgar 2004).

Partitioning of the datasets into regions of similar substitution patterns allows for between-site variation in rates to be accounted for (Yang & Rannala 2012). Partitions were defined using NCBI annotation for NC_005044.2: tRNA, rRNA, the first and second codon positions (C1+2), third codon positions (C3), D-loop, and the remainder of the molecule (Appendix Table 3.1). Misspecification of partitions can lead to overfitting or underfitting, in cases when the defined partitions were too fine or broad respectively. The appropriateness of these partitions was tested using PartitionFinder (Lanfear et al. 2012), testing all models, linking branch lengths, performing model selection using the Bayesian information criterion statistic and using a greedy search algorithm. The best models determined were HKY+I for C1+2, TRN for C3, HKY+I+G for the D-loop, and TRN+I for a combined partition of tRNA genes, rRNA genes, and the remainder of the mitogenome.

For the BEAST analysis, site and clock models for each partition were unlinked, with the tree linked across partitions. To replicate the TRN model in BEAST, TN93 with estimated base frequencies was used instead. For HKY, base frequencies were also set to estimated. Clocks for each partition were set to strict. Priors for sample age were set as Normal distributions (an assumption which is incorrect (Reimer et al. 2004) but not a major issue (Ho & Phillips

2009)), mean equal to the midpoint of the radiocarbon 95% CI, sigma equal to one quarter of the length of the 95% CI. Clock priors were set to Log Normal, M=-18.42068, S=1.5. Kappa priors were set to Log Normal, and gamma shape priors set to exponential. A Coalescent Bayesian Skyline model was used as the tree model.

Four independent runs of 100 million chains were performed, with 10% burn-in, and assessed using Tracer (Rambaut A, Suchard MA, Xie D & Drummond AJ 2014). Assuming each independent run converged with all ESS >3000, runs were merged. A Maximum Clade Credibility tree was then constructed with median heights, using TreeAnnotator (Drummond et al. 2012).

To estimate a mutation rate for the mitogenome with the D-loop for the purpose of modelling, the above was repeated using the same settings except for the clock models, which were linked across the non D-loop partitions.

Analysis of molecular variance (AMOVA)

The analysis of variance framework, developed by Ronald Fisher (Fisher 1921), has been extended to molecular haplogroup in AMOVA (Excoffier et al. 1992), and is implemented in Arlequin package (Excoffier & Lischer 2010). AMOVA tests if the variance observed within groups, between populations which make up these groups, and the populations themselves are significant using a permutation approach. AMOVA calculates an inter-individual squared distance matrix, which can be partitioned into submatrices representing different subsets of individual comparisons. Rows and the corresponding columns are randomly swapped during permutation to generate a matrix reflecting a null model of panmixia *i.e.* variance observed between groups or populations is not statistically distinguishable from the variance observed within populations.

Partitioning of genetic diversity was calculated using Arlequin v3.5 (Excoffier & Lischer 2010). Populations and Groups were defined as in Appendix Tables 3.2 and 3.3; Groups refer to broad regions such as the Levant, Iran+Turkmenistan, Turkey+southeast Europe, while populations refer to specific sites or several small sites combined. Maximum missing data per site was set at 0.05. Significance of variance components and Fixation Indices were computed using permutation tests (1000 permutations).

Demographic modelling of population histories - mtDNA

Demographic modelling using mtDNA was performed primarily by Pierpaolo Maisano Delser, and is reproduced here (Daly et al. 2018). Whole mitochondrial genomes from 23 samples were analysed and all sites were called as described above, and the D-loop removed (positions 15,431-16,643). Considering the heterogeneous level of missing data across samples (ranging from 0 to 23%, Appendix Table 3.4), a dataset including only sites shared across all samples would have not had sufficient information. Therefore, each summary statistics was calculated using individual pairwise comparisons both within and between populations. Each within-population summary statistic was calculated as average across all the individual pairwise comparisons between all samples belonging to that specific population. Each between-population summary statistic was calculated as average across all individual pairwise comparisons between samples belonging to the two populations under study. Following this approach, both nucleotide diversity (π) per population and Hudson's pairwise F_{ST} (Hudson et al. 1992) were calculated with an in-house R script v3.2.3 (R Core Team 2015) (Appendix Tables 3.5 and 3.6).

We developed an approximate Bayesian computation (ABC) (Beaumont et al. 2002) framework to estimate parameters and compare models. Two demographic models were designed to investigate the demographic histories of samples belonging to the Western, Eastern and Levantine populations: model SINGLE_MT and model MULTIPLE_MT (Figure 3.1). Model SINGLE_MT represents a single domestication process shared by the three populations. An ancestral population (Nanc2) goes through a bottleneck from 11,000 to 10,500 years ago representing the domestication process before splitting in the three ancestral population which give rise to the Neolithic western, eastern and Levantine populations (represented by Nneow, Nneoe and Nneol respectively). Model MULTIPLE_MT describes a scenario with multiple domestications. From an ancestral population (Nanc3), the Levantine branch splits before going through a bottleneck from 11,000 to 10,500 years ago and then exponentially expands from 10,500 to 8,000 years ago (Neolithic Levantine population). Subsequently, the ancestral population (Nanc2) splits into the ancestral population for the Western and Eastern samples (Nanc1w and Nanc1e respectively) before going through a bottleneck at 11,000 years ago and then exponentially expand up to 8,000 years ago (Neolithic Western and Eastern populations). Prior distributions for all parameters of the two models are presented in Appendix Table 3.7.

Simulations were built to have the same configuration as the observed data (to conform with sequence length and pattern of missing data). Specifically, we first recorded the exact position of each missing nucleotide across all sequences in the real dataset ("missing data layer"). Then, the maximum number of base pair (15,429) was simulated and subsequently the "missing data layer" was applied to each simulated dataset. In this way we were able to recreate the exact pattern of missing data in terms of percentage and position observed in the real dataset in each simulated dataset.

We performed 100,000 simulations under each model using fastsimcoal 2 v.25221 (Excoffier et al. 2013). Generation time was assumed at 2.5 years (Wilson & Mittermeier 2011). The mutation rate was calibrated as described above and a value of 1.411×10^{-7} per site per generation was used. The following summary statistics were used: nucleotide diversity per population (π_3E, π_3L) and pairwise Hudson's F_{ST} for the following comparisons (Fst_3W_3E, Fst_3L_3W, Fst_3L_3E). Model posterior probabilities were calculated by a weighted multinomial logistic regression (Beaumont 2008) for which we retained the best 25,000 and 50,000 simulations. Parameters under the most supported model were estimated from the 5,000 simulations closest to the observed dataset using the neuralnet algorithm (Csilléry et al. 2012). Analyses were performed in the R environment (R Core Team 2015) with the library abc (Csilléry et al. 2012).

In addition to the model described above and presented in Daly et al. 2018, an additional set of scenarios were tested using western and eastern populations. Levantine populations were excluded due to the high missingness of Neolithic Levant mtDNA (Appendix Table 3.4 and 3.5). The aim of this modelling exercise was to evaluate whether these populations best fit a panmictic population (Model BOT), two distinct populations (Model SPLIT), or two populations which experienced gene flow following the Neolithic (MODEL SPIT-MIG). These three models are graphically represented in Appendix Figure 3.1; samples used and population designations are shown in Appendix Table 3.4; prior distributions for each parameter are presented in Appendix Table 3.8. Modelling was performed using samples from the Neolithic up to the Iron Age and Medieval periods (Appendix Table 3.4). Summary statistics used were nucleotide diversity (π) for all populations and timepoints, and pairwise Hudson's F_{ST} for the following comparisons (Fst_3W_3E, Fst_2W_2E, Fst_1W_1E, Fst_3E_2E, Fst_2E_1E, Fst_3W_2W, Fst_2W_1W). Summary statistics, model posterior probabilities, and parameters estimation for the most supported model were calculated as above.

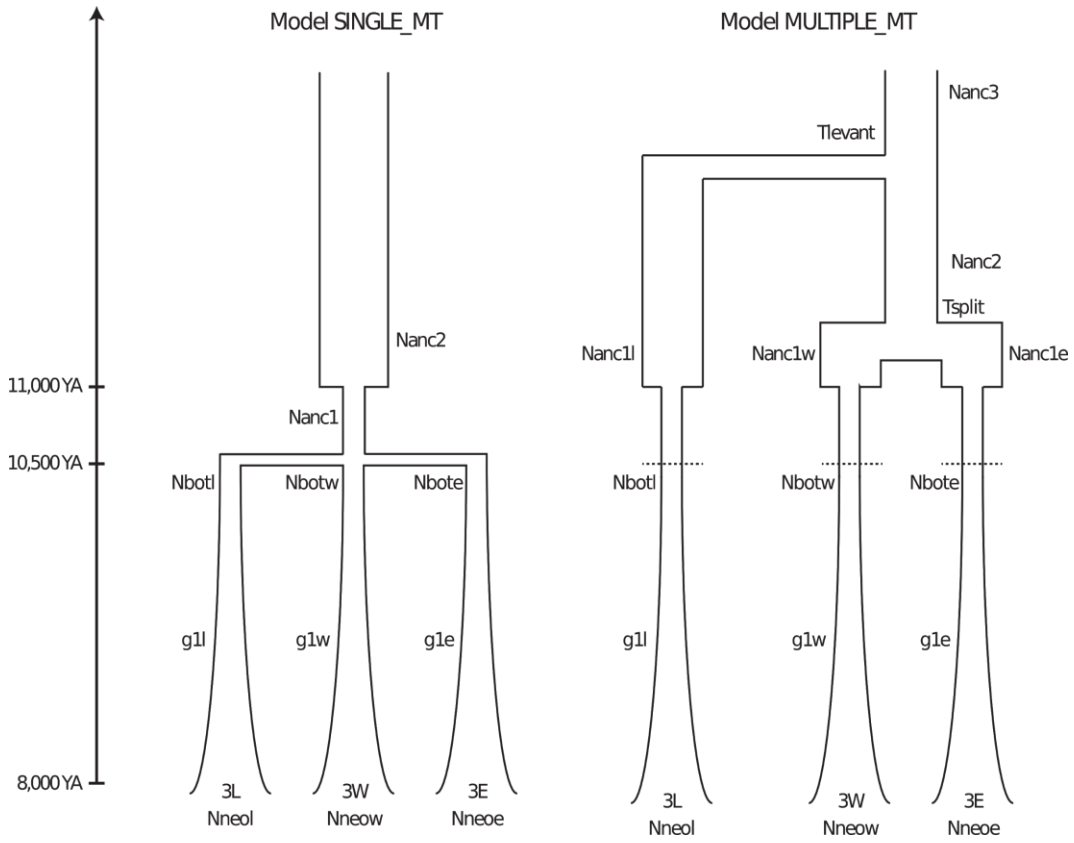


Figure 3.1 - Demographic models tested with whole mitochondrial genomes.

3L=Neolithic Levant, 3W=Neolithic West, 3E=Neolithic East, YA=years ago.

Results

Maximum likelihood phylogeny and haplogroup distribution

The maximum likelihood phylogeny constructed using ancient and reference modern sequences (Figure 3.2) indicates that while some bezoar from pre-domestic contexts (*i.e.* Direkli Cave) had mtDNA haplogroups not observed in modern goat (T, see below), the large majority of ancient samples had haplogroups observed in modern populations (F, B, G, D, A, and C). In most cases the available bezoar ibex haplogroup representatives were sister branches to ancient sequences from domestic contexts, with some exceptions (see below). Similarly, a phylogeny constructed using a large dataset of modern goat mitochondria and the ancient sequences generated here indicates that in most cases, ancient and modern goat form clades to the exclusion of bezoar (Figure 3.3). Bootstrap support for major/deep nodes were mainly high (>0.6), while shallower nodes tended to be less well supported.

The distribution of goat haplogroups across southwest Asia shows apparent geographic segregation of lineages in the Neolithic (Figure 3.4.A). Haplogroup A and C are distributed solely in western Turkey and southeast Europe. Haplogroups G, B, and D were observed only in sites in modern-day Iran and Turkmenistan, while in the Levant only haplogroup F was detected. Following the Neolithic, haplogroup A is the predominant haplogroup observed in all parts of southwest Asia (Figure 3.4.B). Some haplogroups observed in Neolithic Iran persist, but at much reduced frequencies. Additional haplogroup distribution changes are observed; in the Levant haplogroup D is detected, while haplogroup G is found in two individuals in southeastern Europe.

Examining each haplogroup in more detail, Haplogroup A appears at a high frequency in the dataset (46/82 after Bulak3 is removed due to its high relatedness with Bulak1). In the Neolithic it is restricted to the western Turkish sites of Ulucak Höyük (site 7 in Figure 2.5) and Aşağı Pınar (site 6), and several sites throughout southeastern Europe (Kovačevo, Ovčarovo-gorata, Čavdar, Uivar, and Blagotin-Poljna). It is not observed in eastern parts of southwest Asia (*i.e.* the Zagros Mountains, eastern Iran, Turkmenistan), or in the Levant (Israel, Jordan). This segregation appears to break-down following the Neolithic, at which point haplogroup A is found at high frequencies in Iran, Uzbekistan, Turkmenistan, Israel, and Georgia. It maintains a high frequency in Turkey and southeast Europe in post-Neolithic time periods.

There are mixed indications of geographic structure within haplogroup A. Six Bronze and Iron Age Levantine sequences (Safi2, Azer4, Yarmut1, Yoqneam2, Yarmut7, Miqne5) form a clade, with the addition of a Bronze Age sample from central Turkey (Acem1) (Figure 3.2). Neolithic sequences from southeast Europe (Blagotin1, Blagotin2, and Kov57) form a clade that branches closely with contemporary and later individuals from the same region (AP49, AP50, Kan19). The majority of the remainder of haplogroup A shows little geographic clustering (*e.g.* a clade containing a Bronze Age British sample, Potterne1, and a Chalcolithic

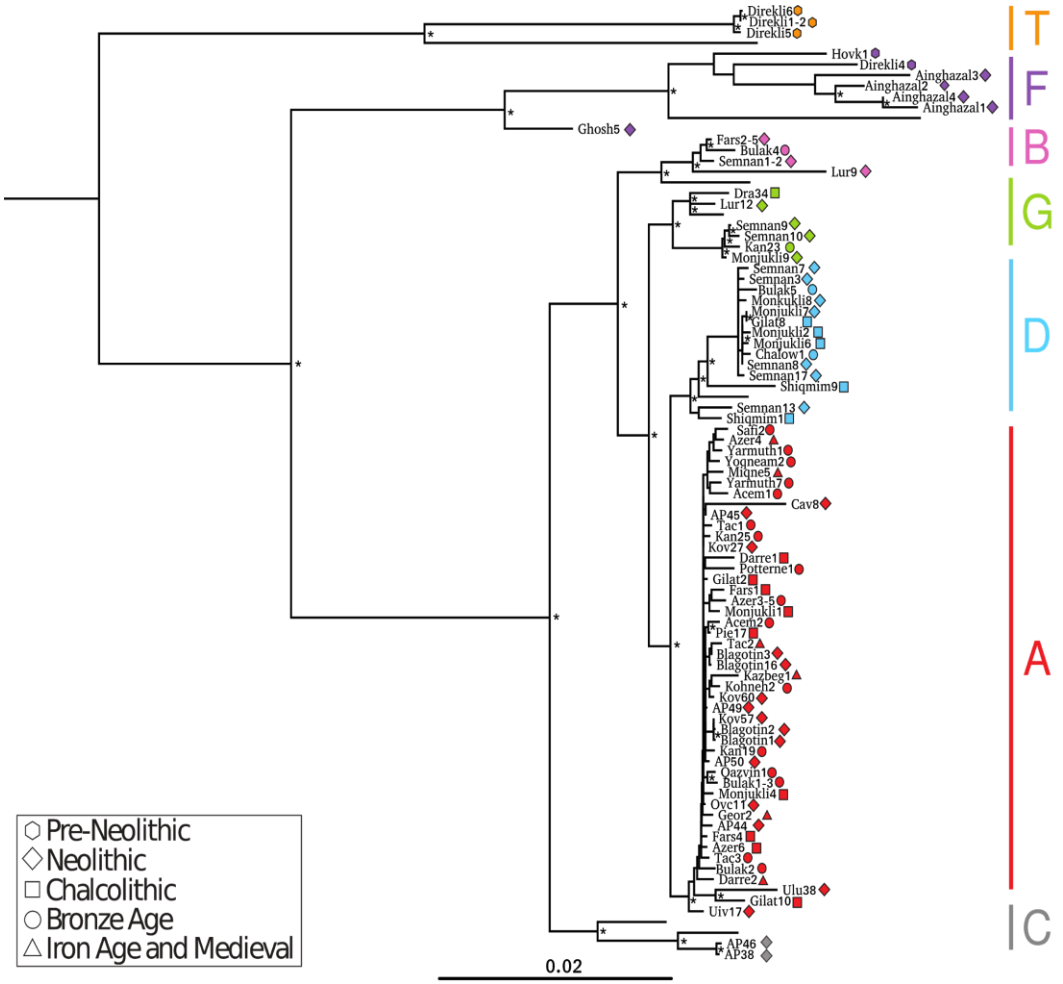


Figure 3.2 - Maximum likelihood tree of ancient goat mtDNA and reference modern sequences. Context of each ancient sample is indicated by symbol shape, haplogroup by symbol colour. Reference individuals are denoted by blank tips. Nubian ibex outgroup is not shown. High bootstrap confidence nodes (>0.6) are indicated by asterisks.

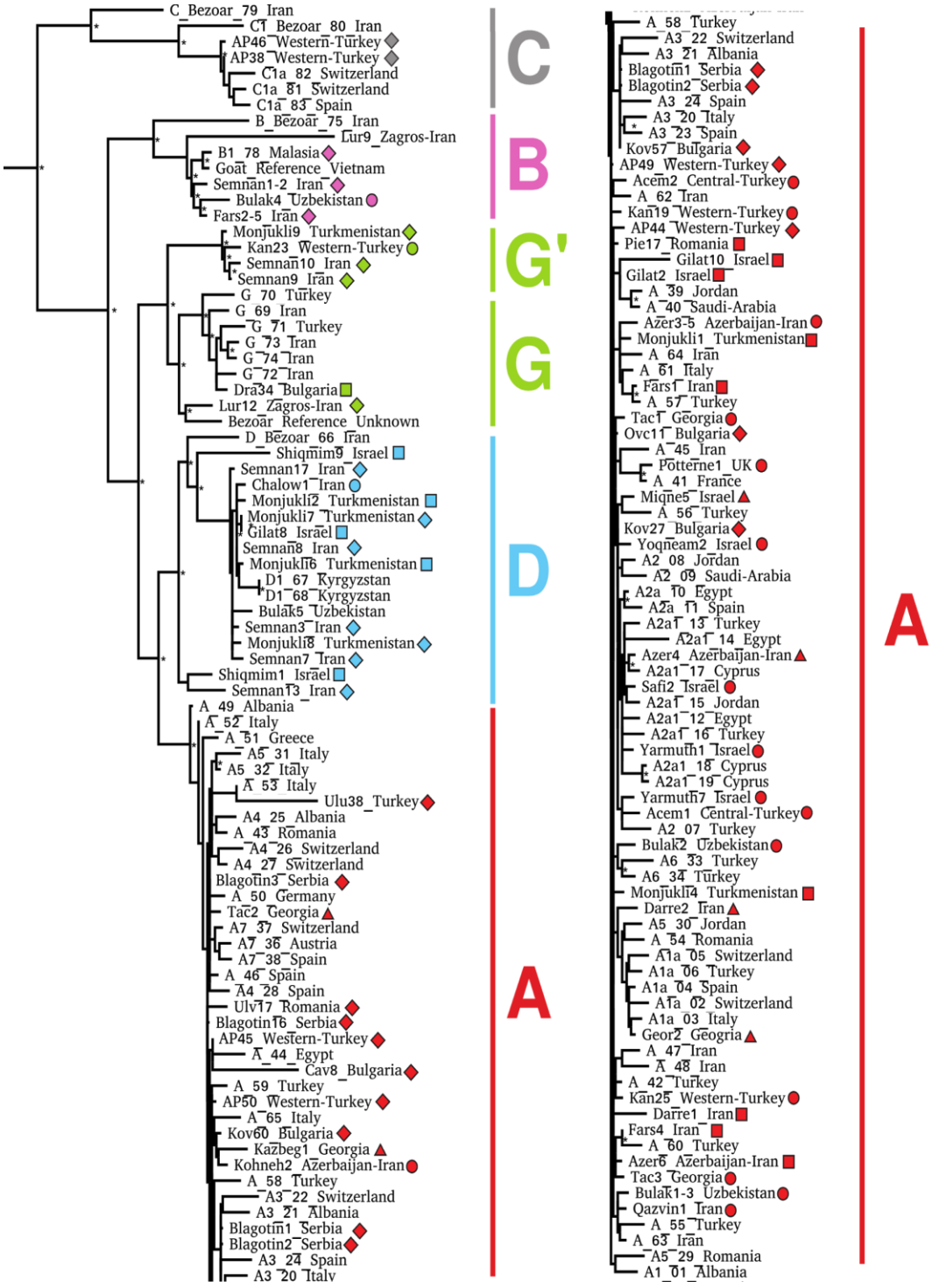


Figure 3.3 - Maximum Likelihood tree constructed using ancient and modern mitochondrial haplogroups A, B, C, D, and G. Ancient samples are indicated by the same contextual symbols as used in Figure 3.2. Nubian ibex was used as an outgroup. High bootstrap confidence node (>0.6) are indicated by asterisks.

Iranian sample, Darre1). This poor differentiation is mirrored in the low bootstrap values for most haplogroup A clades. In the context of modern goat mtDNA, the lack of geographic clades is even more striking (Figure 3.3), but some are observed. For example, a clade is

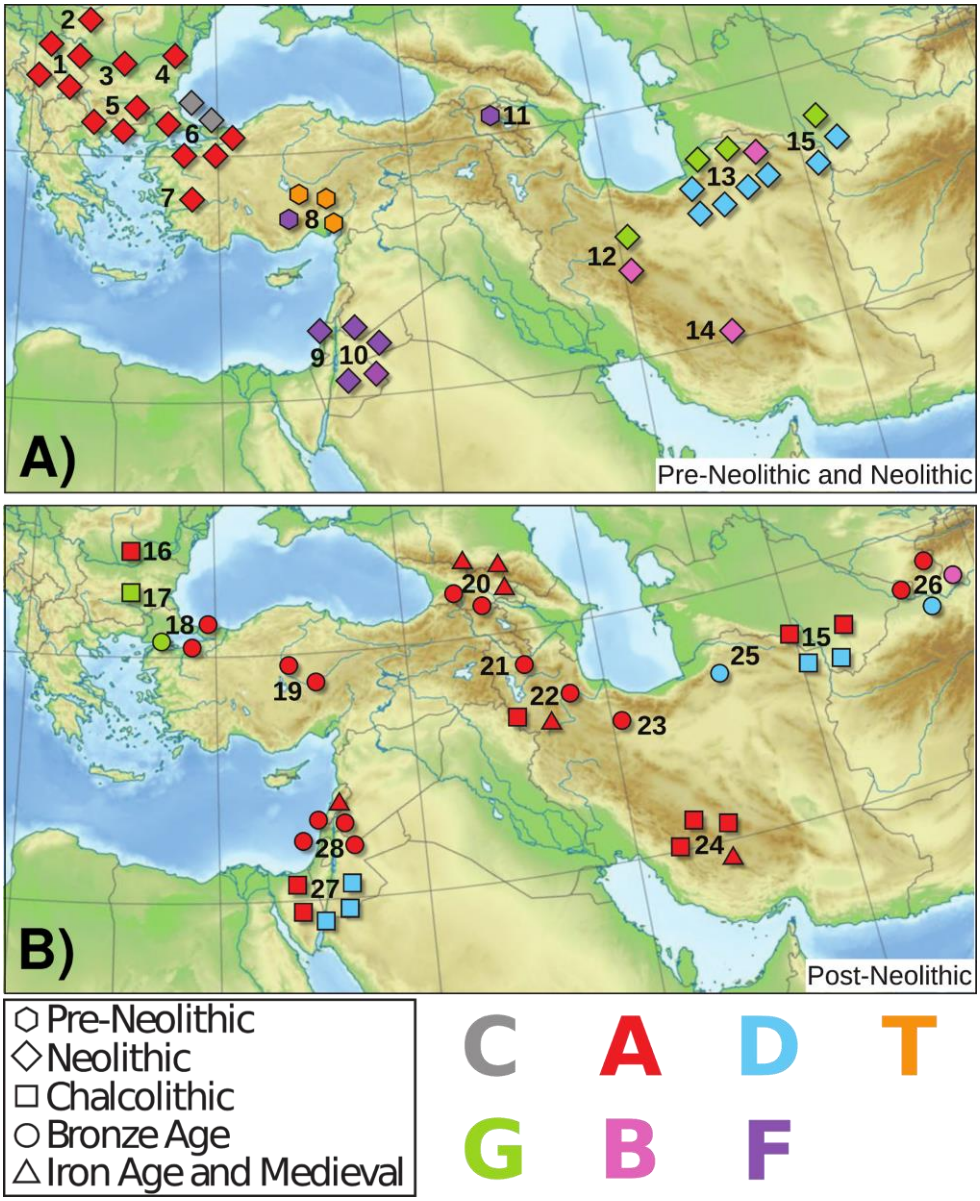


Figure 3.4 - Distribution of ancient goat mtDNA haplogroups, A) up to and B) after the Neolithic. Sites are indicated by numbers specified in Table 2.2. Time period of samples are indicated by point shape, haplogroup by symbol colour. The transition to the Neolithic and Post-Neolithic is defined as 5,300 to 5,000 BC, which overlaps samples from Neolithic Aşağı Pınar (site 6) Chalcolithic Mianroud (site 24). Not shown: a single haplogroup A from Bronze Age Potterne, UK. Map adapted from Викtор 2010.

formed of two Chalcolithic goat from Israel (Gilat2 and Gilat10), and two modern goat from Jordan and Saudi Arabia. Haplogroup C is observed in two individuals from Neolithic Aşağı Pınar, in Turkish Thrace. Two modern bezoar sequences are outgroups to this pair (Figure 3.2). In the context of modern mitochondrial sequences, the two ancient goat form a clade with the available domestic modern sequences, excluding the modern bezoar ibex (Figure 3.3). Haplogroup C is not observed elsewhere during the Neolithic, nor anywhere later in the time series.

Haplogroup D is found only in two eastern sites in the Neolithic: Sang-e Chakmaq (site 13) and Monjukli Depe (site 15), both dated to *c.* 6,000 cal BC. Following the Neolithic, it is observed in eastern sites at Chalcolithic Monjukli Depe, Bronze Age Chalow (site 25), and Bronze Age Tilla Bulak, Uzbekistan (site 26). Additionally, it appears at two Israeli sites: Shiqmim (site 27A) and Gilat (27B). Two of these (Shiqmim1 and Shiqmim9), along with a single sample from Sang-e Chakmaq (Semnan13) and a modern bezoar sequence, branch as a sister clade to all other haplogroup D sequences. These remaining mitochondria form a clade containing the two available modern D sequences from Kyrgyzstan (Figure 3.3).

Haplogroup G sequences form two distinct clades. The first consists of a modern bezoar; an early Neolithic individual (Lur12) from Tepe Abdul Hosein in the Zagros Mountains (site 12) dated to *c.* 8,000 cal BC (Table 2.1); and an individual from Merdžumekja (site 17), Chalcolithic Bulgaria (Dra34). The second clade is composed of three individuals from Neolithic Sang-e Chakmaq and Monjukli Depe, in addition to a single sequence from Kanlıgeçit, western Turkey (site 18). This clade will be referred to here as G'. Similar to haplogroup D, this haplogroup is found only in eastern sites in the Neolithic, but also appears at a low frequency in more western sites in subsequent time periods. When modern goats are included in the phylogeny, the bezoar reference sequences and Lur12 form a clade that is sister to all modern sequences and Dra34 (Figure 3.3). No modern sequences in the dataset fall in the G' clade.

Only four B haplogroup sequences were detected, all in eastern sites: the Pre-Pottery Neolithic part of Sang-e Chakmaq (site 13), the Pre-Pottery Neolithic site of Kelek Asad Morad (site 12; sample ID is Lur9), the Pottery Neolithic site Rahmat Abad (site 14), and Bronze Age Tilla Bulak (site 26). The available modern bezoar sequence is an outgroup to these four individuals. Bootstrap support for this clade and the previous (G) are in general

higher than haplogroup A. Modern goat sequences indicate the bezoar sequence and then Lur9 are outgroup to all other ancient and modern B sequences (Figure 3.3).

Haplogroup F was observed in two samples from pre-domestication contexts: Hovk1, from Armenia and dated beyond the C_{14} radiocarbon limit, and Direkli4, from southern Turkey and dated to *c.* 12,000 cal BC. All mitochondrial sequences recovered from Middle PPNB levels of 'Ain Ghazal (site 10) are haplogroup F, as is a single sequence from Middle PPNB Abu Ghosh (site 9). Although several of these sequences are of relatively-poor quality (Table 2.1), the 'Ain Ghazal samples form a well-supported clade. In comparison to the previously-described haplogroups, haplogroup F shows little shallow branching with substantial divergence between sequences.

Three mitochondrial sequences from the Epipaleolithic site of Direkli Cave (site 6), located in the Taurus Mountains of central/southern Turkey, form a well-supported sister clade to *Capra caucasica*, the West Caucasian Tur. The Direkli sequences show little divergence among each other, but are deeply diverged from the Tur reference mitochondria. As this clade has not been yet reported, it is labelled here as haplogroup T.

Bayesian analysis

As each individual BEAST run resulted in parameter Effective Sampling Sizes (ESS) of >>3,000 (*i.e.* were converged), runs were combined and clock rates for different partitions obtained (Table 3.1). When four partitions were used (codon positions 1+2, codon position 3, D loop, tRNA+rRNA+remainder of the molecular), similar rate estimates were obtained whether or not *Capra falconeri* and *Capra caucasica* were included, although slightly lower standard deviations were observed when they were. Mutation rate estimates were highest in the D loop (mean 8.27×10^{-7} mutations per site per year, using *Capra hircus*/*aegagrus* only), and decreased sequentially for the third codon position (1.13×10^{-7}), the tRNA+rRNA+remainder partition (2.52×10^{-8}), and codon positions 1+2 (1.84×10^{-8}). When partitioned between the D loop and the remainder of the molecule, the D loop had an overall higher rate (8.35×10^{-7} versus 5.57×10^{-8}).

Bayesian phylogenies constructed without *Capra falconeri/caucasica* (Figure 3.5) and with *Capra falconeri/caucasica* (Appendix Figure 3.2) result in a similar topology as to that obtained using maximum likelihood. Divergence time estimates for the main phylogeny

nodes are displayed in Table 3.2. The divergence time between the Tur mitogenome and haplogroup T is 137-201 kya (95% HPD - highest posterior density); this clade is estimated to diverge from the ancestor of *Capra hircus/aegagrus* and *Capra falconeri* 269-369 kya (95% HPD). The markhor *Capra falconeri* is then estimated to split from the ancestor of *Capra hircus*/*aegagrus* 251-347 kya. Divergence times of the main goat haplogroups fall robustly before the Neolithic: F (209-298 kya), C (61-88 kya), B (38-56 kya), G (30-43 kya), and finally the D|A split (23-34 kya). Within each haplogroup, domestic sequences diverged roughly in the early Holocene: C (7.8-10.6 kya), B (10.5-16.5 kya), G (8.2-14.7 kya - excluding G' sequence Kan23), D (8.4-11.6 kya), and A (10-14.5 kya). The estimated divergence time of G and the sister clade observed only in ancient (G') is pre-Neolithic and wide, 19.8-31.1 kya. Node support is high outside of haplogroup A, in which approximately half of nodes have <0.6 posterior support.

Table 3.1 - Clock-rate estimates for BEAST runs, using *Capra hircus/aegagrus* only, and also for *Capra falconeri* and *Capra caucasica*. S.D = Standard Deviation. HPD = Highest Posterior Density.

Partitions	Mean	Median	SD	95% HPD	ESS
clockRate.D_loop	8.32E-07	8.27E-07	1.02E-07	[6.4522E-7, 1.0392E-6]	4855
clockRate.tRNA+rRNA+ remainder	2.52E-08	2.51E-08	2.98E-09	[1.9473E-8, 3.1007E-8]	9971
clockRate.C3	1.13E-07	1.12E-07	9.40E-09	[9.4415E-8, 1.3111E-7]	5545
clockRate.C1-2	1.84E-08	1.83E-08	1.95E-09	[1.4649E-8, 2.2274E-8]	8628
Two Partitions: *Capra hircus/aegagrus* only					
clockRate.D_loop	8.35E-07	8.29E-07	1.01E-07	[6.4359E-7, 1.0344E-6]	5943
clockRate.Non-D_loop	5.57E-08	5.55E-08	4.59E-09	[4.6446E-8, 6.4438E-8]	5285
Four Partitions: *Capra hircus, aegagrus, falconeri, caucasica*					
clockRate.D_loop	8.28E-07	8.22E-07	9.61E-08	[6.4907E-7, 1.0222E-6]	5766
clockRate.tRNA+rRNA+ remainder	2.70E-08	2.68E-08	2.82E-09	[2.1602E-8, 3.2578E-8]	8690
clockRate.C3	1.18E-07	1.18E-07	9.39E-09	[1.0052E-7, 1.371E-7]	5162
clockRate.C1-2	1.88E-08	1.88E-08	1.82E-09	[1.5284E-8, 2.2388E-8]	7454

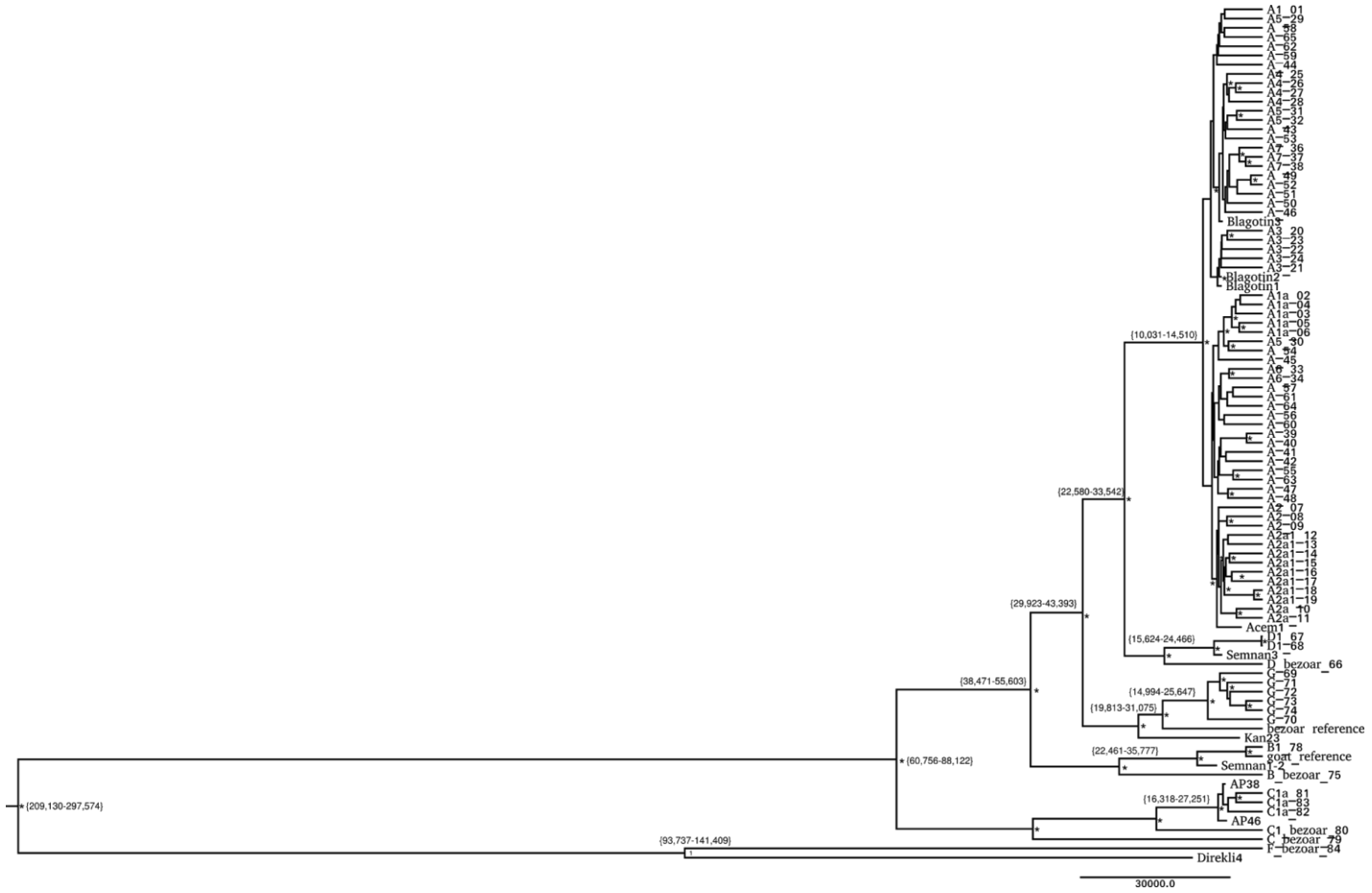


Figure 3.5 - Phylogeny constructed using BEAST, using *Capra hircus/aegagrus* only. Posterior support for high confidence (>0.6) nodes are marked with an asterix. 95% highest posterior density for key nodes are displayed in braces, and are rounded to whole numbers.

Analysis of molecular variance (AMOVA)

Analysis of Molecular Variance indicates that during the Neolithic, variation between groups (*i.e.* significant fixation index between western Turkey/southeast Europe, the Levant, and Iran/Turkmenistan) was statistically significant and explained 80.57% of mtDNA variation (Table 3.3). Variation within populations within groups was also substantial, comprising 19.17% variation. In contrast, among population within group variation was non-significant and close to zero (0.26%).

Chalcolithic and Bronze Age mtDNA variation is mostly explained by within population within group components (92.78%). Among group (≈ 0%) and among populations within groups (7.87%) is low in comparison, and in both cases have non-significant variance components. mtDNA is similarly distributed in the Iron Age/Medieval/Modern grouping: among group (≈ 0%) and among population within group (8.5%) variation is non-significant, whereas within population within group variation is significant and large (94.12%).

Table 3.2 - BEAST estimation of node age using *Capra hircus/aegagrus* only, and additionally *Capra falconeri* and *Capra caucasica*. HPD = Highest Posterior Density.

Tree Node	Median Age (years ago)	Age 95% HPD (years ago)
Capra hircus/aegagrus only		
F|CBGDA	250,213	209,130-297,574
C|BGDA	73,616	60,756-88,122
B|GDA	46,678	38,471-55,603
G|DA	36,207	29,923-43,393
G|G’	28,849	19,813-31,074
D|A	27,770	22,580-33,542
C internal split	8,964	7,783-10,577
B internal split	13,155	10,505-16,468
G internal split	11,041	8,118-14,672
D internal split	9,760	8,386-11,630
A internal split	11,993	10,031-14,510
Capra hircus, aegagrus, falconeri, caucasica		
T|MFCBGDA	315,976	268,736-368,761
M|FCBGDA	297,042	250,619-346,741
T|Direkli1-2	167,548	137,231-201,478

mtDNA modelling

Model posterior probabilities suggest MULTIPLE_MT as the most supported model using two different thresholds for simulations retained (25,000 and 50,000) (Table 3.4). Subsequently, parameter estimations were determined under model MULTIPLE_MT for Tsplit and Tlevant (Figure 3.4 and Table 3.5). The mode for Tsplit was 12.1 kya (95% credible interval 11.1-18.4 kya), while the mode for Tlevant was 138 kya (95% credible interval 38.5-195.2 kya). This pre-domestic divergence time for the Levantine mtDNA supports a maternal contribution from a bezoar lineage distinct from those found in western and eastern parts of southwest Asia.

The additional set of mtDNA models strongly supported the two populations with admixture between eastern and western goats, model SPLIT-MIG (Appendix Table 3.9; summary statistics are displayed in Appendix Table 3.10). A substantially larger for Tsplit was obtained (mode 30.4 kya), with a wider 95% credible interval (13.8-56.3 kya, Appendix Table 3.11). The estimation of migration time was also highly uncertain (3.2-8.5 kya) suggesting that admixture happened between the two populations but this is difficult to data accurately.

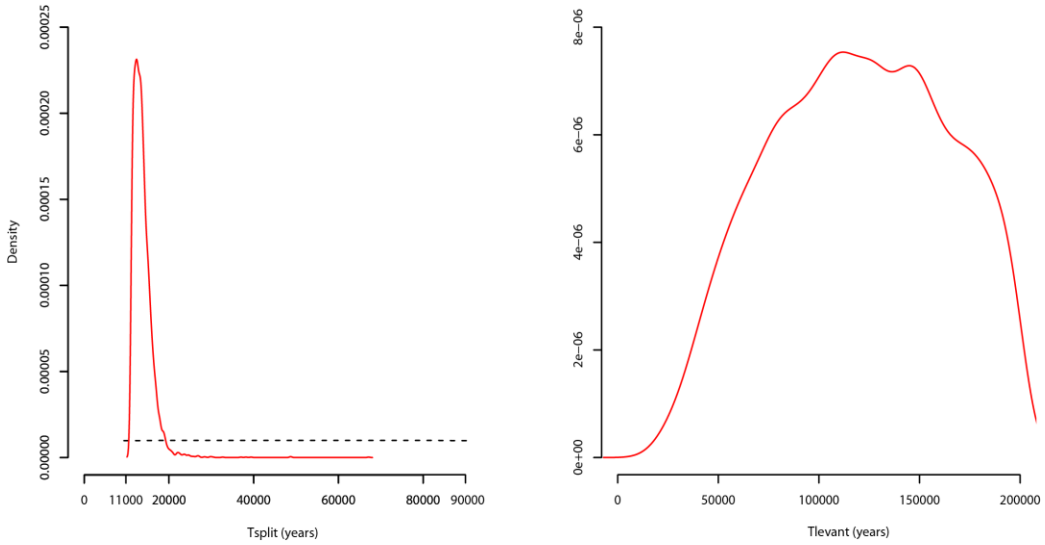


Figure 3.6. Posterior distribution of Tsplit and Tlevant estimated under model MULTIPLE_MT. Black dotted line: prior distribution; red line: posterior distribution calculated using a neuralnet algorithm.

Table 3.3 - Arlequin AMOVA results. d.f = degrees of freedom. P values calculated by permutation.

Partition	d.f.	Sum of Squares	Variance Components	Percentage of Variation	Fixation Indices	P value
Neolithic						
Among groups	2	675.938	49.58747 Va	80.57	0.806	0.028+-0.006
Among populations	3	37.089	0.16038 Vb	0.26	0.013	0.413+-0.017
Within populations	17	200.625	11.80147 Vc	19.17	0.808	0.000+-0.000
Chalcolithic and Bronze Age						
Among groups	2	35.64	-0.08976 Va	-0.65	-0.006	0.359+-0.013
Among populations	5	86.978	1.08891 Vb	7.87	0.078	0.235+-0.015
Within populations	19	372.193	12.83424 Vc	92.78	0.072	0.082+-0.008
Iron Age, Medieval and Modern						
Among groups	3	102.017	-0.57519 Va	-2.62	-0.026	0.463+-0.018
Among populations	2	85.113	1.86819 Vb	8.5	0.083	0.047+-0.009
Within populations	71	1468.22	20.67915 Vc	94.12	0.059	0.009+-0.003

Table 3.4 - Model posterior probabilities. Calculated by a weighted multinomial logistic regression using whole mitochondrial genomes.

Number of simulations retained	Model SINGLE_MT	Model MULTIPLE_MT
25,000	0.2	0.8
50,000	0.19	0.81

Table 3.5 - Parameter estimation under the most supported model (MULTIPLE_MT). Prior distributions and estimates of both Tsplit and Tlevant have been converted in years using a generation time of 2.5 years.

Model MULTIPLE_MT	Prior	Median	Mode	95% Credible Interval
Tsplit	Uniform (11,000-90,000)	13,309	12,083	11,132-18,421
Tlevant	Uniform (Tsplit-200,000)	121,674	138,370	38,482-195,210

Discussion

mtDNA structure in Neolithic goat

Mitochondrial DNA of goat in the Neolithic shows strong geographic partitioning. The three regions of the Levant, western Turkey/southeast Europe, and Iran/Turkmenistan show distinct haplogroup compositions which in the current data do not co-occur in the Neolithic. AMOVA results indicate that this partitioning of variance is significant and substantial (80.57%) (Table 3.3).

In the western region, haplogroup A appears at a high frequency (15/17, or 88%), with some observations of haplogroup C (2/17, 12%). This is remarkably similar to the frequency observed in modern European goat (Colli et al. 2015), and supports previous indications from aDNA that the mtDNA gene pool of European goat was established early (Fernández et al. 2006). The geographic proximity to Turkey and early nature of the Neolithic sites suggests that these haplogroups were present in managed goat populations in Turkey prior to their introduction to Europe. This is consistent with the suggestions of (Naderi et al. 2008) that the goat domestication process occurred in eastern Turkey, where both haplogroups are found in modern bezoar. However, here this explanation only fits for Neolithic goat from western regions. From the Neolithic phylogeographic structure and modern mtDNA distributions, it appears as if the ancestors of goat from western Turkey and southeast Europe derived from a population similar to those observed in bezoar ibex from eastern Turkey today.

Zooarchaeological data is consistent with this hypothesis. The earliest evidence of young male kill-off of goat occur roughly in the 8th millennium BC at the sites of Ganj Dareh in the Zagros Mountains, Beidha and Tell Aswad in the Levant, and importantly Çayönü in southeastern Turkey (Arbuckle & Atici 2013). Additionally, the absence of haplogroup A or C in the data presented here from more southerly and easterly regions (Figure 3.4) restricts possible sources of both haplogroups to the unsampled region of central or eastern Turkey/northern Iraq/northeast Syria (see below). The practice of goat herding was likely introduced from east Turkey to central Turkey by the late 8th millennium BC and the west by early 7th millennium BC (Arbuckle & Atici 2013; Arbuckle et al. 2014). There have been previous suggestions that haplogroup C was incorporated into domestic or managed herds separately (Naderi et al. 2008; Colli et al. 2015), and it is possible that either haplogroup was introduced to managed herds in Turkey via introgressive capture (Larson & Fuller 2014).

In the eastern part of the study region, a diversity of haplogroups are observed. B, G, and D were observed in sites across the Zagros and as far east as Turkmenistan, while haplogroups A and C were not detected. This is despite reports of haplogroup A in the Neolithic Zagros site of Chia Sabz (Mazdarani et al. 2014). The failure here to detect haplogroup A in the Zagros can be considered as additional evidence against this publication's findings, although a broader range of samples is required before a definitive conclusion is reached. In contrast to the western region, the haplogroup frequencies of eastern Neolithic goat mtDNA are quite different to that of modern goat, where haplogroup A is now found at a frequency of 93.99% (Colli et al. 2015). The haplogroups B, D, and G, common in Neolithic goat from the region, appear at a very low rate in modern Iran populations, with only G occurring in >1% of animals (Naderi et al. 2007). The absence of haplogroup C in eastern Neolithic goat is somewhat perplexing, but highlights the issue of using modern wild distributions to infer those of the past.

As the samples from Iran and Turkmenistan are of a wider range of ages than those from the Levant or the western Turkey/southeastern Europe, some information regarding their origin can be derived by the pattern of clade formation. Two samples from some of the earliest Pre-Pottery Neolithic sites in the Zagros, Lur9 and Lur12, are roughly older (Table 2.1) than the earliest phases of Ganj Dareh where the earliest evidence of goat management is found (Zeder 2005). Although it is likely that the practice existed prior to the earliest observed evidence, it is also possible that the samples are hunted individuals, distinct from the population first managed. The two sequences do not show the shallow branching observed in later and modern samples; the low coverage of the Lur9 mtDNA and bootstrap support challenges interpretation in this regard. The single sample from the earliest phase of Sang-e Chakmaq in eastern Iran (Semnan1-2, dating to 7,454-6,850 cal BC) groups in the same clade as Lur9 (with a later Neolithic sample from Iran, Fars2-5, and a Bronze Age samples from Uzbekistan, Bulak4) but these ancient sequences form a clade with the modern B domestic individuals (Figure 3.3). In contrast, D haplogroup sequences from *c*. 6,000 BC from Sang-e Chakmaq and Monjukli Depe (Turkmenistan) show shallow branching and clearly group together, along with later ancient and modern mtDNA sequences. This is suggestive of a restricted maternal lineage ancestral to these samples. An explanation for the difference in patterns may be that the earliest stages of management did not result in as strong a maternal bottleneck that were the consequences of later stages of domestication.

In the Levant, a distinctly different mtDNA lineage is observed in Neolithic goat. Four samples from ‘Ain Ghazal, Jordan, and one from the nearby Abu Ghosh, Israel, carry haplogroup F. This haplogroup has only been reported in bezoar and three goats from Sicily (Sardina et al. 2006), and as such its designation as a “domestic” haplotype is tenuous. Its absence elsewhere in the Neolithic, and deep divergence from other mtDNA lineages (noting the issue of mtDNA divergence estimates discussed in Chapter 1), suggests that haplogroup F was incorporated into herds by early Levantine farmers. Demographic modeling also supports a distinct, pre-Neolithic origin of this Levantine population (Figure 3.6). The Mid PPNB goat from Abu Ghosh have been interpreted as being managed or in the early phase of domestication, based on morphometric similarity to wild goat, evidence of foddering, and selective culling resulting in 30% of goat surviving to adulthood (Ducos & Horwitz 1995; Horwitz 2003; Makarewicz & Tuross 2012). Hunting of wild caprines is thought to have been practised, based on DNA evidence of *Capra ibex* from the site (Bar-Gal et al. 2002).

Similarly, the ‘Ain Ghazal Middle PPNB goat are thought to be have been under some degree of management, with evidence of the persistence of hunting (Von den Driesch & Wodtke 1997; Wasse 2002; Martin & Edwards 2013). Although it is possible that the five sequences here are derived from solely-hunted populations, the clade of ‘Ain Ghazal sequences is suggestive of a restricted maternal ancestry of that group. The evidence of management at the two sites and distinct mtDNA profile can be interpreted as the application of developing, incipient animal management strategies to local goat populations. In the absence of data from the nuclear genome, there is no evidence from mtDNA that these Levantine managed goat contributed to modern mitochondrial diversity, as haplogroup F has been found in just three present-day animals (Sardina et al. 2006).

A major limitation in interpreting these Neolithic distributions is the absence of sequences from central/eastern/southeastern Turkey, north Iraq, and northeast Syria. The lack of these samples may contribute to the discontinuous distribution of mtDNA in goat observed here. Samples from PPNA Körtik Tepe and Epipaleolithic Hallan Çemi were screened, but were either sheep or did not have preserved endogenous DNA (Appendix Table 2.1). A time series of goat from such sites, and also more west in the Upper Euphrates-Tigris basin, would address the question as to when and where the diversity of mtDNA haplogroups arose. If only haplogroups A and C were observed, it would support distinct maternal origins of eastern and Levantine goat populations in the Neolithic (ignoring the question of wild restocking or independent domestication). A greater diversity of haplogroups might suggest that some were

sequentially gained or lost as goat herding spread throughout southwest Asia. The overlapping distributions for the domestic lineages node coalescent times (Table 3.2) also demonstrates the limitation in such inferences using mtDNA alone, as such divergence estimates have large variances (Hudson et al. 2003). This challenges their use in determining the timing of the introduction of haplogroups into the domestic population, if haplogroups are treated as equivalent to populations.

The mtDNA data here suggests that the goat kept by the earliest farmers of Iran and Turkmenistan derived from a distinct population to that of Neolithic Turkish/European and Levantine goat. Alternatively, they may share a common ancestral population from the beginning of the Neolithic and herd management, but this was obfuscated by the later addition of local wild goat into herds as the practice spread throughout southwest Asia. Differentiating between these two hypotheses is difficult using mtDNA alone, and will be addressed using nuclear genome data in Chapter 4. A third possibility is that the domestication process was applied to a single wild population which contained the mtDNA haplogroups observed here (A, B, C, D, G, and possibly F), some of which were lost due to genetic drift as the goat herding spread throughout the Fertile Crescent. Demographic modelling using an ABC framework did not support this explanation (Table 3.4), rather suggesting inputs from distinct populations into the goat mtDNA pool.

mtDNA homogeneity following the Neolithic

The geographic structuring of mtDNA haplogroups changes drastically from the Chalcolithic onwards. Haplogroup A appears in Chalcolithic goat from all areas of study: south in Levant at Shiqmim and Gilat; north at Soha Chay Tepe in Iranian Azerbaijan; in central Iranian sites such Rahmat Abad and Darre-ye Bolāghi; and at Chalcolithic levels at Monjukli Depe, Turkmenistan. This pattern persists in the Bronze Age and later periods, including sites in the Levant (Tel Yarmuth and Tel Yoqne'am), the Caucasus region (Tachti Perda, Tepe Hasanlu), Iran (Chalow and Tepe Chizar), and Uzbekistan (Tilla Bulak). Although there is some persistence of the haplogroups observed in the Neolithic in Iran and Turkmenistan, that is not the case in the Levant. There is no evidence of haplogroup F following the Mid PPNB, with a complete replacement of mtDNA haplogroups in the intervening ~3,000 years (7,500-4,500 BC).

While apparent from visual inspection of the haplogroup distributions (Figure 3.4), AMOVA analysis also supports homogeneity in time periods following the Neolithic, when molecular variation is primarily explained by within population differences rather than between group (Table 3.3). Modelling strongly supports a model in which the Neolithic structure of mtDNA breaks down due to migration between regions (Appendix Table 3.9). The relative haplogroup frequencies are similar to those observed in the present, with haplogroup A predominating in all geographic regions. Continuity of the mtDNA gene pool appears to begin in the Chalcolithic time-period and extends to the present-day.

As haplogroup A was not observed in Levantine or Iranian sites prior to the Chalcolithic, a natural question to ask is from where haplogroup A was introduced to Iran and the Levant. The potential sources of this haplogroup are the sampled and unsampled regions west of the Zagros: Syria, Iraq, and Turkey, where haplogroup A is found in westerly sites. As Neolithic goat from the Caucasus region are reported to have haplogroup A (Al-Araimi et al. 2017), that area is also a possible source. It is also possible that multiple regions were involved *i.e.* haplogroup A being introduced to the Levant from Turkey, and to Iran from the Caucasus region. These more complex scenarios are difficult to distinguish with mtDNA evidence alone, and will be discussed in the context of nuclear data in Chapter 4.

Evidence from human aDNA supports population movement and admixture in the periods following the Neolithic. The genomes of early farmers were highly differentiated, with a large divergence between Iranian farmers and Anatolian/Levantine (Lazaridis et al. 2016; Broushaki et al. 2016). This differentiation decreases in the Chalcolithic when sampled Iranians (4,500-3,500 BC) are best modelled as a mixture of Neolithic Iranians, Caucasus hunter-gatherers, and Levantine farmers, indicating a large degree of population movement and admixture at or before that period (Lazaridis et al. 2016). Consistent with this, Bronze Age Levantine humans can be modelled as a combination of Neolithic Levantine and Chalcolithic Iranian genomes, though the admixture date estimates for this are wide (Haber et al. 2017). In the absence of Chalcolithic Levant and Mesopotamian human genomes it is difficult to determine the chronological order of this admixture, if indeed there is an order. It is plausible that the movement of humans was accompanied by the movement of goat, whether by demographic expansion or trade. The asymmetric nature of the goat mtDNA turnover (*i.e.* increase in frequency of A) suggests this movement may have originated in a region close to or in Turkey, where A is inferred to have been at high frequency.

There is some movement of the "eastern" haplogroups B, D, and G westwards, based on their haplogroup distributions. Haplogroup D is observed in two individuals from Chalcolithic Shiqmim (4,300-3,700 BC), and single individual from Chalcolithic Gilat (4,500-4,200 BC). Haplogroup G is found in Pietrele, Romania, *c.* 4,450-4,250 BC. This would suggest far-ranging exchange and movement of goat by humans; alternatively, both haplogroups could have been introduced to these regions via wild admixture. Interestingly, if wild admixture is not the source of the non-A haplogroups in the Levant, then Chalcolithic Levant goat populations are the product of at least two demographic events: admixture between western and eastern populations, and replacement of the Neolithic Levantine goat population. As mtDNA represents only one genetic locus because of its non-recombining nature, it does not contain enough information to accurately infer admixture events (Larson & Fuller 2014), this will be addressed more thoroughly in Chapter 4.

Epipaleolithic bezoar from Southern Turkey show affinity with *Capra caucasica*

The discovery of mtDNA with high affinity to the West Caucasian Tur, *Capra caucasica*, in Direkli Cave in the Taurus Mountains of southern Turkey was unexpected. This clade was labelled haplogroup T for convenience. The surprise of this finding is due to the fact that *Capra caucasica* and its sister species *Capra cylindricornis* (the East Caucasian Tur) are now found only in the Caucasus Mountains (Wilson & Mittermeier 2011). Direkli Cave, thought to have been seasonally occupied by Epipaleolithic hunters and used as a slaughter site for wild goat, is approximately 800km from the Greater Caucasus (Arbuckle & Erek 2012). Two of the four samples have been directly dated to 13th-12th millennium BC (Table 2.1), making contamination from a later or modern time period unlikely.

Three of the four mtDNA recovered from the site form a clade with the *Capra caucasica* reference sequence, which is well supported in maximum likelihood and Bayesian phylogenies (Figures 3.2 and 3.5). There is no *Capra cylindricornis* mitogenome sequence currently available, so it is unclear how the Direkli sequences relate to it compared to the *Capra caucasica* reference mitochondria. The large divergence time estimated between the T clade and the *Capra caucasica* sequence (Table 3.2) is also worth noting. The T clade may coalesce more recently with the *C. cylindricornis* mitochondria; alternatively, it may be an outgroup to the two Tur mitogenomes. In the latter case, the divergence time range of 137-201 kya would put the split of the two mtDNA lineages as securely prior the Last Glacial Maximum.

Assuming the veracity of these mtDNA results and their ages, two explanation are possible. The first is that the distribution of *Capra* species differed greatly in the (relatively recent) past compared to now. It is possible, if unlikely, that a *Capra* population related to the Tur existed in the Taurus Mountains in the late Pleistocene. There have been reported of *Capra caucasica* remains in Late Pleistocene deposits in France, but these results are difficult to assess (Rivals 2006). There is no fossil/ancient evidence of Tur remains outside of the Caucasus region in the late Pleistocene (Uerpmann 1987; Weinberg 2002), supporting at best a discontinuous distribution. It is possible that these *Capra* species could have co-existed with *Capra aegagrus*, which is also found in the Taurus Mountains; bezoar have a different altitude preference than Tur in the Caucasus Mountains, where there habitats are close and occasionally overlap (Groves & Grubb 2011). Hybridization may have occurred, with a mixing of mtDNA lineages between the species.

The second explanation is that our knowledge of *Capra aegagrus* and *Capra caucasica/ cylindricornis* genetic diversity is poor. Many of the bezoar control region sequences reported (Naderi et al. 2008) are undefined and have not had complete mitochondria sequenced. It may be the case that modern-day bezoar from the Taurus Mountains also have haplogroup T, and the mitogenome has simply not been sequenced yet. Additionally, only a single Tur mitogenome has been reported as of 2018 (Hassanin et al. 2012). Tur can form hybrids with bezoar and may back- cross (Weinberg 2002; Groves & Grubb 2011), and so may be admixed with bezoar. The declining population size for both Tur species would exacerbate this introgression issue (Weinberg 2008). In this case, the available Tur mitogenome would represent an introgressed *Capra aegagrus* mitochondria. Alternatively, this may be a case of incomplete lineage sorting: the T haplogroup may be specific to bezoar, but appears more closely related to Tur due to random chance. These samples will be addressed further in Chapters 4 and 6.

Limitations

Significant limitations are inherent to mtDNA analyses, and have been discussed in Chapter 1. In brief, single loci analyses are less robust than those which utilize multiple loci, due to genetic drift; estimates of divergence times between populations using mtDNA can be affected by the large variances in coalescence times between loci; and the possibility of failure to capture admixture events. All inferences in this chapter should be considered with

these in mind; mtDNA were targeted to best exploit the available modern data and poor preservation of some samples.

There are considerable paucities to the available bezoar mitochondrial sequences. The absence of a more complete dataset of bezoar from across their distribution is problematic when attempting to interpret how early Neolithic goat mitochondria relate to those of modern wild goat. Additional bezoar sequences may indicate if there are indisputable domestic clades, which was claimed in Colli et al. (2015) but based on a handful of wild sequences. The observation here of a novel goat clade G', and the Tur-like clade T, would benefit from the context of a geographical-broad sampling of bezoar mitochondrial genomes. This applies also to low number of *Capra caucasica/cylindricornis* mitogenomes.

There are some technical limitations in the analyses presented. The use of ancient samples to calibrate the mutation rate would have benefited from a greater number of time points, across a broad range of time depths and branches of the tree (Rieux & Balloux 2016). It may have been correct to down-sample haplogroup A sequences to avoid tree imbalance, which can bias date estimates downwards in heterochronous dataset employing tip date calibration (D. Duchêne et al. 2015). Testing for a temporal signal in the dataset, for example using date randomization (S. Duchêne et al. 2015), would have confirmed that using the radiocarbon-dated samples for rate calibration was appropriate. If no temporal signal is detected then tip date calibration is inappropriate for the dataset, and would lead to erroneous rate estimation.

Bootstrap support and posterior probabilities for the ML and Bayesian phylogenies were low in many cases. This was particularly observed in haplogroup A, in which very few clades were well supported in any phylogeny. Bootstrap values were not reported in the publication containing the majority of the modern mitochondrial sequences used, making it difficult to assess if a difference in phylogeny reconstruction was responsible (Colli et al. 2015). It is possible that residual error from deamination resulted in incorrect clade formation during bootstrapping, but this should have affected non-A haplogroups also, which here show higher support than haplogroup A. Low bootstrap values within haplogroup A were reported in a ML phylogeny using a similar dataset (Cassidy et al. 2017), suggesting that it was not the presence of ancient sequences resulting in low support. The low degree of differentiation among haplogroup A sequences that likely contributes to the low bootstrap indicates a rapid population expansion in their past (Watson et al. 1997), which is observed in the spread of the haplogroup (Figure 3.4).

Conclusion

A disjointed distribution of goat mtDNA haplogroups is observed during the Neolithic. Iran, the Levant, and western Turkey/southeast Europe have distinct mtDNA profiles with no overlap. The likely origin of haplogroup A is central/eastern Turkey, from which goat and animal economies likely spread westward. In Iran and Turkmenistan, haplogroups now rare in domestic goat are found at high frequencies in the Neolithic. And in Neolithic Levantine sites, a haplogroup effectively absent in modern domestic goat was found, suggesting a genetic ancestry not present in modern populations. These measurably-different Neolithic goat herds are suggestive of unique maternal histories, and supports a model of goat domestication that is diffuse with strong local contributions.

A dramatic change in mtDNA frequencies is observed from the Chalcolithic onwards (*c.* 4,500 BC). Haplogroup A appears in a high proportion of goat remains in the Iran eastwards, in the Caucasus region, and in the Levant. Some "eastern" haplogroups appear at a low frequency in the Levant and in southeast Europe, whereas the Levantine F haplogroup mtDNA are not observed in any region. The predominance of haplogroup A is broadly consistent with modern haplogroup frequencies, suggesting that the processes responsible for this mtDNA turnover are responsible for the establishment of the modern goat mtDNA gene pool.

Finally, a mitochondrial lineage similar to the Tur (*Capra caucasica*) is found in remains from the Epipaleolithic site of Direkli Cave in the Taurus Mountains of southern Turkey. This finding highlights the deficit in our understanding of the genetics of wild *Capra*, and the limits to inference with mitochondrial data alone. Although mtDNA enrichment has allowed poorly-preserved samples to be analyzed in the context of relatively large global datasets, the process has raised many questions that are best addressed with multi-locus or whole genome data, which will be the focus of the following chapters.

Chapter 3 References

Akis, I., Onar, V., et al., 2014. Ancient DNA Analysis of Anatolian Goat Remains Excavated from a Urartian Castle in Eastern Turkey. *International Journal of Osteoarchaeology*. Available at: http://dx.doi.org/10.1002/oa.2415.

Akis, I., Oztabak, K., et al., 2014. Mitochondrial DNA diversity of Anatolian indigenous domestic goats. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 131(6), pp.487–495.

Al-Araimi, N.A. et al., 2017. Genetic origin of goat populations in Oman revealed by mitochondrial DNA analysis. *PloS one*, 12(12), p.e0190235.

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), pp.2906–2910.

Arbuckle, B. et al., 2014. Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PloS one*, 9(6), p.e99845.

Arbuckle, B.S. & Atici, L., 2013. Initial diversity in sheep and goat management in Neolithic south-western Asia. *Levantina*, 45(2), pp.219–235.

Arbuckle, B.S. & Erek, C.M., 2012. Late Epipaleolithic hunters of the central Taurus: Faunal remains from Direkli Cave, Kahramanmaraş, Turkey. *International Journal of Osteoarchaeology*, 22(6), pp.694–707.

Awotunde, E.O. et al., 2015. Mitochondrial DNA sequence analyses and phylogenetic relationships among two Nigerian goat breeds and the South African Kalahari Red. *Animal biotechnology*, 26(3), pp.180–187.

Azor, P.J. et al., 2005. Phylogenetic relationships among Spanish goats breeds. *Animal genetics*, 36(5), pp.423–425.

Bar-Gal, G., Ducos, P. & Horwitz, L., 2003. The application of ancient DNA analysis to identify neolithic caprinae: a case study from the site of Hatoula, Israel. *International Journal of Osteoarchaeology*, 13(3), pp.120–131.

Bar-Gal, G.K. et al., 2002. Ancient DNA Evidence for the Transition from Wild to Domestic Status in Neolithic Goats: A Case Study from the Site of Abu Gosh, Israel. *Ancient biomolecules*, 4(1), pp.9–17.

Beaumont, M.A., 2008. Joint determination of topology, divergence time, and immigration in population trees. In Matsumura, S., Forster, P. and Renfrew, C, ed. *Simulation, genetics and human prehistory*. Cambridge: McDonald Institute, pp. 134–154.

Beaumont, M.A., Zhang, W. & Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics*, 162(4), pp.2025–2035.

Bouckaert, R. et al., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, 10(4), p.e1003537.

Broushaki, F. et al., 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science*. Available at: http://dx.doi.org/10.1126/science.aaf7943.

Carroll, R.L., 1988. *Vertebrate Palaeontology and Evolution*, New York: W.H. Freeman and Company.

Cassidy, L.M. et al., 2017. Capturing goats: documenting two hundred years of mitochondrial DNA diversity among goat populations from Britain and Ireland. *Biology letters*, 13(3). Available at: http://dx.doi.org/10.1098/rsbl.2016.0876.

Colli, L. et al., 2015. Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC genomics*, 16(1), p.1115.

Csilléry, K., François, O. & Blum, M., 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution / British Ecological Society*.

Daly, K.G. et al., 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, 361(6397), pp.85–88.

Drummond, A.J. et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, 29(8), pp.1969–1973.

Drummond, A.J. et al., 2006. Relaxed phylogenetics and dating with confidence. *PLoS biology*, 4(5), p.e88.

Drummond, A.J. & Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7, p.214.

Duchêne, D., Duchêne, S. & Ho, S.Y.W., 2015. Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. *Molecular ecology resources*, 15(4), pp.785–794.

Duchêne, S. et al., 2015. The Performance of the Date-Randomization Test in Phylogenetic Analyses of Time-Structured Virus Data. *Molecular biology and evolution*, 32(7), pp.1895–1906.

Ducos, P. & Horwitz, L.K., 1995. Pre-Pottery Neolithic B fauna from the Lechevallier excavations at Abu-Ghosh. In H. Khalaily & O. Marder, eds. *The Neolithic Site of Abu Ghosh: The 1995 Excavations*. Israel Antiquities Authority Reports. Jerusalem: Israel Antiquities Authority, pp. 103–120.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–1797.

E, G.-X. et al., 2018. Genetic diversity of the Chinese goat in the littoral zone of the Yangtze River as assessed by microsatellite and mtDNA. *Ecology and evolution*, 8(10), pp.5111–5123.

Excoffier, L. et al., 2013. Robust demographic inference from genomic and SNP data. *PLoS genetics*, 9(10), p.e1003905.

Excoffier, L. & Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology*

resources, 10(3), pp.564–567.

Excoffier, L., Smouse, P.E. & Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), pp.479–491.

Felsenstein, J., 1978. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. *Systematic zoology*, 27(4), pp.401–410.

Felsenstein, J., 2004. *Inferring phylogenies*, Sunderland, MA: Sinauer Associates.

Fernandez, H. et al., 2005. Assessing the origin and diffusion of domestic goats using ancient DNA. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow, pp. 50–54.

Fernández, H. et al., 2006. Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Sciences of the United States of America*, 103, pp.15375–15379.

Fisher, R.A., 1921. On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron. International Journal of Statistics*, 1, pp.3–32.

Gouy, M., Guindon, S. & Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*, 27(2), pp.221–224.

Groves, C. & Grubb, P., 2011. *Ungulate Taxonomy*, Baltimore: JHU Press.

Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–321.

Haber, M. et al., 2017. Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *American journal of human genetics*, 101(2), pp.274–282.

Han, L. et al., 2010. Mitochondrial DNA analysis provides new insights into the origin of the Chinese domestic goat. *Small ruminant research: the journal of the International Goat Association*, 90(1–3), pp.41–46.

Hansen, H.B. et al., 2017. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PloS one*, 12(1), p.e0170940.

Hassanin, A. et al., 2010. Comparisons between mitochondrial genomes of domestic goat (Capra hircus) reveal the presence of numts and multiple sequencing errors. *Mitochondrial DNA*, 21, pp.68–76.

Hassanin, A. et al., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes rendus biologies*, 335(1), pp.32–50.

Hofmanová, Z. et al., 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United*

States of America, 113(25), pp.6886–6891.

Horwitz, L.K., 2003. The Neolithic fauna. In H. Khalaily & O. Marder, eds. *The Neolithic Site of Abu Gosh. The 1995 Excavations.* Jerusalem: Israel Antiquities Authority Reports, pp. 87–101.

Ho, S.Y.W. et al., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular biology and evolution*, 22(7), pp.1561–1568.

Ho, S.Y.W. et al., 2011. Time-dependent rates of molecular evolution. *Molecular ecology*, 20(15), pp.3087–3101.

Ho, S.Y.W. & Larson, G., 2006. Molecular clocks: when times are a-changin'. *Trends in genetics: TIG*, 22(2), pp.79–83.

Ho, S.Y.W. & Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic biology*, 58(3), pp.367–380.

Hudson, R.R., Slatkin, M. & Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132(2), pp.583–589.

Hudson, R.R., Turelli, M. & Whitlock, M., 2003. Stochasticity Overrules The "Three-Times Rule": Genetic Drift, Genetic Draft, And Coalescence Times For Nuclear Loci Versus Mitochondrial DNA. *Evolution*, 57(1), pp.182–190.

Hughes, S. et al., 2012. A dig into the past mitochondrial diversity of Corsican goats reveals the influence of secular herding practices. *PloS one*, 7(1), p.e30272.

Joshi, M.B. et al., 2004. Phylogeography and origin of Indian domestic goats. *Molecular biology and evolution*, 21(3), pp.454–462.

Kadowaki, S. et al., 2016. Mitochondrial DNA Analysis of Ancient Domestic Goats in the Southern Caucasus: A Preliminary Result from Neolithic Settlements at Göytepe and Hacı Elamxanlı Tepe. *International Journal of Osteoarchaeology*. Available at: http://dx.doi.org/10.1126/science.aaf7943.

Keane, T.M. et al., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology*, 6, p.29.

Kibegwa, F.M. et al., 2016. Mitochondrial DNA variation of indigenous goats in Narok and Isiolo counties of Kenya. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 133(3), pp.238–247.

Lanfear, R. et al., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, 29(6), pp.1695–1701.

Larson, G. & Fuller, D.Q., 2014. The Evolution of Animal Domestication. *Annual review of ecology, evolution, and systematics*, 45(1), pp.115–136.

Lazaridis, I. et al., 2014. Ancient human genomes suggest three ancestral populations for

present-day Europeans. *Nature*, 513(7518), pp.409–413.

Lazaridis, I. et al., 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617), pp.419–424.

Lin, B.Z. et al., 2013. Molecular phylogeography and genetic diversity of East Asian goats. *Animal genetics*, 44(1), pp.79–85.

Luikart, G. et al., 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America*, 98, pp.5927–5932.

Luikart, G. et al., 2006. Origins and diffusion of domestic goats inferred from DNA markers. *Documenting Domestication: New Genetic and Archaeological Paradigms*, pp.294–305.

Makarewicz, C. & Tuross, N., 2012. Finding Fodder and Tracking Transhumance: Isotopic Detection of Goat Domestication Processes in the Near East. *Current anthropology*, 53(4), pp.495–505.

Martin, L. & Edwards, Y., 2013. Diverse Strategies: evaluating the appearance and spread of domestic caprines in the southern Levant. In S. Colledge et al., eds. *Origins and Spread of Domestic Animals in Southwest Asia and Europe*. Left Coast Press, pp. 49–82.

Mazdarani, F.H. et al., 2014. Molecular Identification of Capra Hircus in East Chia Sabz, an Iranian Pre-Pottery Neolithic Site, Central Zagros, based on mtDNA. *The Journal of Animal & Plant Sciences*, 24(3), pp.945–950.

McCracken, K. & Sorenson, M., 2005. Is homoplasy or lineage sorting the source of incongruent mtdna and nuclear gene trees in the stiff-tailed ducks (Nomonyx-Oxyura)? *Systematic biology*, 54(1), pp.35–55.

Molak, M. et al., 2013. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Molecular biology and evolution*, 30(2), pp.253–262.

Naderi, S. et al., 2007. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PloS one*, 2, p.e1012.

Naderi, S. et al., 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 105, pp.17659–17664.

O'Sullivan, N.J. et al., 2016. A whole mitochondria analysis of the Tyrolean Iceman's leather provides insights into the animal sources of Copper Age clothing. *Scientific reports*, 6, p.31279.

Pereira, F. et al., 2005. The mtDNA catalogue of all Portuguese autochthonous goat (Capra hircus) breeds: High diversity of female lineages at the western fringe of European distribution. *Molecular ecology*, 14, pp.2313–2318.

Pereira, F. et al., 2009. Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. *Molecular biology and*

evolution, 26, pp.2765–2773.

Porter, V. & Tebbit, J., 1996. *Goats of the World*, London: Farming Press Ltd.

Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* , 16(4), pp.395–399.

Rambaut, A., 2009. FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Suchard MA, Xie D & Drummond AJ, 2014. Rambaut et al. 2014. *Tracer v1.6, Available from http://tree.bio.ed.ac.uk/software/tracer*. Available at: http://beast.community/tracer.

R Core Team, 2015. *R: A language and environment for statistical computing*, Vienna, Austria. Available at: https://www.R-project.org/.

Reimer, P.J. et al., 2004. Intcal04 Terrestrial Radiocarbon Age Calibration, 0–26 Cal Kyr BP. *Radiocarbon*, 46(3), pp.1029–1058.

Rieux, A. et al., 2014. Improved calibration of the human mitochondrial clock using ancient genomes. *Molecular biology and evolution*, 31(10), pp.2780–2792.

Rieux, A. & Balloux, F., 2016. Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular ecology*, 25(9), pp.1911–1924.

Rivals, F., 2006. Découverte de Capra caucasica et d'Hemitragus cedrensis (Mammalia, Bovidae) dans les niveaux du Pléistocène supérieur de la Caune de l'Arago (Tautavel, France): implication biochronologique dans le contexte du Bassin Méditerranéen. *Geobios . Memoire special*, 39(1), pp.85–102.

Royo, L.J. et al., 2009. Analysis of mitochondrial DNA diversity in Burkina Faso populations confirms the maternal genetic homogeneity of the West African goat. *Animal genetics*, 40(3), pp.344–347.

Sardina, M.T. et al., 2006. Phylogenetic analysis of Sicilian goats reveals a new mtDNA lineage. *Animal genetics*, 37(4), pp.376–378.

Scheu, A. et al., 2012. The arrival of domesticated animals in South-Eastern Europe as seen from ancient DNA. In E. Kaiser, J. Burger, & S. Wolfram, eds. *New Approaches Using Stable Isotopes and Genetics*. Berlin, Boston: De Gruyter, pp. 45–54.

Schlumbaum, A. et al., 2010. Ancient DNA, a Neolithic legging from the Swiss Alps and the early history of goat. *Journal of archaeological science*, 37(6), pp.1247–1251.

de Smet, K. et al., 2008. Capra aegagrus. *The IUCN Red List of Threatened Species*. Available at: http://www.iucnredlist.org/details/3786/0.

Sultana, S., Mannen, H. & Tsuji, S., 2003. Mitochondrial DNA diversity of Pakistani goats. *Animal genetics*, 34(6), pp.417–421.

Tarekegn, G.M. et al., 2018. Mitochondrial DNA variation reveals maternal origins and demographic dynamics of Ethiopian indigenous goats. *Ecology and evolution*. Available at: http://dx.doi.org/10.1002/ece3.3710.

Thirstrup, J.P. et al., 2009. Population viability analysis on domestic horse breeds (Equus caballus). *Journal of animal science*, 87(11), pp.3525–3535.

Uerpmann, H.-P., 1987. *The ancient distribution of ungulate mammals in the Middle East: fauna and archaeological sites in Southwest Asia and Northeast Africa*, Wiesbaden: L. Reichert Verlag.

Von den Driesch, A. & Wodtke, U., 1997. The fauna of 'Ain Ghazal, a major PPN and early PN settlement in central Jordan. *The Prehistory of Jordan II. Perspectives from*, 1997, pp.511–556.

Wasse, A., 2002. Final Results of an Analysis of the Sheep and Goat Bones from Ain Ghazal, Jordan. *Levantina*, 34(1), pp.59–82.

Watson, E. et al., 1997. Mitochondrial footprints of human expansions in Africa. *American journal of human genetics*, 61(3), pp.691–704.

Weinberg, P., 2008. Capra caucasica. The IUCN Red List of Threatened Species 2008. *IUCN*. Available at: http://www.iucnredlist.org/details/3794/0.

Weinberg, P.J., 2002. Capra cylindricornis. *Mammalian Species*, pp.1–9.

Wilson, D.E. & Mittermeier, R.A. eds., 2011. *Handbook of the Mammals of the World. Vol 2. Hoofed mammals*, Barcelona: Lynx Edicions.

Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends in ecology & evolution*, 11(9), pp.367–372.

Yang, Z. & Rannala, B., 2012. Molecular phylogenetics: principles and practice. *Nature reviews. Genetics*, 13(5), pp.303–314.

Zeder, M.A., 2005. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 125–146.

Zeder, M.A. & Hesse, B., 2000. The initial domestication of goats (Capra hircus) in the Zagros mountains 10,000 years ago. *Science* , 287, pp.2254–2257.

Zhao, Y. et al., 2014. Genetic diversity and molecular phylogeography of Chinese domestic goats by large-scale mitochondrial DNA analysis. *Molecular biology reports*, 41(6), pp.3695–3704.

Виктор, В., 2010. *Relief Map of Middle East*, Available at: https://commons.wikimedia.org/wiki/File:Relief_Map_of_Middle_East.jpg.

Chapter 4: Ancient goat nuclear DNA

Introduction

Modern goat nuclear genetic diversity

Our understanding of goat autosomal variation lags behind that of other livestock animals such as cattle or pig. Microsatellites were initially used as the primary tool for assessing nuclear genome patterns of diversity, and suggested greater geographic structure was present in the nuclear genome compared to mtDNA. From microsatellite data, European goat show a cline of decreasing genetic diversity from southeast to northwest, consistent with their introduction from southwest Turkey in the Neolithic (Cañón et al. 2006). Studies of Chinese (Yang et al. 1999; Xiang-Long & Valentini 2004), southeast Asian/Australian (Barker et al. 2001), Indian (Ganai & Yadav 2001), Swiss (Saitbekova et al. 1999), European and Middle Eastern (Cañón et al. 2006) breeds/populations indicated that individuals could often be assigned to their source population correctly based on a small number of microsatellite markers. Microsatellites remain a common method of population analysis of goat (Nafti et al. 2016; Lenstra et al. 2017).

The publishing of the first version of the goat genome (Dong et al. 2013), CHIR_1.0, was an important milestone for goat genomics. Prior to this, goat genetic variability in the nuclear genome was poorly characterized. Studies that did not use microsatellite variation instead relied on a handful of SNP sites (Cappuccio et al. 2006; Pariset et al. 2006; Pariset et al. 2009), some of which were identified by gene-specific sequencing and as such were not appropriate as neutral markers (Chessa et al. 2003; Criscione et al. 2007). CHIR_1.0 was used to design a 50K SNP chip, incorporating 50,000 variable sites found across six different breeds (Tosser-Klopp et al. 2014). Despite SNPs being ascertained from a small number of regions and breeds, the 50K chip has been applied to studies of regional diversity; no global dataset has been published as of writing. These breed- or regional-level analyses have indicated genetic differentiation between regional goat populations, for example between South African breeds (Mdladla et al. 2016; Lashmar et al. 2016), Spanish breeds (Mdladla et al. 2016), Swiss (Burren et al. 2016), Canadian and Australian (Kijas et al. 2013; Brito et al. 2017), and between regional Angora breed populations (Visser et al. 2016). This relatively recent structure, evidence for admixture (*e.g.* between Andalusian and African goat), and

SNP ascertainment bias complicates the use of the available SNP chip data to infer past demographic events.

Modern goat nuclear genomes

A limited number of goat whole genomes have been published. A large survey of Moroccan goat by whole genome sequencing indicated a small degree of differentiation between breeds indicated by low F_{ST} values, diverse breed ancestries as modelled by ADMIXTURE, and a lack of clear breed separation in Principal Component Analysis (Benjelloun et al. 2015). The lack of geographic structure observed in Moroccan goats likely reflects intra-regional gene flow between breeds.

A more recent study generated a whole genome dataset of European, Iranian, Moroccan, and Australian domestic goat, as well as Iranian bezoar, that is better suited to addressing goat domestication (Alberto et al. 2018). Surprisingly, modern Iranian bezoar show lower nucleotide diversity and higher inbreeding that domestic Iranian and Moroccan goat, as well as a higher genetic load then any domestic group. Estimation of the effective population size through time using MSMC (Schiffels & Durbin 2014) indicated that the different goat and bezoar groups have distinct demographic histories, diverging roughly at the time of domestication. Iranian bezoar and all domestic goat formed sister monophyletic clades in a Treemix analysis (Pickrell & Pritchard 2012), consistent with a single population origin of goat. No admixture events were detected between domestic and wild goat, based on Treemix residuals and non-negative f_3 values between all individuals, noting that drift since admixture can result in a failure to detect older gene flow using f_3 statistics (Patterson et al. 2012; Reich et al. 2009).

Despite lacking wild and domestic genomes from some of the most important regions relevant to the dissection of goat domestication (i.e. Turkey; the Levant), these genomes are the best resources currently available for ancient-modern whole genome comparisons.

Aims

The aims are this chapter are:

1. To investigate how Neolithic goat populations relate to each other, and to pre-domestic bezoar populations.
2. To test if post-Neolithic goat populations show different genetic affinities to Neolithic goat from similar geographic regions.
3. To estimate an autosomal *Capra hircus* mutation rate.
4. To build a framework of the population history of the domestic goat, *Capra hircus*.

Material and Methods

Material

Ancient data

62 *Capra* specimens identified in Chapter 2 were selected for high-throughput nuclear genome sequencing, and produced genomes with mean coverages in the range 0.001× to 14.89×. 52 were of >0.01× mean coverage, and median ancient genome coverage was 1.05×. Sample identifiers, geographic and contextual information, and nuclear genome coverage are displayed in Table 4.1. Two of the ancient samples sequenced, Bulak1 and Bulak3, displayed the criteria for having come from the same individual (see Chapter 2) despite deriving from petrous bones of the same orientation. As such, Bulak3 was excluded from all analyses.

Modern data

In addition to the ancient genomes sequenced and aligned in Chapter 2, modern wild and domestic goat from France, Iran and Morocco were selected for inclusion in analyses (Benjelloun et al. 2015; Alberto et al. 2018). Due to the high number of Moroccan goat genomes, a random subset of nine were chosen. To include an East Asian goat genome, the individual used to generate CHIR_1.0 was selected for realignment to itself (Dong et al. 2013).

To supplement this dataset, two modern goat from Ireland and Togo were selected for whole genome sequencing. These were IOG, an individual from an Irish feral *Capra hircus* population, and Tog, a Togolese village goat. Sequencing was performed on an Illumina HiSeq 4000, pair end, read length 2x150bp.

To generate an ancestral genome sequence, reads used to generate the Yak reference genome BosGru_v2.0 were downloaded. A summary of the modern genomes selected for alignment is displayed in Appendix Table 4.1.

Methods

Whole genome data processing - modern samples

Modern samples were aligned to reference genome CHIR_1.0 (Dong et al. 2013) using bwa mem (Li 2013), with mate information of paired end reads filled in using samtools fixmate (Li et al. 2009). Duplicates were marked and removed using MarkDuplicates function of Picard Tools (The Broad Institute 2018). Indel realigned was performed using GATK (McKenna et al. 2010). Reads with mapping quality less than 30 were then removed using samtools.

Mean genome coverage for each individual was calculated using the GATK DepthOfCoverage tool (McKenna et al. 2010).

Variant calling - high coverage individuals (>8×)

All modern goat samples, and ancient samples with average coverage >8×, were included in a "high confidence" variant call set. Samtools mpileup (Li et al. 2009) was used to call variants (-C 50 -q 30 -Q 20 -s -O -u -t SP,AD,INFO/AD,ADF,ADR,DP,INFO/DPR) and bcftools to generate vcf files (-v -mO z -f GQ,GP). Protein coding regions and repeat regions as defined by the Genbank annotation and RepeatMasker files (Smit et al. 2013-2015) were not called. An additional 50kb was added to both sides of protein coding regions and not called. Indels and any variants within 3bp of them were removed using bcftools filter (Li 2011). Tri- and quad-allelic sites were removed. For each variant, individuals were marked as missing ("./.") if coverage at that site was below 2 or twice the mean coverage, or if SP (strand bias) was above 13. Heterozygous variants present in a single individual or more than 75% of individuals were removed. Variant positions with missing data in any individual were then removed. Finally, LD pruning was performed using PLINK v1.07 (Purcell et al. 2007) with the settings --indep-pairwise 50 5 0.2.

Variant calling - low coverage individuals (<8×)

For all other ancient samples, the "high confidence" sites defined above were called. The same initial sites were called using samtools mpileup with the same options, except without recalibration (samtools mpileup -B) and without filtering for variant sites with bcftools. Tri- and quad- allelic sites were removed, indels and sites within 3bp of indels were removed. For

samples with >2× mean coverage, sites were set to missing as above (<2 reads, > twice mean coverage, >13 SP). For samples with less than 2× mean genomic coverage, no minimum coverage filter was imposed, and a maximum of 4 read coverage per site was permitted. Individuals with mean genome coverage >0.01× were then pseudo-haploidized by randomly sampling a read at each site and setting that individual homozygous for that allele.

Screening for related individuals

All pairs of individuals were assessed using lcMLkin (Lipatov et al. 2015). lcMLkin estimates kinship coefficient using genotype likelihoods, based on a probabilistic relationship between sites that are identical by state and also identical by descent (shared between two individuals and derived from a common ancestor). lcMLkin best practices were followed by thinning of the vcf to reduce LD (lcMLkin assumes site independence), using the following command: vcftools --vcf input.vcf --thin 100000 --remove-indels --maf 0.05 --recode --recode-INFO-all --out output. 0.9 was taken as a pi-HAT of threshold.

Principal components analysis

A frequent first approach in the study of populations using genetic data is to summarize the genetic variance of all individuals into synthetic variables (Principal Components, or PCs), and examining those variables/components for evidence of population structure. Briefly, a correlation matrix is computed from the genetic data (*e.g.* a SNP matrix), from which eigenvectors (the synthetic variables summarizing the genetic variance) and eigenvalues (the amount of variance each eigenvector explains) are derived. The first eigenvector/principal component explains the greatest amount of genetic variance (have the largest corresponding eigenvalue), while the second principal component will explain the second largest amount of variance, etc. Principal Component Analysis was first applied to population genetic data (10 loci or 38 independent alleles) by Cavalli-Sforza (Menozzi et al. 1978). The approach was extended to individuals represented by thousands of loci by Patterson, Price, and Reich, who laid out the statistical framework to test the significance of the computed components (Patterson et al. 2006). When significant, principal components can represent population structure, isolation-by-distance, migration events involving populations of distinct ancestry, or technical bias affecting certain samples (Novembre & Stephens 2008).

Several PCA approaches have attempted to account for missing data (i.e. sites at which certain individuals may not have been genotyped). Missing data is a recurring issue in aDNA studies, due to the low endogenous content, relatively low sequencing depth, and limited raw material characteristic of such studies. Smartpca implements a least-squares-based method to project samples with missing data onto PCs computed on samples with little-to-no missing data (Patterson et al. 2012). PCAngsd computes a correlation matrix based on individual allele using Genotype Likelihoods, thus accounting for the uncertainty of low coverage data (Meisner & Albrechtsen 2018). Initially developed to exploit off-target reads in targeted sequencing experiments, LASER computes Principal Components from a reference panel of individuals onto which more poorly sequenced individuals are projected by Procrustes analysis (Wang et al. 2014; Wang et al. 2015). Briefly, reads are simulated based on the reference individuals' genotypes and sample individuals' coverage, and a Principal Components Analysis performed on the low coverage simulated reference data. The simulated reference individuals are then projected back onto the reference PCA space by Procrustes analysis. This transformation is then applied to the sample individual data, placing them in the context of the reference PCA space.

PCA reference space and projection transformation were constructed using LD-pruned diploid genotype calls. All other (pseudo-haploid) samples were then projected onto the PCA space, and then filtered for individuals covered by <500 loci. To reduce the effect of simulation stochasticity, ten repetitions were performed, and the mean value of sample coordinates used in plotting. Other settings were left at default. This was repeated with and without modern bezoar.

To maximize the number of ancient samples projected, LASER projection was repeated using non LD-pruned diploid calls with modern bezoar removed as the reference set, and filtering for individuals <1,000 loci covered. Both plots were visually inspected to ensure that the use of non LD-pruned sites did not bias principle component calculation.

Population analyses using ANGSD

When working with low coverage sequencing data, it is preferable to take into account the inherent uncertainty in calling genotypes from the few reads covering a given site (Nielsen et al. 2011). For example, in cases where a single read is covering a variable site, it is difficult to state with certainty whether the site is homozygous for the observed allele or heterozygous.

Modelling of genotype probability can be computed in the form of genotype likelihoods. For a given biallelic site, the possible genotypes (heterozygous, homozygous reference or homozygous alternative) can be expressed in the form P(X|G), where X is the data present in a single read and G is the genotype of interest. Assuming the independence of reads, the probability of each genotype for all reads at a site can be multiplied to produce a genotype likelihood for all data at the site. A failure to address genotype uncertainty in low coverage data can bias allele frequency estimates and skew the site frequency spectrum in favour of singleton variants (Nielsen et al. 2011), which can confound population-level analyses.

Due to the high proportion of low coverage genomes in this dataset, we used a genotype likelihood framework in ANGSD (Korneliussen et al. 2014) to avoid explicit genotype calls. For all analyses using ANGSD, the following settings were used: -minQ 20 -minMapQ 30. This results in ANGSD ignoring bases with read quality less than 20 (1% chance of base error) and reads with mapping quality less than 30 (0.1% chance of misalignment). Analyses were restricted to the autosomes. Yak (*Bos grunneins*) was used to define the ancestral allele. For ANGSD analyses involving modern populations, these were subsampled randomly to ten individuals (Appendix Table 4.1).

To generate an ancestral genome, consensus sequences were generated using the ANGSD doFasta option, using the following options: -minQ 20 -minMapQ 30 -setMinDepth X -setMaxDepth Y, where X and Y is half and twice the mean genome coverage respectively. Yak was selected as the outgroup due to the possibility of hybridization and ancestral admixture between sheep and goat (Mine et al. 2000; Pauciullo et al. 2016).

Identity-By-State (IBS) nuclear genome phylogeny

As an alternative approach to assess how ancient and modern *Capra* species relate to one another, an identity-by-state (IBS) matrix computed using ANGSD (Korneliussen et al. 2014). As implemented in ANGSD, at each site a single read is sampled from each individual, minimizing the effect of reference bias that may have a large effect at poorly-covered sites. When a pair of individuals share the same allele at a given site, they are scored as 1; otherwise they are scored as 0. A pairwise distance score is computed for all individuals by summing over all sites for each pair, and normalizing by the number of sites shared between that pair.

The maximum missing individuals per site was set as half the number of individuals in the analysis rounded up. The following settings were used in IBS calculation: -minFreq 0.05 -GL 1 -doCounts 1 -doMajorMinor 5 -doCov 1. Only samples with >0.01× mean genome coverage were included.

An unrooted neighbor-joining tree was constructed from the pairwise distance matrix using the R package ape (Paradis et al. 2004). The tree was aesthetically modified using Figtree (Rambaut 2009), branches coloured based on location and time period, and rooted on *Bos grunniens*.

D statistics

The *D* statistic, also known as the ABBA/BABA test, was first introduced in the analysis of the draft Neanderthal genome (Green et al. 2010). Essentially, the *D* statistic is test of treeness. If a supposed sister group/individual (H3) shows a greater degree of allele sharing with one of two populations/individuals (H1 and H2), a simple tree model is rejected (Figure 4.1). At biallelic sites, a known outgroup (H4) of H1, H2 and H3 is used to define the ancestral (or more accurately the outgroup allele) allele A, and H3 is used to define the derived allele B (more accurately the non-outgroup allele), discarding sites where the outgroup and H3 are in concordance. The number of sites at which H2 and H3 share allele B, while H1 has allele A, is counted and represented by nABBA (the number of ABBA sites); the corresponding nBABA value is also calculated. When the purposed tree is true, the difference between nABBA and nBABA is expected to be zero, or statistically indistinguishable from zero; small deviations are expected even if the tree is true, due to incomplete lineage sorting. To account for this, a standard error is computed using a jackknife approach, and the *D* statistic expressed as a Z score. If the tree is not true, *D* will deviate from zero more than is expected by chance, usually taken as an absolute Z score of 3 or greater.

Assuming the outgroup H4 is true, treeness can be rejected in several scenarios. If admixture from a population related to H3 into the ancestors of either H1 or H2 (with the exclusion of the other) has occurred. For example, if a population related to H3 has admixed exclusively with the ancestors of H2 (or has admixed with the ancestors of H2 to a greater extent than the ancestors of H1), H2 will share more derived alleles with H3 than H1 and H3 will, leading to an excess of ABBA sites relative to BABA. Treeness will also be rejected in cases where tree

is simply topologically incorrect i.e. if H1 and H3 share a more recent ancestor than H1/H2 or H2/H3, and no admixture has occurred, treeness will be rejected as H1 and H3 will share more derived alleles resulting in an excess of BABA sites. In this manner, the *D* statistic can be used as a simple measure of relatedness among three population.

As implemented in ANGSD (Korneliussen et al. 2014), a single read is sampled from each individual at each site in order to determine nABBA and nBABA for each set of four individuals. Transitions (when the A/B pair are with either C/T or G/A, in any order) can be removed due to the well- documented issue of transition errors in aDNA (Chapter 1). *D* is then calculated by the equation:

$$D = \frac{nABBA - nBABA}{nABBA + nBABA}$$

such that cases in which H2 and H3 share more derived alleles will result in significantly

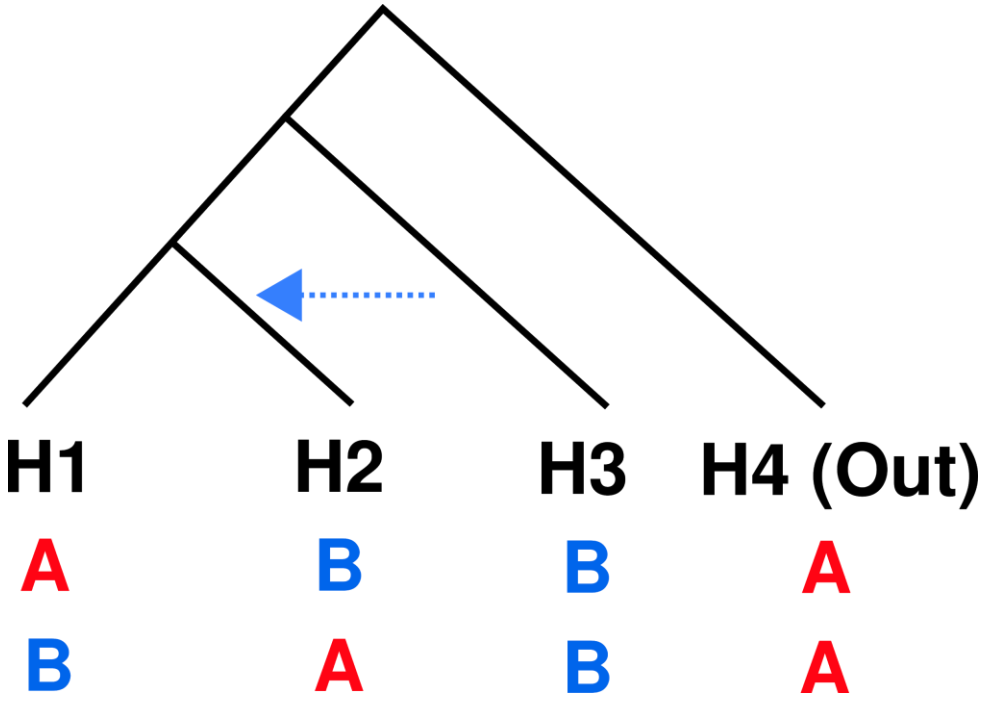


Figure 4.1: Example tree tested using *D* statistics. In the proposed tree, a known outgroup (H4) defines the ancestral allele A, and an individual/population (H3) sister to two other individuals/ populations (H1 and H2) defines the derived allele B. In this example, gene flow has occurred from a population ancestral to H3 into the ancestors of H2, resulting in a greater number of ABBA sites than BABA sites, and *D* is expected to deviate from zero.

positive *D* values, while negative *D* values will be observed when H1 and H3 share more derived alleles. The approach has been extended to groups of individuals (Soraggi et al. 2018), randomly sampling reads from each individual while weighting each individual based on their sequencing depth at a given site.

To investigate population relatedness and to test for admixture between populations, the *D* statistic (Green et al. 2010) was calculated at an individual level. Transitions were ignored in the analysis, to reduce the effect of residual DNA damage on calculations. An absolute Z score of 3 or greater was taken to be significant. At each position and sample, a random base was sampled (-doAbbabbab 1) to reduce the potential effect of reference bias.

Treemix

The relationship between different populations are often represented by simple bifurcating trees, where tips represent sampled populations and nodes represent divergence between ancestral populations. However, this approach does not well model gene flow events between populations (admixture), a well-documented phenomenon (Green et al. 2010; Park et al. 2015). Several methods have been developed to model populations (or individuals) as a phylogenetic tree, determined by pairwise differences in allele frequencies, while allowing admixture between branches. These varying in the degree of automation; MixMapper (Lipson et al. 2013) requires that unadmixed populations be first defined and used to construct a scaffold neighbour-joining tree, onto which admixed populations are added. In contrast, Treemix (Pickrell & Pritchard 2012) constructs a maximum likelihood tree of all populations; residuals are calculated between the model and the real data, and migration edges iteratively added to explain the largest residuals and most increase the model likelihood.

TreeMix was used to construct a model of population splits and admixture events, based on the High Coverage Ancients and Moderns dataset, with CHIR_1.0 removed due to it being the reference individual. Only ancient genomes with >8× mean coverage were included (Direkli1-2, Semnan3, Blagotin3, and Acem2). Samples were grouped based on Appendix Table 4.1 for modern individuals and Appendix Table 4.2 for ancient. Migration events were varied from 0 to 5. The following settings were used: -root Yak -k1000 --noss. Bootstrapping was performed using blocks of 1000 contiguous SNPs and repeated for 500 iterations, and a consensus tree generated using PHYLIP version 3.697 (Felsenstein 1989). Confidence of

nodes is given as the proportion of bootstrap iterations supporting that grouping, when that proportion was not one.

Ancestry estimation

First introduced in the STRUCTURE approach (Pritchard et al. 2000), ancestry estimation models individuals as being composed of different proportions of *K* distinct ancestral groups or populations. Both STRUCTURE and maximum likelihood-based variants (Alexander et al. 2009) estimates allele frequencies for each hypothesized ancestral population jointly with the admixture proportions of sample individuals. NGSadmix (Skotte et al. 2013) extends the ADMIXTURE likelihood approach to genotype likelihoods rather than explicit genotype calls, making it more appropriate for datasets including low coverage individuals. NGSadmix is more accurate compared to other approaches when applied to low and in particular variable coverage datasets, and can estimate admixture proportions as low as 5% in low coverage (~2× mean coverage) individuals.

NGSadmix (Skotte et al. 2013) was used to estimate ancestry proportions using genotype likelihoods. The following settings were used for the analysis: -GL 1 -doGlf 2 -doMaf 1 -SNP_pval 1e-6 -minInd, with -minInd set to half the number of individuals in the analysis, rounded up. A further filter of -minMaf 0.05 was used in the ancestral component estimation. *K* was set to 2 for all runs. Ancestry estimation was repeated a total of fifty times, and the iteration with the highest best likelihood retained. A minimum genome coverage of 0.01× was employed. NGSadmix ancestry estimation was performed on two datasets: one including all modern and ancient samples above the coverage threshold of 0.01×, and the second featuring only ancient samples >0.01× mean coverage.

f statistics

f statistics are a group of related statistics introduced and formalized by Reich and Patterson (Patterson et al. 2012) which utilize allele frequency correlations to formally test for admixture between populations and infer mixture proportions. The allele frequency correlation between a pair of populations can be thought of as their shared drift, and represented by the branches common to both in a phylogeny (Peter 2016), Figure 4.2. The f_2 statistic represents the shared drift or branch length common between two populations *e.g.*

f_2(A, B) measures the branch length or shared drift between populations A and B. All succeeding *f* statistics can be decomposed into a linear sum of f_2 statistics.

The statistic f_3(A, B; C) corresponds to the branch length between population C and the internal node common to all three populations. In most cases, a given f_3 value should be positive i.e. the underlying phylogenetic tree has positive branch lengths. This null hypothesis is used to test for cases of admixture; when population C is the result of an admixture event between populations A and B, the corresponding f_3 value will be negative as alleles will be able to contribute to C along paths of the phylogeny that contribute negatively to f_3. However, a non-negative f_3 is not proof of that admixture has not occurred, as false negatives can occur due to excessive drift exclusive to C (i.e. admixture occurred many generations in the past or C has drifted due to a bottleneck, etc), or when the admixing populations are not long diverged.

The f_4 statistic can be exploited in cases of population mixing to estimate the admixture proportions. In simple two populations cases, these proportions are labelled as α and β, where β=1-α. The proportion α can be estimated using the following equation (Patterson et al. 2012):

$$\alpha = \frac{F_4(P_o,\ P_1; P_X, P_I)}{F_4(P_o,\ P_1;\ P_2, P_I)}$$

where P_O is an outgroup, P_X is the introgressed population, P_1 and P_2 are the potential "parent" and P_I is a population that did not introgress to form P_X but is related to either P_1 or P_2. α can be considered as a measure of how much closer the common ancestor of P_X and P_I is to the common ancestor of P_I and P_1 and the common ancestor of P_I and P_2 , or on average where P_X merges with the underlying genealogy (Peter 2016). Care must be made when interpreting α, as it is defined in cases where no admixture has occurred.

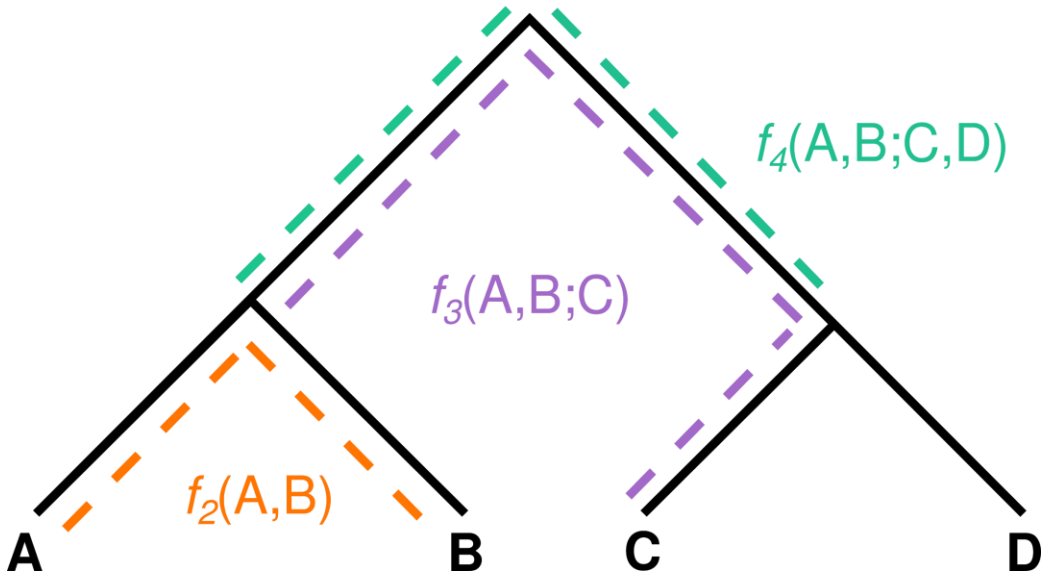


Figure 4.2: *f* statistics expressed as shared drift between populations in a phylogeny. Dashes indicate the branch lengths measured in a given *f* statistic. Adapted from (Peter 2016).

Outgroup f_3 analysis

The f_3 statistic can be exploited by use of an outgroup to measure the relative degree of shared drift between populations. In the form f_3(A, B; O), where O is an outgroup common to A and B, B can be varied to measure levels of drift between A and a range of different populations. The primary use of this is to determine which modern population and ancient population shares the most genetic drift i.e. ancestry. By setting A to an ancient population of interest and B varied among modern populations, the present-day group sharing the greatest amount of drift with the ancient population can be determined.

To investigate shared drift between Neolithic populations and other domestic populations, outgroup f_3 statistics were calculated using ADMIXTOOLS (Reich et al. 2009; Patterson et al. 2012), using individuals with greater than 0.01× mean coverage combined into populations, in the form of f_3(X, Neolithic; Outgroup), where X is a population as defined in Appendix Tables 4.1 and 4.2. Neolithic was varied between western, eastern, and Levantine Neolithic groups (Appendix Table 4.2). The outgroup selected was the Qazvin Bezoar population, due to the nonsignificant *D*(Eastern Neolithic, Western Neolithic, Qazvin Bezoar, Yak) (Appendix Table 4.11). For visualization purposes, f_3 values were plotted on a map of

southwest Asia (Виктор 2010). To investigate if the shared drift of Neolithic Levant and Modern Africa is independent to the drift shared between Neolithic West and Modern Africa, f_3 for all pairwise combinations of Neolithic population were then plotted with a linear regression and associated confidence interval using the ggplot (Wickham 2009) function geom_smooth; studentized residuals were computed and tested outlier status using a two-tailed test, 16 degrees of freedom, alpha of 0.05. In addition, shared drift between the two modern goat genomes generated here, IOG and Tog representing feral Old Irish Goat and Togolese village goat respectively, was computed. Finally, f_3 outgroups values were calculated for all pairs of goat populations, and used to construct a heatmap using the heatmap.2 function and default hclust clustering algorithm. For all tests, samples dated to prior to the Neolithic and bezoar were excluded.

f_4 ratio estimation

To estimate the contribution of Ancient Turkish Wild bezoar to Levant Neolithic and Western Neolithic goat, f_4 ratios in the form (Yak, Direkli5+Dirkeli6; Neolithic X, Neolithic Iran)/(Yak, Direkli5+Dirkeli6; Direkli1-2, Neolithic Iran) were constructed using ADMIXTOOLS (Reich et al. 2009; Patterson et al. 2012), where X is Neolithic Levant or Neolithic West. Ancient Turkish Wild were divided into two haploid genomes (Direkli5 and Direkli6) and one diploid genome (Direkli1-2) in order to satisfy the requirements of the ratio. This was repeated to additionally test the proportion of:

1. Ancient Turkish Wild ancestry in Neolithic Levant.
2. Eastern Neolithic ancestry in Bronze Age Turkey.
3. Eastern and western Neolithic ancestry in post-Neolithic Levant.
4. Western Neolithic ancestry in eastern post-Neolithic populations.

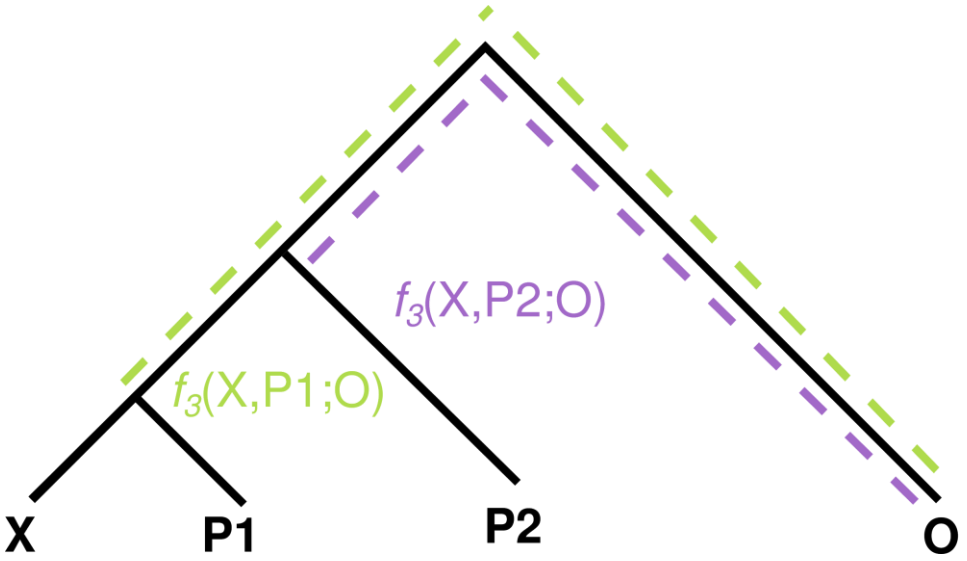


Figure 4.3 - f_3 outgroup comparisons. By calculating the f_3(X, Pop; Outgroup), a relative measure of shared drift between X (our population or individual of interest) and Pop can be obtained. By varying the identity of Pop (*e.g.* between P1 and P2 above), the degree of shared drift with X among groups can be compared. In this example, P1 and X split more recently and have a greater degree of shared genetic drift as expressed by f_3(X, P1; Outgroup) compared to P2, which diverged earlier from X.

Admixture graph construction

f statistics have been implemented in a user-defined, model-based framework in the form of qpGraph (Patterson et al. 2012). The *f* statistics expected from a user-inputted graph topology are computed, and a score calculated between the expected and observed statistics; the score is then maximized to estimate admixture proportions and branch lengths. The appropriateness of the proposed phylogenetic topology for the data can then be assessed by examining which *f* statistics showed large discrepancies between their expected and observed values.

To build a model of the population history of domestic goats, admixture graphs were fitted using qpGraph included in the ADMIXTOOLS package (Reich et al. 2009; Patterson et al. 2012) which uses *f*-statistics based on allele frequency correlations between samples to assess whether a fitted admixture graph of population history is consistent with the data.

Population groupings were as defined in Appendix Tables 4.1 and 4.2. Yak was used as an outgroup. qpGraph was run using default settings with a Z score=3 as a cutoff for outlier f-statistics.

The process taken for iteratively adding populations and fitting different models to the data is described in detail in Appendix Text 4.1. Accompanying figures are displayed in Appendix Figures 4.1 and 4.2. The number of SNPs used in each graph is presented in Appendix Table 4.4.

Estimation of the nuclear genome mutation rate

To estimate the goat autosomal mutation rate, the F(A|B) method described in (Green et al. 2010; Skoglund et al. 2015) was used. Heterozygous positions in modern individual B are identified, and the proportion of times the derived allele is randomly sampled in an ancient individual A at those positions was recorded, F(A|B). The Neolithic Serbian sample Blagotin3 was selected as the ancient individual A and the modern Old Irish Goat (IOG) as individual B. To control for genetic drift in the lineage specific to B, a calibration curve was constructed using PSMC (Li & Durbin 2011) to estimate past population demography. Sites in IOG were not considered if coverage was less than one third the mean coverage or more than twice the mean coverage. PSMC was performed on IOG using the following settings: -N25 -t15 -r5 -p '4+50*1+4+6'. msHOT-lite (Hudson 2002; Hellenthal & Stephens 2007) was used to simulate 800mb of sequence data under the estimated demography, while varying the mutation rate and divergence time of A and B (measured in generations). The F(A|B) ratio for each simulation and blagotin3 was then estimated using POPSTATS (Skoglund 2015), and calibration curves constructed using the ggplot2 function geom_smooth (Wickham 2009). The plot was then examined to identify which mutation rate curve overlapped with the known age of Blagotin3 (6,096-5,892 cal BC) at the observed F(A|B) value.

Nuclear genome modelling

Nuclear genome modelling and ABC model selection was performed primarily by Pierpaolo Maisano Delser, and is reproduced here from (Daly et al. 2018). To model the demographic history of goat using ancient genomes, an additional call set was generated. Samples from Turkey, the Balkans, and Iran with coverage >2.5× were included (Table 4.1; Appendix Table 3.4). The same calling pipeline was used except that the samtools mpileup recalibration

option was disabled (-B). Sites were also filtered for linkage disequilibrium and only variants at least 100kb apart were retained. Site were finally filtered for 0% missing data in the dataset to remove any additional source of uncertainty. The final call set was composed of 9,385 variants, which were pseudo-haploidized by random read sampling.

Both nucleotide diversity per population and Hudson's pairwise F_{ST} (Hudson et al. 1992) were calculated using R script R v3.2.3 (R Core Team 2016). Population groupings used are displayed in Appendix Table 4.1. Nucleotide diversity calculated on a pre-selected subset of variant sites does not correspond to the nucleotide diversity calculated across the whole genome. In order to take this bias into account, simulated data was generated in the same way that the variant sites in the real dataset were preselected. The first 9,385 variant sites were subset and the nucleotide diversity per population calculated for this subset. In this way, the nucleotide diversity calculated on both the simulated and real data are comparable.

An approximate Bayesian computation (ABC) (Beaumont et al. 2002) framework was developed to estimate parameters and compare models. Two demographic models were designed to investigate the demographic histories of samples belonging to the Western and Eastern populations: model SINGLE_AU and model BINARY_AU (Figure 4.4). Model SINGLE_AU describes an ancestral population (Nanc2) that goes through a bottleneck (Nanc1) from 11,000 to 10,500 years ago representing the domestication process. After the bottleneck, Nanc1 branches into the ancestral populations (Nbotw and Nbtoe) of the Neolithic Western and Eastern samples respectively. Both populations exponentially increase in size from 10,500 to 8,000 years ago (Nneow and Nneoe). Model BINARY_AU describes an ancestral population that at the time Tsplit branches into the two ancestral populations to the Western and Eastern samples (Nanc1w and Nanc1e respectively). Each of these two populations goes through a bottleneck from 11,000 to 10,500 years ago representing independent domestications (Nbotw and Nbote). Afterwards, both populations exponentially increase in size from 10,500 to 8,000 years ago (Nneow and Nneoe). Prior distributions for all parameters of the two models are reported in Appendix Table 4.5.

50,000 simulations under each model were performed using fastsimcoal 2 v.25221 (Excoffier et al. 2013). The mutation rate was calibrated as described above and a value of $1.3x10^{-8}$ per site per generation was used along (see Results) with a generation time of 2.5 years (Don E. Wilson 2011). The following summary statistics were used: nucleotide diversity per population (π_3E, π_3W) and pairwise Hudson's F_{ST} (Fst_3W_3E). Model posterior

probabilities were calculated by a weighted multinomial logistic regression (Beaumont 2008) for which the best 25,000 and 50,000 simulations were retained. Parameters under the most supported model were estimated from the 5,000 simulations closest to the observed dataset using the neuralnet algorithm (Csilléry et al. 2012). Analyses were performed in the R environment (R Core Team 2016) with the library abc (Csilléry et al. 2012).

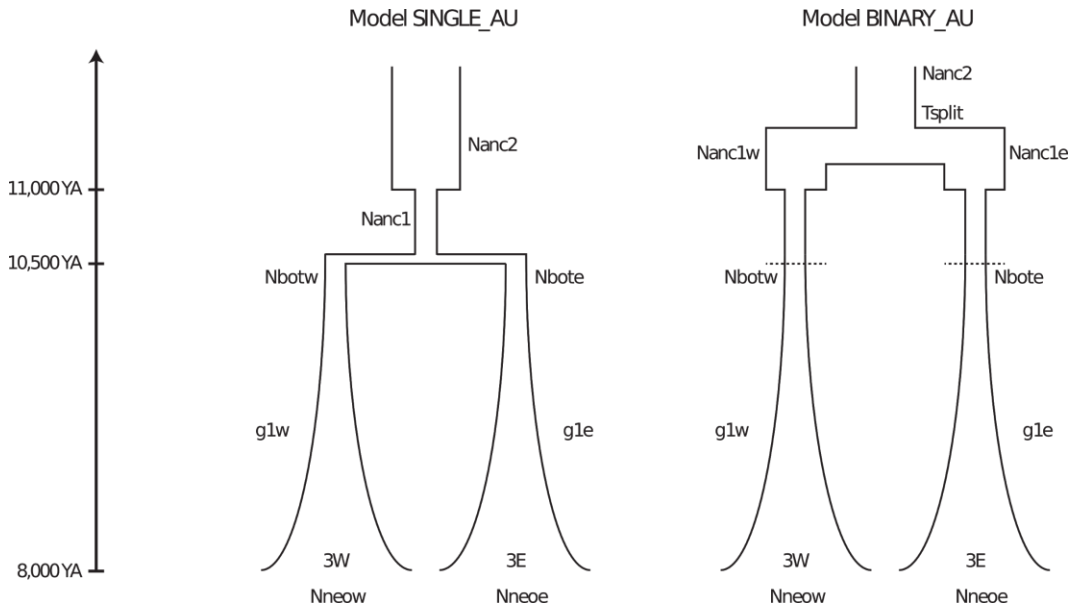


Figure 4.4 - Demographic models tested with whole genome sequences. YA: years ago. 3L: Neolithic Levant; 3W: Neolithic West and 3E: Neolithic East.

Results

Modern genome sequencing and alignment

High coverage genomes were generated for two modern goats, and are summarized within Appendix Table 4.1. Both genomes were of a high mean coverage following quality control, 36.5× and 41.9× respectively. Additionally, publicly available data from 55 modern *Capra aegagrus/hircus* were aligned to CHIR_1.0. This set of genomes was relatively high coverage, with a median of 12.9× coverage. The yak individual aligned to CHIR_1.0 produced a high coverage genome of 20.83×.

Variant calling

Prior to LD pruning, 3,003,233 biallelic sites were defined using the high coverage modern and ancient goat genomes. LD pruning resulted in a final call set of 726,401 sites. Pseudo-haploid genotypes were generated for 47 individuals; call rates are displayed in Table 4.1. Median call rate for pseudo-haploidized samples was 0.53 (386,667 sites).

Screening for related individuals

lcMLkin results for the top 30 pi-HAT values are presented in Appendix Table 4.6. Three pairs of samples were previously identified as likely being from the same individual (Chapter 2), and this was confirmed by lcMLkin pi-HAT estimations; the pairs Direkli1/Direkli2, Azer3/Azer5, and Semnan1/Semnan2, all produced had pi-HAT values >0.9. Additionally, the pair Bulak1/Bulak3 had a pi-HAT of 0.938, but these samples derived from petrous bones of the same orientation. To exclude the risk that Bulak1 and Bulak3 were cross-contaminated, Bulak3 was removed from all analyses.

Principal components analysis

The plot of PC1 vs PC2 (Figure 4.5) using all samples LD-pruned shows that PC1 differentiates modern and ancient wild bezoar from modern and ancient domestics. Bezoar from Azerbaijan and Iranian Azerbaijan fall on the most extreme end of PC1. As these represent 10 of the 61 genomes used to compute the reference PCs, sampling bias may explain their plot location. Domesticates shows some small variation on PC1, with a single low-coverage Levant genome (Shiqmim1) and modern African samples falling on the

extreme end of the PC. Among domesticates, there is a weak ancient-to-recent trend along PC1. Ancient pre-domestic bezoar from Turkey (Direkli Cave) and Armenia (Hovk-1 Cave) are found between modern bezoar and domestic goat on PC1, which may again be explained by the relatively-large sample size of modern bezoar in this dataset.

PC2 better differentiates domestic east (Asian) and west (European) samples; bezoar from Hamedan, west Iran fall on one extreme of PC2, with modern Europeans falling on the other extreme. Within the domestic group, Neolithic West (western Turkey and southeast Europe) group apart from eastern Neolithic (Iran and Turkmenistan). These show affinity with modern representatives of the same continental regions: Neolithic Iranian samples show greater affinity to modern Iranian goat than Neolithic Serbian goat, with the reverse being true for modern European goat. Modern Africans fall apart from both modern clusters but show greater affinity with modern Europeans. A Bronze Age sample from Potterne, Britain, groups closely with a modern Irish Old Goat, and Neolithic samples from Blagotin. Samples from ‘Ain Ghazal, Jordan, are found between the western Neolithic and modern Africa clusters, with substantial variation in their relative affinities with other samples. Later Levantine samples do not consistently fall in the same cluster and also show substantial variation.

Within the eastern group, a difference is observed between Neolithic ancients and post-Neolithic ancients, with the later showing greater affinity with modern samples from Iran. The reference individual from China, CHIR_1.0, clusters with this eastern group, close to Chalcolithic and Bronze Age goat from Iran and Turkmenistan. Later samples from Iran and Georgia fall closer to modern Iranian goat. Samples from Bronze Age Turkey (Acemhöyük) are found between western/Levantine Neolithic samples and the post-Neolithic/modern eastern samples, displaying greater affinity with the later. Other bezoar show some variation along this axis, with a wild goat from the Markazi Province falling closest to domestic individuals.

Plotting PC1 vs PC3 reveals greater variation among the domesticates (Appendix Figure 4.3). As PC1 still poorly differentiated domestic samples, the PCs computed with modern bezoar removed were instead examined to better explore possible population structure (Figure 4.6).

There was strong agreement between the PCAs constructed using LD-pruned (Appendix Figure 4.4) and non LD-pruned data (Figure 4.6). The largest difference between the two

plots is the Turkish Ancient Wild samples (Direkli Cave) who show greater affinity with western Turkish and European samples in the LD-pruned plot compared with the non LD-pruned plot. Additionally, the sample Blagotin3 shows greater divergence from European moderns and ancient when the LD-pruned dataset is used. The non LD-pruned dataset allowed several low coverage samples from Israel to be plotted (Gilat10, Shiqmim9, Miqne5), as they did not pass the covered loci filtering step when the LD-pruned dataset was used. These samples fall in a similar position as other post-Neolithic goats from Israel, and as such the PCA constructed from this dataset was selected to be displayed as Figure 4.6. The PCA of the LD-pruned dataset is shown in Appendix Figure 4.4.

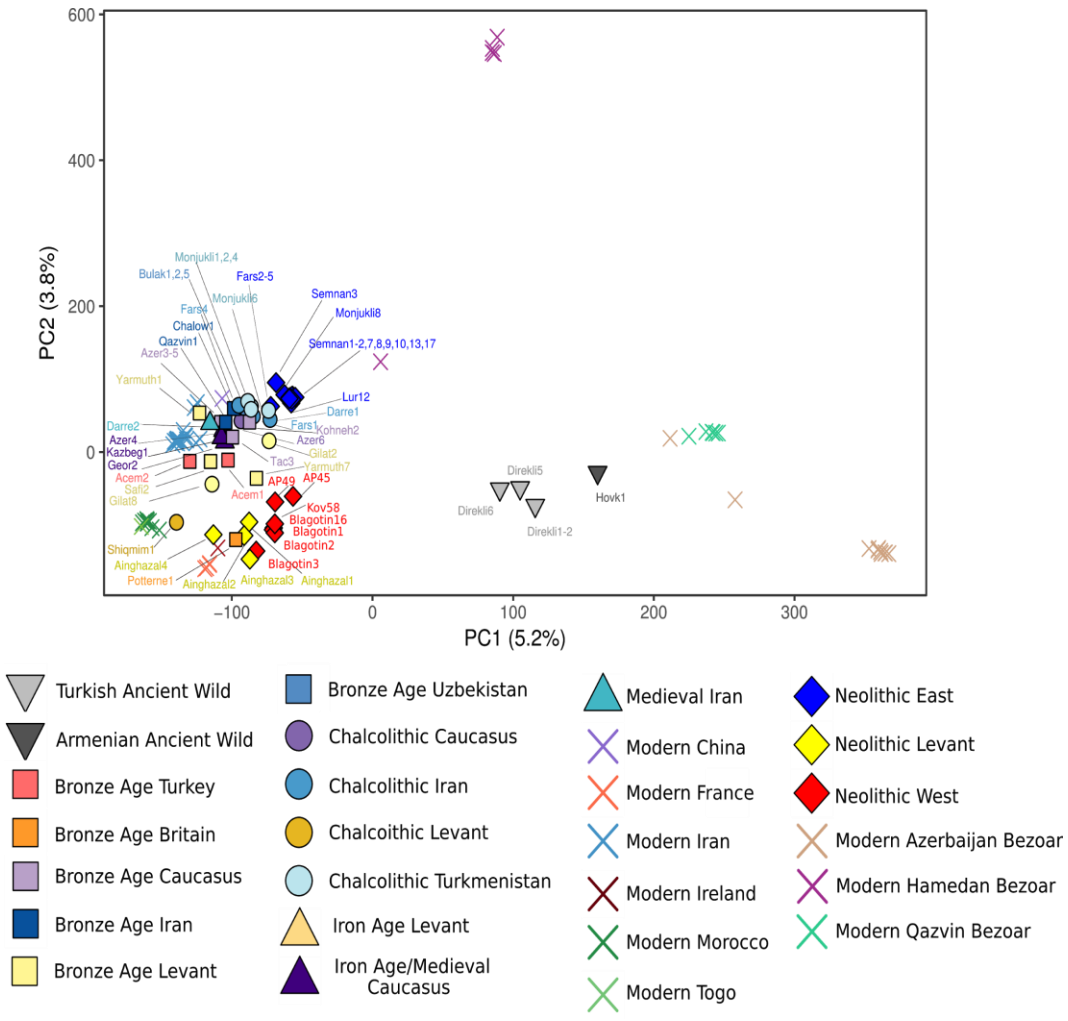


Figure 4.5 - LASER projection PCA of all ancient samples and modern goat/bezoar, plotting PC1 vs PC2, using LD-pruned sites. Values in parenthesis represent the percentage of variance explained by a given PC, as estimated by LASER.

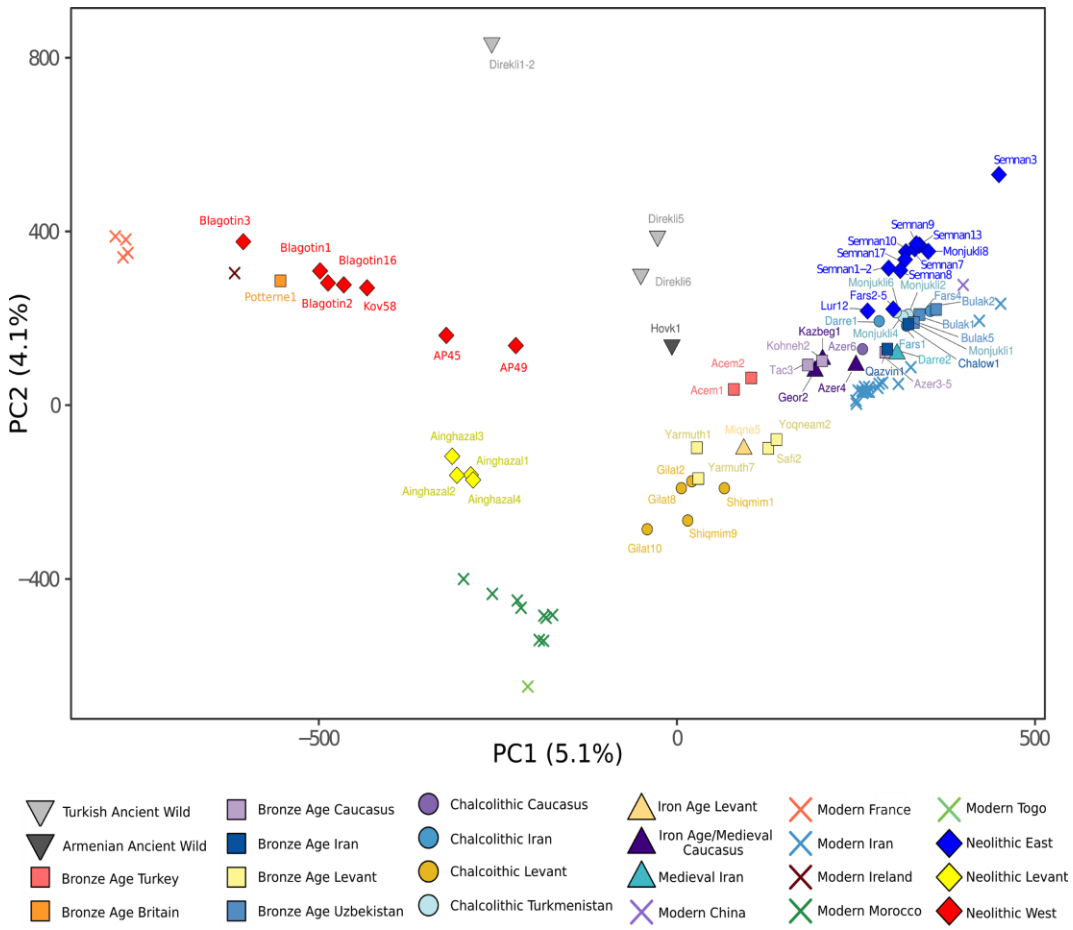


Figure 4.6 - LASER projection PCA of all ancient samples and modern domestic goat, plotting PC1 vs PC2, using non LD-pruned sites. Values in parenthesis represent the percentage of variance explained by a given PC, as estimated by LASER.

The general pattern of diversity observed in Figure 4.6 is similar to that detected when modern bezoar are included:

- Modern European goat and Blagotin3 fall on one extreme of PC1, while a few modern Asian (Iranian and Chinese) and Semnan3 fall on the other extreme.
- African goat are found at one end of PC2, and the ancient wild Direkli1-2 on the other; other samples from Direkli Cave are less extreme in their PC2 position.
- Hovk1, a wild goat from Armenia dating to beyond the C^{14} limit (>47,074 BP), falls close to the centre of the PCA
- Goat from Neolithic western Turkey, Serbia, and Bulgaria are found on the "European/western" end of PC1, while Neolithic goat from Iran and Turkmenistan are found on the opposite "Asian/eastern" end.
- Neolithic Levantine samples are found between western Neolithic and modern African samples.
- Post-Neolithic Levantine samples show greater affinity with eastern samples than their Neolithic counterparts.
- Goat from Bronze Age Acemhöyük, central Turkey, fall closer with Post-Neolithic Levant and later Georgian/eastern samples than Neolithic goat from western Turkey.
- Chalcolithic, Bronze Age and Medieval samples from Iran, Turkmenistan, and Uzbekistan show greater affinity to modern Iranian samples than Neolithic Iranian.

Ancient diploid Neolithic genomes used in the reference space calculation (Blagotin3, Semnan3, and Direkli1-2) occupy the extremes of the PCA; the unique variation in these ancient genomes may explain their outlying positions. Neolithic samples from western Iran (Lur12 and Fars2-5) show greater affinity to post-Neolithic samples than those from Neolithic eastern Iran and Turkmenistan, and do not fall in the cluster of samples from Neolithic Sang-e Chaqmak and Monjukli Depe. The Chinese reference individual CHIR_1.0 falls between these two clusters on PC1 but somewhat apparent on PC2. The majority of Bronze Age goat from the Caucasus region, along with Caucasus Iron Age and Medieval samples, fall further in the "western" direction of PC1 than Iranian goat from similar time periods. Finally, the Bronze Age goat from Britain, Potterne1, displays a close affinity to a modern day feral individual from Ireland.

A plot of PC3 vs PC4 demonstrates that PC3 captures variation specific to the Direkli samples, while PC4 captures variation present in Neolithic eastern Iran and Turkmenistan (Appendix Figure 4.5). Modern French and Iranian samples also show variation along PCs 3

and 4 respectively. Other modern and ancient domestic samples are poorly differentiated along these axes.

Identity-By-State

An Identity-By-State (IBS) matrix was constructed using modern and ancient samples with >0.01× coverage. This was then converted into a distance matrix and used to construct a Neighbour-Joining (NJ) phylogeny (Figure 4.7).

Bezoar, both ancient and modern, are outgroups to all modern domestic and ancient samples. Modern bezoar populations did not form clades to the exclusion of other subpopulations. The Direkli wild goat formed a clade with no other members. Hovk1, an Armenian wild goat at least 47,000 years old (Table 4.1), was not an outgroup to all bezoar/goat, suggesting structure observed within modern wild goat to have origins prior to the Last Glacial Maximum (~31,000-17,000 BC) (Clark et al. 2009).

Within ancient and modern domestic goat, modern Iranians form a clade that is sister to all other domestic samples, suggesting that they are either outgroups to other goat or a product of admixture. The next bifurcation within domestics separates western (Turkish, European), Levantine, and African goat from eastern (Iranian, Georgian, Uzbekistan, Chinese). Within the "western" branch, Bronze Age Turkish and Israeli goat split first, followed by modern Africans and all other samples. Neolithic samples from western Turkey form their own clade, as do Neolithic samples from southeast Europe. Modern French goat form a clade that is sister to the Irish modern IOG and Potterne1, the British Bronze Age sample. Goat from 'Ain Ghazal, Neolithic Jordan, form a clade that is sister to modern and ancient European goat.

Within the "eastern" clade, the majority of Neolithic goat form a distinct monophyletic group, with Lur12 (an early Neolithic sample from the Zagros) as an internal outgroup. The remaining Neolithic sample, Fars2-5 from southwest Iran *c*. 7,000-6,800 cal BC, falls within a sister clade to the other Neolithic samples, which itself is composed of Chalcolithic and Bronze Age goat from Iran, Turkmenistan, and Uzbekistan, in addition to CHIR_1.0. Goat from Georgia, as well as the remaining Bronze Age and Medieval Iranian samples, form various branches which are outgroups to the two previously-described large clades.

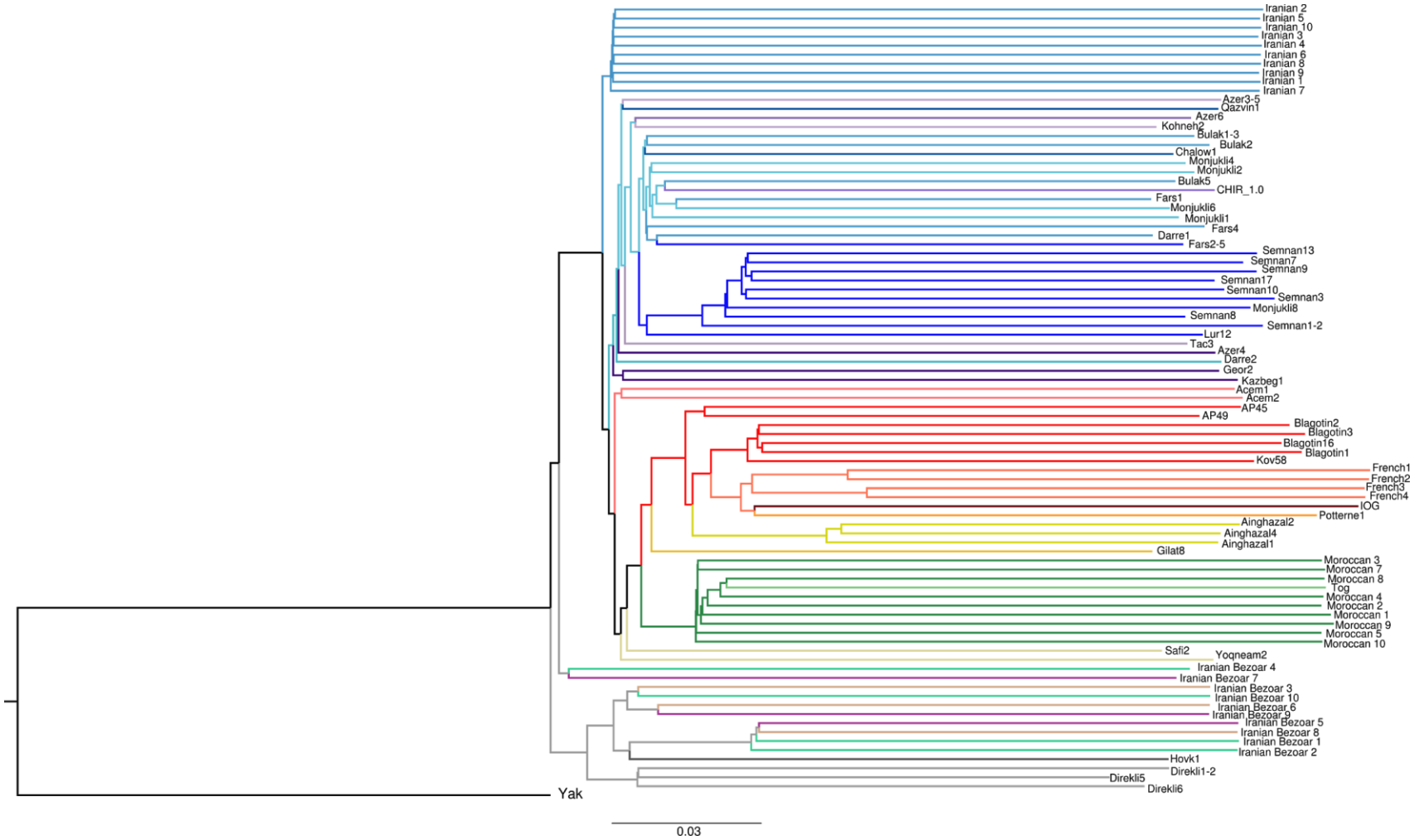


Figure 4.7 - Neighbour-Joining tree of IBS matrix of ancient and modern goat/bezoar. Yak is used as an outgroup. Branches are coloured according to sample context, matching colours used in Figure 4.5.

Table 4.1 - Summary of goat nuclear genomes. Radiocarbon-dated samples are indicated with *; ages are otherwise a contextual estimate.

Sample	Site	Region	Context	Age	×-fold Coverage	Sites called
Acem1	Acemhöyük	Aksaray Plain, Turkey	Bronze Age	*2,346-2,040 cal BC	4.76	641426
Acem2	Acemhöyük	Aksaray Plain, Turkey	Middle Bronze Age	~1,700 BC	8.67	726401
Ainghazal1	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	0.03	19050
Ainghazal2	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	0.06	36200
Ainghazal4	'Ain Ghazal	Amman, Jordan	Middle Pre-Pottery Neolithic B	Late 8th-Early 7th mil. BC	0.01	4774
AP45	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,500-5,000 BC	0.02	10678
AP49	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	5,500-5,200 BC	0.02	11849
Azer3-5	Tepe Hasanlu	Azerbaijan, Iran	Early Bronze Age	3,000-2,100 BC	4.66	617934
Azer4	Tepe Hasanlu	Azerbaijan, Iran	Iron Age	550-330 BC	2.57	447448
Azer6	Soha Chay Tepe	Azerbaijan, Iran	Chalcolithic	~4,200 BC	0.28	156425
Blagotin1	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,398-6,098 cal BC	6.99	674257
Blagotin2	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,379-6,078 cal BC	4.02	562646
Blagotin3	Blagotin-Poljna	Trstenik, Serbia	Neolithic	*6,096-5,892 cal BC	11.47	726401
Blagotin16	Blagotin-Poljna	Trstenik, Serbia	Neolithic	~6,100 cal BC	3.51	524882
Bulak1	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	0.87	386667
Bulak2	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	2.67	417905
Bulak5	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age	2,000-1,700 cal BC	0.27	149209

Chalow1	Chalow	Khorasan, Iran	Bronze Age	2,300-2,000 BC	0.05	31198
Darre1	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	5th mil. BC	0.04	24852
Darre2	Darre-ye Bolāghi	Fars, Iran	Chalcolithic	*1,473-1,641 cal AD	3.93	590505
Direkli1-2	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	*11,351-11,166 cal BC	11.55	726401
Direkli5	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	~9,500 cal BC	0.27	148250
Direkli6	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic	~9,500 cal BC	1.93	441895
Fars1	Rahmat Abad	Fars, Iran	Early Chalcolithic	~4,600 BC	0.02	12525
Fars2-5	Rahmat Abad	Fars, Iran	Pottery Neolithic	*7,047-6,772 cal BC	0.03	20515
Fars4	Mianrud	Fars, Iran	Early / Mid Chalcolithic	*5,460-5,211 cal BC	1.05	423597
Geor2	Tamara Fort	Kazbegi, Georgia	Medieval	11th-15th Century AD	1.5	495343
Gilat8	Gilat	Northern Negev, Israel	Chalcolithic	4,500-4,200 BC	0.02	9541
Hovk1	Hovk-1 Cave	Tavush, Armenia	Paleolithic	*>47,074 BP	3.08	529378
Kazbeg1	Tamara Fort	Kazbegi, Georgia	Medieval	10th Century AD	3.84	612533
Kohneh2	Kohneh Tepesi	Azerbaijan, Iran	Early Bronze Age	3,300-3,000 BC	0.04	25380
Kov57	Kovačevo	Blagoevgrad, Bulgaria	Neolithic	6,200-5,600 BC	0.07	42794
Lur12	Tepe Abdul Hosein	Luristan, Iran	Pre-Pottery Neolithic	*8,171-7,745 cal BC	1.05	412386
Monjukli1	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	0.24	131278
Monjukli2	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	0.21	120137
Monjukli4	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	0.6	282758
Monjukli6	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic	~5,100-4,500 cal BC	0.03	16945
Monjukli8	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic	~6,400-5,900 cal BC	2.57	526069
Potterne1	Potterne	Wiltshire, UK	Bronze Age	2,040-990 BC	3.67	505700

Qazvin1	Tepe Chizar	Qazvin, Iran	Middle Bronze Age	2,400-1,900 BC	3.16	498130
Safi2	Tel es-Safi	Ashkelon, Israel	Early Bronze Age	2,570-2,900 BC	0.04	26694
Semnan1-2	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic	*7,454-6,850 cal BC	6.85	692060
Semnan3	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	*6,214-6,004 cal BC	14.89	726401
Semnan7	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	3.28	535794
Semnan8	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	0.21	114726
Semnan9	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	3.05	509837
Semnan10	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	1.43	447491
Semnan13	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	2.54	393688
Semnan17	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	~6,000 BC	0.12	71717
Tac3	Tachti Perda	Kakheti, Georgia	Late Bronze Age	1,400-1,000 BC	0.13	76874
Yoqneam2	Tel Yoqne’am	Haifa, Israel	Middle Bronze Age	~1650-1550/1540 BC	2.2	561895

D statistics

D statistics, as implemented by ANGSD (Korneliussen et al. 2014), were computed to address specific questions. For clarity, each question will be addressed within a distinct subsection. For tests relating to a Neolithic geographic group (Levant, Eastern, Western), the highest coverage genome from that group was chosen as a representative; these were Ainghazal2, Semnan3, and Blagotin3 respectively. $|Z| \geq 3$ was taken as significant.

How do Neolithic goat from different sampled regions relate to each other?

To test how Neolithic goats relate to each other, the *D* statistic *D*(H1, H2, Test, Yak) was computed, where H1 and H2 were pairs of Ainghazal2, Blagotin3, or Semnan3, and Test was a Neolithic goat genome.

All Neolithic goat from western and Levantine regions showed greater affinity to the Neolithic Serbian genome Blagotin3 than to the Neolithic Iranian genome Semnan3, based on positive significant values of the statistic *D*(Blagotin3, Semnan3, Test, Yak) (Figure 4.8). In contrast, genomes from Iran and Turkmenistan showed greater affinity with Semnan3, with two samples from early Neolithic sites in western Iran (Lur12 and Fars2-5) showing relatively lower derived allele sharing than those from more eastern locales. A similar pattern was observed using the test *D*(Ainghazal2, Semnan3, Test, Yak). Western and Levantine Neolithic genomes showed greater affinity with Ainghazal2 than with Semnan3. Eastern Neolithic genomes showed greater derived allele sharing with Semnan3 than Ainghazal2. Likely due to the low coverage of Ainghazal2, there was a greater number of non-significant results.

Genomes from both western and eastern Neolithic regions showed greater allele sharing with Blagotin3 than with Ainghazal2, by the positive values observed in using the test *D*(Ainghazal2, Blagotin3, Test, Yak). Two of the three tested genomes from ‘Ain Ghazal gave non-significant results, a likely consequence of the test containing two low coverage samples. Ainghazal1 produced the only significant result, with a greater degree of derived allele sharing with Ainghazal2 than with Blagotin3.

Neolithic genomes from the same region tend to share more alleles with a representative modern genome from the same or similar region. The most unexpected result was the greater

degree of derived allele sharing between eastern Neolithic genomes and western Neolithics, than with Neolithic Levant. This can be explained by a greater degree of shared ancestry with eastern and western Neolithic goat, or with an input of ancestral alleles into the ancestors of Neolithic Levant.

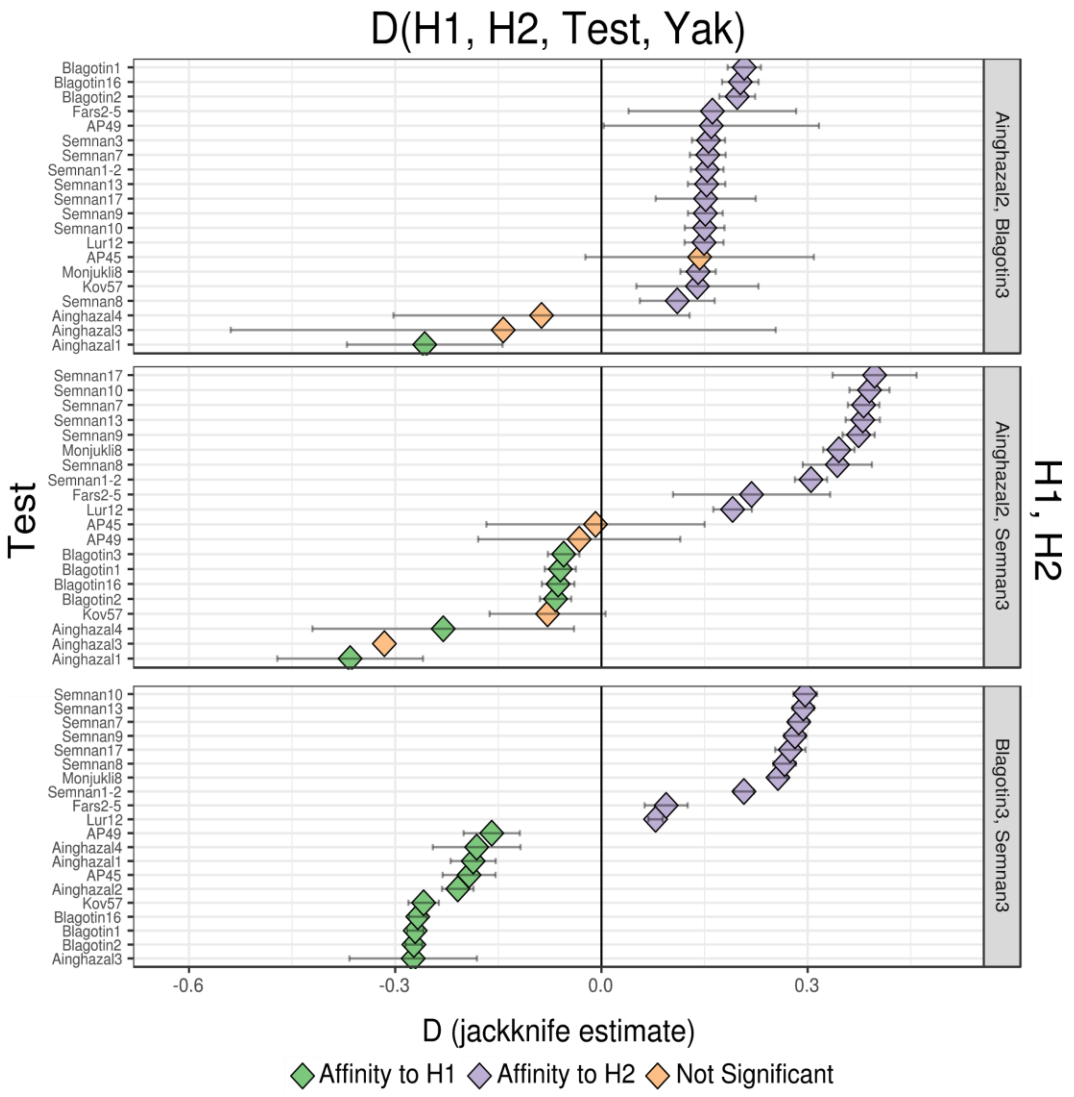


Figure 4.8 - *D* statistic test for relative affinity (derived allele sharing) of a Test Neolithic goat to members of pairs of Neolithic genomes, H1 and H2. Colour indicates Test affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange. Test genomes are ordered based on their *D* value for each set of pairs (H1 and H2)

How do pre-domestic wild goat relate to goat following domestication?

To test if Neolithic goat showed differing affinities to pre-domestic wild goat, the *D*(Semnan3, Test, Ancient Wild, Yak) was computed, were Test was a Neolithic genome and Ancient Wild was one of Direkli1-2, Direkli5, Direkli6, and Hovk1. Results for Direkli1-2 and Hovk1 are plotted in Figure 4.9; all ancient wilds are plotted in Appendix Figure 6.

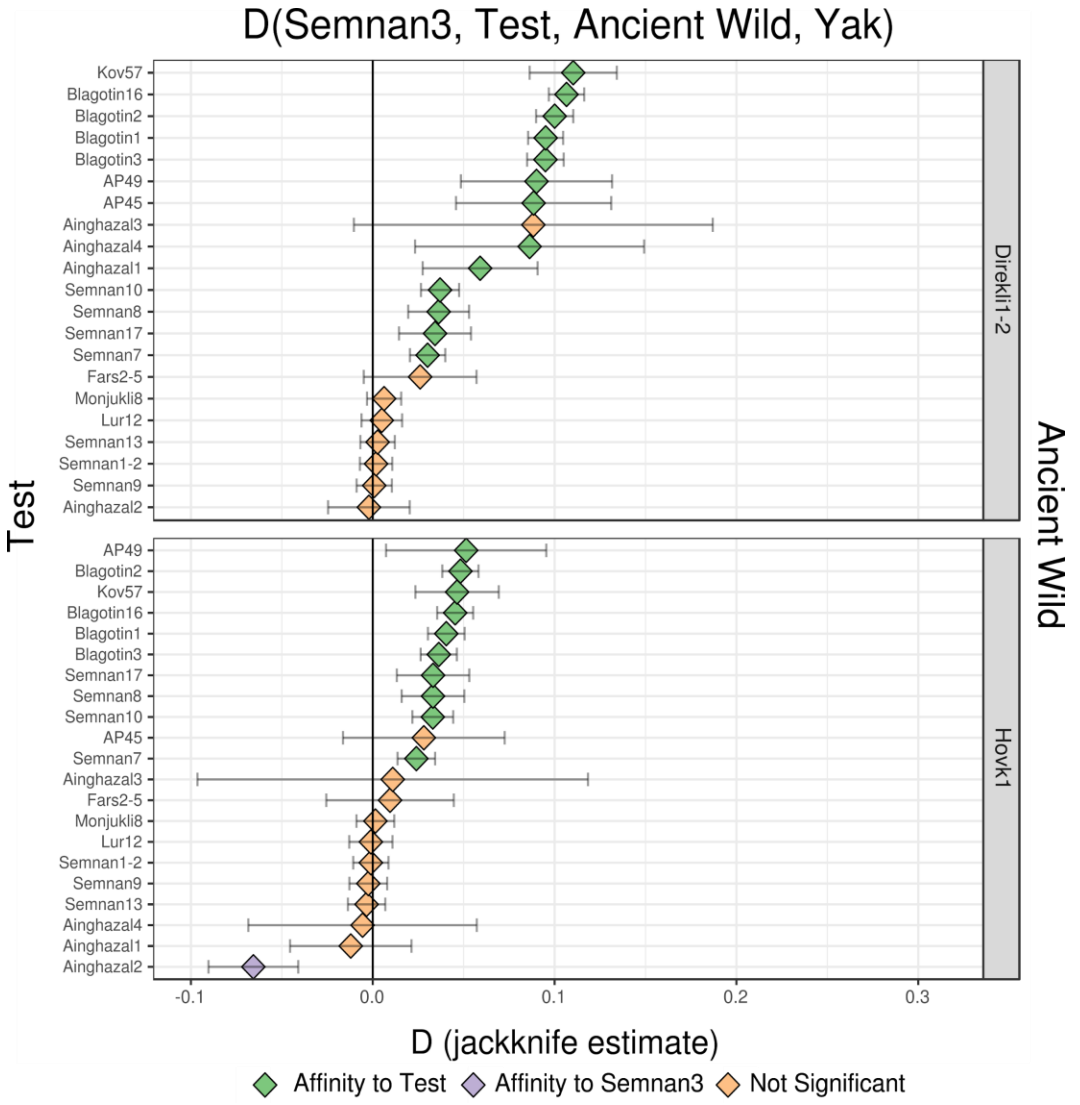


Figure 4.9 - *D* statistic test for relative affinity (derived allele sharing) of a Test genome to an Ancient Wild individual, relative to an Ancient Eastern (Semnan3). Colour indicates relative affinity of the Ancient Wild to either Semnan3 or Test, with non-significance (|Z| < 3) indicated by the colour orange.

When Ancient Wild was one of the Direkli Cave samples, a general pattern was observed. Western Neolithic genomes consistently have large positive significant *D*(Semnan3, Test, Direkli, Yak) values, indicating greater derived allele sharing between Direkli samples and western Neolithic genomes, relative to Semnan3. Neolithic Levantine genomes show mixed evidence of greater derived allele sharing, with positive values varying in significance between tests. Eastern Neolithic genomes either do not share significantly more alleles with Direkli samples than Semnan3 does, or slightly more.

When Hovk1 was Ancient Wild, greater affinity is observed between Hovk1 and western Neolithic genomes than between Hovk1 and Semnan3. Similar to tests involving Direkli samples, a subset of eastern Neolithic genomes also show significant positive *D* scores. The majority of tests involving a Levantine Neolithic sample are nonsignificant, with the *D*(Semnan3, Ainghazal2, Hovk1, Yak) being the exception: Hovk1 shows greater derived allele sharing with Semnan3 than with Ainghazal2. This may indicate an excess of ancestral alleles in Ainghazal2; tests involving this sample showed negative values in most cases. Alternatively, it may reflect shared ancestry between Hovk1 and the eastern Iranian genome Semnan3 that is not present in Levantine Neolithic genome, which is only detected in tests involving Ainghazal2 due to its high coverage compared to other ‘Ain Ghazal genomes.

Does the genetic affinities of goat from western regions (Turkey and Europe) change with time?

To test if the genomes of western sampled regions (Turkey, Europe) show a change in affinities over time, the test *D*(Blagotin3, Test, Neolithic Reference, Yak) was utilized (Figure 4.10). Test was varied among western genomes, Neolithic Reference varied between Semnan3 and Ainghazal2. When admixture from eastern Neolithics (relative to the baseline of Blagotin3) is explicitly tested using the *D*(Blagotin3, Test, Semnan3), two Neolithic samples (Blagotin16 and AP49), the British Bronze Age sample Potterne1, and the two Bronze Age Acemhöyük goat show significantly more allele sharing with Semnan3. When Ainghazal2 is set as the Neolithic Reference all samples, except those from Bronze Age Acemhöyük (Acem1 and Acem2), show no significant difference in derived allele sharing compared to Blagotin3. The two Acemhöyük show significantly less derived allele sharing with Ainghazal2 compared to between Blagotin3 and Ainghazal2.

As a different measure of genetic affinity, the *D*(H1, H2, Ancient Western, Yak) was computed, where H1 and H2 were a pair of Blagotin3, Semnan3, or Ainghazal2 (Figure 4.11). All samples showed greater allele sharing with Blagotin3 relative to Semnan3, based on negative values of the test *D*(Blagotin3, Semnan3, Ancient Western, Yak). The two

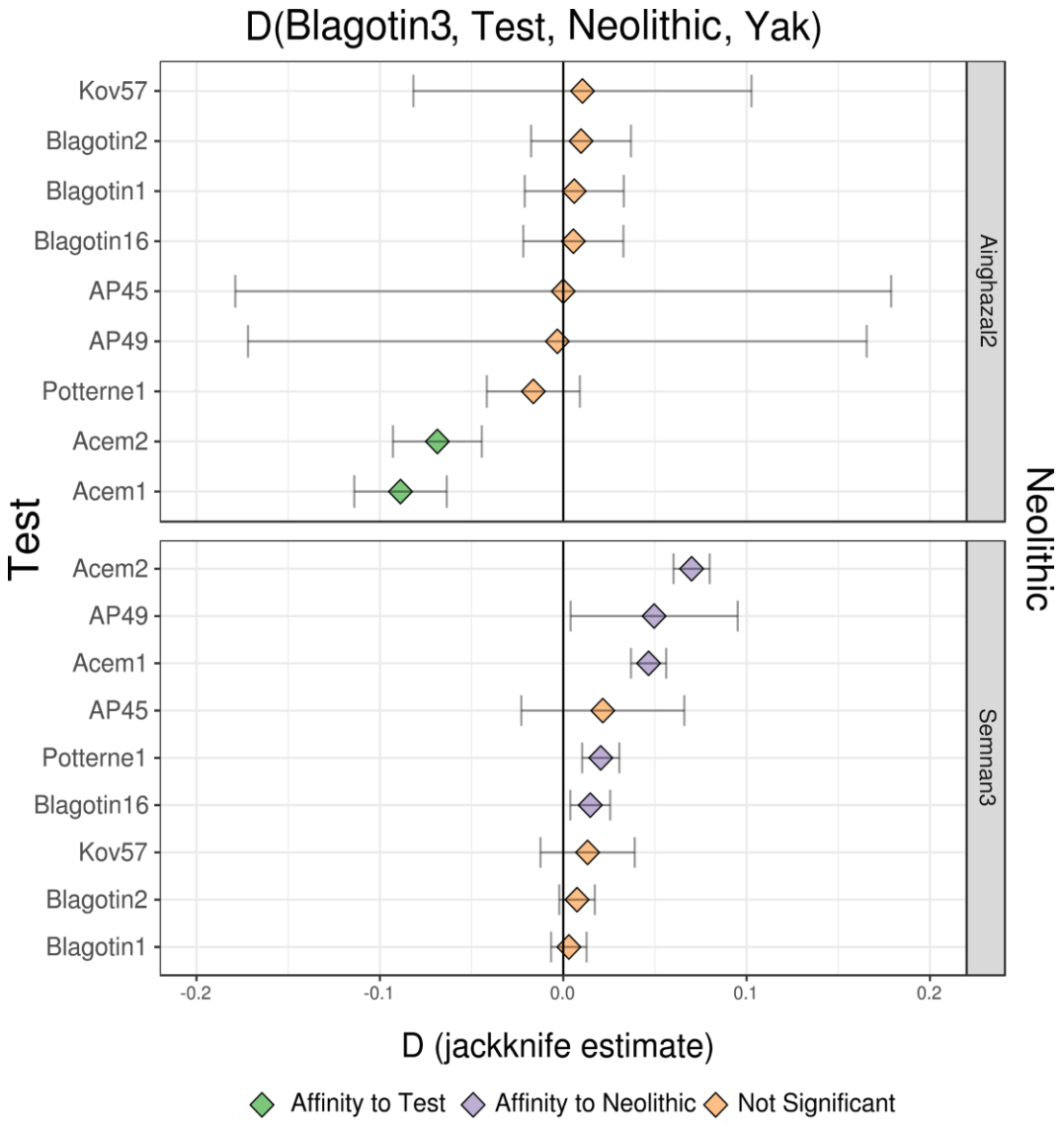


Figure 4.10 - *D* statistics testing the change in affinity of an ancient western Test genome relative to a high coverage western Neolithic, Blagotin3. Positive values indicate greater derived allele sharing with a Neolithic reference (Ainghazal2 or Semnan3) and a Test genome, relative to sharing between the Neolithic reference and Blagotin3. Colour indicates affinity of the Neolithic reference to either Blagotin3 (H1) or the Test (H2), with non-significance (|Z| < 3) indicated by the colour orange.

Acemhöyük samples showed substantially lower *D* values compared to other western genomes. These samples also were outliers in the test *D*(Ainghazal2, Semnan3, Ancient

Western, Yak), in which both have significant positive values, indicating greater derived allele sharing with the Iranian Neolithic than the Levantine Neolithic. Other western genomes produced either nonsignificant or negative values. Finally, all western samples showed greater affinity with Blagotin3 than with Ainghazal2, from positive *D*(Ainghazal2, Blagotin3, Ancient Western, Yak) results; one sample, AP45, gave a nonsignificant value.

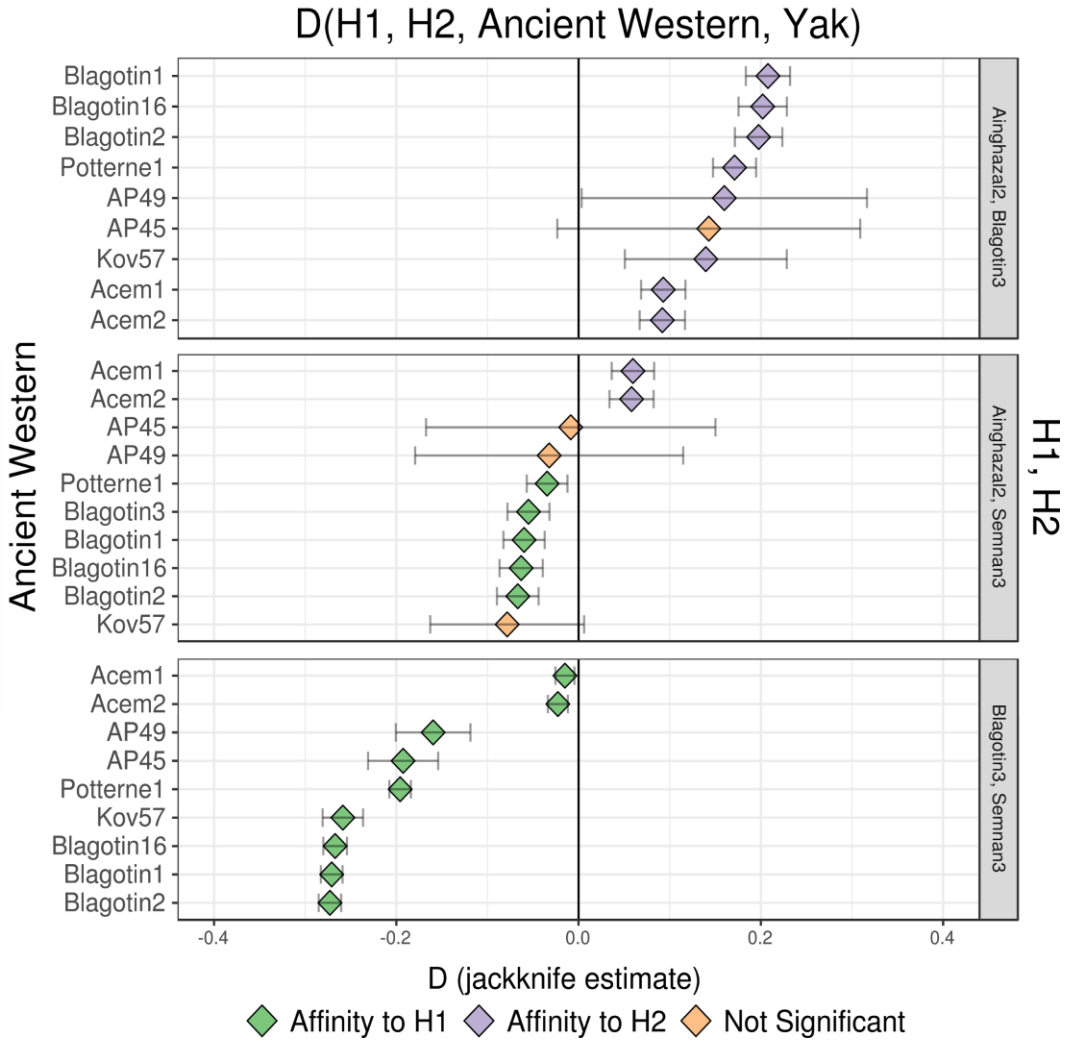


Figure 4.11 - *D* statistic test for relative affinity (derived allele sharing) of an Ancient Western goat to members of pairs of two Neolithic genomes, H1 and H2. Colour indicates affinity of the Ancient Western goat to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour blue.

These results indicate a change in ancestry in Bronze Age central Turkish samples relative to the Neolithic Serbian genome Blagotin3, as measured by *D* statistics. There are indications

that this change in ancestry is related to greater eastern Neolithic ancestry, from the pattern of *D* values involving the Acemhöyük samples and Semnan3.

Does the genetic affinities of goat from eastern regions (Iran, Georgia, Turkmenistan, Uzbekistan) change with time?

To test if post-Neolithic genomes showed changes in through time relative to the Neolithic genome Semnan3, the test *D*(H1, H2, Post-Neolithic Eastern, Yak) was computed, were H1 and H2 are a pair of Blagotin3, Semnan3, or Ainghazal2 (Figure 4.12). In all cases of the test *D*(Blagotin3, Semnan3, Post-Neolithic Eastern, Yak), significantly more derived allele sharing was observed with the Semnan3, indicating a degree of continuity with Neolithic populations. The same pattern was observed in the tests *D*(Ainghazal2, Semnan3/Blagotin3, Post-Neolithic Eastern, Yak) indicating a tier of relatedness (Neolithic Iran > Neolithic Serbia > Neolithic Levant). One sample (Kohneh2) failing to reach significance in both cases.

In the case of *D*(Blagotin3, Semnan3, Post-Neolithic Eastern, Yak), there was a clear temporal signal of decreasing *D* values from Chalcolithic Turkmenistan (Monjukli) and Bronze Age Uzbekistan (Tilla Bulak), Chalcolithic and Bronze Age Iran, Iranian Azerbaijan samples from the Chalcolithic to Iron Age, and finally Bronze Age and Medieval Georgian samples. This trend roughly held for the two other sets of tests. To explicitly test for greater allele sharing between post-Neolithic individuals and the western Neolithic genome Blagotin3, the statistic *D*(Semnan3, Test, Blagotin3, Yak) was computed (Figure 4.13). Non-significant values were obtained for Chalcolithic and Bronze Age samples from Iran and Turkmenistan (Monjukli1,2,4 and 6; Fars1 and Fars4; Chalow1) and a Bronze Age sample from Uzbekistan (Bulak1). Significantly more allele sharing between Blagotin3 and the remaining samples was observed, with the highest *D* values found in Bronze Age, Iron Age, and Medieval samples from Iran (Darre2) and the Caucasus region (Tac3, Kohneh2, Tac3, Kazbeg1, and Geor2).

Does the genetic affinities of goat from the Levant (Israel) change with time?

To investigate the genetic affinities of Levantine goat, the *D*(H1, H2, Ancient Levant, Yak) was computed, where H1 and H2 were pairs of high coverage representative Neolithic genomes (Figure 4.14). Many of these samples were of very low coverage, resulting in high standard errors and nonsignificant Z values; as such, there was limited power in detecting

genetic turnover from the Neolithic. Samples from ‘Ain Ghazal showed greater affinity with Blagotin3 than Semnan3, based on negative values of the *D*(Blagotin3, Semnan3, Ancient Levant, Yak). One Bronze Age sample from Tel Yoqne’am, Israel, showed affinity towards Semnan3. This sample also deviated in affinity with Ainghazal2, the representative genome

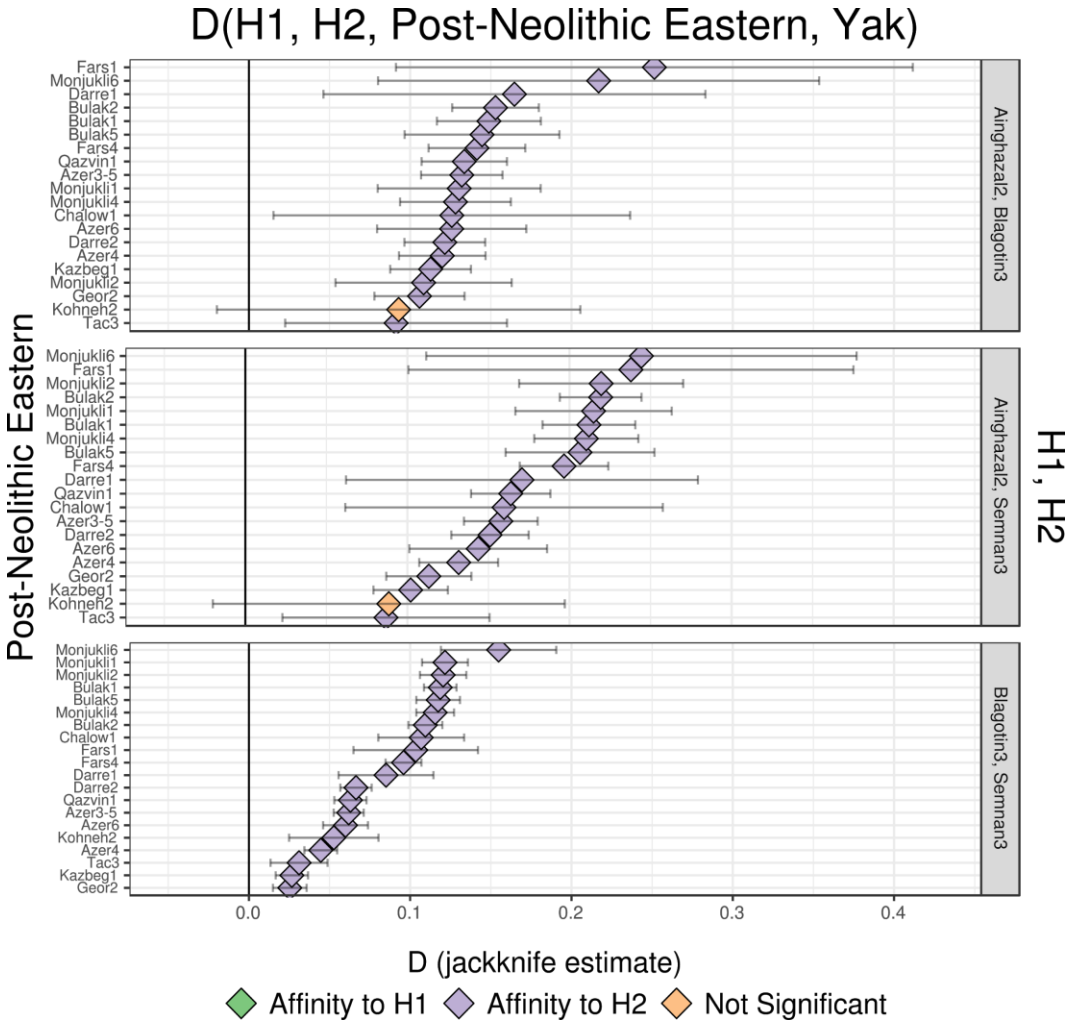


Figure 4.12 - *D* statistic test for relative affinity (derived allele sharing) of a Post-Neolithic Eastern goat to members of pairs of Neolithic genomes, H1 and H2. Colour indicates affinity of the Post-Neolithic Eastern genome to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour blue.

of Neolithic Levant, in the tests *D*(Ainghazal2, Blagotin2/Semnan3, Ancient Levant, Yak). In these cases, Yoqneam2 showed greater derived allele sharing with either Blagotin2 or Semnan3, than with the Neolithic Levantine Ainghazal2, indicating a degree of ancestry change in Israeli goat populations. Other 'Ain Ghazal genomes, and a Chalcolithic individual Gilat8 in one case, showed greater derived allele sharing with Ainghazal2 than either of the other Neolithic genomes.

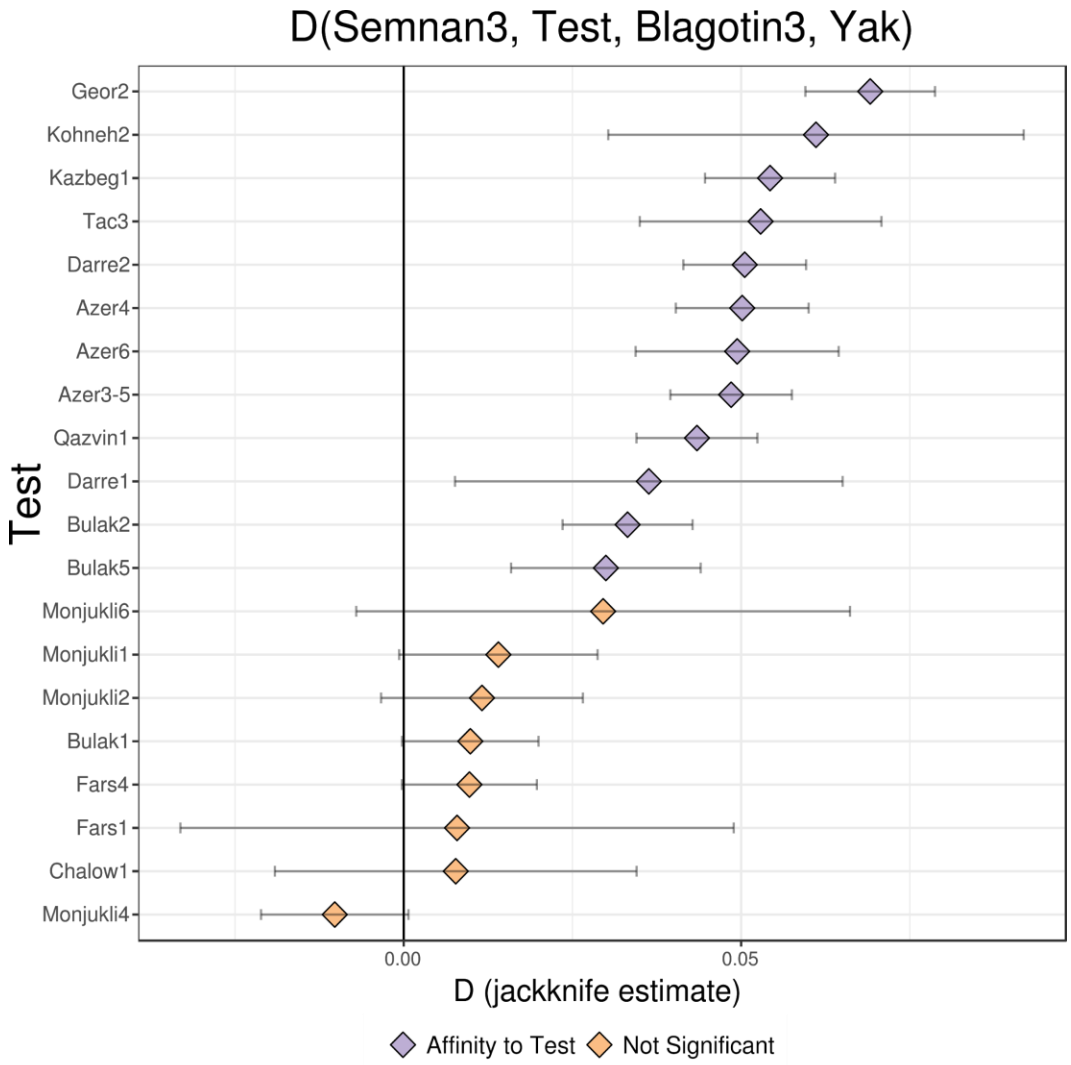


Figure 4.13 - *D* statistics testing the change in derived allele sharing between a high coverage western Neolithic Blagotin3 and Post-Neolithic genomes, relative to an eastern Neolithic reference. Colour indicates derived allele sharing of the Blagotin3 to Semnan3 (H1) or Test (H2), with non-significance ($|Z| < 3$) indicated by the colour orange.

To explicitly test for admixture following the Neolithic, the *D*(Ainghazal2, Test, Neolithic Reference, Yak) was calculated (Figure 4.15). The majority of tests produced nonsignificant results. Yoqneam2, the Bronze Age sample showing shifted affinities relative to Ainghazal2, showed an excess of derived alleles shared with Semnan3 compared with Ainghazal2, consistent with greater eastern ancestry of Yoqneam2.

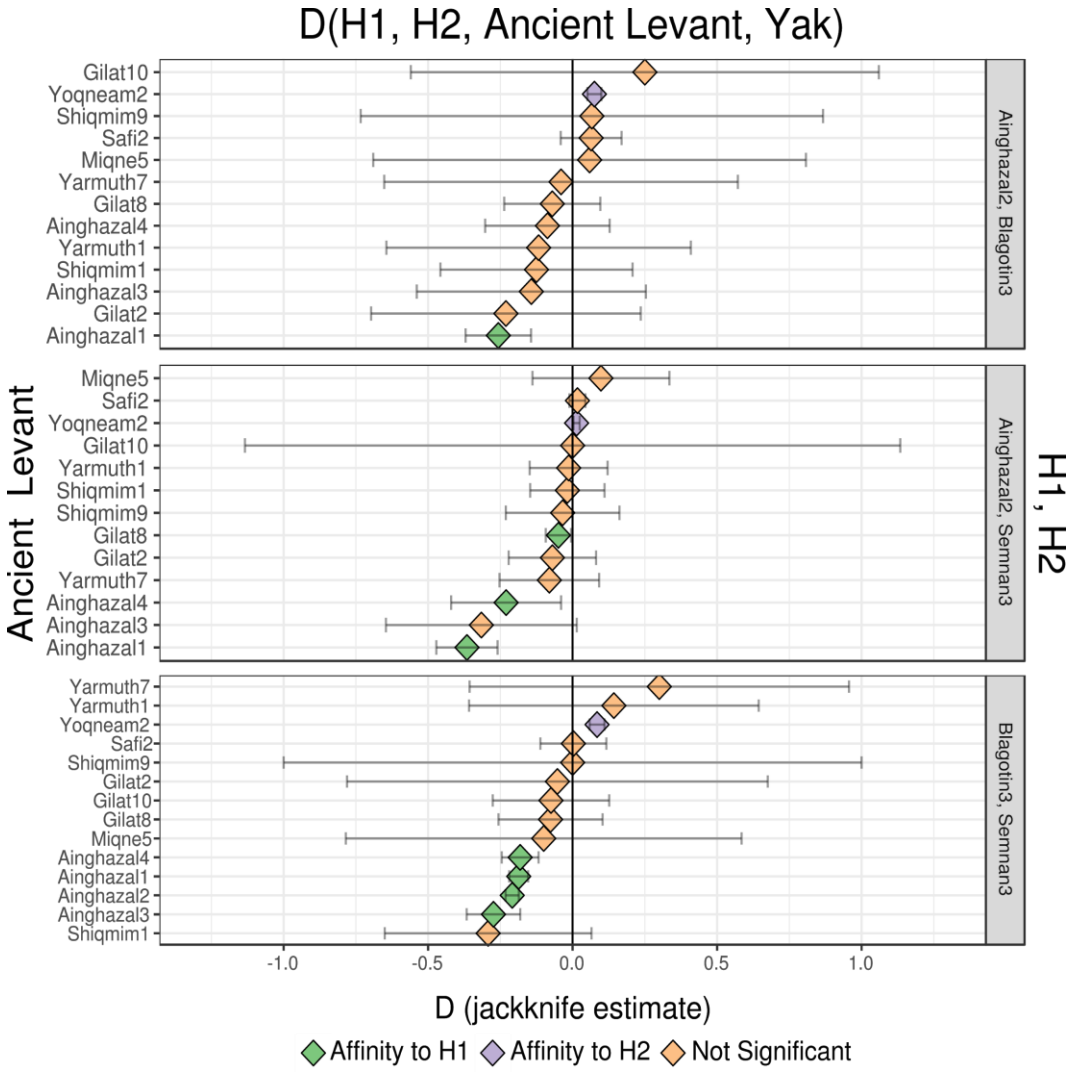


Figure 4.14 - *D* statistic test for relative affinity of a Post-Neolithic Levantine goat to members of pairs of Neolithic genomes, H1 and H2. Colour indicates Test affinity (derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.

In addition, it shared significantly fewer derived alleles with Blagotin3 than Ainghazal2 did. Another Bronze Age Israeli goat, Safi2, also produced significant positive results in the *D*(Ainghazal2, Safi2, Semnan3, Yak). One Chalcolithic sample, Shiqmim1, showed significant allele sharing with Ainghazal2 in both versions of *D*(Ainghazal2, Shiqmim1, Blagotin3/Semnan3, Yak), suggesting an excess of ancestral alleles in Shiqmim1.

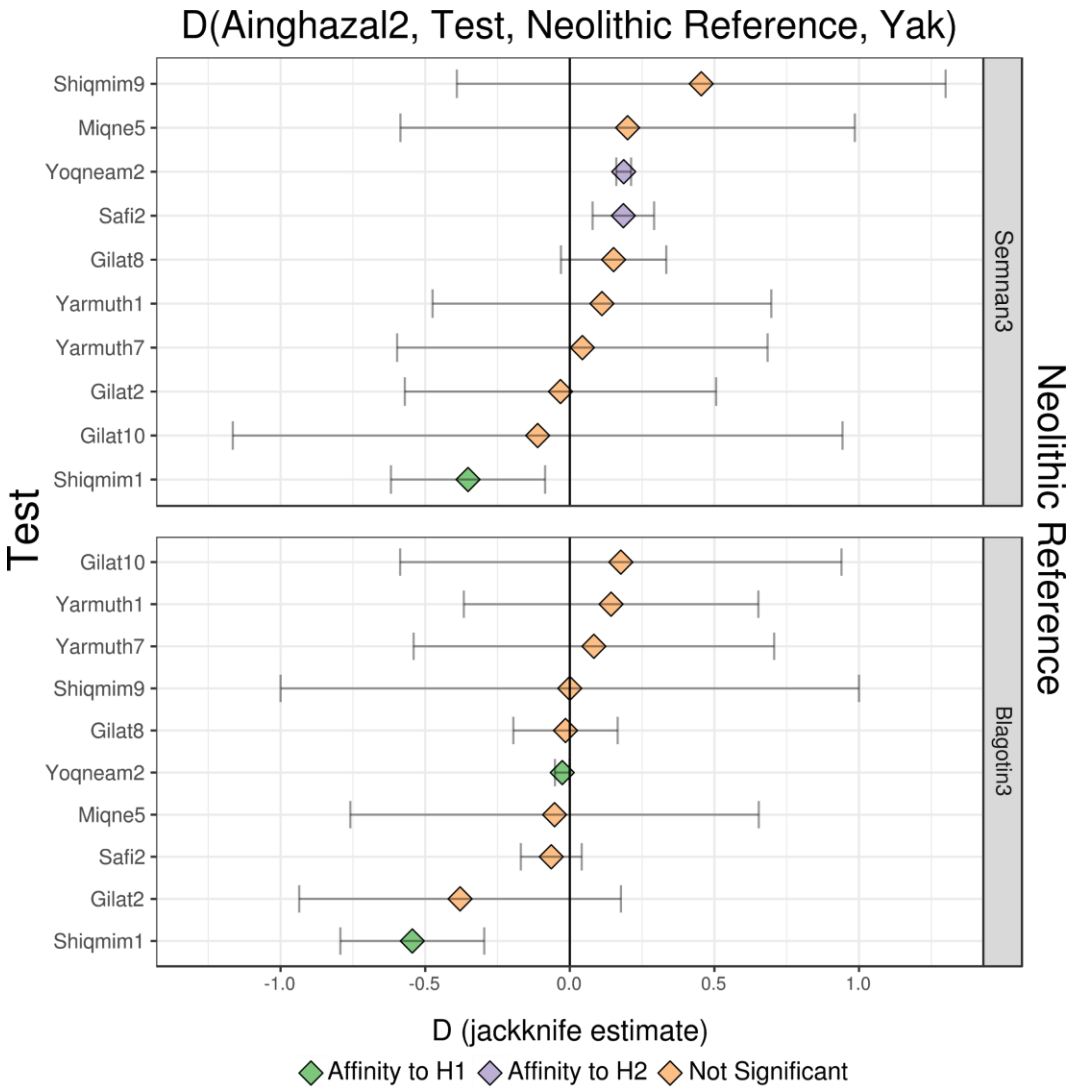


Figure 4.15 - *D* statistics testing the change in affinity of Post-Neolithic Levantine genomes relative to a Neolithic Levant genome Ainghazal2. Colour indicates relative affinity of the Neolithic Reference to Ainghazal2 H1 or the Test H2, with non-significance (|Z| < 3) indicated by the colour orange.

Treemix

Treemix was used to compute population splits and migration events on high coverage (>8×) ancient and modern genomes. Migration events were varied between 0 and 6; models with 0 and 6 migration events are displayed in Figure 4.16. All migration events are displayed in Appendix Figure 4.7.

Under a model of no migration edges, bezoar are modelled as an outgroup to all domestic goat. Within domestics, Neolithic East (Semnan3) first branches out, followed by Bronze Age Turkish (Acem2) and modern Iranian domestics as a clade. African goat form a sister clade to European modern and ancients. A model of a single migration edge results in an admixture event from Ancient Turkish Wilds (Direkli1-2) to the common node of modern and ancient European goat. A second migration edge is modelled as an admixture event from Neolithic East (Semnan3) into modern Iranian Domestics, suggesting that the bifurcating tree model is not sufficient to explain how the two populations relate to each other. When three migration events are modelled, an additional migration edge from Ancient Turkish Wild (Direkli1-2) to Neolithic West (Blagotin3) suggests that different modern and ancient goat populations have differing degrees of Ancient Turkish Wild ancestry (Direkli1-2). The larger amount of shared ancestry observed in Neolithic West (Blagotin3) than the modern European goat population implies that the Neolithic West population alone is insufficient to explain modern European ancestry. Migration edges five and six are modelled as admixture within Africa, from Togolese goat to Moroccans, and between wild goat populations, from a population related to Hamedan bezoar to Qazvin bezoar, suggesting that genetic exchange between domestic populations, and between wild populations, has occurred in addition to wild-domestic admixture.

Ancestry Estimation

Ancestry proportions were calculated using the genotype likelihood-based NGSadmix (Skotte et al. 2013). Ancestral components were calculated with (Figure 4.17) and without (Figure 4.18) modern genomes.

Estimation of ancestry proportions using all modern and ancient samples (Figure 4.17) resulted in Iranian bezoar being modelled as a blue ancestral component, and modern domesticates modelled as a second red component. Some modern individuals (*e.g.* modern

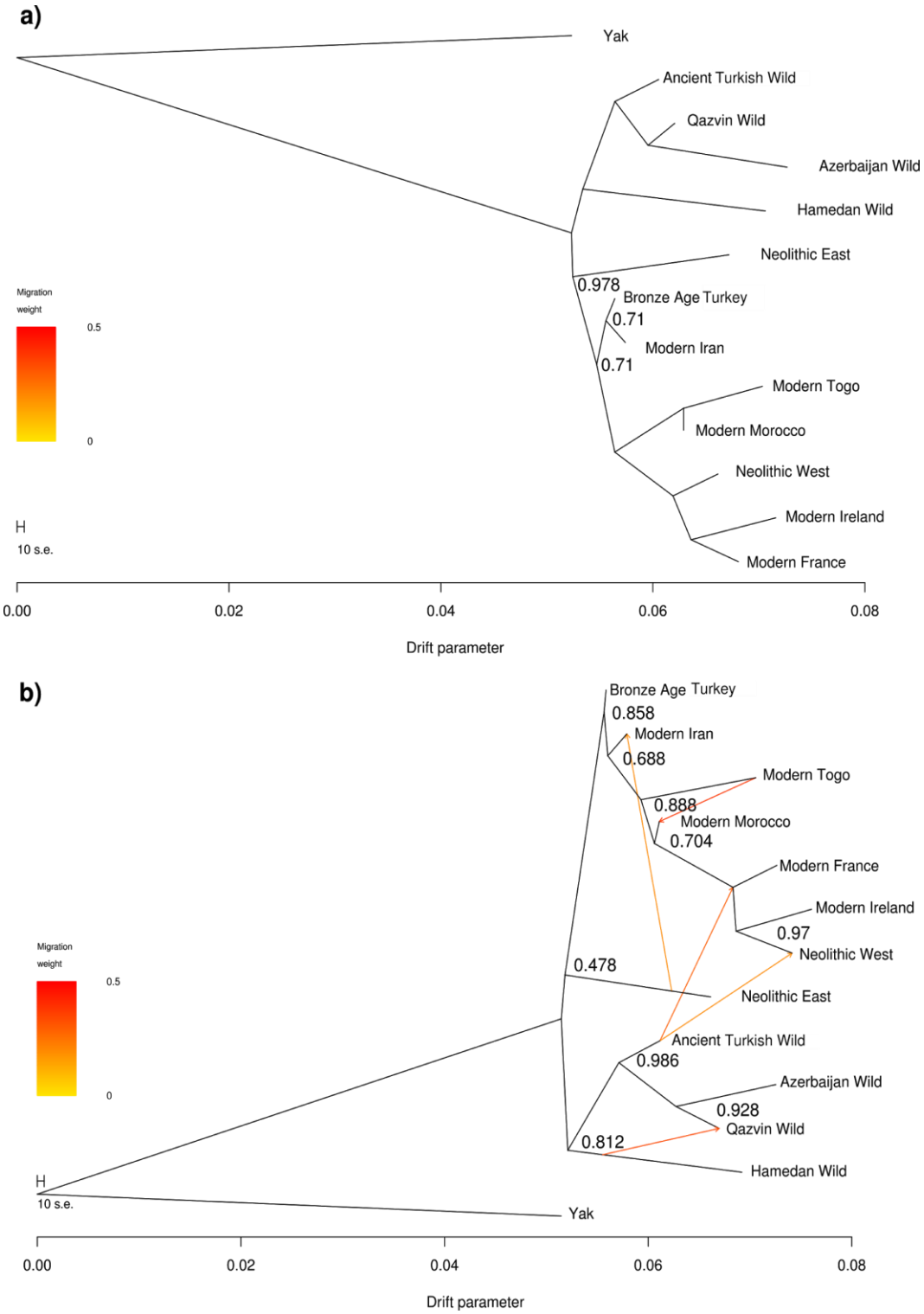


Figure 4.16 - Treemix analysis of high coverage (>8×) samples. Bootstrap support (500 iterations) for branches is displayed with the bootstrap score was less than 1. Migration edges were varied from 0 to 5, shown in a) no migration edge and b) five migration edges.

Europeans) are modelled as having a small proportion of “bezoar” ancestry. Ancient bezoar are modelled as being >50% of the red “bezoar” component. The remaining domestic goat are modelled as predominantly the red “domestic” component, with varying low levels of the “bezoar” which declines slightly through time.

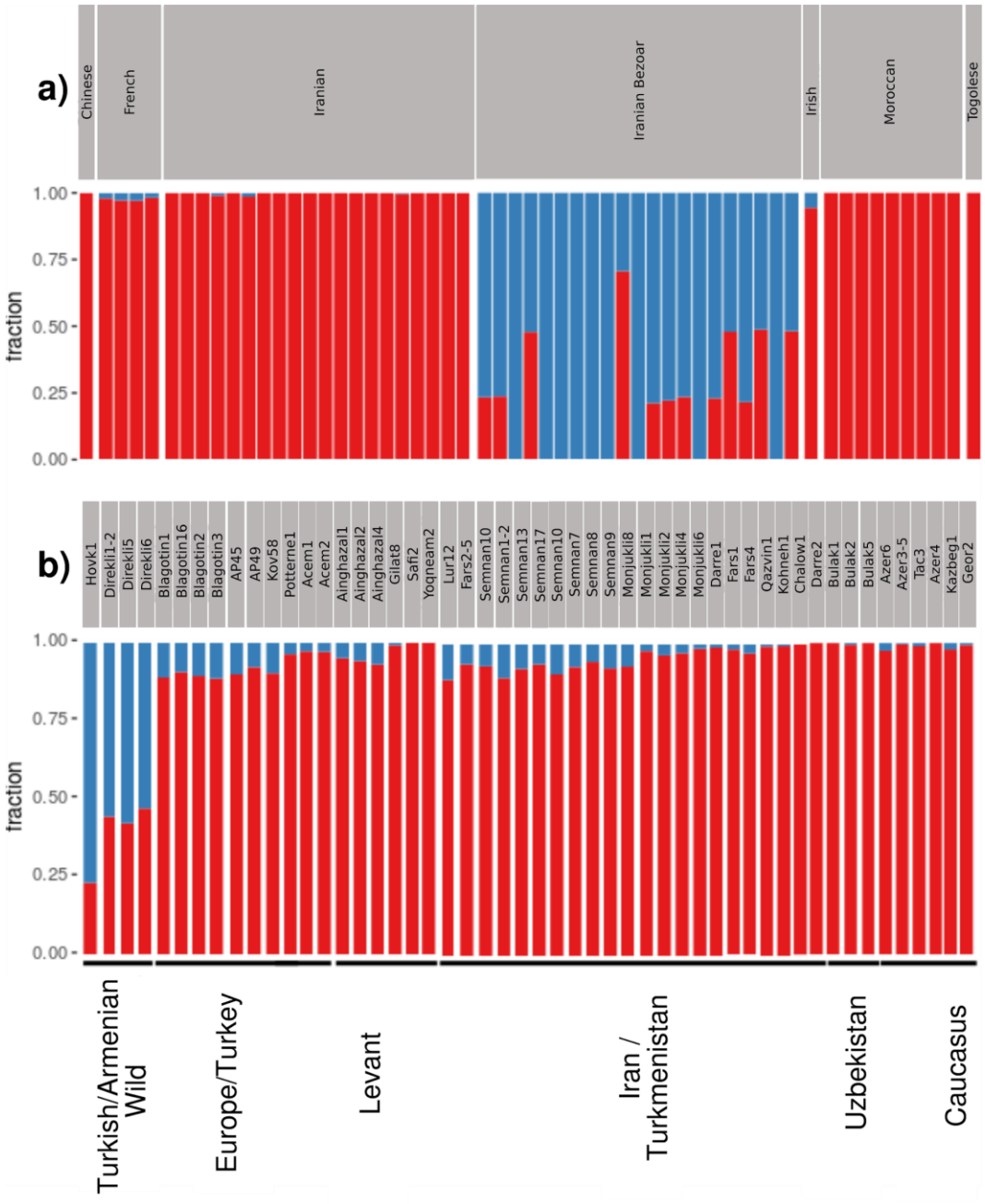


Figure 4.17 - NGSadmix of a) modern and b) ancient wild and domestic goat. Ancestral allele frequencies and genome proportions of a) and b) were calculated together. Sample cutoff of 0.01× mean coverage. K=2. Samples within geographic region ordered by descending age.

Using ancient genomes only, pre-domestic bezoar (excluding Hovk1), Neolithic goat from Serbia, western Turkey and the Levant, and a goat from Bronze Age Britain are modelled as entirely a red component (Figure 4.18). Hovk1, an Armenian sample at least 47,000 thousand years old, is described by predominantly the red "western" component. Eastern Neolithic samples and the majority of those from eastern post-Neolithic contexts are modelled as a single blue component. A subset of post-Neolithic eastern samples are modelled as a mixture of both the blue "eastern" and red "western" components, including samples from the

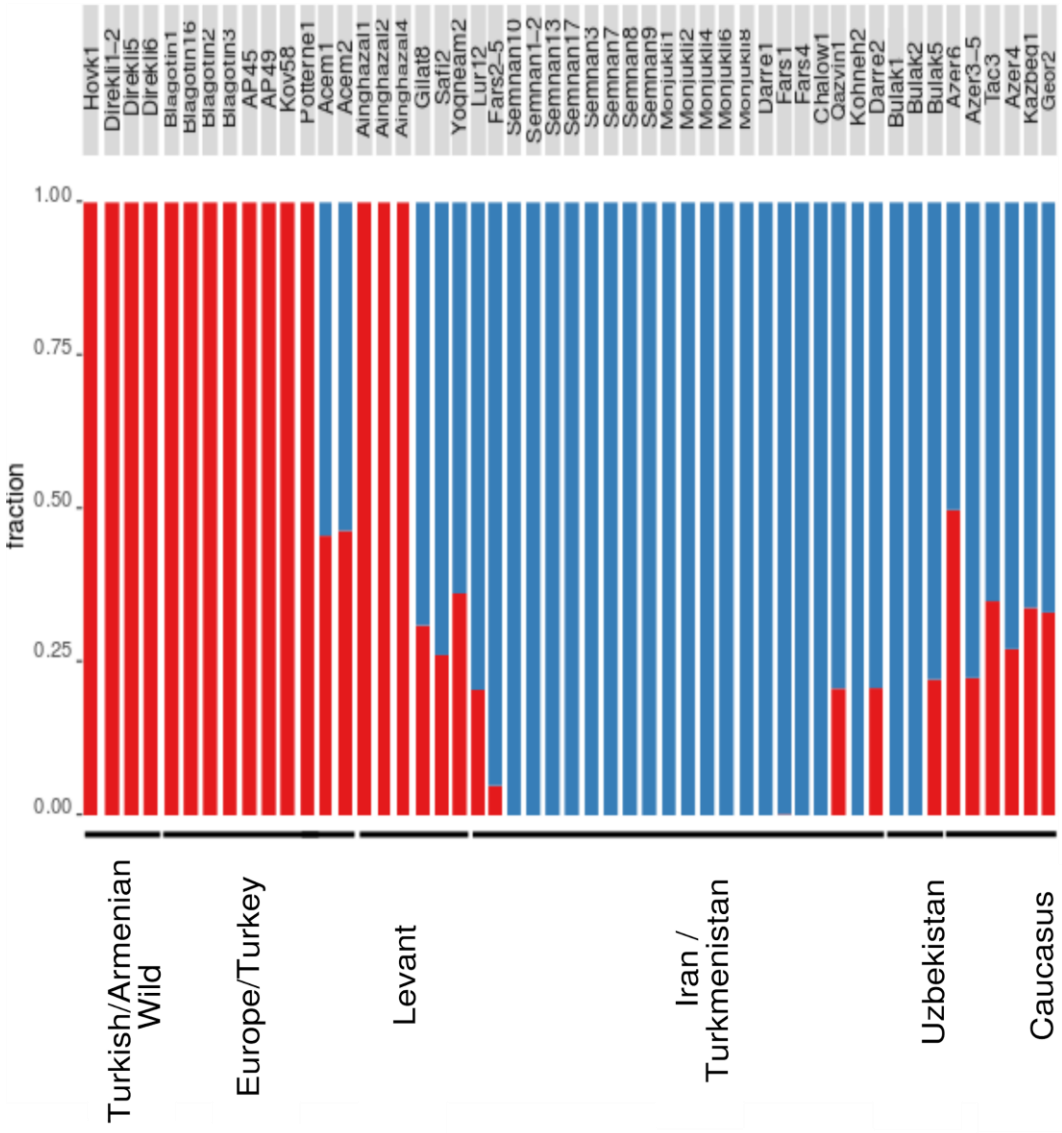


Figure 4.18 - NGSadmix of ancient wild and domestic goat. Sample cutoff of 0.01× mean coverage. K=2. Samples within geographic region ordered by descending age.

Caucasus region (Georgia, Iranian Azerbaijan) and do not appear to have the red "western" component. In contrast, Bronze Age goat from Tilla Bulak, Uzbekistan, also do not show the red component. Samples from 'Ain Ghazal (Neolithic Jordan) are modelled as entirely the red "western" component, while Chalcolithic and Bronze Age samples from several sites in Israel are modelled as a mixture of both, but primarily of the blue "eastern" component.

Outgroup f_3

Shared drift between Neolithic goat groups and ancient/modern goat populations was computed using the outgroup f_3(Neolithic, X; Qazvin Bezoar), where X was an ancient or modern goat group. f_3 values are displayed in Appendix Table 4.7. To visualize how the degree of shared drift with Neolithic populations varied with geography and time, f_3 values were plotted on a relief map of southwest Asia (Figure 4.19) (Виктор 2010).

A strong difference was observed between eastern and western Neolithic populations. Shared drift with Neolithic East genomes was high in later eastern populations and low in more western population; the opposite was observed for Neolithic western shared drift. Shared drift with Neolithic West was highest in Bronze Age British and modern European goat, and also high with Neolithic Levant and Modern Africa. Shared drift between Levantine goat and Neolithic West decreased with time, while affinity between western Neolithic and Iranian populations increases closer to the present. A correlation between western and Levantine Neolithic shared drift was also observed: populations with low shared drift with Neolithic West tended to also have low shared drift with Neolithic Levant. A greater degree of shared drift was observed between Neolithic Levant and Modern Africa, compared to between Neolithic West and Modern Africa. Neolithic Iran shows high levels of genetic affinity with post-Neolithic Iranian, Caucasus, and Central Asian populations. A change in shared drift with Neolithic Iran is observed in the Levant; low eastern genetic affinity in Neolithic Levant is followed by greater affinity in Chalcolithic and Bronze Age Levantine goat.

The observed affinity of Neolithic Levant to modern African samples may be confounded by Neolithic West-like ancestry in modern Africa. To investigate if the shared drift of Neolithic Levant and Modern Africa is independent to the drift shared between Neolithic West and Modern Africa, f_3 for all pairwise combinations of Neolithic population were then plotted with a linear regression and associated confidence interval using the ggplot (Wickham 2009) function geom_smooth (Figure 4.20). Three populations show an excess of Neolithic Levant

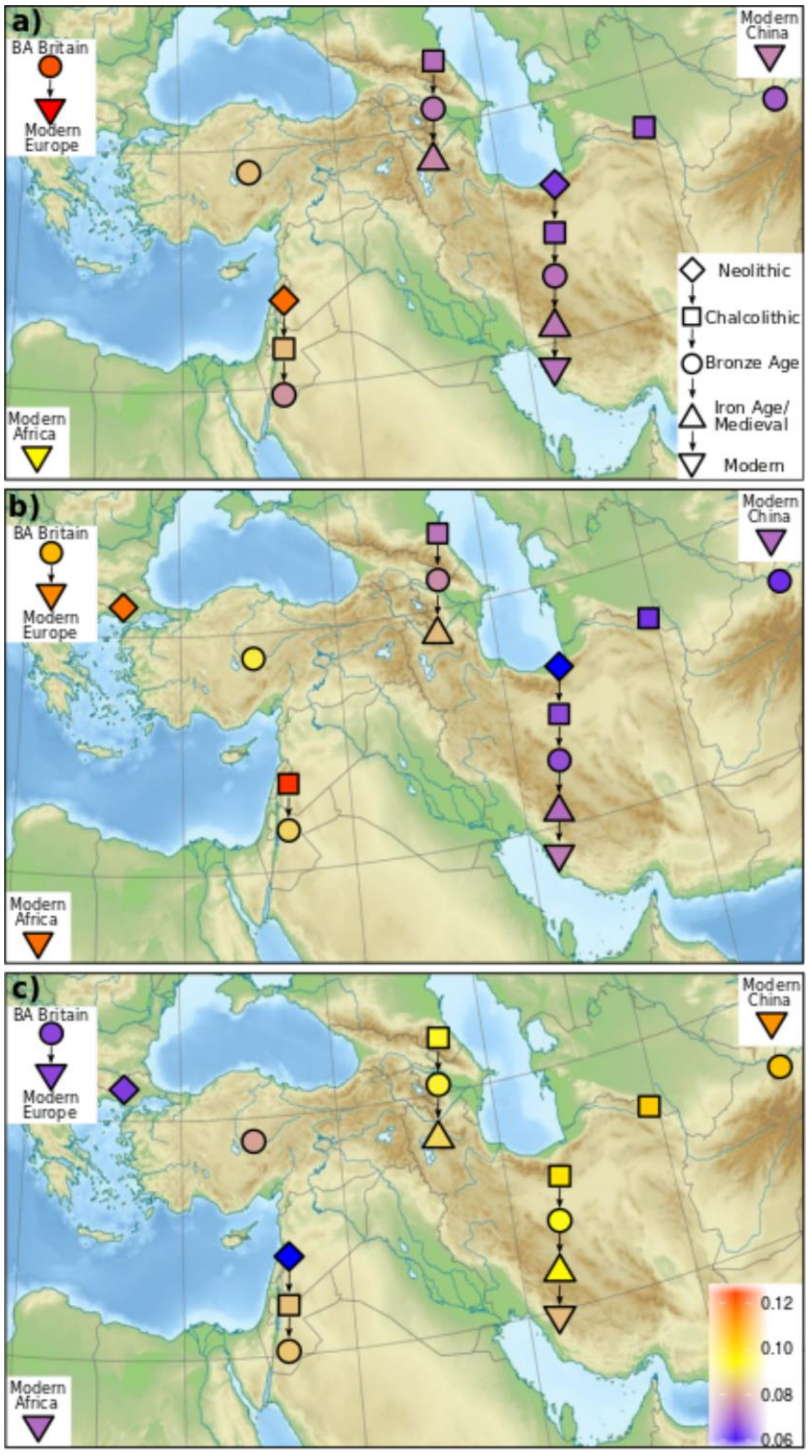


Figure 4.19 - Plot of outgroup f_3 values of ancient and modern domestic goat, measuring relative affinities with a) Neolithic West, b) Neolithic Levant, and c) Neolithic East. Qazvin Bezoar were selected as an outgroup due to the equal affinity to Neolithic East and Neolithic West, based on the *D* statistic *D*(Neolithic East, Neolithic West, Qazvin Bezoar, Yak), Z=-0.8 (Appendix Table 4.11).

shared drift relative to their drift with Neolithic West: Chalcolithic Israel, Modern Togo, and Modern Morocco, suggesting ancestry shared with Neolithic Levant, but not Neolithic West, is present in these populations. Of these, only Chalcolithic Israel was significantly different from expectation, under the t-distribution (two tailed test, alpha=0.05, 16 degrees of freedom; see Appendix Table 4.8).

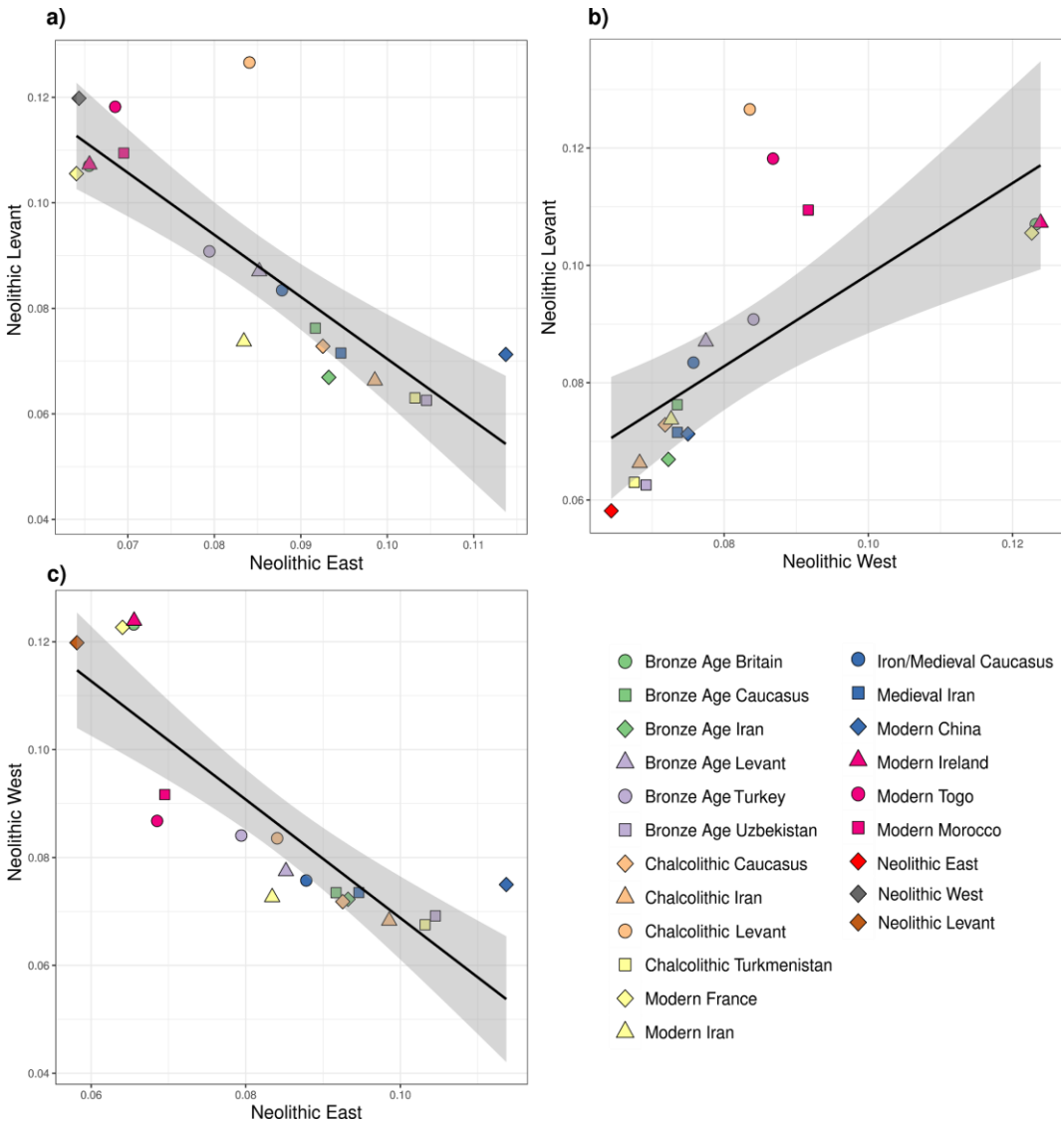


Figure 4.20 - Pairwise shared drifts (outgroup f_3) plots, with linear regression and 95% confidence interval. Each domestic population shared drift with two Neolithic populations is plotted against one another: a) Neolithic Levant versus Neolithic East, b) Neolithic Levant versus Neolithic West, and c) Neolithic West versus Neolithic East. Outgroup used is Qazvin Bezoar.

Shared drift between the modern genomes presented here, IOG and Tog, was also determined (Appendix Table 4.7). Modern Togo shows highest shared drift with Modern Morocco, in line with their geographic proximity in Western Africa. Ancient Levantine and modern European populations show the next highest degree of shared drift with Modern Togo, suggesting that modern goat from western Africa share ancestry with a population related to European goat, Levantine goat, or a mixture of both. Lowest shared drift is observed with eastern populations (Iran, Turkmenistan, China). The highest shared drift between Modern Ireland is Bronze Age Britain (f_3 =0.151, SE=0.002) rather than Modern France (f_3=0.139, SE=0.002). High drift was also observed with Neolithic West, Neolithic Levant and modern African populations, while low shared drift was observed with eastern populations.

A heatmap was constructed using f_3 outgroup shared drift between all modern and ancient goat groups (Figure 4.21). Two primary clusters were observed: the first consisting of Neolithic Levant, European, and African populations, while the second was composed of eastern (Iran, Caucasus, modern China and Iran), Bronze Age Turkish, and post-Neolithic Levantine groups. Of these, Chalcolithic Israel was somewhat of an outlier, showing higher shared drift with Modern Togo and Neolithic Levant than the Bronze Age Levant population.

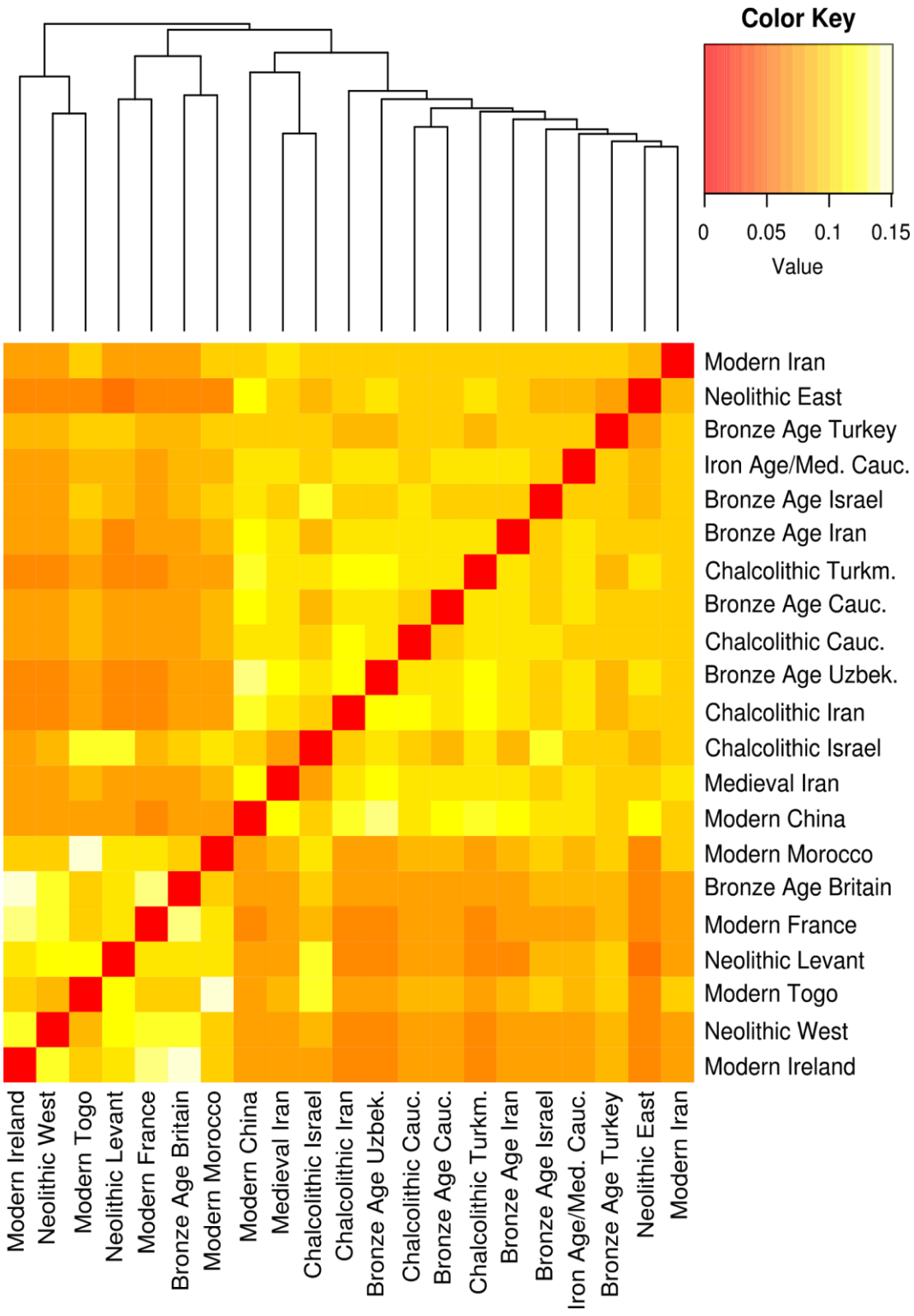


Figure 4.21 - Heatmap of f_3 outgroup (shared drift) values. Qazvin Bezoar was selected as the outgroup.

f_4 ratios

f_4 ratios were used to estimate the proportion of ancestry of one ancient population attributable to another population. Population-level results are displayed in Table 4.2. Individual level results are displayed in Appendix Table 4.9.

The ratio f_4(Yak, Ancient Turkish Wild; Neolithic West/Levant, Neolithic East) / f_4(Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East) was used to estimate the proportion of Ancient Turkish Wild-like ancestry present in pooled populations of Neolithic Levant and Neolithic West. Both western Neolithic and Levantine Neolithic groups show approximately 50% of their ancestry as deriving from the Ancient Turkish Wild population, with Neolithic Levant showing a higher proportion (0.56) but a greater standard error (0.07).

The ratio f_4(Yak, Neolithic East; post-Neolithic Levant, Neolithic Levant) / f_4(Yak, Neolithic East; Semnan3, Neolithic Levant) and f_4(Yak, Neolithic West; post-Neolithic Levant, Neolithic Levant) / f_4(Yak, Neolithic West; Blagotin3, Neolithic Levant) were employed to estimate the proportion of post-Neolithic ancestry that could be attributed to eastern or western Neolithic genomes. In both cases, nonsignificant results were obtained, likely due to high standard error.

To test if a significant degree of the Bronze Age Turkish goat ancestry could be explained by eastern Neolithic goat, the ratio f_4(Yak, Neolithic East; Bronze Age Turkey, Neolithic West) / f_4(Yak, Neolithic East; Semnan3, Neolithic West) was computed. Significantly negative values were obtained (Z=-5.023).

To test if the genomes of post-Neolithic eastern genomes had a greater proportion of western Neolithic-like ancestry, the ratio f_4(Yak, Neolithic West; post-Neolithic East, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East) was calculated across different geographic and temporal groupings. Negative or nonsignificant values were obtained for Chalcolithic Iran and Turkmenistan, Bronze Age Iran and Caucasus, with the later approaching significance (Z=2.862). Significant ratios were obtained for Bronze Age Uzbekistan (0.08), Iron Age/Medieval Caucasus (0.1), and for Modern Iranian goat (0.11). When f_4 ratios were computed for Bronze Age Uzbekistan individuals, a significant positive ratio was obtained for only a single individual, Bulak1 (0.2).

Table 4.2 - f_4 ratio results for goat populations. The two underlined populations (X, Y) indicate that test estimates the proportion of Y ancestry in X.

Test	Ratio (α)	Error	Z
f_4(Yak, Ancient Turkish Wild; Neolithic West, Neolithic East) / f_4(Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East)	0.500783	0.016381	30.571
f_4(Yak, Ancient Turkish Wild; Neolithic Levant, Neolithic East) / f_4(Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East)	0.556317	0.06707	8.295
f_4(Yak, Ancient Turkish Wild; Bronze Age Britain, Neolithic East) / f_4(Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East)	0.461403	0.026913	17.145
f_4(Yak, Neolithic East; Bronze Age Turkey, Neolithic West) / f_4(Yak, Neolithic East; Semnan3, Neolithic West)	-0.166694	0.033185	-5.023
f_4(Yak, Neolithic East; Bronze Age Israel, Neolithic Levant) / f_4(Yak, Neolithic East; Semnan3, Neolithic Levant)	-0.076243	0.139717	-0.546
f_4(Yak, Neolithic West; Bronze Age Israel, Neolithic Levant) / f_4(Yak, Neolithic West; Blagotin3, Neolithic Levant)	-3.72947	2.96882	-1.256
f_4(Yak, Neolithic West; Chalcolithic Iran, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	-0.112181	0.022049	-5.088
f_4(Yak, Neolithic West; Chalcolithic Turkmenistan, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	-0.11759	0.022542	-5.217
f_4(Yak, Neolithic West; Bronze Age Uzbekistan, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	0.081125	0.015157	5.352
f_4(Yak, Neolithic West; Bronze Age Iran, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	0.018827	0.021046	0.895
f_4(Yak, Neolithic West; Bronze Age Caucasus, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	0.055295	0.019323	2.862
f_4(Yak, Neolithic West; Iron Age/Medieval Caucasus, Neolithic East) / f_4(Yak, Neolithic West; Blagotin3, Neolithic East)	0.095781	0.015674	6.111

qpGraph

Population models were iteratively compared to the ancient goat genomes in order to find models of their history which best described/fit the data. A large number of models were tested against the data, in order to accommodate an increasing number of populations. These consisted of models featuring both ancient and modern goat populations. That process is described fulling in Appendix Text 4.1, with accompanying figures in Appendix Figures 4.1 and 4.2. The general features of these graphs are described below, with the final model fitting pre-Neolithic, Neolithic, and modern goat displayed in Figure 4.22.

The final admixture graph models a primary division between Neolithic goat and wild pre-Neolithic bezoar. These two lineages then contribute ancestry to Neolithic populations: Levantine (26% : 74%), western (41% : 59%) and eastern (78% : 22%). The source of the "domestic-like" ancestry is different between eastern Neolithics (nE), and both western (nW) and Levantine Neolithics (nL); the "domestic-like" contributor to the later pair has experienced substantial drift (317 drift units / 0.317 F_{ST}). The "wild-like" ancestor of these populations is sister to the Direkli bezoar (D) from central Turkey, which themselves form a sister clade with ancient Armenian genomes Hovk1 (aAR).

The modern Chinese goat reference (mCH) is modeled as received 98% of its ancestry from the "eastern domestic-like" population. Surprisingly, the remaining 2% of modern Chinese goat ancestry is donated from Neolithic Levant (nL). Modern European goat (mEU) are modeled as primarily (96%) descending from a sister group of western Neolithic goats, with a small contribution (4%) from an outgroup to goat and bezoar. Modern Africans are modelled as a complex mixture of modern European (35.5% final), the ancestors of western (7.5% final) and eastern goat (48%), Neolithic Levant (7.0%), and a minor contribution from ancient Armenian bezoar (2%).

This final model produced three outlier f_4 statistics (Z=3.03-3.22), suggesting that this model is imperfect and requires additional sampling of certain populations to fully resolve. These outliers indicated unmodelled affinity between Neolithic East and Modern Europe, Modern Chinese and Neolithic Levant, Neolithic Levant and Modern Africa, or Ancient Armenian Wild and Neolithic Levant. As this analysis was restricted to 11,740 SNPs (Appendix Table 4.4) and featured several populations represented by a small number of pseudo-haploid genomes, this graph should be taken as a tentative initial attempt to model the population

history of goat, which will benefit greatly from higher quality data and a broader sampling of geographic regions.

Additional graphs were generated to fit post-Neolithic ancient goat populations as mixtures of Neolithic and pre-Neolithic groups (Appendix Figures 4.1 and 4.2). Bronze Age Levant is modelled as a mixture between the ancestors of Neolithic Levant and Neolithic East (24% : 76%). Bronze Age Turkey is modelled as a mixture between the groups ancestral to eastern and western Neolithics (44% : 56%). Bronze Age Britain is modelled as a mixture of "western domestic-like" and "Direkli-like" ancestries (49% : 51%), with western Neolithics requiring an additional 13% contribution from the "Direkli-like" source. Chalcolithic and Bronze Age Caucasus populations are best fit as a mixture between groups related to eastern and western Neolithic goat (77% : 23%); Iron Age and Medieval Caucasus groups require a further 11% ancestry from a western Neolithic-like source. Chalcolithic Iranian and Turkmen goat genomes are fit as a mixture of eastern and western populations, (66% : 33%) and (54% : 46%) respectively. Bronze Age Iranian and Uzbek goat are also described as east-west mixture, with successive contributions from a Neolithic western-like ancestor (21% and 16%). Finally, the single Medieval Iranian genome Darre2 is fit as a simple mixture between the ancestors of eastern and western Neolithic goat (52% : 48%).

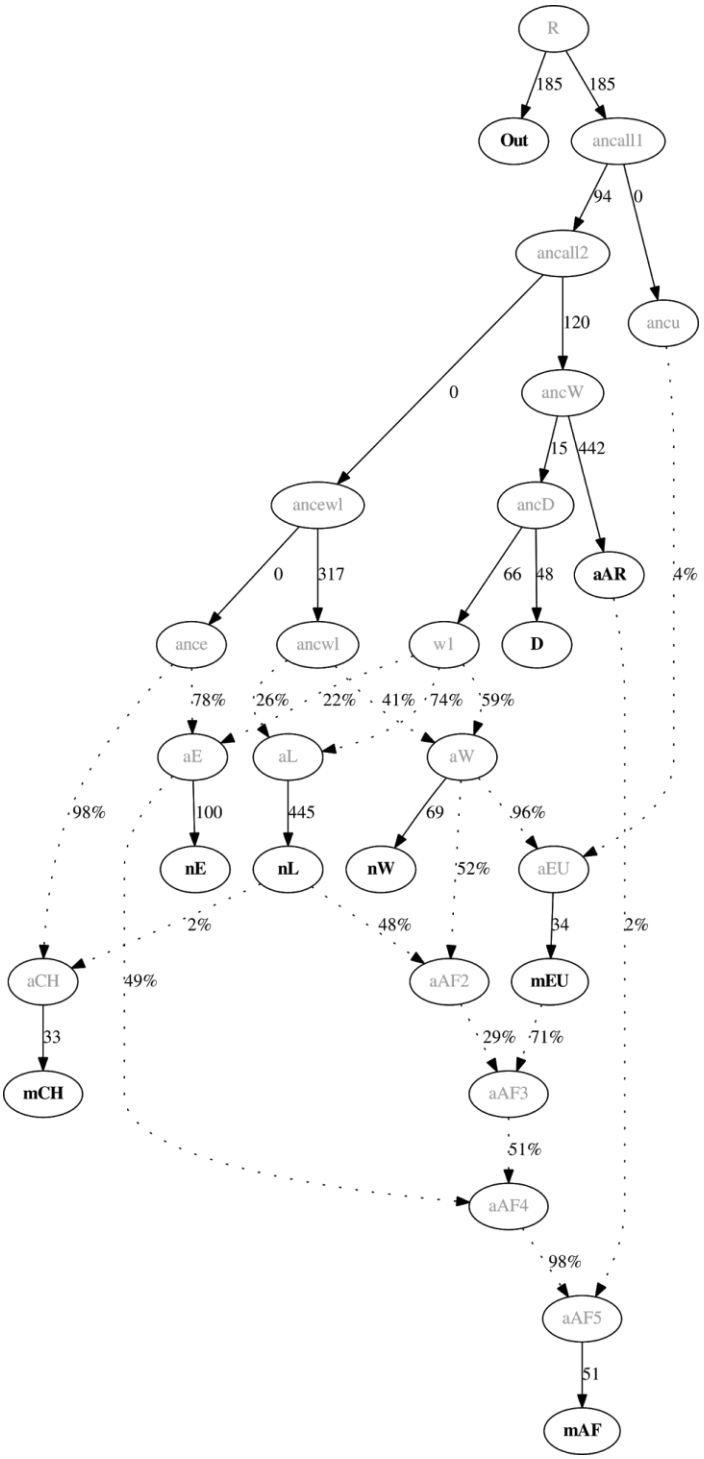


Figure 4.22 - Final admixture graph model for ancient and modern domestic goats. Intermediate, theoretical populations are denoted in grey. Edge drift values = F_{ST} x 1000. nW=Neolithic West, nE=Neolithic East, nL=Neolithic Levant, D=Turkish Ancient Wild (Direkli), aAR=Armenian Ancient Wild (Hovk1), mEU=Modern Europe, mCH=Modern China, AF=Modern Africa.

Autosomal mutation rate estimation

The autosomal mutation rate was estimated using the F(A|B) method, which exploits a sample of known age and the demographic history of a related modern to calibrate the autosomal mutation rate. The calibration curve constructed (Figure 4.23) indicates that the rate of 1.3×10^{-8} per site per generation (5.2×10^{-9} per site per year using a generation time of 2.5 years) fits best with the observed F(A|B) and the radiocarbon age of Blagotin3.

Nuclear Genome Modelling

The nucleotide diversity estimated for both western and eastern Neolithic groups (π_3W, π_3E) was 0.15 and 0.16 respectively, and their F_{ST} calculated as 0.17 (Table 4.3; Appendix Table 4.10). Model posterior probabilities suggest model BINARY_AU as the most supported model by the data using two thresholds of simulations retained (25,000 and 50,000) (Table 4.4). This suggests that the divergence between these Neolithic goat populations cannot be simply modelled as a shared bottleneck and recent divergence, and instead a pre-bottleneck divergence better fits the data.

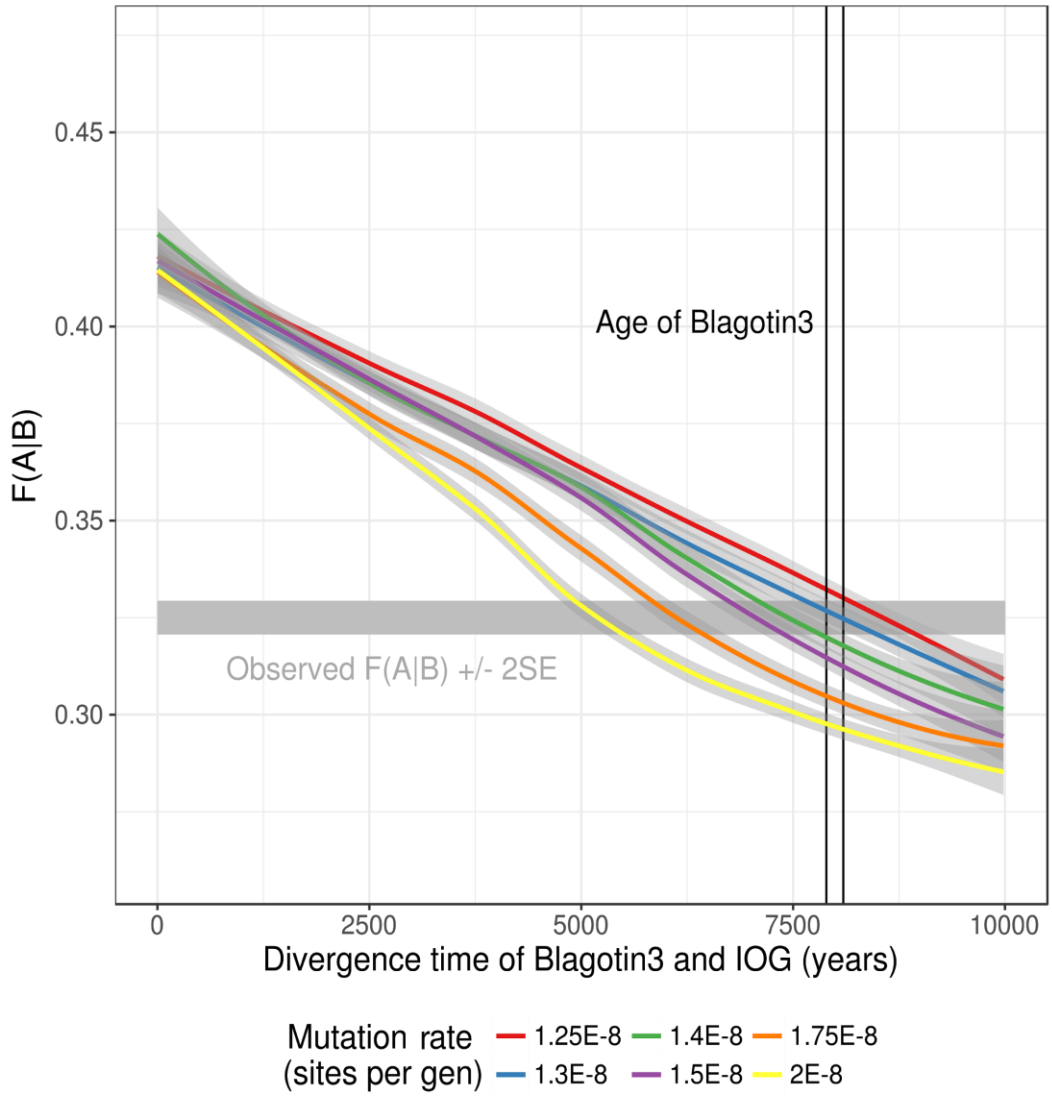


Figure 4.23 - Estimation of the *Capra hircus* genome mutation rate, using Blagotin3 and IOG. Based on the overlap between the observed F(A|B) value, the radiocarbon age of Blagotin3 and the estimated F(A|B) ratio at different mutation rates, $1.3x10^{8}$ sites per generation was chosen as the mutation rate for subsequent analyses.

Table 4.3 - Genetic diversity based on whole genome sequences. π: nucleotide diversity

Population	Acronym	Sample size	π per site
Neolithic West	3W	4	0.15
Chalcolithic and Bronze Age West	2W	2	0.17
Neolithic East	3E	5	0.16
Chalcolithic, and Bronze Age East	2E	2	0.17
Iron Age, Medieval and Modern East	1E	3	0.17

Table 4.4 - Hudson's pairwise F_{ST} based on whole genome sequences. Acronyms used are the same as in Table 4.3.

	1E	2E	2W	3E	3W
1E	-	0	0.01	0.07	0.12
2E	0	-	0.02	0.08	0.13
2W	0.01	0.02	-	0.09	0.1
3E	0.07	0.08	0.09	-	0.17
3W	0.12	0.13	0.1	0.17	-

Table 4.5 - Model posterior probabilities calculated by a weighted multinomial logistic regression using whole genome sequences. Models are presented in Figure 4.4.

Number of simulations retained	Model SINGLE_AU	Model BINARY_AU
25,000	0	1
50,000	0.26	0.74

Discussion

Nuclear genome structure in Neolithic goat

A tripartite geographic structure is observed in the Neolithic whole genomes sampled for this project. This is remarkably consistent with the distribution of mtDNA haplogroups observed in Chapter 3. Goat sampled from the Levant (Jordan), Iran, and Serbia fall in disparate regions in PCA space (Figure 4.6). Principal component 1, explaining 5.1% of the variation, appears to differentiate western (Europe, west Turkey) and eastern (Iran, Turkmenistan, Uzbekistan, Georgia) Neolithic samples. Levantine Neolithic samples fall closer to the "western" end of PC1, and themselves are differentiated by their position on PC2. This structure is also observed in the Neighbour-Joining tree constructed using IBS information (Figure 4.7), in which these three groups form distinct clades or sister clades; an exception to this are two Neolithic genomes from western Turkey (AP45 and AP49), which are form an outgroup clade to Neolithic European and Levantine genomes.

These Neolithic genomes from western Turkey fall in a less extreme position on the "western" end of PC1, which primarily divides samples from western (Turkey, Europe) and eastern (Iran, Turkmenistan, Georgia, Uzbekistan) regions. Similarly, two samples (Lur12 and Fars2-5) from the some of the earliest Neolithic Zagros sites are less extreme in their PCA position than later goat from more easterly sites in Iran and Turkmenistan, but are still distinct from western and Levantine Neolithic genomes. This may be a result of genetic bottlenecks in the history of European and east Iran/Turkmenistan, a consequence of their introduction into the regions by early farmers. Such losses of genetic diversity would result in these samples falling in more extreme locations in the PCA space, compared to goat populations which did not experience such strong population bottlenecks. Serial founding events are thought to have occurred as goat were introduced to Europe (Gerbault et al. 2012), and likely also happened as farmers introduced them from western to eastern Iran and Central Asia. This explanation is consistent with the mixed ancestry profile of Fars2-5 and Lur12, compared to those from Neolithic Sang-e Chakmaq and Monjukli Depe (Figure 4.18); the strong drift associated with bottlenecks is expected to influence ancestry estimation, modelling bottlenecked populations as a single component (Lawson et al. 2018). Additionally, these western Iranian Neolithic genomes do not show an excess of Neolithic Western derived alleles (i.e. European, western Turkish) compared to eastern Iranian genomes, using the explicit test *D*(Semnan3, Lur12/Fars2-5, Blagotin3, Yak) (Appendix

Table 4.3), despite the presence of the blue "western Neolithic" component in Fars2-5 and Lur12 (Figure 4.18), and other affinity measure results (Figure 4.8).

The distinction in "western", "eastern", and "Levantine" genetic backgrounds correlated with the mtDNA haplogroup distribution previously inferred. Goat samples from Neolithic "western" regions overwhelmingly have haplogroup A sequences, observed in Neolithic France (Fernández et al. 2006) and at ~90% frequency in modern worldwide populations (Naderi et al. 2007). The Neolithic genomes from southeast Europe show strong affinity with modern European populations, based on PCA and IBS clustering (Figures 4.6 and 4.7). In contrast, the Neolithic Levantine goat have both mtDNA F haplogroups and nuclear genome diversity not well represented in modern populations. Finally, the mtDNA and nuclear genomes profiles of Neolithic goat from Iran and Turkmenistan are somewhat similar to later goat from the same region, possibly having admixed with a western Neolithic-like population (see below), but have undergone a substantial mtDNA turnover, with some persistence of local haplotypes. A preliminary reading of this would be that the two sets of loci (i.e. the nuclear and mitochondrial genomes) are correlated in Neolithic. For example, haplogroup A (and C) appears to co-occur with the "western" genomic background in the Neolithic, while B, D, and G are found only in goat with the "eastern" profile. Whether this holds for later time periods - that the increase in haplogroup A frequency was accompanied by an increase in "western Neolithic" affinity - will be examined below.

The relationships between Neolithic goat populations

The simplest and most crucial observation of Neolithic goat genomes sampled here is the single clade they form with respect to the available bezoar genomes (leaving the discussion of wild admixture until later). This pattern occurs in the NJ phylogeny constructed using IBS data of all genomes >0.01× mean coverage (Figure 4.7), admixture graphs using the same genomes (Appendix Text 4.1, Appendix Figure 4.1), and the Treemix graph constructed using genomes with mean coverage >8× (Figure 4.16). Additionally, the first principal component computed using modern and ancient bezoar differentiates all wild goat from domestic samples (Figure 4.5). Together, this evidence supports a monophyly of domestic goat, and a single domestication in the strict sense (Larson & Burger 2013).

How these Neolithic populations relate to one another was explored using several analyses. *D* statistics indicate that western Neolithic genomes share more derived alleles with Levantine

Neolithics than with eastern Neolithics (Figure 4.8). The close affinity between western and Levantine Neolithic goat, already observed in PCA (Figure 4.6) and IBS analyses (Figure 4.7), are also supported by ancestry estimation using genotype likelihoods (Figure 4.18), which models the two groups as having similar ancestry profiles, if not their own ancestral component (Figure 4.8). Shared drift as measured using outgroup f_3 values appear to be highly correlated with between the two groups (Figures 4.20), and they are both members of the same major population group when hierarchical clustering is performed on outgroup f_3 values (Figure 4.21). Finally, admixture graph analyses better fit a model in which western and Levantine Neolithic shared a greater degree of drift than either with eastern Neolithic goat (Appendix Text 4.1). These two populations likely also share ancestry from a related wild population (see discussion below, and Figure 4.22).

Taken together, there is ample evidence that the Levant and western Neolithic populations shared ancestry that was distinct from that of eastern Neolithic goat. The exact nature of this relationship is more difficult to untangle, and beginning to do so requires contextual information. The domestic status of the Levantine samples here is pertinent to the question; the western Neolithic goat are unquestionably from populations far into the domestication process, due to their geographic origins far outside the natural range of bezoar. The ‘Ain Ghazal samples are more contentious (see Chapter 3 discussion), and fall within the bezoar Holocene distribution. If Neolithic Levant represent wild or managed goat that did not contribute to the modern domestic goat gene pool, then the shared drift with western Neolithics may represent the wild ancestry related to the Levantine samples that is present in western Neolithic goat due to admixture (see discussion below). Alternatively, Neolithic Levantine goat may have been seeded from a population related to the ancestors of western Neolithics, before mixing with local goat recruited in managed herds (Munro et al. 2018). Thus, their shared drift may be due to a “domestic” ancestry common to both groups.

The data fit both of these models (Appendix Text 4.1); however, bezoar from central Turkey share more derived alleles with western Neolithics rather than with Levantine Neolithics - *D*(Yak, Pre-domestic Turkish Wild, Neolithic Levant, Neolithic West, Z=-16.8). Multiple admixture events between domestic and managed populations would confound this test, and make distinguishing these two scenarios difficult. The observation that Neolithic Levant falls within domestic goat diversity using IBS data (Figure 4.7) is evidence in favour of their domestic ancestry, as are the ancestry compositions estimation by NGSadmix (Figure 4.18), which models the Levantine genomes as primarily a “domestic” red component rather than

the “bezoar” blue component. Principal Component Analyses also group the Neolithic Levantine goat with domestic goat, apart from pre-Neolithic and modern bezoar (Figure 4.5). These analyses are by no means definitive, and should be reassessed using additional Neolithic genomes from the Levantine coast to the Zagros Mountains of western Iran.

The relationship between eastern Neolithic goat and western/Levantine populations is also not straightforward. The eastern goat genomes sampled here are not equally related to western and Levantine Neolithic goat, sharing more derived alleles with western goat (Figure 4.8). This would be expected if the Levantine Neolithic goat have a greater degree of ancestry from a “wild” source distinct from the “domestic” ancestry shared among the populations. Both admixture graphs (Figure 4.22) and directly-calculated f_4 ratios (Table 4.2) support a larger “wild” content in the genomes of Levantine goat populations compared to western Neolithics, although these estimations come with large errors due to the low coverage of the ‘Ain Ghazal samples.

Wild gene flow into Neolithic goat populations

Though the data are supportive of a domestic goat monophyly, there is evidence of asymmetric relationships between certain goat and ancient bezoar populations. Specifically, Neolithic goat from western and Levantine regions show greater derived allele sharing with wild goat from the Epipaleolithic site of Direkli Cave, central Turkey, than goat from eastern regions do. This is evident from significant *D* statistics in the form of *D*(Neolithic East, Neolithic West/Levant, Direkli1-2, Yak) (Figure 4.9, Appendix Figure 4.6). Manually constructed admixture graphs fit the data when admixture from a Direkli-like population into the ancestors of western/Levantine Neolithic goat is incorporated (Figure 4.22, Appendix Text 4.1). The presence of Direkli-like ancestry is not restricted to ancient western genomes, and is detected in modern populations also. The automated graph building approach Treemix also models admixture from pre-Neolithic Turkish bezoar (represented by Direkli1-2) into both the population ancestral to all European goat, and to Neolithic Serbian goat (Blagotin3).

The extent of the Direkli-like contributions into both western and Levantine Neolithic populations are 0.50 ± 0.05 (SE * 3) and 0.56 ± 0.20 respectively, based on f_4 statistics (Table 4.2), which are broadly consistent with the values estimated from qpGraph (Figure 4.22). This admixture is also present in ancient and modern European goat genomes, based on Treemix (Figure 4.16), qpGraph (Appendix Figures 4.1 and 4.2), and *D* statistics (Appendix

Table 4.3), although possibly to a lesser degree (see discussion below). Admixture with wild Turkish goat appears to have a profound effect on the genomes of modern European goat, accounting for 0.43 ± 0.04 of their ancestry (Table 4.2)

Neolithic goat populations in Iran and Turkmenistan may also have a history of admixture with specific bezoar populations. More complicated admixture graphs require a wild input into the ancestors of eastern Neolithic goat (Figure 4.22); admixture graphs modelling later eastern goat populations also require this additional wild contribution to Neolithic groups (Appendix Figure 4.2). Bezoar are known to have been found in eastern Iran and beyond in the early Holocene (Uerpmann 1987), presenting many opportunities for admixture or introgressive capture. An ideal way to test this would be pre-Neolithic bezoar genomes from western to eastern Iran; the statistic *D*(western Neolithic, eastern Neolithic, Pre-Neolithic Iranian Wild, Yak) should give positive values if such admixture did occur.

A model of the early stages of goat domestication

Given the evidence discussed above, the following model for the early domestication history of goat is proposed (Figure 4.23A). The Neolithic goat sampled here likely share an ancestral population, based on their monophyly, which may be traced to one of the earliest attempts at goat management. This proposed population likely occurred within the region where goat herd management (Young Male Kill-Off) is first detected, an area ranging from the Ganj Dareh in the Zagros Mountains to Beidha in southwest Jordan (Arbuckle & Atici 2013). The descendants of this managed population were introduced to other regions of southwest Asia, were they were supplemented by capture of local wild goat or unintended admixture. A localized domestication process for goat in the Levant has been suggested before (Horwitz et al. 1999), a model which has recently been modified to incorporate importation of herds (Munro et al. 2018). In Turkey, admixture with local bezoar resulted in goat populations with an excess of Direkli-like ancestry compared to those introduced into Iran. The same or a different admixture event may explain the Direkli-like ancestry in Levantine Neolithics; wild recruitment or unintended admixture may have also occurred in the Levant. Such localized admixture with the wild has previously reported in pigs from western Turkey, prior to their introduction to Europe (Ottoni et al. 2013). In eastern regions, admixture with local bezoar may have contributed to their differentiation from western herds. Serial founding events may have also increased their genetic distance, due to strong genetic drift as goat herds were introduced to eastern Iran, western Turkey and southeast Europe.

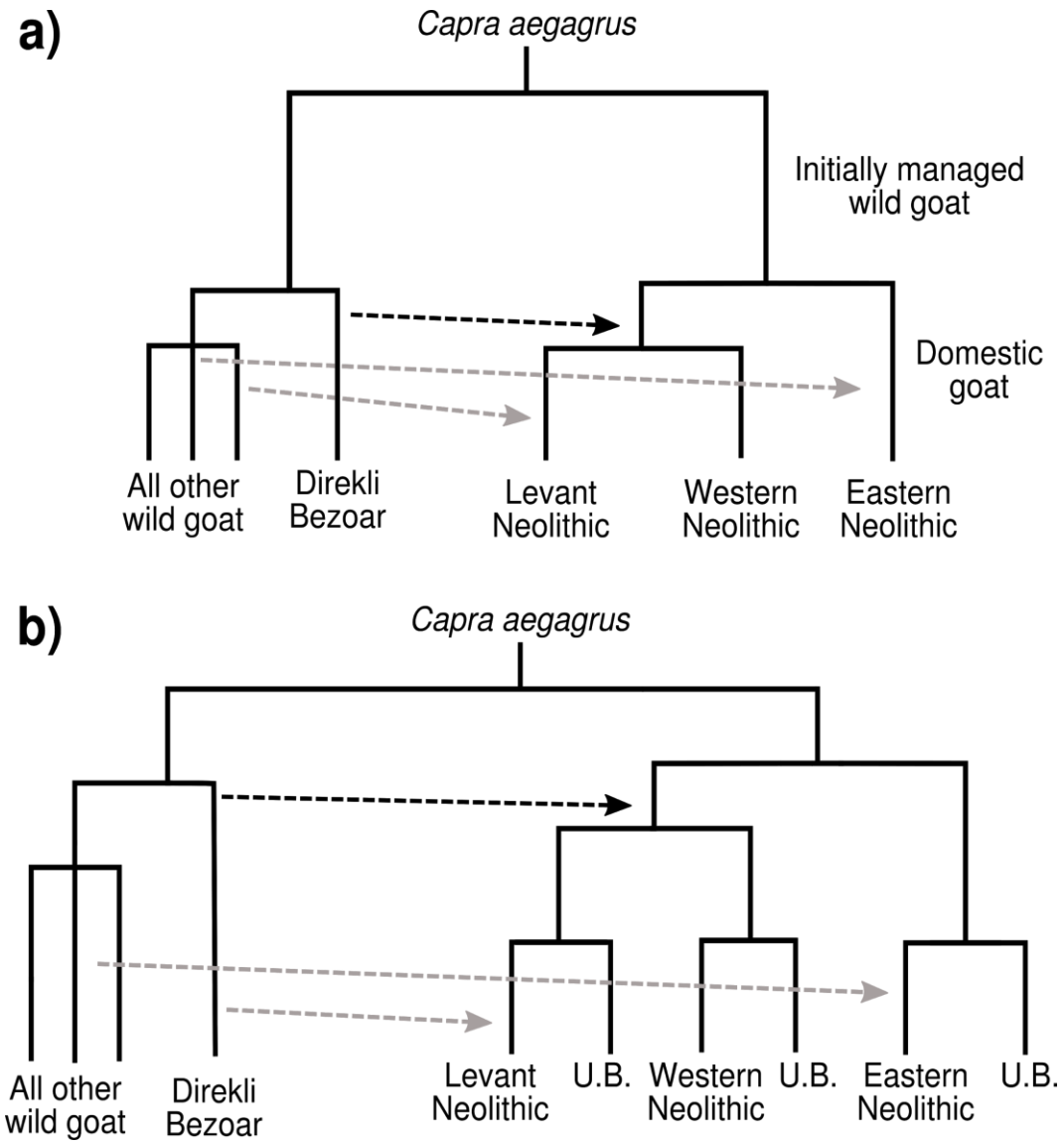


Figure 4.24 - Proposed models for the early stages of goat domestication. a) Simple model in which domestic goat form a monophyly b) More complex model in which multiple domestications of unsampled bezoar populations occurred, leading to paraphyly of domestic goat. Black arrows are directly detected admixture events. Grey arrows are inferred or alternative admixture events. U.B.=Unsampled bezoar.

This model is consistent with the mtDNA data presented in Chapter 3, but requires clarification as to the origin of the various haplogroups. The proposed initial goat population may have had some or all of the haplogroup lineages now observed in modern goat. Serial founding events may have resulted in the loss of haplogroups in certain populations and regions. Introgressive capture events may also have introduced certain haplogroups to the sampled Neolithic populations; this may have occurred instead of or concurrently with serial founding events.

It is possible that the bezoar populations used here are not sufficient; the appearance of the monophyly of goat may be a product of bezoar sampling bias. If one or several unsampled bezoar populations formed clades with certain domestic goat populations, making the *Capra hircus* clade a paraphyly, then multiple domestications of goat may have occurred (Figure 4.24B). This would also explain the early differentiation of Neolithic goat populations. The fact that the bezoar used here come from either side (central Turkey and east of the Zagros) of the proposed region of domestication is evidence against this proposal. It must be noted that the phylogeography of present-day bezoar populations may be substantially different to that of the early Holocene. Similarly, bezoar populations that existed in the past may not persist today. A thorough survey of present-day bezoar genomic variability throughout southwest Asia, preferably with pre-Neolithic bezoar samples, would allow the two proposed models to be differentiated.

Hovk1 reveals pre-glacial genetic structure in bezoar

The pre-domestic sample Hovk1 is not an outgroup to all Neolithic goat (Figure 4.9), or an outgroup to all genomes analyzed here (Figure 4.7). This sample is older than the limit for radiocarbon dating, and can only be ascribed a minimum age of 47,000 BP. From stratigraphic dates, its maximum age is 104,000 +/- 9800 BP (Pinhasi et al. 2011). This makes Hovk1 a representative of bezoar genetic diversity prior to the Last Glacial Maximum (LGM), during which ice sheets reached their maximum ~31,000-24,500 BC, which lasted to ~17,000-18,000 BC (Clark et al. 2009). Climate changed was likely associated with increased glaciation of mountains in southwestern Asia (Ebrahimi & Seif 2016), which may have promoted habitat disruption of bezoar. If wild goat were restricted to a smaller range of geographic regions, it might be expected that a pre-LGM wild individual would be equally related to post-LGM individuals. The asymmetry in genetic affinities instead suggests that

some of the differentiation between Neolithic goat is due to ancestries which diverged from each other prior to the LGM.

The genetic makeup of Bronze Age Turkish goat

The genetic history of two goat (Acem1 and Acem2) from the Bronze Age site of Acemhöyük, central Turkey, is unclear. Projection PCA suggests they share more affinity with Bronze Age and later goat from the Caucasus region and Iran than with Neolithic samples from western Turkey (Figure 4.10). The two samples fall close to Chalcolithic and Bronze Age samples from Israel. NGSadmix models both genomes as having ~0.55 of their ancestry explained by a "blue" component maximized in eastern Neolithic genomes (Figure 4.18). f_3 outgroup clustering placed the two genomes with eastern populations (Figure 4.21). f_4 ratios produce negative values when western Neolithics are used as the non-admixed reference (Appendix Table 4.9); when eastern Neolithic goat are used as the unadmixed group, western Neolithic ancestry in Bronze Age Turkey is estimated as 0.27 ± 0.04 (Tabel 4.2). In contrast, qpGraph models the Bronze Age Turkish goat as receiving 56% of their ancestry from a western Neolithic-like population (Appendix Figure 4.2), and *D* statistics show greater derived allele sharing with western Neolithic genomes than eastern (Figure 4.11). This later result may be due to reference bias, as the reference genome CHIR_1.0 is a Chinese goat (Dong et al. 2013), which shows high affinity with eastern Neolithic genomes (Figure 4.6).

Both Acem1 and Acem2 appear to be the product of admixture between populations related to western and eastern Neolithic goat. The ancestry estimation, f_4 ratios, and admixture graphs mentioned above model these samples as a mixture of Neolithic ancestries. Additionally, both test statistics *D*(Eastern Neolithic, Bronze Age Turkey, Western Neolithic, Yak) and *D*(Western Neolithic, Bronze Age Turkey, Eastern Neolithic, Yak) give positive significant results (Appendix Table 4.3 and Figure 4.10), supporting an admixed population history. Treemix, however, does not model the Acem2 as admixed (Figure 4.16). This may be due to the absence of a high coverage central Turkish Neolithic genome, that would be a better representative of Turkish goat than the Serbian genome Blagotin3.

Ancient humans from Turkey also appear to have undergone admixture between the Neolithic and Chalcolithic. Genomes from Chalcolithic and Bronze Age individuals show greater shared ancestry with Caucasus Hunter Gatherers and Iranian Chalcolithic individuals,

although this may be driven exclusively by admixture with a population high in Caucasus Hunter Gatherer ancestry (Lazaridis et al. 2016; de Barros Damgaard et al. 2018). The Kura-Araxes culture of the southern Caucasus region extended westwards in eastern Turkey (Kohl & Trifonov 2014) during the mid-to-late 4th millennium, and have been a vector for gene flow between different human and animal populations. Ancient human genomes from this and related cultures indicate genetic interaction between Turkish, Caucasus, and Iranian populations (Wang et al. 2018). The 4th millennium Uruk phenomena in Mesopotamia, ranging from southeastern Turkey to central Iran, may also have been involved (Philip 2002; Oates 2014). This period saw the beginning of urbanization in southwest Asia, and likely co-occurred with greater long range movement and trade of resources and animals, promoting genetic exchange. The preceding Halafian and 'Ubaid periods also show cultural links between eastern Turkey, northern Levant, and Iraq (Özdoğan 2014). As ancient genomes - human or goat - from these cultures are not available at the time of writing, it is unknown how Mesopotamian individuals of either species related to other ancient southwest Asian populations.

The genetic makeup of European goat

The small number of ancient European goat genomes available indicate relative stability of populations through time. Outgroup f_3 statistics show a high degree of shared drift between goat from Neolithic southeast Europe/western Turkey, Bronze Age Britain, and modern populations (Figures 4.19 and 4.21). A neighbour-joining tree (Figure 4.7) and Treemix analysis (Figure 4.16) show a high affinity between ancient and modern populations. NGSadmix models all ancient European goat as entirely a single red component (Figure 4.18). Although *D* statistics appear to suggest that British Bronze Age goat are admixed with an eastern Neolithic-like source (Figure 4.10), the aforementioned analyses do not support this, nor does qpGraph (Appendix Figure 4.2). The majority of approaches indicate that the genetic makeup of European goat was formed entirely during their introduction during the Neolithic in the late 6th millennium BC (Chapman 2014).

This apparent genetic continuity is exemplified by two samples from Britain and Ireland. The Bronze Age British sample Potterne1 shows a highest amount of shared drift with IOG, the feral Irish goat genome sequenced here; the converse is true for IOG (Appendix Table 4.7). IBS (Figure 4.7) and projection PCA (Figure 4.6) also indicate that the two genomes show high affinity with each other. Although inference is limited to those available modern

European genomes, this suggests a degree of genetic continuity between ancient (Bronze Age) and modern British and Irish goat populations, supporting modern and historic mitochondrial evidence of an insular goat population across the isles (Cassidy et al. 2017). Potterne1 also shows high shared drift with modern French goat and the western Neolithic population (Figure 4.21).

As discussed above, various analyses indicate that the ancestors of European goat admixed with a population related to the Direkli Cave bezoar. Intriguingly, there is evidence that different European goat populations have differing amounts of Direkli-like ancestry. When Treemix models three or more migrations events, there are two Direkli-like contributions to European goat: one to the ancestors of all ancient and modern genomes, and a second to Neolithic Serbia (Blagotin3). It is possible that residual ancient damage results in a false signal of greater affinity between Direkli1-2 and Blagotin3 than between Direkli1-2 and modern European genomes, despite the softclipping of ancient samples. However, ancient European goat also appear to have less Direkli ancestry than Neolithic Serbian. qpGraph models Potterne1, a Bronze Age goat from Britain, as having less wild ancestry than Neolithic Serbians (Appendix Figure 4.2), which are estimated to have an additional 13% of their ancestry from a Direkli-like source. f_4 ratios also suggest that Potterne1 has a small amount (0.46 ± 0.08) of Direkli-like ancestry (Table 4.2), but these values overlap with the confidence bounds for the western Neolithics individually and as a group (Appendix Table 4.9, Table 4.2). The *D* statistics in the form *D*(Blagotin3, Potterne1, Direkli1-2, Yak), gives a significant and negative value (Appendix Table 4.3), indicating that the Bronze Age British goat has fewer Direkli derived alleles than Blagotin goat.

A possible confounder of this particular *D* statistic test is if the ancestors of Potterne1 admixed with a source of ancestral alleles, increasing the number of observed BABA sites and resulting in Blagotin and Direkli sharing more derived alleles. In line with this, qpGraph models modern European goat as being admixed with an outgroup to wild and domestic goat, and does not model an additional wild input to Blagotin samples (Figure 4.22). The possibility of admixture between European ibex and domestic goat populations, which could cause this excess of ancestral alleles, will be addressed in Chapter 6.

Assuming that admixture with ibex is not the explanation for the deficit of Direkli-like derived alleles in Potterne1 compared to Blagotin goat, there are several other explanations. The first is that the Neolithic Serbian goat sampled here are not the direct ancestors of other

European goat; it may happen that this population is an outlier with regards to European Neolithic goat. The affinity of Potterne1 with Blagotin genomes on the projection PCA (Figure 4.6), the observation that European Neolithic goat form a sister clade to later European goat in the IBS-NJ tree (Figure 4.7), and the similar ancestry profiles computed by NGSadmix (Figure 4.18) are not evidence against this, as the unsampled and hypothetical "true" Neolithic ancestors of European goat may show greater affinity with later goat. One possibility is that the ancestors of goat that spread throughout Europe via the Mediterranean and Danubian routes were distinct. Under this scenario, Danubian-route goat were descended from a population with more Direkli-like ancestry than those spread via the Mediterranean route. This explanation would require post-Neolithic goat of Britain and Ireland to have more ancestry from the Mediterranean route, with the same expectation for modern French goat. It also assumes that goat herds would have mixed as the two cultural traditions interacted. This hypothesis is testable with goat genomes from Neolithic Mediterranean sites: if those genomes also have less Direkli-like ancestry than the Danubian Neolithic goat, then it suggests that Neolithic goat in Europe descended from at least two populations of related but distinct ancestries.

A third hypothesis is that the Blagotin goat do reflect Neolithic European goat as a whole, but there has been a turnover in goat populations between then and the Bronze Age. Human European populations are known to have undergone substantial turnover between the Neolithic and the Bronze Age (Allentoft et al. 2015; Haak et al. 2015). This demographic change is associated with the mass migration of people from the Eurasian Steppe, which stretches from China and Russia to Romania and Hungary. The Steppe is thought to have been a corridor facilitatory to population movement throughout the last 6,000 years (Damgaard et al. 2018). Whether this turnover in European human populations was in parallel with a turnover in livestock is unknown, and would have required migrants to transport herds as they spread to the western edges of Europe (Cassidy et al. 2016). A time series of livestock genomes (i.e. goat) in a specific geographic area would allow for testing of this hypothesis. If no turnover was observed at the time Steppe-related ancestry is known to appear in regional human populations, it would support a model of acquisition of local animal herds by migrating populations. A change in genetic affinities of herds at that time would alternatively support long-range movement of animal herds from the Steppe.

The changing affinities of Levantine goat

Goat from the Neolithic site of ‘Ain Ghazal, despite having divergent mitochondria not frequently observed in modern populations, show affinity with western Neolithic populations. Projection PCA (Figure 4.6), the IBS-NJ phylogeny (Figure 4.7), ancestry estimation (Figure 4.18), outgroup f_3 statistics (Figure 4.19 and 4.21), and admixture graphs (Figure 4.22) all indicate a close relationship between the two populations. Despite this, the Levantine Neolithic goat are distinct in many regards, falling in a unique position along principal components (Figure 4.6) and potentially containing more wild ancestry than western Neolithics, based on f_4 ratios (Table 4.2) and admixture graphs (Figure 4.22).

Based on zooarchaeological evidence the case for autochthonous development of goat management in the southern Levant has been made, most recently from changes in their abundance in the Early (~8,450 - 8,050 cal. BC) and Middle PPNB (~8,050–7,450 cal. BC) with a corresponding decrease in hunting intensity (Munro et al. 2018). This model includes the possibility of goat importation in the Middle PPNB, which overlaps with the contextual age of the ‘Ain Ghazal genomes (Table 4.1). The genomes presented here are in concordance with this model for local-extralocal origin of Neolithic goat in the southern Levant.

The turnover signal in mtDNA which appeared to occur in the Levant between the Neolithic and Chalcolithic (Chapter 3) is partially reproduced in their nuclear genomes. Projection PCA indicates that both Chalcolithic and Bronze Age Israeli goat are displaced towards eastern genomes relative to the Neolithic Jordan samples (Figure 4.6). Measures of shared drift with eastern Neolithic genomes increase substantially in the Chalcolithic and Bronze Age Levant genomes compared to the Neolithic (Figure 4.19). *D* statistics in the form *D*(Ainghazal2, Test, Semnan3, Yak), give significant and positive results for two Bronze Age Israeli genomes, Safi2 and Yoqneam2 (Figure 4.15). The remaining samples tend to give positive but non-significant results. The majority of these Levantine samples have poor endogenous content and consequently low coverage (Table 4.1), limiting the power of analyses involving them. One of the two genomes, Yoqneam2, has relatively high mean coverage at 2.2×, lending some support to the similar results of lower coverage samples. For example, the two Bronze Age genomes and an additional Chalcolithic genome (Shiqmim1) all show mixed ancestry profiles, compared to the entire “red” component of Neolithic genomes (Figure 4.18). Despite the preponderance of haplogroup A in these post-Neolithic genomes (Chapter

3), there is undoubtedly an increase in eastern Neolithic related ancestry in their nuclear genomes.

There are striking parallels between human and goat ancient genomes from the Levant at this time. Bronze Age Canaanite from the Levant (*c.* 1,700 BC) show an increased affinity to Neolithic and Chalcolithic Iranian people relative to Neolithic Levantine farmers (Haber et al. 2017). Approximately 50% of Bronze Age Canaanite genomes can be attributed to Iranian Chalcolithic individuals, with timing of this admixture event to *c.* 3,000 ± 1,500 BC. Later published genomes from Chalcolithic Levant indicate that post-Neolithic genomes from the region also have ancestry related to the Neolithic Anatolians, and that the influx of Chalcolithic Iranian-like ancestry had occurred by the late Chalcolithic (4,500-3,900 cal BC) (Harney et al. 2018). If the shift in genetic affinities of the two species are correlated, then human admixture event likely occurred in earlier end of this admixture time range; the earliest Chalcolithic goat genomes here which show a shift in PCA space towards "eastern" populations (Figure 4.6) are attributed to the early Ghassulian culture *c.* 4,500-4,200 BC. The Halafian culture of Mesopotamia is known to have spread throughout the Levant in the Early Chalcolithic, *c.* 5,800-5,300 BC (Garfinkel 2014). In contrast, interactions between the fourth millennium Mesopotamian Uruk civilizations and the Levant are thought to be limited (Oates 2014). The Transcaucasian Kura-Araxes complex spanned from northern Georgia to the Jordan Valley, and from eastern Turkey to western and central Iran (Iserlis et al. 2010). The culture is thought to have developed in eastern Turkey and southern Caucasus region in the early half of the fourth millennium, and rapidly diffused along the Zagros Mountains, into northwest Iran, and also along Levantine coast, in a process beginning *c.* 3,000 BC (Kohl & Trifonov 2014). This dispersal into southern Levant is thought to have involved the movement of people (Greenberg et al. 2014). The co-occurrence of post-Neolithic Turkish, Caucasus, and Israeli samples in the same region of PCA space (Figure 4.6) is interesting to note in this context. It is possible interactions with these cultures contributed to the change in the genetic makeup of Levantine goat herds.

The relationship between Neolithic Levantine and African goat

The Neolithic Levantine goat sequenced here appear to share ancestry with modern African goat. Projection PCA (Figure 4.6) and f_3 outgroup statistics (Figure 4.19) indicate a degree of shared drift and relatedness with modern genomes from Togo and Morocco. Modern African goat are genetically related to western Neolithic and modern European populations based on

Treemix (Figure 4.16), IBS clustering (Figure 4.7), and admixture graph analyses (Figure 4.22). The affinity of Neolithic Levant goat with modern African populations may therefore be due to ancestry common to Neolithic Levant and Neolithic West. Plotting shared drift of populations with Neolithic Levant versus shared drift with Neolithic West demonstrates an excess of Levantine ancestry in both modern African populations relative to their Neolithic Western ancestry; however, these were not significant (Student's *t*-test, 16 degrees of freedom, alpha of 0.05). Evidence for a Neolithic Levant contribution to modern African goat is therefore mixed.

As discussed earlier in Chapter 1, goat were introduced to northern Africa 5,000 - 4,000 BC, likely via the Levant (Mason 1986; Newman 1995). This would be approximately 2-3,000 years after the 'Ain Ghazal goat sequenced here lived, so it is unclear exactly how the earliest goat introduced to Africa would have related to the Neolithic or post-Neolithic goat populations of the Levant. Admixture graphs suggest a contribution to the ancestors of the African goat analyzed here, approximately 7% of their current genetic makeup (Figure 4.22). The civilizations of northern Africa were important trade centres for thousands of years, and there is genetic evidence that goat were among the goods exchanged (Pereira et al. 2009). The relative degree of influence of these interacting cultures are unclear (Zeder 2008), but qpGraph suggests that African goat are of diverse ancestries (Figure 4.22).

It is possible that the shared drift between Neolithic Levant goat and modern African goat is not due to a direct relationship, but via an intermediary population which contributed to the ancestry of both. One potential candidate is the Nubian ibex, which today inhabits both the Levant and northern Africa (Pidancier et al. 2006). This caprid likely had a similar range in the early Holocene (Uerpmann 1987), and so would have overlapped with both hunted and managed goat populations. Goat and Nubian ibex are known to hybridize in captivity (Groves & Grubb 2011), making this explanation biologically plausible. This hypothesis should be addressed with genomic data from Nubian ibex in Chapter 6.

Caucasus goat: on the border of East and West

Ancient goat genomes presented here from the Caucasus region (Georgia, Iranian Azerbaijan) show greater genetic affinity with Iranian Neolithic goat than those from western Turkey and southeast Europe. This is evident from Principal Components Analysis (Figure 4.6), the IBS-NJ phylogeny (Figure 4.7), measures of shared drift (Figures 4.19 and 4.21) and

derived allele sharing (Figure 4.12), and ancestry profiles (Figure 4.18). This Iranian affinity is intriguing given the reports of haplogroup A in Neolithic goat remains from west Azerbaijan (Kadowaki et al. 2017). If the ancestry detected in western genomes is correlated with haplogroup A in the Neolithic, then it implies there has been substantial genetic turnover in Caucasus goat populations. Assuming this correlation - which can be tested with Neolithic genome data from central and eastern Turkey, and the Caucasus region itself - the turnover would have occurred between the ages of the Neolithic Azerbaijan goat (6,000-5,000 BC) and the earliest Caucasus genome in this dataset: Azer6, from Soha Chay Tepe in Iranian Azerbaijan, dating to *c.* 4,200 BC.

The early Chalcolithic Shulaveri-Shomu complex like parallels with northern Mesopotamian and Iranian cultures (Kohl & Trifonov 2014). Late Chalcolithic Caucasus cultures also appear to have links with the late 'Ubaid (Early Chalcolithic) and early Uruk (Late Chalcolithic) cultures of Mesopotamia. Chalcolithic human genomes from the northern side of the Greater Caucasus show mixed ancestries of Caucasus Hunter Gatherers/Iranian Neolithic farmers, and Anatolian farmers (Wang et al. 2018), supporting regional connections in southwest Asia at this time.

Later goat genomes from the Caucasus region show similar but slightly altered genetic backgrounds. A cluster of Bronze Age (Kohneh2, Tac3) and Medieval (Geor2, Kazbeg1) genomes appear in PCA analyses (Figure 4.6), shifted towards western and Turkish samples. This cluster is apart from the Chalcolithic Azer6 genome and two other Iranian Azerbaijan samples (the Bronze Age Azer3-5 and Iron Age Azer4), which show greater affinity with post-Neolithic genomes from Iran. This distinction is also observed in the test statistic *D*(Blagotin3, Semnan3, Test, Yak), which indicate that greater allele sharing between later Caucasus genomes and the western Neolithic Blagotin3 (Figure 4.11), supported by explicit tests of western admixture (Figure 4.10). f_3 outgroup measures of shared drift give concordant signals of increased western ancestry (Figure 4.19), and admixture graphs best fit models of increased western Neolithic ancestry in later genomes (Appendix Figure 4.2). Human populations in the Caucasus region are known to have interacted with Steppe populations for millennia, and also show increased western/European farmer ancestry from the Bronze Age onwards (Lazaridis et al. 2016; Wang et al. 2018). The implied population flux presents many opportunities for the introduction of animal herds of differing genetic makeup. Additionally, long-range trade networks such as the Silk Road (Frachetti et al. 2017) and successive

empires and dynasties may have also promoted movement of animal and goat populations and gradual changes in their ancestries.

The genetic makeup of goat from Iran and beyond

The contrast between the mtDNA and whole genomes results of Iran and other eastern regions is conspicuous and requires discussion. Mitochondrial haplogroup frequencies exhibit a discontinuity between the Neolithic and post-Neolithic periods (Chapter 3). Haplogroups common in Neolithic Iran and Turkmenistan were largely replaced in the Chalcolithic and Bronze Age by haplogroup A, first observed in Neolithic western Turkey and southeast Europe (Figure 3.4). The inference previously made was a goat population replacement event occurred, and was supported by ABC model testing which favoured a migratory scenario (Appendix Figure 3.1 and Appendix Table 3.9).

The nuclear genome data generated here does not as strongly suggest such extensive admixture, but is not wholly incompatible with it. The time series of goat genomes from Monjukli Depe, Turkmenistan, does indicate a change in ancestries of samples between the Neolithic (Monjukli8) and Chalcolithic/Eneolithic (Monjukli1, 2, 4, and 6) (Figure 4.6). These post-Neolithic samples show greater affinity with contemporaneous and later samples from Iran and Uzbekistan, suggested also by f_3 outgroup clustering (Figure 4.21). Greater shared drift with western Neolithic genomes (and a corresponding decrease in eastern Neolithic shared drift) is observed when f_3 outgroup are plotted (Figure 4.19), but the change is minor. Directly testing for increased western ancestry using the statistic *D*(Semnan3, Test, Blagotin3, Yak) give positive values indicating greater western allele sharing for most Test genomes, but significant results only for some (Figure 4.13). Ancestry profiles show western ancestries in just a subset of post-Neolithic genomes (Figure 4.18), and as do f_4 ratios (Appendix Table 4.9).

These tests may be limited in power due to the low mean coverage of many of the genomes. For example, the only genome from Bronze Age Tilla Bulak (Bulak1) with a positive f_4 ratio (0.20 ± 0.05 western Neolithic ancestry) is not estimated to have any of the red "western" component by NGSadmix (Figure 4.18), and gives a non-significant *D* statistic test for admixture (Figure 4.13) . In comparison Bulak2 and Bulak5 are modelled as 0% and ~20% "western" respectively but do not give positive f_4 ratios (Appendix Table 4.9); both produce significant signals of increased western Neolithic derived alleles with *D* statistics (Figure

4.13). The extent of western ancestry in nuclear genomes of post-Neolithic "eastern" goat may be low such that the low-coverage genomes generated here do not provide sufficient power to detect that ancestry.

The proposed population model is supported by admixture graph analyses (Appendix Figures 4.1 and 4.2) but accommodates the difference between the nuclear and mtDNA genomes results. Simulations indicate that mtDNA gene flow from local to migrant populations should actually be greater than autosomal loci (Currat et al. 2008), making the disparity even more unusual. A possible explanation is if gene flow was heavily female-biased; selective castration of incoming bucks may have facilitated local male continuity. In this case, "western" mtDNA would be more likely to introgress than a random locus. The extent of the difference in nuclear and mitochondrial genome turnover makes this explanation unlikely, particularly as domestic goat herds tend to be disproportionately female (Davis 1987); maternally-inherited loci should admix at similar rates as nuclear loci. This can be addressed directly with ancient Y chromosome haplogroup distributions. Y chromosome studies in goats have been limited to regional analyses, but appear to be highly structured (Waki et al. 2015; Cinar Kul et al. 2015; Vidal et al. 2017). If Y chromosome haplotype frequencies were relatively stable from the Neolithic to Chalcolithic and Bronze Age, it would support a female-biased process. Alternatively, male-mediated influx from the wild may have "topped-up" eastern ancestry in domestic herds, explaining how western mtDNA may be present at a high frequency despite distinctly "eastern" nuclear genomes.

Limitations

A major limitation of the nuclear analyses here is the sampling bias due to the heterogenous levels of DNA preservation between archaeological sites. Certain sites produced a relatively large number of samples with excellent endogenous content, which could then be sequenced to a high ×-fold nuclear coverage. In particular the Neolithic sites of Blagotin-Poljna (Serbia) and Sang-e Chakmaq (eastern Iran) each yielded several well preserved samples (Table 4.1), which were sequenced extensively. Inferences made primarily using these sites raises the possibility that they are not fully representative of their respective regions. Additionally, kinship analyses indicated that genomes from these sites shared a degree of ancestry (Appendix Table 4.6). The absence of other high-quality samples, from the region and time period but a different archaeological site, makes this issue unavoidable. Although it is reassuring in both instances that other sites in the region give similar PCA affinities (*e.g.*

Kov57 and Monjukli8, Figure 4.6), implying that these results are not driven entirely by single sites.

A related issue is the lack of genomes from particular areas (*e.g.* Neolithic central and eastern Turkey, central Iran, the Caucasus), which leads to extrapolation from available data. For instance, the PCA positions of Neolithic samples from western Turkey and southeast Serbia suggest a bottleneck effect into Europe; a similar pattern is seen between Neolithic goat from the Zagros Mountains and eastern Iran (Figure 4.6). Genomic data from the intervening regions would allow this potential loss of genetic diversity to be directly quantified and described, rather than suggested.

The paucity of modern goat genomes from a broad geographic range also limited analyses. The absence of domestic genomes from Turkey and the Levant hinders any attempts at tracing regional Neolithic ancestry to the present-day. The limited number of bezoar genomes is also unfortunate; inferences relied on modern Iranian bezoar and ancient bezoar from Turkey. Although ancient bezoar genomes would be a preferable resource for the study of domestication and wild admixture, modern samples are much more accessible and cheaper to sequence. A dataset of genomes from across the modern *Capra aegagrus* distribution would also allow the ancient bezoar genomes generated here to be placed within a wider context.

An additional limitation is the reference genome, CHIR_1.0. This reference was constructed using genomic reads from a single individual from China (Dong et al. 2013). When this individual is aligned to itself, it appears to be genetically closer to eastern Neolithic genomes (Figure 4.6). As an aim of this chapter was to test the relative affinities between eastern and other Neolithic populations, this creates a potential source of bias. Reference allele bias is a well-documented phenomenon, in which there is a bias towards alleles which are similar to the reference genome (Brandt et al. 2015). In the context of the analyses here, this may have increased the number of "eastern" derived alleles reported in *D* statistics, for example. Residual ancient damage may also increase the difference between ancient reads and the reference sequence, exacerbating the problem. Additionally, as CHIR_1.0 is female, the reference has no Y chromosome assembly.

The availability of an ideal outgroup was a further limitation to these analyses. Although the outgroup used here, *Bos grunniens* (yak), is indeed an outgroup to all *Capra* species, it is estimated to have diverged from the ancestors of goat between 18.3-28.5 mya (Benton &

Donoghue 2007). This long divergence time increases the possibility of multiple mutations occurring at the same site across lineages, leading to an increase in false negative and positive statistical tests. The net effect is reduced sensitivity of both *D* and *f* statistics as the divergence time of the outgroup increases (Zheng & Janke 2018). In f_3 outgroup analyses, bezoar ibex from Qazvin, Iran, was chosen based on *D* statistics between eastern and western Neolithic groups. The use of *Ovis aries* (the domestic sheep) to define ancestral alleles may have been more appropriate, in spite of the possibility of goat-sheep hybridization (Pauciullo et al. 2016). Within the *Caprini* tribe, the *Ovis* and *Capra* genera diverged between 3.2-5.7 mya (Jiang et al. 2014). Ideally a member of the *Capra* genus that had not admixed with ancestors of any *Capra hircus* population would have been used, but these are currently not available. This will be addressed in Chapter 6.

Conclusion

Whole genome data from Neolithic goat obtained from throughout southwest Asia and adjacent regions concur with mtDNA results, showing early goat herds to be highly structured. Neolithic populations from the Levant, southeast Europe and western Turkey, and both western and eastern Iran are highly distinct, with Levantine and western goat being the closest-related pair. Despite this, the available wild goat genomes, ancient and modern, indicate that all Neolithic goat form a clade to the exclusion of bezoar. In the absence of other evidence, this suggests that goat were primarily domesticated once, and rapidly differentiated to create localized patterns of diversity.

This early differentiation was likely intensified by restocking of introduced managed herds from the wild. The influx of wild local genetic diversity would promote the divergence observed among these populations. This is explicitly detected in the case of pre-domestic Turkish goat from Direkli Cave in the Taurus Mountains; a related population likely admixed with the ancestors of the ancient Levantine and western goat sequenced here. The genetic legacy of this event persists today, with modern European genomes showing greater Direkli-like ancestry than Neolithic goat. Fitting of admixture graphs also suggest that this occurred in the ancestor of eastern goat genomes, but this remains to be directly tested.

Goat genomes from time periods following the Neolithic show reduced genetic structure. These nuclear genomes suggest that the apparent mtDNA gene pool turnover does not completely represent the demographic event inferred by mitochondrial sequences alone; this disparity may be a consequence of sex-biased admixture in favour of females. Domestic goat from Bronze Age central Turkey have a substantial increase in ancestry related to eastern Neolithic goat, compared to Neolithic goat from western Turkey. Several post-Neolithic genomes from Iran, Turkmenistan, and Uzbekistan also show an increase in western Neolithic-like ancestry, and mixing of the two ancestral strands is clearly observed in the Caucasus region. Genetic turnover is most striking in the Levant, where Neolithic goat of unique ancestry were succeeded by samples showing high eastern affinity.

The overall impression is that the goat domestication process was highly regionalized and dispersed, influenced by the genetic architecture of local goat populations throughout southwest Asia, the consequences of which can be still observed.

Chapter 4 References

Alberto, F.J. et al., 2018. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), pp.1655–1664.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Arbuckle, B.S. & Atici, L., 2013. Initial diversity in sheep and goat management in Neolithic south-western Asia. *Levantina*, 45(2), pp.219–235.

Barker et al., 2001. Genetic variation within and relationships among populations of Asian goats (Capra hircus). *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 118(4), pp.213–233.

de Barros Damgaard, P. et al., 2018. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science*, p.eaar7711.

Beaumont, M.A., 2008. Joint determination of topology, divergence time, and immigration in population trees. In Matsumura, S., Forster, P. and Renfrew, C, ed. *Simulation, genetics and human prehistory*. Cambridge: McDonald Institute, pp. 134–154.

Beaumont, M.A., Zhang, W. & Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics*, 162(4), pp.2025–2035.

Benjelloun, B. et al., 2015. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Frontiers in genetics*, 6, p.107.

Benton, M.J. & Donoghue, P.C.J., 2007. Paleontological evidence to date the tree of life. *Molecular biology and evolution*, 24(1), pp.26–53.

Brandt, D.Y.C. et al., 2015. Mapping Bias Overestimates Reference Allele Frequencies at the HLA Genes in the 1000 Genomes Project Phase I Data. *G3* , 5(5), pp.931–941.

Brito, L.F. et al., 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC genomics*, 18(1), p.229.

Burren, A. et al., 2016. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Animal genetics*, 47(6), pp.727–739.

Cañón, J. et al., 2006. Geographical partitioning of goat diversity in Europe and the Middle East. *Animal genetics*, 37(4), pp.327–334.

Cappuccio, I. et al., 2006. Allele frequencies and diversity parameters of 27 single nucleotide polymorphisms within and across goat breeds: TECHNICAL NOTE.

Molecular ecology notes, 6(4), pp.992–997.

Cassidy, L.M. et al., 2017. Capturing goats: documenting two hundred years of mitochondrial DNA diversity among goat populations from Britain and Ireland. *Biology letters*, 13(3). Available at: http://dx.doi.org/10.1098/rsbl.2016.0876.

Cassidy, L.M. et al., 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences of the United States of America*, 113(2), pp.368–373.

Chapman, J., 2014. Early Food Production in Southeastern Europe. In C. Renfrew & P. Bahn, eds. *The Cambridge World Prehistory*. Cambridge: Cambridge University Press, pp. 1803–1817.

Chessa, S. et al., 2003. Short Communication: Simultaneous Identification of Five κ-Casein (CSN3) Alleles in Domestic Goat by Polymerase Chain Reaction-Single Strand Conformation Polymorphism. *Journal of dairy science*, 86(11), pp.3726–3729.

Cinar Kul, B. et al., 2015. Y-chromosomal variation of local goat breeds of Turkey close to the domestication centre. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 132(6), pp.449–453.

Clark, P.U. et al., 2009. The Last Glacial Maximum. *Science*, 325(5941), pp.710–714.

Criscione, A. et al., 2007. Characterization of biodiversity in six goat breeds reared in Southern Italy by means of microsatellite and SNP markers. *Italian journal of animal science*, 6(sup1), pp.95–97.

Csilléry, K., François, O. & Blum, M., 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in ecology and evolution / British Ecological Society*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00179.x/full.

Currat, M. et al., 2008. The hidden side of invasions: massive introgression by local genes. *Evolution; international journal of organic evolution*, 62(8), pp.1908–1920.

Damgaard, P. de B. et al., 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature*. Available at: https://doi.org/10.1038/s41586-018-0094-2.

Davis, S.J.M., 1987. *The archaeology of animals*, London: Batsford.

Don E. Wilson, R.A.M. ed., 2011. *Family Bovidae (Hollow-horned Ruminants)*, Lynx Edicions.

Dong, Y. et al., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nature biotechnology*, 31(2), pp.135–141.

Ebrahimi, B. & Seif, A., 2016. Equilibrium-Line Altitudes of Late Quaternary Glaciers in the Zardkuh Mountain, Iran. *Geopersia*, 6(2), pp.299–322.

Excoffier, L. et al., 2013. Robust demographic inference from genomic and SNP data. *PLoS genetics*, 9(10), p.e1003905.

Felsenstein, J., 1989. PHYLIP: Phylogeny Inference Package. Version 3.2. *Cladistics*,

(5), pp.164–166.

Fernández, H. et al., 2006. Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Sciences of the United States of America*, 103, pp.15375–15379.

Frachetti, M.D. et al., 2017. Nomadic ecology shaped the highland geography of Asia's Silk Roads. *Nature*, 543(7644), pp.193–198.

Ganai, N.A. & Yadav, B.R., 2001. Genetic variation within and among three Indian breeds of goat using heterologous microsatellite markers. *Animal biotechnology*, 12(2), pp.121–136.

Garfinkel, Y., 2014. The Levant Pottery Neolithic and Chalcolithic Periods. In C. Renfrew & P. Bahn, eds. *The Cambridge World Prehistory*. Cambridge: Cambridge University Press, pp. 1439–1461.

Gerbault, P., Powell, A. & Thomas, M.G., 2012. Evaluating demographic models for goat domestication using mtDNA sequences. *Anthropozoologica*, 47(2), pp.64–76.

Greenberg, R., Shimelmitz, R. & Iserlis, M., 2014. New evidence for the Anatolian origins of "Khirbet Kerak Ware people" at Tel Bet Yerah (Israel), ca 2800 BC. *Paléorient*, 40(2), pp.183–201.

Green, R.E. et al., 2010. A draft sequence of the Neandertal genome. *Science*, 328(5979), pp.710–722.

Groves, C. & Grubb, P., 2011. *Ungulate Taxonomy*, Baltimore: JHU Press.

Haak, W. et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555), pp.207–211.

Haber, M. et al., 2017. Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *American journal of human genetics*, 101(2), pp.274–282.

Harney, É. et al., 2018. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nature communications*, 9(1), p.3336.

Hellenthal, G. & Stephens, M., 2007. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics*, 23(4), pp.520–521.

Horwitz, L.K. et al., 1999. Animal Domestication in the southern Levant. *Paléorient*, 25(2), pp.63–80.

Hudson, R.R., 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, 18(2), pp.337–338.

Hudson, R.R., Slatkin, M. & Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132(2), pp.583–589.

Iserlis, M. et al., 2010. Bet Yerah, Aparan III and Karnut I: Preliminary observations on Kura-Araxes homeland and diaspora ceramic technologies. *TUBA-AR*, 13(13), pp.245–262.

Jiang, Y. et al., 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, 344(6188), pp.1168–1173.

Kadowaki, S. et al., 2017. Mitochondrial DNA Analysis of Ancient Domestic Goats in the Southern Caucasus: A Preliminary Result from Neolithic Settlements at Göytepe and Hacı Elamxanlı Tepe: Ancient DNA Analysis of Domestic Goats in the Southern Caucasus. *International Journal of Osteoarchaeology*, 27(2), pp.245–260.

Kijas, J.W. et al., 2013. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal genetics*, 44(3), pp.325–335.

Kohl, P.L. & Trifonov, V., 2014. The Prehistory of the Caucasus: Internal Developments and External Interactions. In C. Renfrew & P. Bahn, eds. *The Cambridge World Prehistory*. Cambridge: Cambridge University Press, pp. 1571–1595.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Larson, G. & Burger, J., 2013. A population genetics view of animal domestication. *Trends in genetics: TIG*, 29(4), pp.197–205.

Lashmar, S.F., Visser, C. & van Marle-Köster, E., 2016. SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. *Small ruminant research: the journal of the International Goat Association*, 136, pp.65–71.

Lawson, D.J., van Dorp, L. & Falush, D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature communications*, 9(1), p.3258.

Lazaridis, I. et al., 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617), pp.419–424.

Lenstra, J.A. et al., 2017. Microsatellite diversity of the Nordic type of goats in relation to breed conservation: how relevant is pure ancestry? *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 134(1), pp.78–84.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]*. Available at: http://arxiv.org/abs/1303.3997.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), pp.2987–2993.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), pp.493–496.

Lipatov, M. et al., 2015. Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data. *bioRxiv*, p.023374. Available at: http://dx.doi.org/10.1101/023374.

Lipson, M. et al., 2013. Efficient moment-based inference of admixture parameters and sources of gene flow. *Molecular biology and evolution*, 30(8), pp.1788–1802.

Mason, I.L., 1986. *Evolution of Domesticated Animals*, John Wiley & Sons, Incorporated.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Mdladla, K. et al., 2016. Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Animal genetics*, 47(4), pp.471–482.

Meisner, J. & Albrechtsen, A., 2018. Inferring Population Structure and Admixture Proportions in Low Depth Next-Generation Sequencing Data. *bioRxiv*, p.302463. Available at: https://www.biorxiv.org/content/early/2018/04/17/302463.

Menozzi, P., Piazza, A. & Cavalli-Sforza, L., 1978. Synthetic maps of human gene frequencies in Europeans. *Science*, 201(4358), pp.786–792.

Mine, O.M. et al., 2000. Sheep-goat hybrid born under natural conditions. *Small ruminant research: the journal of the International Goat Association*, 37(1-2), pp.141–145.

Munro, N.D. et al., 2018. The Emergence of Animal Management in the Southern Levant. *Scientific reports*, 8(1), p.9279.

Naderi, S. et al., 2007. Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PloS one*, 2, p.e1012.

Nafti, M., Khaldi, Z. & Haddad, B., 2016. Genetic relationships and structure among goat populations from southern Tunisia assessed using microsatellites. *Journal of New Sciences*, 27.

Newman, J.L., 1995. *The Peopling of Africa: A Geographic Interpretation*, Yale University Press.

Nielsen, R. et al., 2011. Genotype and SNP calling from next-generation sequencing data. *Nature reviews. Genetics*, 12(6), pp.443–451.

Novembre, J. & Stephens, M., 2008. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics*, 40(5), pp.646–649.

Oates, J., 2014. The Rise of Cities in Mesopotamia and Iran. In C. Renfrew & P. Bahn, eds. *The Cambridge World Prehistory*. Cambridge: Cambridge University Press, pp. 1462–1497.

Ottoni, C. et al., 2013. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Molecular biology and evolution*, 30(4), pp.824–832.

Özdoğan, M., 2014. Anatolia: From the Pre-Pottery Neolithic to the End of the Early Bronze Age (10,500 - 2000 BC). In C. Renfrew & P. Bahn, eds. *The Cambridge World Prehistory*. Cambridge: Cambridge University Press, pp. 1508–1544.

Paradis, E., Claude, J. & Strimmer, K., 2004. APE: Analyses of Phylogenetics and

Evolution in R language. *Bioinformatics* , 20(2), pp.289–290.

Pariset, L. et al., 2006. Assessment of population structure by single nucleotide polymorphisms (SNPs) in goat breeds. In *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. pp. 117–120.

Pariset, L. et al., 2009. Geographical patterning of sixteen goat breeds from Italy, Albania and Greece assessed by Single Nucleotide Polymorphisms. *BMC ecology*, 9, p.20.

Park, S.D.E. et al., 2015. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome biology*, 16, p.234.

Patterson, N. et al., 2012. Ancient admixture in human history. *Genetics*, 192(3), pp.1065–1093.

Patterson, N., Price, A.L. & Reich, D., 2006. Population structure and eigenanalysis. *PLoS genetics*, 2(12), p.e190.

Pauciullo, A. et al., 2016. Characterization of a very rare case of living ewe-buck hybrid using classical and molecular cytogenetics. *Scientific reports*, 6, p.34781.

Pereira, F. et al., 2009. Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Molecular biology and evolution*, 26(12), pp.2765–2773.

Peter, B.M., 2016. Admixture, Population Structure, and F-Statistics. *Genetics*, 202(4), pp.1485–1501.

Philip, G., 2002. Contacts between the 'Uruk world and the Levant during the fourth millennium BC: evidence and interpretation. In J. N. Postgate, ed. *Artefacts of Complexity: Tracing the Uruk in the Near East*. Iraq Archaeological Reports. Cambridge: British School of Archaeology in Iraq, pp. 207–235.

Pickrell, J.K. & Pritchard, J.K., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics*, 8(11), p.e1002967.

Pidancier, N. et al., 2006. Evolutionary history of the genus Capra (Mammalia, Artiodactyla): discordance between mitochondrial DNA and Y-chromosome phylogenies. *Molecular phylogenetics and evolution*, 40(3), pp.739–749.

Pinhasi, R. et al., 2011. Middle Palaeolithic human occupation of the high altitude region of Hovk-1, Armenia. *Quaternary science reviews*, 30(27), pp.3846–3857.

Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–959.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

Rambaut, A., 2009. FigTree. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

R Core Team, 2016. *R: A Language and Environment for Statistical Computing*,

Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Reich, D. et al., 2009. Reconstructing Indian population history. *Nature*, 461(7263), pp.489–494.

Saitbekova, N. et al., 1999. Genetic diversity in Swiss goat breeds based on microsatellite analysis. *Animal genetics*, 30(1), pp.36–41.

Schiffels, S. & Durbin, R., 2014. Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, 46(8), pp.919–925.

Skoglund, P. et al., 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current biology: CB*, 25(11), pp.1515–1519.

Skoglund, P., 2015. *POPSTATS*, Github. Available at: https://github.com/pontussk/popstats.

Skotte, L., Korneliussen, T.S. & Albrechtsen, A., 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), pp.693–702.

Smit, A.F.A., Hubley, R. & Green, P., 2013-2015. RepeatMasker Open-4.0. *RepeatMasker Web Server*. Available at: http://www.repeatmasker.org.

Soraggi, S., Wiuf, C. & Albrechtsen, A., 2018. Powerful Inference with the D-Statistic on Low-Coverage Whole-Genome Data. *G3* , 8(2), pp.551–566.

The Broad Institute, 2018. *Picard Tools*, The Broad Institute. Available at: https://broadinstitute.github.io/picard/.

Tosser-Klopp, G. et al., 2014. Design and characterization of a 52K SNP chip for goats. *PloS one*, 9. Available at: http://dx.doi.org/10.1371/journal.pone.0086227.

Uerpmann, H.-P., 1987. *The ancient distribution of ungulate mammals in the Middle East: fauna and archaeological sites in Southwest Asia and Northeast Africa*, Wiesbaden: L. Reichert Verlag.

Vidal, O. et al., 2017. Differential distribution of Y-chromosome haplotypes in Swiss and Southern European goat breeds. *Scientific reports*, 7(1), p.16161.

Visser, C. et al., 2016. Genetic Diversity and Population Structure in South African, French and Argentinian Angora Goats from Genome-Wide SNP Data. *PloS one*, 11(5), p.e0154353.

Waki, A. et al., 2015. Paternal phylogeography and genetic diversity of East Asian goats. *Animal genetics*, 46(3), pp.337–339.

Wang, C. et al., 2014. Ancestry estimation and control of population stratification for sequence-based association studies. *Nature genetics*, 46(4), pp.409–415.

Wang, C. et al., 2015. Improved Ancestry Estimation for both Genotyping and Sequencing Data using Projection Procrustes Analysis and Genotype Imputation. *American journal of human genetics*, 96(6), pp.926–937.

Wang, C.-C. et al., 2018. The genetic prehistory of the Greater Caucasus. *bioRxiv*, p.322347. Available at: https://www.biorxiv.org/content/early/2018/05/16/322347.

Wickham, H., 2009. *Ggplot2: Elegant Graphics for Data Analysis* 2nd ed., Springer Publishing Company, Incorporated.

Xiang-Long, L. & Valentini, A., 2004. Genetic diversity of Chinese indigenous goat breeds based on microsatellite markers. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie*, 121(5), pp.350–355.

Yang, L. et al., 1999. Determination of genetic relationships among five indigenous Chinese goat breeds with six microsatellite markers. *Animal genetics*, 30(6), pp.452–455.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105, pp.11597–11604.

Zheng, Y. & Janke, A., 2018. Gene flow analysis method, the D-statistic, is robust in a wide parameter space. *BMC bioinformatics*, 19(1), p.10.

Виктор, В., 2010. *Relief Map of Middle East*, Available at: https://commons.wikimedia.org/wiki/File:Relief_Map_of_Middle_East.jpg.

Chapter 5: Selection in Neolithic Goat

Introduction

Phenotypic diversity of goat in the past

The domestic goat shows remarkable phenotypic variation, for which there is early historical evidence (Porter & Tebbit 1996). This includes variation in coat colour, fibre type, hairlessness, horn shape, the polled (hornless) phenotype, dwarfism, facial profile, and ear morphology. This is in addition to the variation of physiological adaptations to specific climate conditions (*e.g.* heat tolerance, resistance to infection/parasites, feeding behaviour), and production traits (*e.g.* meat and milk yields) (Maijala 1991). This phenotypic diversity appears early in human records: on pre-dynastic Egyptian papyrus, goat are depicted as having black, white, and piebald coats (Zeuner 1963). The long lop-eared phenotype, now characteristic of many breeds, appeared in Egypt, north Sudan and Libya in the 3rd millennium BC; polled goat also appear in Egypt at this time (Porter et al. 2016).

Much attention has been paid to the appearance of corkscrew or twisted horn morphology (as opposed to the bezoar-like "scimitar" horn shape) in the archaeological record. Twisted or spiral-horned goat are depicted in art from Bronze Age (3rd millennium BC) Mesopotamia and Egypt, with twisted horn-cores found at a similar time (Zeuner 1963), from which point they predominated over scimitar horns. This is often described as a "separate lineage" of domestic goat, and attributed to domestication of or gene flow from either the markhor, *Capra falconeri*, or a hypothesized and putatively-extinct European wild goat, *Capra prisca*. However the evidence for *Capra prisca* is weak (see Chapter 1), and domestic goat horns usually twist in a homonymous fashion (both horns spiral in a clockwise fashion) (Mason 1986); bezoar occasionally show homonymous twisting, while markhor horns are heteronymous (right horn spirals clockwise, the left counter-clockwise).

Domestic phenotypes in goat

As discussed in Chapter 1, there is a set of phenotypes characteristic of many mammalian domesticates that are commonly referred to as domestication syndrome. These include reduced brain size, retention of juvenile traits, curled tails, floppy/lop ears, increased coat colour diversity and piebaldism, and reduction in aggressiveness (Zeder 2012). The

biological link shared by these domestication syndrome phenotypes has been suggested to be neural crest cells, which are involved in the development of the adrenal medulla, cartilaginous structures, large portions of the skull, teeth, and also pigmentation cells (Wilkins et al. 2014). Selection for altered development of neural crest cells (*i.e.* reduced size or activity of the adrenal medulla, and reduced aggressive behaviours) may have pleiotropic effects. This hypothesis is strengthened by the observation that domestication animals frequently have a reduced limbic system. This endocrine system, consisting of the pituitary gland, amygdala, hippocampus, and hypothalamus, is reduced in size by 30-40% in a variety of livestock relative to their wild progenitor (Kruska 1988). The limbic system is a major control point for the autonomic nervous system and hormones which regulate the "fight-or-flight" response. Early farmers may have unintentionally selected for animals with reduced responses to external stressors (such as contact with humans), affecting the development of both neural crest cells and the limbic system.

Domestic goat show many of these domestic syndrome features. *Capra hircus* has a reduced brain size compared to *Capra aegagrus* (Hemmer 1990). They show a large degree of coat colour variation, including piebaldism or the spotted phenotype (Porter & Tebbit 1996). Drooping ears are also a characteristic phenotype of certain goat breeds, such as Anglo-Nubian. Domestic goat display other phenotypic and behavioural changes compared to bezoar ibex. They have reduced sexual dimorphism compared to wild bezoar, particularly in the body and horn size of males (Zeder 2006). Goat even show sensitivity to human gaze in problem-solving tasks (Nawroth et al. 2016), a phenomenon well established in dogs and thought to be the product of selection for "human-like" communication behaviours (Miklósi et al. 2003), and may also be perceptive of human emotional states (Nawroth et al. 2018).

Selection signals in modern goat

These changes in domestic goat compared to the wild bezoar ibex have genetic underpinnings that have yet to be fully understood. As discussed in Chapter 1, these genetic changes may be the result of a relaxation of selection due to the animals living under the supervision and protection of human farmers, or positive selection for certain traits due to intentional or unintentional selection by those farmers. These are not mutually exclusive; some traits may have been affected by one phenomenon exclusively, while other traits may have been affected by both. For example, reduced selection for cryptic coat colouration may have resulted in greater pigmentation diversity, some of which were subsequently selected for by

farmers. The two processes are expected to have different effects on genomic diversity: relaxation of selection is expected to produce an excess of nonsynonymous mutations, while positive selection may result in a few novel variants rising in frequency in the domestic genome.

Research so far has focused on identifying signals of positive selection in domestic goat, rather than testing for an excess of nonsynonymous mutations. Using the bezoar reference genome, (Dong et al. 2015) detected CNV regions in domestic goat. This indicated that ASIP gene copy number was associated in variation of coat colour between breeds, and that several immune gene losses may have occurred in domestic goat relative to bezoar. ZH_p (Z-transformed pooled Heterozygosity statistic) and *di* (differentiated regions among groups) have been applied to pooled-sequencing of Chinese populations in order to identify likely selected genes in specific traits (Wang et al. 2016). This identified known or candidate genes involved in coat colour (*ASIP*, *KITLG*, *GNA11*, *OSTM1*), body size (*TBX15*, *DGCR8*, *CDC25A*, *RDH16*), high-altitude adaptation (*CDK2*, *NOXA1*, *SOC2*, *ENPEP*), and hair fibre (*LXH2*, *FGF9*, *WNT2*). *di* in tandem with F_{ST} has also been used to identify genes involved in altitude adaptation from pooled exome data of Tibetan cashmere goat (Song et al. 2016).

A different approach, the cross-population composite likelihood ratio test (XP-CLR) (Chen et al. 2010) has identified genomic regions likely selected, or underwent a change in allele frequency too fast to be explained by drift, in Moroccan goat (Benjelloun et al. 2015). However, this analysis was hampered by the quality of the CHIR_1.0 annotation, limiting GO enrichment analysis. An alternative method was applied in Alberto et al. 2018. The authors used hapFLK, a haplotype differentiation test that takes into account hierarchical structure within the test population by incorporation of a kinship matrix (Fariello et al. 2013), applied in a stratified False Discovery Rate (FDR) framework to detect regions under convergent selection in sheep and goat. They identified genes including *KITLG* and several associated with immunity, milk composition, meat, hair, and fertility, likely under selection in both sheep and goat, some of which are thought to have pleiotropic effects potentially important during early domestication.

Association studies have been performed on various goat phenotypes, using the GoatSNP50 BeadChip (Tosser-Klopp et al. 2014). A non-synonymous variant at the *TYRP1* locus - associated with black-to-brown colouration in several domestic species - has been linked to the unique coat colour phenotype of the Valais Copperneck (Becker et al. 2015). Two other

pigmentation traits, the “pink” and “pink neck” phenotypes undesirable in the Saanen breed, have been associated with two genomic regions, one of which contains the classic pigmentation gene *ASIP* (P. M. Martin et al. 2017). Throat wattles have been linked to a region in chromosome 10 which contains two genes (*FMN1* and *GREM1*) involved in limb development, although the causative allele was not identified (Reber et al. 2015). Exonic mutations in *DGAT1*, a gene established to affect milk fat content in cattle, have been shown to explain substantial variation in the trait in goat (P. Martin et al. 2017). Finally the hornless or polled phenotype has been shown to be controlled by a deletion in the polled intersex syndrome region of chromosome 1, which may have a common origin across breeds (Pailhoux et al. 2001; Kijas et al. 2013).

As described, the bulk of research in selection signatures within the domestic goat genome has been on recent production phenotypes or breed-specific phenotypes. Although some have attempted to address changes associated with the initial phases of domestication (Alberto et al. 2018), few have attempted to date their occurrence or rise in frequency. Despite the phenotypic diversity pictorially recorded in ancient times, comparatively little attention has been paid to their genetic origin. In addition, the nature of genetic differences which may have distinguished managed goat from wild bezoar has not been investigated.

Aims

The aims of this chapter are:

1. To identify regions which are most differentiated in Neolithic goat populations compared to modern bezoar ibex, which also show reduced variation in the Neolithic goat.
2. To use pigmentation as a case study, investigating if genes associated with early phenotypic variation in domestic goat show differentiation from bezoar.

Material and Methods

Material

Two sets of ancient Neolithic genomes were selected for this analysis: eastern Neolithic (Iran and Turkmenistan) and western Neolithic (Serbia). Samples with a mean genomic coverage of <2× were excluded, leaving five and four genomes in each respective group (Table 5.1).

Modern Iranian bezoar ibex genomes were initially screened based on PCA position. Samples from the Hamedan population (Appendix Table 4.1) were removed due to their proximity to Neolithic West (Appendix Figure 4.2). The remaining 16 bezoar samples were retained in the selection analysis as representatives of wild diversity.

Table 5.1 - Ancient Neolithic genomes included in selection analysis, with mean genomic coverage. Mean genomic ×-fold coverage (after mapping quality 30 filter) is presented.

Neolithic East	Genome Coverage	Neolithic West	Genome Coverage
Semnan1-2	6.85	Blagotin1	6.99
Semnan3	14.89	Blagotin2	4.02
Semnan7	3.28	Blagotin3	11.47
Semnan9	3.05	Blagotin16	3.51
Semnan13	2.54		
Monjukli8	2.57		

Methods

F_{ST} and θ Outlier Regions

The approach taken here was to identify regions of Neolithic genomes which were most differentiated with respect to wild goat, and also showed reduced genetic diversity in the Neolithic group. The outlier regions identified were also filtered so that only regions that also displayed low absolute diversity within Neolithic genomes were included, to identify regions which may have been subject to a selective sweep. The statistic used to quantify molecular diversity in each group was Watterson's estimator, θ_W, which expresses the number of segregating sites in the population (Ewens 1974; Watterson 1975).

This analysis aimed to identify signatures of positive selection due to domestication, as opposed to a reduction in selection as described above. The rationale for this approach was that if strong selection acted on the genomes of Neolithic goat due to their domestication history, the genetic consequences should be identifiable in regions most different from wild goat. Additionally, strong selection may have reduced diversity in these regions in the domestic population compared to the wild.

ANGSD (Korneliussen et al. 2014) was used to compute 50kb sliding windows of F_{ST} using 10kb steps. Yak was used to define the ancestral allele. For bezoar, the following settings were used to calculate the site frequency spectrum: -setMaxDepthInd 20 -HWE_pval 0.01 -minIndDepth 2 -minInd 2 -doMajorMinor 1 -C 50. For Neolithic goat, the following setting were used: -setMaxDepthInd 20 -minIndDepth 2 -minInd 2.

Watterson's θ was then calculated in the same sliding 50kbp windows and 10kbp steps, for each of the three populations using the same filters as above. For each window in Neolithic populations, genetic diversity was expressed in terms of the diversity observed in the same window in modern bezoar:

$$\log(\theta_{Bezoar}/\theta_{Neolithic}),$$

so that Neolithic windows which show less diversity than in the bezoar ibex will have a negative value. If a window had an observed theta of 0, it was replaced with a value of 0.000001 to avoid divisions by zero.

Outlier windows were then identified using the following criteria:

1. F_{ST} with bezoar in the 99.9th percentile .
2. $\log(\theta_{Bezoar}/\theta_{Neolithic})$ in the 5th percentile.
3. $\theta_{Neolithic}$ in the 5th percentile.

Outlier windows as defined above were iteratively combined with adjacent windows which had a F_{ST} score in the top 1% quantile, in order to construct outlier regions. Gene overlapping outlier regions were determined using the GenBank annotation of CHIR_1.0 . For regions with no overlapping genes, the nearest genes were identified.

Gene Ontology enrichment analysis of F_{ST} outliers

To determine if there was a enrichment in Gene Ontology categories in genes detected in outlier scan, the g:GOSt function of the g:Profiler web server was employed (Reimand et al. 2007; Auffray & De Meulder 2016). Enrichment was determined on genes identified in Neolithic East and Neolithic West separately. As the *Capra hircus* genome was not available, *Ovis aries* was instead used, using sheep homologs when different gene names were used between the species. Default settings were used.

Nonsynonymous variants of outlier genes

Following identification of outlier region, overlapping genes were examined for non-synonymous variants at a high frequency in a Neolithic population and low frequency in modern bezoar ibex. 3'UTR and non-genic variants were not considered due to the difficulty in assessing phenotypic importance. Samtools mpileup (Li et al. 2009) was used to call variants in the bezoar and the two Neolithic genomes with mean genomic >8× (Semnan3 and Blagotin3), restricting the calling to exons of the identified genes (Table 5.2). 2bp was added to either side of each exon to detect possible splice donor/acceptor mutations. Sites within 3bp of indels were removed. Sheep and yak outgroups were also called to polarize variants as ancestral or derived. Sites in which both outgroups shared fixed alleles were retained and the allele set as ancestral; sites that in either outgroup were heterozygous or were not in consensus were discarded. Sites were then filtered for homozygous status in either Neolithic genome, with a corresponding maximum frequency filter of 0.2 in bezoar. Synonymous and 3'UTR variants were removed. Allele frequencies of these sites in both Neolithic East and Neolithic West populations used in the selection analysis were then estimated using ANGSD under the following settings: -doMaf 1 -doMajorMinor 5 -GL 1 -trim 4.

The genetic structure of *KIT*

Following on from the identification of the *KIT* gene as an outlier in both Neolithic groups (see Results), the genetic structure of *KIT* among ancient samples was investigated. ANGSD was used to construct an IBS matrix as described in Chapter 4, using all ancient and modern genomes >1×. The analysis was restricted to the union of the outlier regions detected around *KIT* (Table 5.2). This matrix was then visualized as a heatmap using the heatmap.2 function of gplots (Warnes et al. 2016).

Pigmentation genes

To investigate if genes associated with pigmentation also showed differentiation in Neolithic genomes with respect to bezoar ibex, F_{ST} values for windows overlapping the following genes were extracted: *KIT*, *KITLG*, *MC1R*, *PMEL17*, *ASIP*, *TYRP1*, and *MITF*. These genes have all been previously associated with coat colouration in mammalian domesticates (Brooks & Bailey 2005; Seitz et al. 1999; Fontanesi et al. 2009; Brunberg et al. 2006; Zhang et al. 2017; Becker et al. 2015; Hauswirth et al. 2012). In the case of *KITLG*, the region upstream identified in the outlier analysis (see results) was used instead of the *KITLG* gene itself. The F_{ST} distribution of all genomic windows was then plotted, on to which either the highest F_{ST} value of windows overlapping each pigmentation gene or the mean F_{ST} of overlapping windows was plotted.

Results

Outlier regions

A total of 21 outlier regions were detected: 7 regions (78 overlapping 50kb windows) in the Neolithic West population and 14 regions (155 windows) in the Neolithic East. Outlier regions are displayed in Table 5.2, along with overlapping genes or the nearest genes when no gene was found to overlap the region. The highest F_{ST} windows in each region are displayed in Appendix Table 5.1.

Of the identified outlier regions, two pairs of regions were common/overlapping in both populations: one overlapping the *KIT* gene, and one non-genic region for which the closest gene is *KITLG*. *KIT* is a tyrosine kinase receptor robustly associated with pigmentation variation in many mammalian species. Such species include mouse (Chabot et al. 1988), humans (Fleischman et al. 1991), pig (Marklund et al. 1998), horses (Marklund et al. 1999), cat (Cooper et al. 2006), and dog (Wong et al. 2013), in which loss-of-function mutations result in a dominant white or roan pigmentation pattern. The *KIT* protein product is responsible for detecting the *KITLG* (KIT ligand) protein, playing an important role in melanoblast development, gametogenesis, and also gastrointestinal motility (Besmer et al. 1993; Rönnstrand 2004; Farrugia 2008). Loss-of-function mutations in *KIT* are thought to affect neural crest-derived melanocyte precursors, which are dependent on KIT-KITLG signaling for migration and survival (Wehrle-Haller & Weston 1997). As the only ligand of the KIT protein, it is unsurprising that mutations affecting *KITLG* have also been implicated in pigmentation phenotypes. These include roan coat colour in Belgian Blue and Shorthorn cattle (Seitz et al. 1999) and Pakistani goat (Talenti et al. 2017).

16 of the regions overlapped at least one annotated gene. Several of these genes have been linked to production traits in other livestock animals or mammals. In the Neolithic West group *SIRT1* is found in an outlier region, a transcription factor which regulates endocrine signaling and metabolism, and may be involved in growth traits in cattle (Li et al. 2013). The same outlier region contains the *HERC4* gene, an E3 ubiquitin ligase which has been shown to have a role in male fertility in mice (Rodriguez & Stewart 2007). This locus also contains *MYPN*, involved in skeletal muscle organization, which has been linked to carcass traits in pig and cattle (Jiao et al. 2010; Braglia et al. 2013). *KIRREL3*, a gene involved in synapse formation and located in a region associated with carcass, production, reproduction, and

Table 5.2 - Outlier regions detected in F_{ST} and θ scan. Regions detected in either Neolithic East or Neolithic West are shown, with overlapping genes or the closest genes to each region. Chrom. = Chromosome.

Population	Chrom.	Region Start	Region End	Genes in region (discontinued entries removed)	Nearest Gene (when none are present within region)
Neolithic West	1	133120000	133210000	*SRPRB, LOC102172205, LOC102172488*	-
	1	143980000	144100000	*PRMT2*	-
	5	18060000	18180000	None	*KITLG*, *DUSP6*
	6	68180000	68340000	*KIT*	-
	17	20730000	20850000	*LOC102170258/WBP11*	-
	28	22170000	22390000	*SIRT1, HERC4, MYPN*	-
	29	28660000	28750000	*KIRREL3*	-
Neolithic East	2	78830000	78940000	*STAT1, STAT4*	-
	2	127950000	128050000	*IL22RA1, MYOM3*	-
	3	51280000	51460000	*GBP6, LOC106501943*	-
	3	102850000	102950000	*MACF1, KIAA0754*	-
	4	73500000	73590000	None	*IGFBP3*
	4	92400000	92600000	*MKLN1*	-
	5	18020000	18180000	None	*KITLG, DUSP6*
	6	68220000	68400000	*KIT*	-
	6	86460000	86580000	*EPGN, EREG*	-
	8	38500000	38600000	*RCL1, AK3*	-
	8	38620000	38770000	*CDC37L1, SPATA6L, PPAPDC2*	-
	9	11560000	11690000	None	*RSPO3*
	10	38440000	38530000	None	*RPS29*
	26	14970000	15110000	*LOC102185708* (cytochrome P450 2C19), *LOC102185056* (cytochrome P450 2C9)	-

behavioural trait QTLs (Quantitative Trait Loci), is also found in an F_{ST} outlier region in Neolithic West (Saatchi et al. 2014). Finally, the protein product of *PMT2* has been shown to play a role in neuronal morphogenesis via methylation of arginine residues of Cobl, an actin nucleator protein responsible for dendritogenesis (Hou et al. 2018).

Genes identified in Neolithic East outlier regions have also been associated with production traits or those relevant to the domestic environment. A single outlier region on the caprine chromosome 2 contains both the *STAT1* and *STAT4* genes, two members of the signal transducer and activator of transcription (STAT) family. *STAT4* has a role in immune system regulation in humans (Jacobson et al. 1995), and the gene is implicated in autoimmune disease risk (Korman et al. 2008). *STAT1* is known to be involved in viral resistance in mice (Durbin et al. 1996), early embryonic survival in cattle (Khatib et al. 2009), and also milk protein and fat content in Holstein cows (Cobanoglu et al. 2006). Another outlier region detected in Neolithic East overlaps with the *EPGN* and *EREG* genes, which have been associated in cattle calving interval variation (Raven et al. 2016) and murine oocyte maturation (Riese & Cullum 2014) respectively. Finally, an outlier region on chromosome 26 overlaps with two Cytochrome P450 genes, 2C9 and 2C19, part of the CYP family responsible for toxin and drug metabolism (Singh et al. 2011). Together, the enzymes produced by these two genes are responsible for the oxidative metabolism of 15% of prescribed drugs (10% and 5%) (Pinto & Dolan 2011). Both also metabolize endogenous compounds; of note is the CYP450 2C19 substrate steroid progesterone (Isvoran et al. 2017), a critical hormone involved in fertility and pregnancy.

Gene Ontology enrichment analysis of F_{ST} outliers

Gene Ontology enrichment was performed using the g:GOSt function of the g:Profiler web server (Auffray & De Meulder 2016), using *Ovis aries* as the organism due to the unavailability of *Capra hircus*. Enrichment results are displayed in Table 5.3.

Both Neolithic East and West showed enrichment for regulation of various kinase activity, for pigmentation (either via the "regulation of developmental pigmentation" biological process term or the "White Forelock" human phenotype term), and ectopic germ cell programmed cell death. In these three cases, *KIT* and *KITLG* were associated with the enriched terms. In addition, outlier regions detected in Neolithic West were significantly enriched for "Positive regulation of smooth muscle cell differentiation" and "Programmed

cell death involved in cell development". Neolithic East outlier regions were enriched for the "Regulation of MAP kinase activity" term.

Table 5.3 - g:GOSTt enrichment results for genes in outlier regions. BP = Biological Process. HP = Human Phenotype.

Group	Term Name	Term ID	Term Type	Term Genes	Query	Corrected p-value	Common Gene List
Neolithic West	Positive regulation of smooth muscle cell differentiation	GO:0051152	BP	6	12	9.61E-03	*KIT, SIRT1*
	Regulation of developmental pigmentation	GO:0048070	BP	13	12	4.99E-02	*KIT, KITLG*
	Programmed cell death involved in cell development	GO:0010623	BP	11	12	3.52E-02	*KIT, KITLG*
	Ectopic germ cell programmed cell death	GO:0035234	BP	7	12	1.35E-02	*KIT, KITLG*
	Regulation of protein serine/threonine kinase activity	GO:0071900	BP	273	12	3.70E-02	*KIT, SIRT1, KITLG, DUSP6*
Neolithic East	Regulation of MAP kinase activity	GO:0043405	BP	184	15	2.24E-02	*KIT, EPGN, KITLG DUSP6*
	Ectopic germ cell programmed cell death	GO:0035234	BP	7	15	2.28E-02	*KIT, KITLG*
	Regulation of protein kinase activity	GO:0045859	BP	447	15	3.84E-02	*KIT, EPGN, EREG, KITLG DUSP6*
	White forelock	HP:0002211	HP	12	4	3.61E-02	*KIT, KITLG*

Nonsynonymous variants in outlier region genes

Nine nonsynonymous sites were found in seven genes: *LOC102172205* (serotransferrin, two variants), *STAT1*, *MYOM3*, *KITLG*, *KIT*, *LOC102185708* (*CYP2C19*), and *SIRT1* (two variants) (Table 5.4).

One *LOC102172205* (serotransferrin) variant was identified in Neolithic East (with frequency of 0.65), while a second variant was identified in both East and West as being fixed for the ancestral allele. The *STAT1* variant matched the ancestral allele, and is at a high frequency in both East (0.91) and West (0.7) Neolithic groups. A *MYOM3* variant is fixed as derived in Neolithic East, the population in which the gene was initially detected using the outlier approach, but was absent in Neolithic West. The *KITLG* nonsynonymous variant is fixed (1.0) in Neolithic East but common in Neolithic West (0.43). The *KIT* variant identified appears at a frequency of 0.75 in the Neolithic West but is absent in Neolithic East. Both *SIRT1* variants are a high frequency (≥0.75) in Neolithic West but low (~0.1) in Neolithic East; *SIRT1* was identified in an outlier region in Neolithic West (Table 5.2). *LOC102185708* (*CYP2C19*), identified originally in Neolithic East, was fixed for an ancestral allele in the same population, and fixed for the derived allele in Neolithic West.

F_{ST} values for pigmentation-associated loci

F_{ST} values were extracted for seven pigmentation-associated genes (*KIT*, *KITLG*, *MC1R*, *PMEL17*, *ASIP*, *TYRP1*, and *MITF*). For each Neolithic grouping the highest F_{ST} value for each gene and the mean F_{ST} of all windows overlapping the gene were calculated, and are displayed in Table 5.5. The highest F_{ST} window values were plotted along with the density plot of F_{ST} values between the Neolithic group and modern bezoar ibex for all 50kb windows, and are shown in Figure 5.1.

The previously detected *KIT* and region upstream of *KITLG* showed high F_{ST} values in both Neolithic groups relative to the total F_{ST} distribution. The highest *KIT* window F_{ST} computed was 0.59 and 0.51 (rounding to two decimals) in Neolithic East and West respectively; for the *KITLG* outlier region, highest F_{ST} values were 0.62 and 0.56. Mean F_{ST} values of all overlapping windows were also high, but in the Neolithic West population, substantially reduced for *KIT*: 0.53 and 0.34 for Neolithic East and West. In comparison, mean values for the *KITLG* outlier region were 0.56 and 0.53 respectively. As this region was defined by nature of it being an F_{ST} outlier, this result is not surprising.

Genetic structure of the *KIT* locus

To investigate the pattern of genetic structure at the *KIT* locus in ancient genomes, an identity-by-state matrix was computed using all ancient and modern genomes >1×, across the union of the outlier regions overlapping *KIT* (Table 5.2). A heatmap was then generated from this matrix, and is displayed in Figure 5.2.

Three main clusters are observed. The first cluster contains mainly ancient and modern eastern goat, in addition to several modern African, Bronze Age Turkish and Levantine genomes, and two goats from the Caucasus region. The second cluster is highly differentiated from the first, and composed of ancient Neolithic Serbian and modern European goat, as well as a Bronze Age goat from Britain. A third distinct cluster is observed on the right end of the plot, consisting of ancient and modern wild goat genomes. A final fourth rough grouping is composed goat from several different populations, including ancient modern Iranian, ancient

Table 5.4 - Nonsynonymous variants in outlier regions. The allele identified as being low frequency (<0.2) in bezoar ibex but fixed in an ancient Neolithic high coverage is indicated in bold. Chrom = Chromosome, Ref = CHIR_1.0 Reference allele, Anc = Ancestral allele, Deriv = Derived allele.

Chrom	Position	Gene	Ref	Anc	Deriv	East Deriv Frequency	West Deriv Frequency	Anc Residue	Deriv Residue
1	133157078	*LOC102172205*	C	C	**T**	0.65	0.00	A	T
1	133167772	*LOC102172205*	G	**G**	A	0.00	0.00	N	K
2	78845804	*STAT1*	T	**T**	G	0.09	0.30	N	T
2	127995419	*MYOM3*	G	A	**G**	1.00	0.00	K	R
5	17673144	*KITLG*	T	T	**A**	1.00	0.43	T	S
6	68332366	*KIT*	T	T	**A**	0.00	0.75	Y	N
26	15057235	*CYP2C19*	G	**G**	C	0.00	1.00	T	R
28	22205610	*SIRT1*	C	A	**C**	0.10	0.81	Q	D
28	22205761	*SIRT1*	T	T	**G**	0.11	0.75	S	A

The *ASIP* and *MITF* genes appear differentiated in both of the Neolithic groups relative to modern Iranian bezoar ibex. *ASIP* highest F_{ST} windows were 0.27 and 0.49 for Neolithic East and West, and 0.28 and 0.47 for *MITF*. Mean F_{ST} values of windows overlapping these genes were lower (as expected), particularly for *MITF* in the Neolithic West population (0.31). The *TYRP1* gene was also somewhat differentiated from bezoar in the Neolithic West, with a highest F_{ST} value of 0.32 and mean F_{ST} of 0.30. This gene was not particularly differentiated from bezoar in the Neolithic East population (highest F_{ST}: 0.14).

The remaining pigmentation genes showed F_{ST} values close to the mean of the F_{ST} distribution. The highest F_{ST} values for windows overlapping *PMEL17* were 0.16 for both Neolithic East and West. For *MC1R*, these values were 0.20 and 0.17. As such, these genes had little evidence of differentiation in early domestic goat populations relative to bezoar ibex.

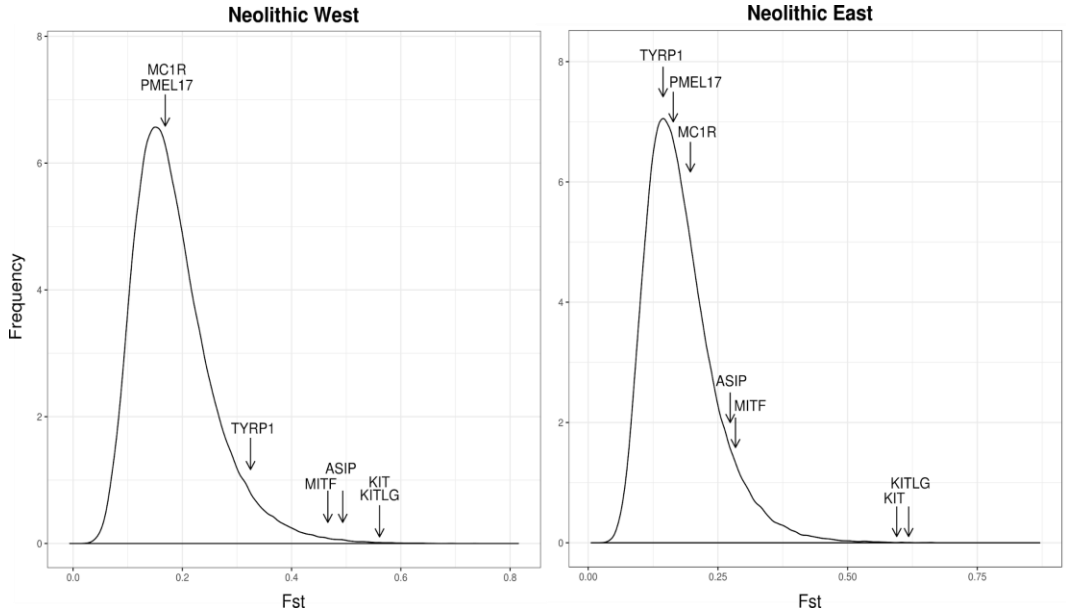


Figure 5.1 - Highest F_{ST} values for windows overlapping seven genes associated with coat pigmentation, with distribution of all F_{ST} values plotted as a density function. F_{ST} is calculated between Neolithic groups and modern bezoar ibex.

Caucasus, Moroccan, and a single modern French goat. Substructure within this cluster is observed. For example, four Modern Iranians, a Caucasus Medieval (Kazbeg1), and an Iranian Medieval (Darre2) for a small cluster which shows lows IBS with the four Neolithic Serbian genomes from Blagotin, but slightly higher allele sharing with the remaining western individuals.

Table 5.5 - F_{ST} values for windows overlapping pigmentation genes. * was calculated on the outlier region upstream of *KITLG*, using the overlap between Neolithic East and West (5:18060000-1817999).

Gene	Chrom	Start	End	Highest F_{ST} Window - East	Mean F_{ST} - East	Highest F_{ST} Window - West	Mean F_{ST} - West
*KITLG**	5	17607323	17713886	0.617585	0.555796	0.561898	0.525580
PMEL17	5	55829082	55845319	0.163902	0.146615	0.16488	0.152095
KIT	6	68316048	68361679	0.594616	0.527418	0.507348	0.340998
TYRP1	8	30671334	30687862	0.144336	0.118669	0.324814	0.297307
ASIP	13	61693104	61698483	0.273657	0.225399	0.493624	0.420402
MC1R	18	14208837	14212670	0.196922	0.175957	0.168685	0.156913
MITF	22	31427864	31659079	0.283904	0.234776	0.466363	0.305813

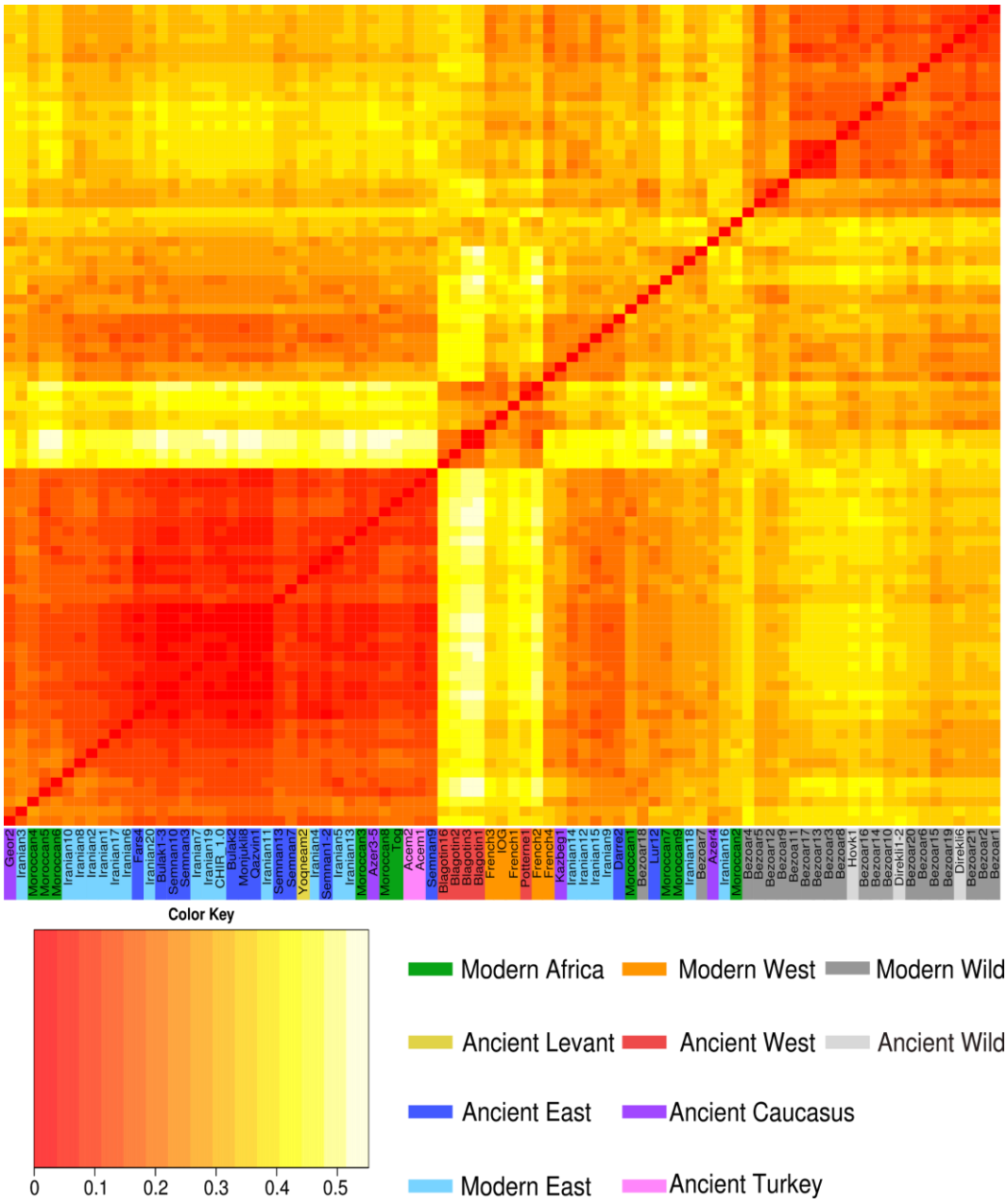


Figure 5.2 - IBS heatmap of outlier region overlapping *KIT*. Modern and ancient genomes of mean coverage >1× were included. East = Iran, Turkmenistan and Uzbekistan. West = Europe.

Discussion

Signals of differentiation at pigmentation-associated loci

KIT and a noncoding region upstream of *KITLG* appear in the only outlier regions found in both Neolithic East and Neolithic West. Both genes are involved in white pigmentation phenotypes, such as roan coat colour in cattle (Seitz et al. 1999) or spotting in rabbits (Fontanesi et al. 2014). There is evidence that these genes also have roles in colouration in goat. Variants near *KIT* were recently found to be associated with white coat colour in the mohair-producing Iranian Markhoz goat (Nazari-Ghadikolaei et al. 2018). Using the Z score-transformed pooled heterozygosity statistic ZH_P and a measurement of breed-specific differentiation, *KITLG* was identified in a selection scan of both Taihang Black and Tibetan goat (Wang et al. 2016). An analysis of Italian and Pakistani goat breeds using Runs Of Homozygosity (ROH) and the Cross Population Extended Haplotype Homozygosity statistic (XP-EHH) also found the region surrounding *KITLG* to be differentiated in roan-coloured breeds compared to non-roan breeds (Talenti et al. 2017).

The results presented here suggest that both Neolithic goat populations were differentiated from wild bezoar at or nearby genes involved in coat colour in goat. This assumes the findings of the association studies above can be validated using functional genetic techniques such as CRISPR-Cas9 (Zhang et al. 2017). If they are, and if the approach taken here identified regions genuinely subject to selection in Neolithic goat, it can be argued that early farmers deliberately selected for or favoured goat with novel coat colours or patterns, different from those observed in wild bezoar. The appearance of coat colours distinct from those observed in the wild is expected due to the relaxation of selection for camouflage colouration, a consequence of human protection against predation (Zohary et al. 1998).

One nonsynonymous variant was identified in *KIT* that was also at a low frequency in bezoar (Table 5.4). This variant was found to be at a frequency of 0.75 in Neolithic West genomes but absent in the Neolithic East population, and is expected to cause a Tyrosine-to-Arginine substitution. This change from an aromatic, partially hydrophobic residue to a positively-charged, polar amino acid would likely not be inconsequential, due to their differing biochemical properties (Betts & Russell 2003). The failure to find a similar nonsynonymous mutation in Neolithic East suggests that if selection did occur on this locus, it did not act on a nonsynonymous variant, or at least not one at a low frequency in modern bezoar.

Furthermore, an identity-by-state heatmap and clustering of *KIT* (Figure 5.2) indicated that the Neolithic East and West populations were also differentiated between each other, suggesting different haplogroups in the two populations were responsible for their genetic distance from bezoar. The genetic structure of *KIT* persists through time; Neolithic West was found to cluster with Bronze Age and modern European samples, while the majority of Neolithic East genomes clustered with modern Iranian and African genomes, as well as ancient samples from Iran, Bronze Age Turkey, and the Caucasus region. It is unclear why genetic structure at a coat colour locus would persist through thousands of years, given that most pigmentation change would be expected to be in the relatively recent past during modern breed formation. It is possible that the changes observed in Neolithic goat are a consequence of selection for trait linked to *KIT* or *KITLG* via pleiotropy. *KITLG* and *KIT* are expressed at many stages of ovarian follicular development in goat (Silva et al. 2006), and contributes to follicular survival and growth *in vitro* (Lima et al. 2011). There have been some evidence that *KITLG* variants affect litter size in goat (An et al. 2012) and pig (Okumura et al. 2008), while *KIT* has been implicated in male infertility in humans (Galan et al. 2006). It is therefore plausible that selection at either locus may have been acting on a reproductive or fertility-related phenotype, rather than colouration.

The possibility that *KIT* and *KITLG* were under selection due to their role in pigmentation is somewhat strengthened by the statistically-significant enrichment for pigmentation-related genes (Table 5.3), and the high F_{ST} of other coat colour-related genes (Figure 5.1). *ASIP* and *MITF*, and to a lesser extent *TYRP1*, fall on the right of the F_{ST} distribution. A duplication in *ASIP* has been shown to cause a dominant white coat phenotype in sheep (Norris & Whan 2008), and the gene has recently been identified in a selection scan of Taihang Black goat (Wang et al. 2016). Coat colour diversity in domestic goat has been recorded on Egyptian papyrus from the predynastic era (Zeuner 1963), suggesting that the genetic variants responsible for different colourations were already present in the goat gene pool.

If genes which affect coat colour variation were indeed under selection in the Neolithic specifically for their coat colour effect (rather than another pleiotropic effect), it would challenge a prevailing concept of the role of the farmer in the early stages of domestication. Domestication of goat and other early livestock animals is thought to have been accidental *i.e.* without intent on the part of the farmer (Larson & Fuller 2014). A decline in resource availability or demographic growth may have resulted in intensification of the hunter-hunted relationship, leading to herd management that required additional energy investment to

sustain (Stiner et al. 2014), generating a positive feedback loop. This does not require intent on the part of the farmer to control the live cycle of the animal but instead to retain a supply of the animal resource. Directed selection (*i.e.* with intent on the part of the farmer) in domestic goat is expected to have been weak or absent until secondary products (wool and milk) became more important (Marshall et al. 2014).

In comparison, selection for coat colour suggests active preference for one type of animal over another, resulting in genomic selection via differential reproduction of the animal. Some have argued that the symbolic worth of animal ownership may have catalyzed the early stages of animal management (Vigne 2011). If herd keeping was indeed a marker of worth or wealth for the first practitioners of animal management, then it is plausible that distinct or unique coat colouration of kept animals would also add to social prestige. Alternatively, coat colours which were different from those of a neighbour's herd or from local wild goat may have had practical advantage, to keep track of one's own herd.

Differentiation of toxin-metabolism cytochrome P450 genes

The occurrence of two genomically-close cytochrome P450 genes, 2C9 and 2C19, in the outlier scan in the Neolithic East population (Table 5.2) is striking. These genes belong to a family of oxidase enzymes with major roles in both exogenous and endogenous substance metabolism in mammals (Seliskar & Rozman 2007). Cytochrome P450 genes are plausible targets for selection in early domesticates, due to the process involving a stark change in the environment of the animal. Such changes can be inferred by the plastic phenotype responses recorded of early animal remains (Zeder 2015). One such change is foddering, the provision of food that would otherwise be unavailable to animals, which can be tracked by the analysis of stable radioisotopes in bones collagen (Makarewicz & Tuross 2012). Such provisioning would provide managed herds with a stable food source year-round, thus ensuring that humans also maintained a predictable food source. Foddering has been demonstrated to help maintain body weight in goats when their usual food source is restricted, such as in summer months in the Mediterranean (Papachristou et al. 1999).

The feed provided to the animals may have been by-products of agriculture, wild vegetation, or grain grown specifically for foddering. This practice would have necessitated the storage of food material between their harvesting and their use as fodder. This time span would allow for the growth of fungus and other food-spoilage organisms, and potentially contamination of

the feed with mycotoxins. Cytochrome P450 proteins have been established to process mycotoxins to active toxic form (He et al. 2006). In particular 2C19, one of the cytochrome P450 genes identified in the outlier scan, has recently been shown to have a role in the biotransformation of Enniatin B, a cytotoxic mycotoxin (Fæste et al. 2011). This mycotoxin is produced by *Fusarium* fungal strains, a frequent contaminant of cereals (Uhlig et al. 2007). If the metabolites of this or other mycotoxins found in cereals have a negative effect on animals accidentally ingesting it, a low-activity variant of cytochrome P450 2C19 may have been selected for in early managed herds. This may have occurred at cytochrome P450 2C19 during the last 10,000 years in humans (Janha et al. 2014). Alternatively, if the mycotoxin itself showed greater toxicity than its breakdown products, a high activity cytochrome variant may have been selected for.

Selection for fertility and reproduction traits

Several of the genes detected here in the outlier F_{ST} scan have been implicated in fertility or reproduction traits. *HERC4*, *KITLG*, *KIT* have been linked to variation in human and mouse male infertility, and litter size in goat. (Galan et al. 2006; An et al. 2015). *EPGN* and *EREG* are thought to have roles in calving intervals and oocyte maturation respectively (Raven et al. 2016; Riese & Cullum 2014). Both cytochrome P450 2C19 and 2C9 have plausible roles in reproductive variation, metabolizing the steroid hormones progesterone and testosterone (Yamazaki & Shimada 1997). The number of oestrous cycles in a breeding season, length of time when oestrous behaviour is expressed (reproductive receptivity), and the timing and length of the breeding season all show variability between breeds, and are also sensitive to environmental conditions (Fatet et al. 2011). As the economies of farmers relied more on the managed herd for supply of meat and other products, it is plausible that selection may have occurred for more frequent estrous cycles, short gestation times, or increased litter size. Domestic goat do indeed reach sexual maturity earlier than bezoar, and twin more often than *Capra ibex* (Tchernov & Horwitz 1991; Castelló et al. 2016). This selection does not need to have been intentional; more fecund goat may have simply outbred the less prolific animals, due to reduced intra-specific competition in the anthropogenic environment (Zohary et al. 1998).

Neural crest cell hypothesis

The F_{ST} outlier scan did not detect genes obviously linkable to neural crest cell hypothesis of domestication syndrome. Although *MITF* and *KIT* are associated with domestication syndrome (Wilkins et al. 2014), their activity in relation to neural crest cells is primarily in their melanocyte derivatives (Reissmann & Ludwig 2013). However, the approach taken here may not have been appropriate to detecting genetic variants which regulate neural crest cell numbers or migratory patterns. The genetic architecture of domestication syndrome is expected to be polygenic, caused by many alleles of small effect sizes (Wilkins et al. 2014). Novel variation can contribute to domestication syndrome, but it is likely that standing variation in wild species is sufficient to recapitulate the phenotype. As such, increases in the frequency of alleles at many loci may be sufficient to produce the phenotype, as opposed to a small number of large chances which the outlier approach should have detected. Repeat element expansion or extraction rather than point mutations may also play a role in the phenotype; this genetic variability would not be detected in the approach taken.

Limitations

The F_{ST} outlier approach taken here has substantial limitations. The premise of the approach - that regions of the genome under selection in early domestic goat should be highly differentiated from wild bezoar - is dependent on having a sample representative of wild goat from which domestic goat are derived. Ideally, this would be temporally matched to when domestication occurred *i.e.* performed using a sample of bezoar genomes from *c*. 8,000 cal BC. At this point in time, this is not feasible, and is complicated by the fact that early managed goat were likely interbreeding with unmanaged goat (Zeder 2005). Ignoring the temporal dimension, the bezoar samples used here may not be geographically appropriate. If the bulk of the genetic ancestry of domestic goat come from a wild population in the Zagros Mountains or in eastern Turkey, we might expect large genetic differences between domestic and modern bezoar genomes. This may be due to genetic structure not yet quantified, or genetic adaptations to distinct environments. The fact that different domestic populations are admixed with different bezoar populations (Chapter 4) further complicates this.

The distribution of F_{ST} values are affected by forces other than selection, confounding outlier tests which rely on the premise that the F_{ST} distribution occurs due to drift alone. However, population bottlenecks can create long haplotypes and changes in local allele frequencies that

are similar to those expected from a selective sweep (Pavlidis et al. 2010). Accounting for the demographic histories of the populations can also better define the null distribution of F_{ST}, assuming the estimated demographic histories are accurate (Lotterhos & Whitlock 2014). Range expansion and isolation-by-distance, relevant in this analysis due to the nature of the Neolithic expansion that these ancient goat genomes derive from, can cause false-positive F_{ST} outliers. To account for this in our analysis, a conservative cut-off was used to define outlier F_{ST} windows (99.9th quantile), although true positives may not be enriched in the tail of the F_{ST} distribution (Pavlidis et al. 2010). In addition, we attempted to identify only regions with a reduced number of segregating sites in the domestic population relative to bezoar by using two θ_W filtering criteria, under the premise that the identified regions should be likely to be the product of a selective sweep in domestic goat. However, in both cases essentially arbitrary cut-off values were taken for the F_{ST} and θ_W statistics. The same applies for the window size and step size used in statistic calculation. Small changes to arbitrarily-defined parameters can drastically change the results of such selection analysis (Pavlidis et al. 2012).

A better approach may have been to account for their demographic history, rather than applying filters on an *ad hoc* basis. A recent study of F_{ST} outlier regions in village dogs relative to wolves accounted for the demographic history of dogs and wolves by performing neutral simulations to create a neutral null model, and identified several genes associated with neural crest cells and neurological function in general (Pendleton et al. 2018). Although that study relied mostly on modern genetic data, here it would have been possible to generate a F_{ST} null distribution of genomic windows, using a demographic history of modern goat estimated by PSMC (Li & Durbin 2011) or similar approaches. The true number of false positives would still be difficult if not impossible to estimate (Pavlidis et al. 2010; Pavlidis et al. 2012), so it is unclear if there is an unambiguously correct approach to have taken.

Several other factors were not considered in this analysis. Utilizing patterns of linkage disequilibrium can help distinguish loci under selection from outlier loci under strong bottleneck scenarios (Jensen et al. 2007), which in the approach taken here is not exploited. An additional confounder is regions of reduced recombination: by breaking up haplotypes less frequently, genetic diversity is reduced in these regions relative to the genomic average (Cruickshank & Hahn 2014). As F_{ST} is affected by within-group levels of genetic diversity, a reduction in recombination at certain genomic regions would inflate the local F_{ST}. As the genomes used here derive mostly from single archaeological sites (see below), any unusual recombination patterns specific to goat from that site or region may produce false positive F_{ST}

outliers. Within-species variation in recombination does occur; cattle breeds show specific recombination patterns, although global recombination patterns are similar (Shen et al. 2018).

There is also the danger here of storytelling: accepting the method as the results can be made to make sense *a posteriori*, and searching the literature for studies that support the narrative of selection for a specific phenotype (Pavlidis et al. 2012). Although the Gene Ontology enrichment analysis did support an excess of pigmentation-related terms detected, these were driven entirely by *KIT* and *KITLG*. In addition, no enrichment was detected in fertility, gonadal, or detoxifying-related terms. As such, the findings here should be validated using a different approach, such as a LD-aware method, and ideally account for the goat demographic history as well. *KITLG* was recently detected as likely under selection in sheep and goat using the LD-sensitive hapFLK (Alberto et al. 2018). They reported increased genetic diversity in domestic goat populations compared to wild, in regions closely flanking the gene. In contrast, in this project reduced diversity and increased differentiation was detected ~300kb upstream of *KITLG*. Gene regulatory elements such as enhancers can be found hundreds of kilobases away from their target gene (Lajoie et al. 2015), so it possible that such a regulatory element for *KITLG* is present in the outlier region. No tests for the presence of transcription factor binding sites were performed, which may have revealed the presence of such regulatory elements.

An additional issue that is difficult to circumvent was previously discussed in Chapter 4: the fact that the majority of genomes from both Neolithic East and Neolithic West come from a single respective site - either Sang-e Chakmaq for Neolithic East or Blagotin for Neolithic West. The goat from these archaeological sites may not be representative of all goat from their respective geographic area, and differences between them and bezoar may not be extrapolatable to the general domestic goat population. For example, the high differentiation observed at pigmentation-associated loci may be entirely local phenomena *i.e.* specific to the Serbia region in the Neolithic. Traditions specific to these sites may have resulted in selection for specific variation at pigmentation loci. Alternatively, each Neolithic population may be somewhat related (see lcMLkin analysis in Chapter 4), and happen to have recombination patterns that produce low sequence diversity, and therefore differentiation, at these loci (Cruickshank & Hahn 2014). However, the strong genetic structure observed at *KIT* is evidence against this. The differentiation at the locus persists through time to the present-day, and is supported by ancient samples from a number of regions, and is therefore unlikely to be purely a local effect. It would be preferable for this differentiation to be validated with

additional genomes from an expanded geographic area and greater number of archaeological sites, to confirm whether these were local events or reflect a broader trend in goat domestication.

Conclusion

The F_{ST} outlier approach taken in this chapter suggest that geographically-distinct populations of early domestic goat were strongly differentiated from wild bezoar at key pigmentation loci. The eastern and western goat populations were also highly structured at the *KIT* locus, based on an identity-by-state analysis, a pattern which continued through time to the present-day. This suggests that early farmers may have favoured, and intentionally selected for, novel or distinct coat colours in goat. Gene Ontology enrichment and F_{ST} values of other pigmentation-associated genes support early selection for coat colour in domestic goat herd. Alternatively, epistatic phenotypes may have been the target of selection, such as the fecundity-related traits of litter size or male fertility. In addition, an adaptation to a novel food source - stored grain - is suggested, due to strong differentiation at the detoxification locus cytochrome P450 2C19.

There are significant limitations to the conclusions that can be drawn, however. A small number of sample sites and an inappropriate bezoar reference population can be overcome with broader sampling of both ancient and modern genomes. Deeper sequencing or imputation of ancient samples would allow LD-aware methods to be employed. Modeling the expected distribution of F_{ST} values based on an estimated demographic history of goat would be preferable, to obtain a better approximation of the null distribution. This would help to distinguish between the false and true positive signals of selection in early domestic goat genomes.

Chapter 5 References

Alberto, F.J. et al., 2018. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

An, X.P. et al., 2015. Association analysis between variants in KITLG gene and litter size in goats. *Gene*, 558(1), pp.126–130.

An, X.P. et al., 2012. Polymorphism identification in the goat KITLG gene and association analysis with litter size. *Animal genetics*, 43(1), pp.104–107.

Auffray, C. & De Meulder, B., 2016. Faculty of 1000 evaluation for g:Profiler-a web server for functional interpretation of gene lists (2016 update). *F1000 - Post-publication peer review of the biomedical literature*. Available at: http://dx.doi.org/10.3410/f.726303978.793520093.

Becker, D. et al., 2015. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the TYRP1 locus on chromosome 8. *Animal genetics*, 46(1), pp.50–54.

Benjelloun, B. et al., 2015. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Frontiers in genetics*, 6, p.107.

Besmer, P. et al., 1993. The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Development . Supplement*, pp.125–137.

Betts, M.J. & Russell, R.B., 2003. Amino Acid Properties and Consequences of Substitutions. In M. R. Barnes & I. C. Gray, eds. *Bioinformatics for Geneticists*. Chichester : Wiley, pp. 289–316.

Braglia, S. et al., 2013. SNPs of MYPN and TTN genes are associated to meat and carcass traits in Italian Large White and Italian Duroc pigs. *Molecular biology reports*, 40(12), pp.6927–6933.

Brooks, S.A. & Bailey, E., 2005. Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mammalian genome: official journal of the International Mammalian Genome Society*, 16(11), pp.893–902.

Brunberg, E. et al., 2006. A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC genetics*, 7, p.46.

Castelló, J.R., Huffman, B. & Groves, C., 2016. *Antelopes, Gazelles, Cattle, Goats, Sheep, and Relatives*, Princeton University Press.

Chabot, B. et al., 1988. The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature*, 335(6185), pp.88–89.

Chen, H., Patterson, N. & Reich, D., 2010. Population differentiation as a test for selective sweeps. *Genome research*, 20(3), pp.393–402.

Cobanoglu, O. et al., 2006. Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *Journal of dairy*

science, 89(11), pp.4433–4437.

Cooper, M.P. et al., 2006. White spotting in the domestic cat (Felis catus) maps near KIT on feline chromosome B1. *Animal genetics*, 37(2), pp.163–165.

Cruickshank, T.E. & Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, 23(13), pp.3133–3157.

Durbin, J.E. et al., 1996. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell*, 84(3), pp.443–450.

Ewens, W.J., 1974. A note on the sampling theory for infinite alleles and infinite sites models. *Theoretical population biology*, 6(2), pp.143–148.

Fæste, C.K., Ivanova, L. & Uhlig, S., 2011. In vitro metabolism of the mycotoxin enniatin B in different species and cytochrome p450 enzyme phenotyping by chemical inhibitors. *Drug metabolism and disposition: the biological fate of chemicals*, 39(9), pp.1768–1776.

Fariello, M.I. et al., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, 193(3), pp.929–941.

Farrugia, G., 2008. Interstitial cells of Cajal in health and disease. *Neurogastroenterology and motility: the official journal of the European Gastrointestinal Motility Society*, 20 Suppl 1, pp.54–63.

Fatet, A., Pellicer-Rubio, M.-T. & Leboeuf, B., 2011. Reproductive cycle of goats. *Animal reproduction science*, 124(3-4), pp.211–219.

Fleischman, R.A. et al., 1991. Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23), pp.10885–10889.

Fontanesi, L. et al., 2009. Missense and nonsense mutations in melanocortin 1 receptor (MC1R) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. *BMC genetics*, 10, p.47.

Fontanesi, L. et al., 2014. The KIT gene is associated with the english spotting coat color locus and congenital megacolon in Checkered Giant rabbits (Oryctolagus cuniculus). *PloS one*, 9(4), p.e93750.

Galan, J.J. et al., 2006. Association of genetic markers within the KIT and KITLG genes with human male infertility. *Human reproduction* , 21(12), pp.3185–3192.

Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables, 2016. *gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1*, Available at: https://CRAN.R-project.org/package=gplots.

Hauswirth, R. et al., 2012. Mutations in MITF and PAX3 cause “splashed white” and other white spotting phenotypes in horses. *PLoS genetics*, 8(4), p.e1002653.

Hemmer, H., 1990. *Domestication: The Decline of Environmental Appreciation*, Cambridge University Press.

He, X.-Y. et al., 2006. Efficient activation of aflatoxin B1 by cytochrome P450 2A13, an enzyme predominantly expressed in human respiratory tract. *International journal of cancer. Journal international du cancer*, 118(11), pp.2665–2671.

Hou, W. et al., 2018. Arginine Methylation by PRMT2 Controls the Functions of the Actin Nucleator Cobl. *Developmental cell*, 45(2), pp.262–275.e8.

Isvoran, A. et al., 2017. Pharmacogenomics of the cytochrome P450 2C family: impacts of amino acid variations on drug metabolism. *Drug discovery today*, 22(2), pp.366–376.

Jacobson, N.G. et al., 1995. Interleukin 12 signaling in T helper type 1 (Th1) cells involves tyrosine phosphorylation of signal transducer and activator of transcription (Stat)3 and Stat4. *The Journal of experimental medicine*, 181(5), pp.1755–1762.

Janha, R.E. et al., 2014. Inactive alleles of cytochrome P450 2C19 may be positively selected in human evolution. *BMC evolutionary biology*, 14, p.71.

Jensen, J.D. et al., 2007. On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*, 176(4), pp.2371–2379.

Jiao, Y. et al., 2010. A novel polymorphism of the MYPN gene and its association with meat quality traits in Bos taurus. *Genetics and molecular research: GMR*, 9(3), pp.1751–1758.

Khatib, H. et al., 2009. Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of dairy science*, 92(12), pp.6186–6191.

Kijas, J.W. et al., 2013. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal genetics*, 44(3), pp.325–335.

Korman, B.D. et al., 2008. STAT4: genetics, mechanisms, and implications for autoimmunity. *Current allergy and asthma reports*, 8(5), pp.398–403.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Kruska, D., 1988. Mammalian Domestication and its Effect on Brain Structure and Behavior. In *Intelligence and Evolutionary Biology*. Springer Berlin Heidelberg, pp. 211–250.

Lajoie, B.R., Dekker, J. & Kaplan, N., 2015. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* , 72, pp.65–75.

Larson, G. & Fuller, D.Q., 2014. The Evolution of Animal Domestication. *Annual review of ecology, evolution, and systematics*, 45(1), pp.115–136.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , 25(16), pp.2078–2079.

Li, H. & Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), pp.493–496.

Li, M. et al., 2013. SIRT1 gene polymorphisms are associated with growth traits in Nanyang cattle. *Molecular and cellular probes*, 27(5-6), pp.215–220.

Lima, I.M.T. et al., 2011. Presence of c-kit mRNA in goat ovaries and improvement of in vitro preantral follicle survival and development with kit ligand. *Molecular and cellular endocrinology*, 345(1-2), pp.38–47.

Lotterhos, K.E. & Whitlock, M.C., 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular ecology*, 23(9), pp.2178–2192.

Maijala, K., 1991. *Genetic resources of pig, sheep, and goat*, Elsevier Science Publishers.

Makarewicz, C. & Tuross, N., 2012. Finding Fodder and Tracking Transhumance: Isotopic Detection of Goat Domestication Processes in the Near East. *Current anthropology*, 53(4), pp.495–505.

Marklund, S. et al., 1999. Close association between sequence polymorphism in the KIT gene and the roan coat color in horses. *Mammalian genome: official journal of the International Mammalian Genome Society*, 10(3), pp.283–288.

Marklund, S. et al., 1998. Molecular basis for the dominant white phenotype in the domestic pig. *Genome research*, 8(8), pp.826–833.

Marshall, F.B. et al., 2014. Evaluating the roles of directed breeding and gene flow in animal domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), pp.6153–6158.

Martin, P. et al., 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. *Scientific reports*, 7(1), p.1872.

Martin, P.M. et al., 2017. Correction: Genome Wide Association Study Identifies New Loci Associated with Undesired Coat Color Phenotypes in Saanen Goats. *PloS one*, 12(10), p.e0186029.

Mason, I.L., 1986. *Evolution of Domesticated Animals*, John Wiley & Sons, Incorporated.

Miklósi, A. et al., 2003. A simple reason for a big difference: wolves do not look back at humans, but dogs do. *Current biology: CB*, 13(9), pp.763–766.

Nawroth, C. et al., 2018. Goats prefer positive human emotional facial expressions. *Royal Society Open Science*, 5(8), p.180491.

Nawroth, C., Brett, J.M. & McElligott, A.G., 2016. Goats display audience-dependent human-directed gazing behaviour in a problem-solving task. *Biology letters*, 12(7). Available at: http://dx.doi.org/10.1098/rsbl.2016.0283.

Nazari-Ghadikolaei, A. et al., 2018. Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat.

Frontiers in genetics, 9, p.105.

Norris, B.J. & Whan, V.A., 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome research*, 18(8), pp.1282–1293.

Okumura, N. et al., 2008. Single nucleotide polymorphisms of the KIT and KITLG genes in pigs. *Animal science journal = Nihon chikusan Gakkaiho*, 79(3), pp.303–313.

Pailhoux, E. et al., 2001. A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature genetics*, 29(4), pp.453–458.

Papachristou, T.G. et al., 1999. Use of deciduous woody species as a diet supplement for goats grazing Mediterranean shrublands during the dry season. *Animal feed science and technology*, 80(3), pp.267–279.

Pavlidis, P. et al., 2012. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular biology and evolution*, 29(10), pp.3237–3248.

Pavlidis, P., Jensen, J.D. & Stephan, W., 2010. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, 185(3), pp.907–922.

Pendleton, A.L. et al., 2018. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC biology*, 16(1), p.64.

Pinto, N. & Dolan, M.E., 2011. Clinically relevant genetic variations in drug metabolizing enzymes. *Current drug metabolism*, 12(5), pp.487–497.

Porter, V. et al., 2016. *Mason's World Encyclopedia of Livestock Breeds and Breeding*, Boston: CABI.

Porter, V. & Tebbit, J., 1996. *Goats of the World*, London: Farming Press Ltd.

Raven, L.-A. et al., 2016. Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. *Mammalian genome: official journal of the International Mammalian Genome Society*, 27(1-2), pp.81–97.

Reber, I. et al., 2015. Wattles in goats are associated with the FMN1/GREM1 region on chromosome 10. *Animal genetics*, 46(3), pp.316–320.

Reimand, J. et al., 2007. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research*, 35(suppl_2), pp.W193–W200.

Reissmann, M. & Ludwig, A., 2013. Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Seminars in cell & developmental biology*, 24(6-7), pp.576–586.

Riese, D.J., 2nd & Cullum, R.L., 2014. Epiregulin: roles in normal physiology and cancer. *Seminars in cell & developmental biology*, 28, pp.49–56.

Rodriguez, C.I. & Stewart, C.L., 2007. Disruption of the ubiquitin ligase HERC4 causes defects in spermatozoon maturation and impaired fertility. *Developmental biology*, 312(2), pp.501–508.

Rönnstrand, L., 2004. Signal transduction via the stem cell factor receptor/c-Kit. *Cellular and molecular life sciences: CMLS*, 61(19-20), pp.2535–2548.

Saatchi, M. et al., 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC genomics*, 15, p.442.

Seitz, J.J. et al., 1999. A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mammalian genome: official journal of the International Mammalian Genome Society*, 10(7), pp.710–712.

Seliskar, M. & Rozman, D., 2007. Mammalian cytochromes P450--importance of tissue specificity. *Biochimica et biophysica acta*, 1770(3), pp.458–466.

Shen, B. et al., 2018. Characterization of recombination features and the genetic basis in multiple cattle breeds. *BMC genomics*, 19(1), p.304.

Silva, J.R.V. et al., 2006. The Kit ligand/c-Kit receptor system in goat ovaries: gene expression and protein localization. *Zygote* , 14(4), pp.317–328.

Singh, D. et al., 2011. Novel advances in cytochrome P450 research. *Drug discovery today*, 16(17-18), pp.793–799.

Song, S. et al., 2016. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). *BMC genomics*, 17, p.122.

Stiner, M.C. et al., 2014. A forager-herder trade-off, from broad-spectrum hunting to sheep management at Aşıklı Höyük, Turkey. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), pp.8404–8409.

Talenti, A. et al., 2017. Genomic analysis suggests KITLG is responsible for a roan pattern in two Pakistani goat breeds. *The Journal of heredity*. Available at: http://dx.doi.org/10.1093/jhered/esx093.

Tchernov, E. & Horwitz, L.K., 1991. Body Size Diminution under Domestication - Unconscious Selection in Primeval Domesticates. *Journal of Anthropological Archaeology*, 10(1)), pp.54–75.

Tosser-Klopp, G. et al., 2014. Design and characterization of a 52K SNP chip for goats. *PloS one*, 9. Available at: http://dx.doi.org/10.1371/journal.pone.0086227.

Uhlig, S., Jestoi, M. & Parikka, P., 2007. Fusarium avenaceum -- the North European situation. *International journal of food microbiology*, 119(1-2), pp.17–24.

Vigne, J.-D., 2011. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. *Comptes rendus biologies*, 334(3), pp.171–181.

Wang, X. et al., 2016. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Scientific*

reports, 6, p.38932.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. *Theoretical population biology*, 7(2), pp.256–276.

Wehrle-Haller, B. & Weston, J.A., 1997. Receptor tyrosine kinase-dependent neural crest migration in response to differentially localized growth factors. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 19(4), pp.337–345.

Wilkins, A.S., Wrangham, R.W. & Fitch, W.T., 2014. The "domestication syndrome" in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics*, 197(3), pp.795–808.

Wong, A.K. et al., 2013. A de novo mutation in KIT causes white spotting in a subpopulation of German Shepherd dogs. *Animal genetics*, 44(3), pp.305–310.

Yamazaki, H. & Shimada, T., 1997. Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Archives of biochemistry and biophysics*, 346(1), pp.161–169.

Zeder, M.A., 2006. A critical assessment of markers of initial domestication in goats (Capra hircus). In M. Zeder et al., eds. *Documenting Domestication: New Genetic and Archaeological Paradigms*. Berkeley: University of California Press, pp. 181–208.

Zeder, M.A., 2005. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 125–146.

Zeder, M.A., 2015. Core questions in domestication research. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), pp.3191–3198.

Zeder, M.A., 2012. The Domestication of Animals. *Journal of anthropological research*, 68(2), pp.161–190.

Zeuner, F.E., 1963. *A History of Domesticated Animals*, London: Hutchinson & Co.(Publishers) Ltd.

Zhang, X. et al., 2017. Alteration of sheep coat color pattern by disruption of ASIP gene via CRISPR Cas9. *Scientific reports*, 7(1), p.8149.

Zohary, D., Tchernov, E. & Horwitz, L.K., 1998. The role of unconscious selection in the domestication of sheep and goats. *Journal of zoology*, 245(2), pp.129–135.

Chapter 6: The *Capra* genus

Introduction

The *Capra* genus

The *Capra* genus is comprised of nine wild species, the designation of which have been somewhat controversial. The taxonomic assignments used here follow those proposed by Heptner et al. 1961. Binomial names and geographic distribution are presented in Table 6.1 and Figure 6.1; images of each *Capra* species are displayed in Figure 6.2. These species are discontinuously distributed, being found in mountainous, alpine and arid regions across Eurasia, with some species encroaching into northern Africa.

Table 6.1 - Capra species, their binomial names and an overview of the countries in which they are currently found in the wild. Species definition are those proposed by Heptner et al. 1961. Species distributions are drawn from Shackleton 1997.

Common Name	Binomial Name	Geographic Distribution
Bezoar Ibex	*Capra aegagrus*	Turkey, Iran, Iraq, Georgia, Armenia, Azerbaijan, Turkmenistan, Pakistan
Markhor	*Capra falconeri*	Afghanistan, Pakistan, India, Tajikistan, Uzbekistan
East Caucasian Tur	*Capra cylindricornis*	Eastern Caucasus Mountains
West Caucasian Tur	*Capra caucasica*	Western Caucasus Mountains
Alpine Ibex	*Capra ibex*	European Alps (Switzerland, France, Germany, Austria, Italy)
Iberian Ibex	*Capra pyrenaica*	Iberian Peninsula
Nubian Ibex	*Capra nubiana*	Algeria, Egypt, Ethiopia, Eritrea, Israel, Jordan, Lebanon, Oman, Saudi Arabia, Sudan, Yemen
Walia Ibex	*Capra walie*	Semien Mountains of Ethiopia
Siberian Ibex	*Capra sibirica*	Russia, Mongolia, China, India, Pakistan, Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan

As discussed in Chapter 1, the phylogeny of *Capra* has not proven easy to resolve, either by morphometrics or molecular data. Skull morphology, horn shape and spiraling/curling, and leg markings are not congruent in the relationships they suggest (Groves and Grubb 2011). Molecular phylogenies are also not consistent, with mtDNA, Y chromosome, and nuclear marker analyses in disagreement (Manceau et al. 1999; Pidancier et al. 2006; Ropiquet & Hassanin 2006). An additional complexity is the capacity of these species to interbreed, both in the wild (Zalikhanov 1967) and in captivity (Mason 1986). Members of the *Capra* genus have the same karyotype of 2n = 60 (Schmitt & Ulbrich 1968), suggesting that incompatible chromosome number is not a postzygotic barrier between species. Hybridization between *Capra* species may contribute to incongruent molecular phylogenies; in zoo populations of markhor, ~36% of those sequenced showed domestic goat mitochondrial haplotypes (Hammer et al. 2008), indicating the ease of interspecific gene flow in certain environments. Alternatively, incomplete lineage sorting or selective advantage of certain mitochondrial haplotypes may explain the phylogenetic inconsistencies.

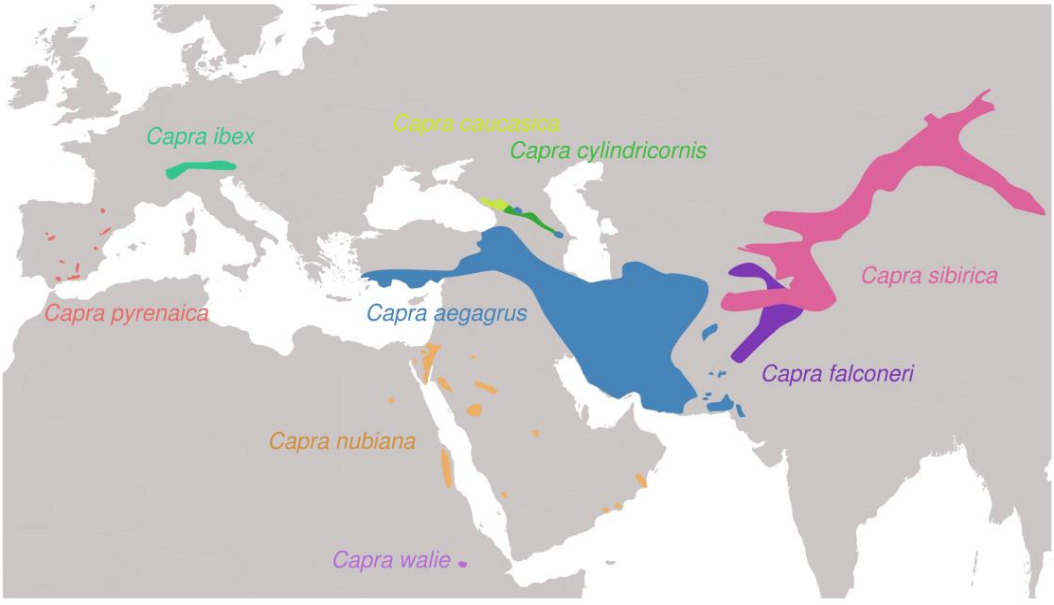


Figure 6.1: Distribution of wild caprid species. Approximate distributions taken from (Shackleton 1997), and inspired by (Pidancier et al. 2006).

The bezoar ibex, *Capra aegagrus*

The wild progenitor of the domestic goat, the bezoar ibex, is found throughout southwest Asia, from western Turkey to Pakistan, India, and Russia (Groves & Grubb 2011). The species is also found in the eastern Caucasus Mountains, where it tends to inhabit forests and not the higher altitude regions where the East Caucasian tur are found (Weinberg 2002). The range of the bezoar also overlaps with that of the markhor in Pakistan, but the two species tend to occupy distinct locales and do not directly compete (Schaller 1977). The horns of the bezoar have a characteristic scimitar or sickle shape, and are the largest of the genus relative to their body size. The habitat range of this species was likely wider in the late Pleistocene and early Holocene, when it extended along the Levantine coast and came in close contact with the Nubian ibex (Uerpmann 1987). Bezoar ibex appear to show less genetic diversity than domestic goat (Alberto et al. 2018), suggesting that decline of their range and the activities of humans have had an impact on their genetic diversity.

The Alpine ibex, *Capra ibex*

The Alpine ibex is currently found in the Alpine Mountains of Europe, and tends to inhabit high-altitude regions above the snowline (Shackleton 1997). The species was reduced to a single population of less than 100 individuals in the Italian Gran Paradiso massif by the early 19th century due to over-hunting; an extensive conservation programme has allowed their numbers to recover to over 40,000 (Stüwe & Nievergelt 1991). The consequences of this population bottleneck - 88 individuals comprised the initial captive breeding population - is observed in the patterns of genetic diversity of the reintroduced ibex populations. Secondary reintroductions have produced a stepwise gradient of decreased genetic diversity and increased substructure (Biebach & Keller 2009; Grossen et al. 2018). Reintroduced populations also have a higher frequency of large runs of homozygosity (>20 Mb), which are not observed in the present-day Gran Paradiso population. Alpine ibex have also experienced admixture with domestic goat, and is the likely source of one of the two alleles at major histocompatibility complex locus *DRB* exon 2 (Grossen et al. 2014).

Unlike *Capra aegagrus*, the horns of the alpine and other species designated ibex do not show a sharp anterior keel, but have a more prominent ridges along the horn length (Schaller 1977).

Figure 6.2: Members of the Capra genus. Top left, clockwise and then the centre: bezoar ibex, Walia ibex, East Caucasian tur, West Caucasian tur, markhor, Iberian ibex, Siberian ibex, Alpine ibex, Nubian ibex. Photos obtained from Wikimedia Commons under Creative Commons Licenses; license details, author attributions, and links to the original images are supplied in Appendix Table 6.1.

The Iberian ibex, *Capra pyrenaica*

The Iberian ibex, also known as the Spanish or Pyrenean ibex, is now found only in fragmented areas across Spain (Groves & Grubb 2011). The horns of the Iberian ibex are distinctly different from those of other ibex, curving outwards and sideways and then backwards and inwards, and show a posterior keen (Schaller 1977). Several subspecies were recognized but some have been hunted to extinction over the past centuries. The Portuguese population *Capra lusitanica* became extinct *c.* 1890, while the Pyrenean population (known as the bucardo) became extinct when the last remaining female died in January 2000 (Perez et al. 2002). The bucardo was also the first case of attempted de-extinction, but the newborn animal died soon after birth due to lung defects (Folch et al. 2009). The levels of genetic diversity in Iberian ibex are comparable to that of the present-day Gran Paradiso alpine ibex population, indicative of the population bottlenecks in the history of this ibex (Grossen et al. 2018).

As described in Chapter 1, a wild goat related to *Capra caucasica/cylindricornis* has been suggested to be the ancestor of the Iberian ibex, based on paleontological remains in the French Pyrenees (Crégut-Bonnoure 1991; Rivals 2006). However molecular evidence supports a monophyly of Alpine and Iberian, consistent across mitochondrial (Hassanin et al. 2012), Y chromosome (Pidancier et al. 2006), and nuclear (Grossen et al. 2018) data sets (Figure 6.3). Based on this, the likely origin of this species is the alpine ibex, which is thought to have inhabited the Pyrenees between the Riss and Würm glaciation periods (~130-115 kya) (Engländer 1986).

The Siberian Ibex, *Capra sibirica*

The Siberian or Asiatic ibex occupy a vast - if fragmented - distribution in central Asia, including the Hindu Kush mountains of Afghanistan, the Altai mountains, and the Sayan Mountains in southern Russia (Schaller 1977). In several regions their range overlaps with that of the markhor, but tend to inhabit higher altitudes above the forest-line during the winter during rutting season. The range of the species was likely greater in the Late Pleistocene, when its distribution may have overlapped with *Capra ibex* (Heptner et al. 1961). The species displays much variation in coat colour, body size, and horn shape across different populations, and has led to many sub-species designations (Fedosenko & Blank 2001).

Hybrids between *Capra sibirica* and domestic goat and both species of Caucasian tur have been reported, and in some cases were confirmed to be fertile (Gray 1954).

Genetic analyses of *Capra sibirica* have been limited and not produced consistent phylogenies (Figure 6.3). Trees constructed from mitochondrial data indicate the species to be the outgroup of all other *Capra* (Manceau et al. 1999; Ropiquet & Hassanin 2006; Pidancier et al. 2006), and also shows a close relationship with the Himalayan Tahr *Hemitragus jemlahicus* (Hassanin et al. 2012). This would be consistent with the genus evolution occurring in central Asia (Pilgrim 1947; Fedosenko & Blank 2001). However, a Y chromosome phylogeny places *Capra sibirica* as a clade within an overall "ibex" monophyly, sister to a bezoar and markhor clade (Pidancier et al. 2006). This led to a complex hybridization model being proposed, in with a Siberian ibex-related ancestor admixed with the ancestors of bezoar and markhor, resulting in the Nubian, European, and Caucasus ibex species. This hypothesis has not been yet tested with nuclear genome data.

The Nubian Ibex, *Capra nubiana*

The Nubian ibex is found in fragmented regions across northeast Africa, the Sinai and Arabian peninsulas, and in the southern Levant (Shackleton 1997). The northern limit of its habitat range may have been more southerly in the Epipaleolithic and Neolithic (Uerpmann 1987), and there is evidence that it overlapped with that of bezoar ibex as far south as Beidha, Jordan, which then had a more southerly range (Hecker 1976). The species inhabits rocky, mountainous parts of arid deserts, and there are thought to be just 2,500 mature individuals in the species (de Smet et al. 2008). The Nubian ibex has the smallest body size of the genus (Groves & Grubb 2011).

The phylogenetic relationship of *Capra nubiana* to other *Capra* species is unclear. Whole mitochondrial sequences place the species the outgroup to all other *Capra*, following their divergence from *Capra sibirica* (Hassanin et al. 2012). Y chromosome genetic data place the species in a clade within an "ibex" monophyly (Pidancier et al. 2006). A small number of autosomal loci suggest instead that the species is sister to the markhor *Capra falconeri* (Ropiquet & Hassanin 2006). No attempt has been made to address its phylogenetic position using whole genome data.

The Walia Ibex, *Capra walie*

The Walia ibex, also known as the Abyssinian ibex, is found only within restricted regions of the Simien Mountains of Ethiopia (Geberemedhin & Grubb 2008). The Walia ibex is the most endangered of the *Capra* genus; recent estimates suggest that the census population is growing, with a mean population in 2009-2011 of ~750 (Ejigu et al. 2017), from 200-250 individuals in 1998 (Nievergelt 1998). This species shares many morphological characteristics with the *Capra nubiana*, but is among the largest of the genus; it is considered to be a hypermorphic variant of the Nubian ibex but is commonly designated as a distinct species (Groves & Grubb 2011). Its habitat has a higher altitude, greater precipitation rate, and lower mean temperature than that of the Nubian ibex (Nievergelt 1981). In contrast to other *Capra* species, the Walia ibex breeds throughout the year rather than in specific rutting seasons, likely due to the stable year-long temperature of the Semien Mountains. Walia ibex are expected to group with Nubian ibex, based on (Gebremedhin et al. 2009).

An analysis of mtDNA, Y chromosome, and microsatellite variation in *Capra walie* established that the 32 sampled individuals were identical at the cyt *b* mtDNA locus, and formed a monophyly within *Capra nubiana* variation (Gebremedhin et al. 2009). Y chromosome loci showed *Capra walie* forming a clade with *Capra ibex* and *Capra cylindricornis* to the exclusion of *Capra aegagrus*, consistent with the ibex/bezoar divide reported elsewhere (Pidancier et al. 2006). Microsatellites also indicate extremely low levels of heterozygosity. Otherwise, no genetic studies have been performed on this species.

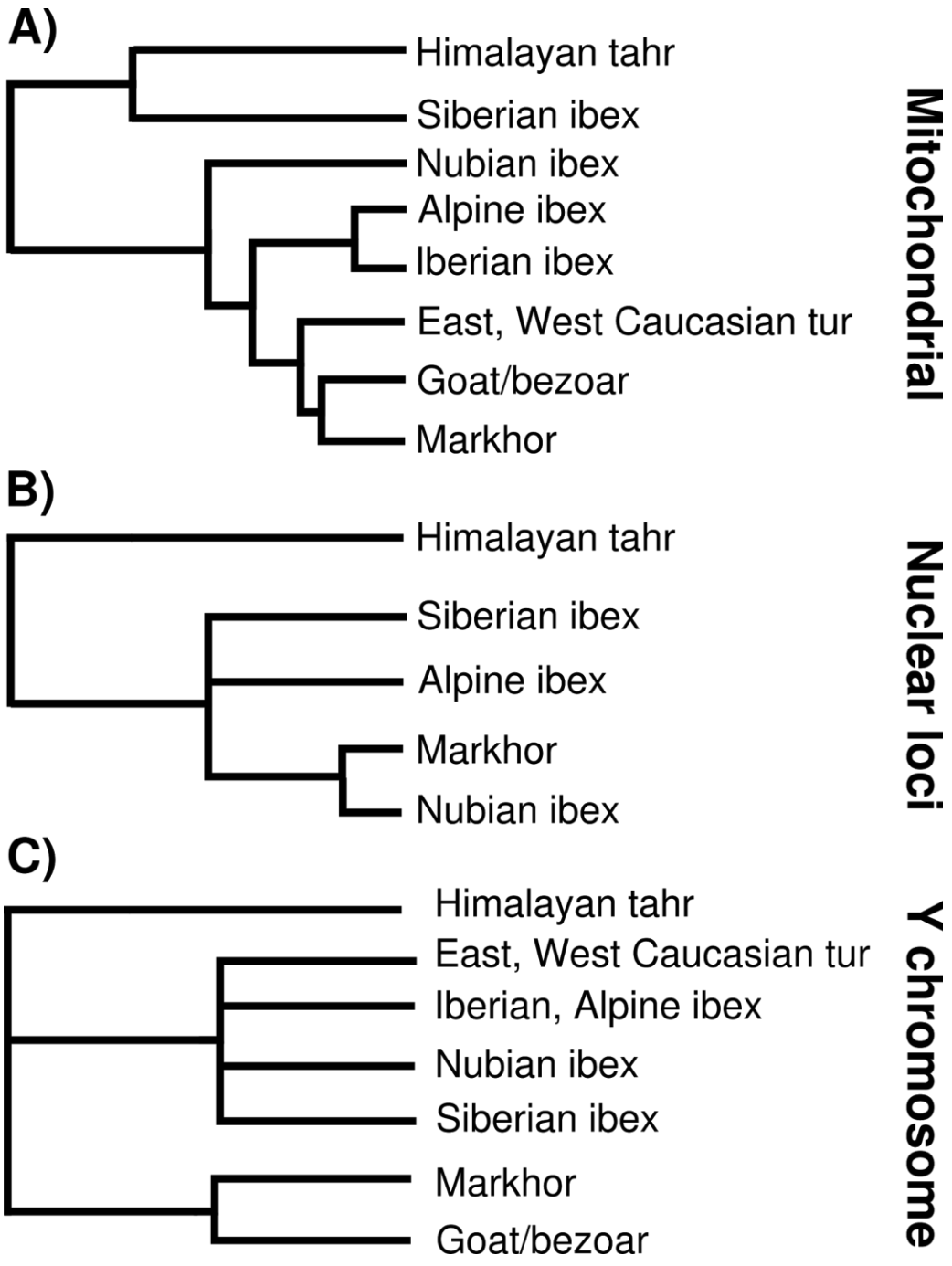


Figure 6.3 - Previous proposed *Capra* phylogenies from molecular data. A) Phylogeny constructed using mtDNA data, drawing from (Pidancier et al. 2006; Hassanin et al. 2012). B) Phylogeny constructed four nuclear loci, presented in (Ropiquet & Hassanin 2006). C) Phylogeny constructed using Y chromosome sequence, reported in (Pidancier et al. 2006).

The West Caucasian Tur, *Capra caucasica*

The West Caucasian tur (or ibex), and referred to as the Kuban tur, is found only in the western one third of the Caucasus Mountains (Groves & Grubb 2011). Their distribution partially overlaps with the East Caucasian tur on the southern slopes, and they occasionally hybridize (Heptner et al. 1961). The horns of males of this species are somewhat similar to those of the ibex, but show some similarities with the distinctive horns of the East Caucasian tur. Its population appears to be declining despite conservation efforts (Weinberg 2008).

Mitochondrial DNA and Y chromosomes suggest that the two Caucasian tur form a clade (Pidancier et al. 2006), although how that clade relates to other *Capra* species is not consistent among phylogenies and there appears to be evidence of gene flow with *Capra aegagrus*. Other mtDNA analysis are also inconclusive as to how the three species relate (Manceau et al. 1999; Kazanskaia et al. 2007).

The East Caucasian Tur, *Capra cylindricornis*

The East Caucasian or Daghestan tur is endemic to the eastern two thirds of the Greater Caucasus (Groves & Grubb 2011). The horns of males of this species are distinctive, curving out, backwards, and then forward, not unlike those observed in some sheep relatives (Schaller 1977). Their habitat partially overlaps with that of *Capra aegagrus* and *Capra caucasica*, and occasionally hybridize (Zalikhanov 1967). They will readily hybridize when artificially introduced into the same region (Heptner et al. 1961). No whole mitochondrial sequence is currently available for the Daghestan tur, but mtDNA fragments suggest a close relationship with *Capra caucasica*, and possible introgression from *Capra aegagrus* (Pidancier et al. 2006). Y chromosome data group the species with *Capra caucasica* in the "ibex" clade (Figure 6.3).

It has been suggested that *Capra caucasica* represent the descendants of a stable hybrid population between *Capra cylindricornis* and relatives of *Capra ibex* (Groves & Grubb 2011). Alternatively, a common ancestor of the species may have been partitioned into two populations by Pleistocene glaciers which spanned the central Caucasus and included both Mount Elbrus in the west and Mount Kazbegi in the east (Weinberg et al. 2010). Climate modelling suggests that two major refugia existed in the Greater Caucasus region during Last Glacial Maximum, which provides an ecological explanation for the distinct tur species.

The markhor, *Capra falconeri*

The markhor *Capra falconeri* is found in southern Central Asia: Afghanistan, south Uzbekistan, north Pakistan, north India, and Tajikistan (Geist 1990). Its habitat range is extremely fragmented, and it prefers lower altitudes than other *Capra* species (Schaller 1977). In some parts, it co-exists with *Capra aegagrus* or *Capra sibirica* (Shackleton 1997), but interactions are limited. The species has several designated subspecies. The markhor has the most distinctive horns of the wild caprids, which twist in a spiral fashion and are flattened (Groves & Grubb 2011). Its name likely derives from the Persian *mar*, meaning "snake", and *khor*, meaning "eating", although it is unclear how this relates to the animal itself (Huffman 2004).

Mitochondrial phylogenies place the markhor as a sister clade to *Capra aegagrus*/*hircus* (Hassanin et al. 2012), as do Y chromosome phylogenies (Pidancier et al. 2006) (Figure 6.3). Contradicting this, the available autosomal data instead shows the markhor having greater affinity with *Capra nubiana* than other ibex (Ropiquet & Hassanin 2006). Complicating phylogenetic reconstruction are reports of introgression of domestic goat mitochondria in markhor zoo populations (Hammer et al. 2008).

Aims

The aims of this chapter are:

1. To investigate the mitochondrial and nuclear genome phylogenies of the *Capra* genus.
2. To test if gene flow occurred between domestic goat *Capra hircus* and other *Capra* species.

Materials and Methods

Materials

Fourteen historic samples from the *Capra* genus were acquired from the Muséum National d'Histoire Naturelle (MNHN) of Paris, France, and are displayed Table 6.2. This set of samples included two of each the major *Capra* species, excluding domestic goat or the bezoar ibex: the markhor (*Capra falconeri*), the Alpine ibex (*Capra ibex*), the Iberian ibex (*Capra pyrenaica*), the Siberian ibex (*Capra sibirica*), the Nubian ibex (*Capra nubiana*). Additionally, the sample set included single representatives of the East Caucasian Tur (*Capra cylindricornis*), the West Caucasian Tur (*Capra caucasica*), the Walia ibex (*Capra walia*), and the domestic goat from the Levant (*Capra hircus*). The time of death of the *Capra* specimens were from the early 20th century BC to early 21st century BC. The skeletal elements obtained for each sample were a variety of non-petrosal materials, and included dental remains. An additional sample from Direkli Cave (Direkli4), Turkey, was included in nuclear genome alignment.

Table 6.2 - Muséum National d'Histoire Naturelle (MNHN) historic *Capra* samples.
*Thought to be a hybrid between an ibex of unknown origin and a Nubian ibex.

Sample	MNHN Identifier	Species	Skeletal Element
Bouc1	MNHN ZM AC 1906-579	*Capra hircus*	Half-crown and root
Falconeri1	MNHN ZM AC 2009-243	*Capra falconeri hepteneri*	Canine/incisor root
Falconeri2	MNHN ZM AC 1988-227	*Capra falconeri*	Canine/incisor root
Ibex1	MNHN ZM AC 1909-113	*Capra ibex*	Tooth root
Ibex2	MNHN ZM AC 1938-1296	*Capra ibex*	Tooth root
Nubiana1	MNHN ZM AC 2010-642	*Capra nubiana*	Molar root
Nubiana2	MNHN ZM AC 1985-45	*Capra nubiana (*Negev*)*	Sesamoid
Pyrenica1	MNHN ZM AC 1946-170	*Capra pyrenaica*	Sesamoid
Pyrenica2	MNHN ZM AC 1947-1	*Capra pyrenaica*	Tooth root
Sibirica1	MNHN ZM AC 1997-1776	*Capra sibirica**	Root and crown
Sibirica2	MNHN ZM AC 1974-98	*Capra sibirica*	Tooth root
Tur1	MNHN ZM AC 1982-1092	*Capra caucasica*	Lower P4 molar root
Tur2	MNHN ZM AC 1945-68	*Capra cylindricornis*	Tooth root
Walie1	MNHN ZM AC 1990-691	*Capra walie*	Hal- crown and root

Methods

Drilling, DNA extraction and NGS library construction

Sample drilling, DNA extraction, and NGS library construction was performed as described in Chapter 2. Uracil-DNA glycosylase/Endonuclease VIII treatment was not performed due to the relatively recent year-of-death of the samples.

Next Generation Sequencing

A multiplexed pool of DNA libraries was initially screened on a single lane of an Illumina HiSeq 2500 platform using single end, read length 1x100bp sequencing (Macrogen Inc., 1002, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 153-781, Republic of Korea). The pooling procedure described in Chapter 2 was followed. After initial endogenous content estimation, *Capra* samples were re-sequenced on eight HiSeq 2500 lanes.

Capra species data processing

Read quality and potential systematic bias was assessed using FastQC. Read trimming and length filtering was performed using cutadapt 1.1, removing reads less than 30bp in length (cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -O 1 -m 30).

Alignment to the improved goat assembly ARS1 (Bickhart et al. 2017) was performed using bwa aln (Li & Durbin 2009). Alignment settings were altered to allow a greater number of gaps and differences from the reference genome, and to disable seeding (Schubert et al. 2012). Alignment parameters were as follows: bwa aln -l 1024 -n 0.01 -o 2.

Bam files were produced and filtered using the same pipeline as Chapter 2, with a single change: a mapping quality filter of 20 was used. This change was made due to the increased number of differences expected between the reference genome and a diverged sample genome; as no other *Capra* reference genomes were available, it was necessary to use the *Capra hircus* reference genome for alignment. As each *Capra* species is expected to have accrued lineage-specific differences to *Capra aegagrus*/*hircus*, the average mapping quality of reads is expected to decrease when a non-*Capra aegagrus*/*hircus* sample is aligned to ARS1, due to a greater number of mismatches between sample and reference genome. By lowering the mapping quality filter used, the unnecessary removal of reads that have a lower

mapping quality due to the diverged reference genome is avoided. Direkli4 was included in the samples filtered this way (see Results), while *Capra hircus* sample Bouc1 reads was filtered by mapping quality of 30.

Sequencing data from the initial screen were combined with the subsequent sequencing using Picard (The Broad Institute 2018) prior to GATK indel realignment. Damage patterns were assessed using mapDamage2 (Jónsson et al. 2013). Bam files were soft-clipped by 4bp at either end of each read. Mean genome coverage for each individual was calculated using the GATK DepthOfCoverage tool (McKenna et al. 2010). Endogenous content was estimated by dividing the number of reads aligned following mapping quality filtering and removal duplicates, divided by the number of reads remaining after cutadapt trimming.

Ancestral genome sequences

To generate ancestral sequences, reads from the yak reference genomes (Appendix Table 4.1) were aligned to ARS1 using the bwa mem pipeline described in Chapter 4, and the bam file filtering settings described for *Capra* species above; in this case, a mapping quality filter of 20 was used. The mean genomic coverage was estimated using GATK's DepthOfCoverage function.

ANGSD was used to generated ancestral fasta sequences for yak, using the following settings: angsd -C 50 -doFasta 2 -doCounts 1 -minQ 20 -minMapQ 20 -setMinDepth X -setMaxDepth Y, where X and Y were half and twice the mean depth per genome respectively.

Realignment to ARS1

The domestic and wild ancient goat samples for which whole genome sequencing was described in Chapter 2 were realigned to the ARS1 assembly. Alignment settings and bam file processing steps were used identical to those described above, except a mapping quality filter of 30 was used. Soft-clipping and genome depth was performed as described.

Modern goat samples listed in Appendix Table 4.1 were also realigned to ARS1, using the pipeline described in Chapter 4 (mapping quality filter of 30). In addition, five modern Italian goat (Alberto et al. 2018) and twenty modern Cashmere goat from China (Li et al. 2017)

were aligned using the same procedure (Appendix Table 6.2). Mean genomic coverage was estimated using GATK.

Molecular sex assignment

Molecular sex was determined using the approach described in Chapter 4, exploring the ratio of reads aligning to each chromosome and the length of each chromosome. As there are no single X chromosome assembled molecule for ARS1, the reads aligned to the two unlocalized X chromosome genomic scaffolds (NW_017189516.1 and NW_017189517.1) were combined.

Mitochondrial alignment and sequence generation

The mitochondrial alignment-realignment approach described in Chapter 2 was used in alignment of *Capra* reads to an appropriate reference genome. Samples were initially aligned to a circularized version of the *Capra hircus* mitochondrial reference NC_005044.2 using bwa aln -l 1024. ANGSD (Korneliussen et al. 2014) was used to generate fasta sequences; these which were then de-circularized, aligned together with a set of reference mitochondria (Appendix Table 6.3) using MUSCLE (Edgar 2004), and used to produce an initial maximum likelihood phylogeny with phyML (Guindon et al. 2010). Fasta generation for poorer quality samples (*i.e.* Nubiana2 and Pyrenica1, see results) was performed using all sites regardless of coverage, while the remaining samples were subject to a per-site minimum coverage of 3 reads. Samples were then realigned to a more appropriate reference sequence (Appendix Table 6.4). A final maximum likelihood phylogeny was constructed using phyML, using a rate model estimated by modelgenerator (Keane et al. 2006) (6 gamma rate categories, Bayesian Information Criteria), performing 100 bootstraps. A set of *Caprini* mitochondrion were included (Appendix Table 6.3), and the yak mtDNA reference was used as an outgroup.

Identity-By-State nuclear genome phylogeny

To determine the nuclear genome phylogeny of the *Capra* genus, an identity-by-state matrix was computed using ANGSD, using the *Capra* genomes generated in this chapter, filtering out genomes of <0.01× mean coverage. A subset of other modern and ancient genomes were included in the IBS matrix generation, which are displayed in Appendix Table 6.5. IBS calculation was restricted to the autosomes. The beginning and end of all reads were trimmed

by 4bp. Sites included must have been present in half of the individuals (47/2 = 24, rounding up). Transitions were ignored, to avoid ancient samples having inflated IBS with one another. The command used was: angsd -doIBS 1 -doCounts 1 -doMajorMinor 5 -doCov 1 -minQ 20 -minMapQ 20 -minFreq 0.05 -GL 1 -makeMatrix 1 -maxMis 24 -trim 4 -rmTrans 1.

D statistics

D statistics were computed using ANGSD (Korneliussen et al. 2014), removing transitions and trimming reads by four base pairs. Only *Capra* samples with mean genomic coverage > 0.01× were included. For each position and sample, a random base was sampled. Yak was used to determine the ancestral state. Analyses were restricted to the autosomes. The command used was as follows: angsd -doAbbaBaba 1-trim 4 -doCounts 1 -minQ 20 -minMapQ 20. A |Z| score of 3 was taken to be significant.

Results

Next Generation Sequencing and endogenous DNA estimation

Sequencing was performed on 14 historic *Capra* specimens, obtained from the Muséum National d'Histoire Naturelle, Paris. A summary of sequencing and alignment results is displayed in Table 6.3; alignment metrics for the historic *Capra* specimen are presented in Appendix Table 6.6. Endogenous content values varied from 0.13% to 59.62%. Two samples, Nubiana2 and Pyrenica1, produced poor quality libraries with a low number of sequenced reads. The mitochondrial sequences of these samples were generated but otherwise were excluded from subsequent analyses.

Table 6.3 - Historic *Capra* genome summary. Nuclear and mtDNA genome coverage are presented as mean ×-fold coverage. Endogenous content is estimated as the number of reads aligning to ARS1 after mapping quality (mapQ) 20 filter, as a percentage of the total reads following adaptor trimming. For Bouc1, a mapQ filter of 30 was used due to it being a domestic goat/*Capra hircus*.

Sample	Morphological Species	Endogenous Content (%)	Nuclear Coverage	mtDNA Coverage	Sex
Bouc1	*Capra hircus*	56.44	1.19	150.76	M
Falconeri1	*Capra falconeri hepteneri*	11.18	0.51	29.03	M
Falconeri2	*Capra falconeri*	3.59	0.05	45.78	M
Ibex1	*Capra ibex*	62.09	3.53	179.03	F
Ibex2	*Capra ibex*	18.44	0.05	21.40	M
Nubiana1	*Capra nubiana*	24.73	1.10	211.32	M
Nubiana2	*Capra nubiana* (Negev)	29.06	0.001	0.18	F
Pyrenica1	*Capra pyrenaica*	0.13	0.0002	0.16	M
Pyrenica2	*Capra pyrenaica*	9.36	0.14	51.20	M
Sibirica1	*Capra sibirica*	11.79	0.03	163.16	M
Sibirica2	*Capra sibirica*	41.6	1.34	205.78	F
Tur1	*Capra caucasica*	57.19	2.48	59.22	M
Tur2	*Capra cylindricornis*	9.08	0.02	4.96	F
Walie1	*Capra walie*	22.87	0.66	103.12	M

Molecular sex for each sample was estimated using the ratio of reads aligning to autosome and the two ARS1 scaffolds assigned to the X chromosome, and their length. Sample sex is displayed in Table 6.3.

Realignment to ARS1

Realignment to the updated goat reference genome was performed using bwa aln and relaxed parameters. Alignment statistics for the ancient genomes described in Chapter 2 displayed in Appendix Table 6.7; coverage information for the modern samples introduced in Chapter 4 are displayed in Appendix Table 6.8. In general, mean ×-fold genomic coverage decreased for all samples relative to their CHIR_1.0 genomic coverage. This is likely due to the longer genome length and better resolution of repeat-rich or constitutive heterochromatic regions (Bickhart et al. 2017).

Realignment of samples included Direkli4, an Epipaleolithic goat from the Taurus Mountains. A mapping quality filter of 20 was used, similar to the filter used for non-*Capra aegagrus*/*hircus* samples (see below). Mean ×-fold genomic coverage for this sample was 0.5.

Molecular damage assessment

Damage patterns and fragment length distributions of the MNHN *Capra* genomes were assessed using mapDamage2.0 (Jónsson et al. 2013). 5'C>T and 3'G>A error as a result of deamination was observed in all genomes (Figure 6.4). However, error rates were comparable to the rates of USER-treated libraries presented in Chapter 2 (Figure 2.6). Fragment length distributions showed a greater degree of periodicity (Figure 6.5) than those generated in Chapter 2 (Figure 2.7). In addition, a greater proportion of reads were of the maximum read length (~100bp), indicative of the shorter period of molecular degradation experienced by these historic samples.

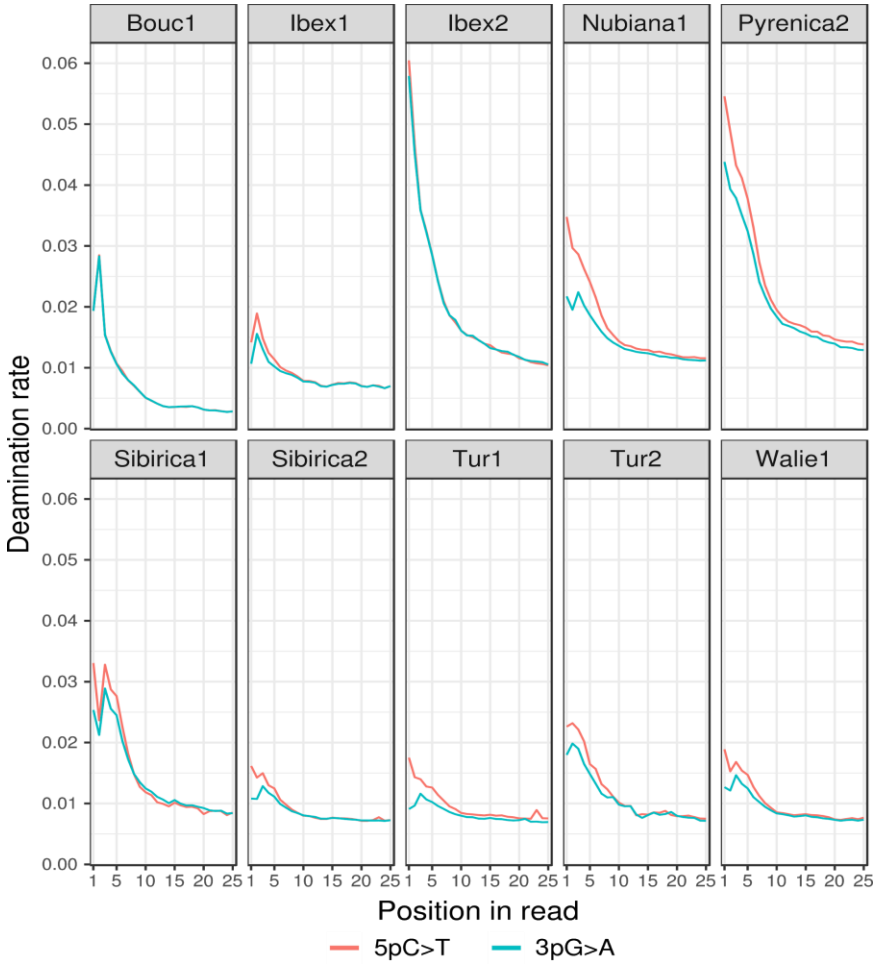


Figure 6.4 - Damage patterns of MNHN *Capra* samples, relative to distance to end of read. Only non USER-treated libraries were constructed for these historic samples.

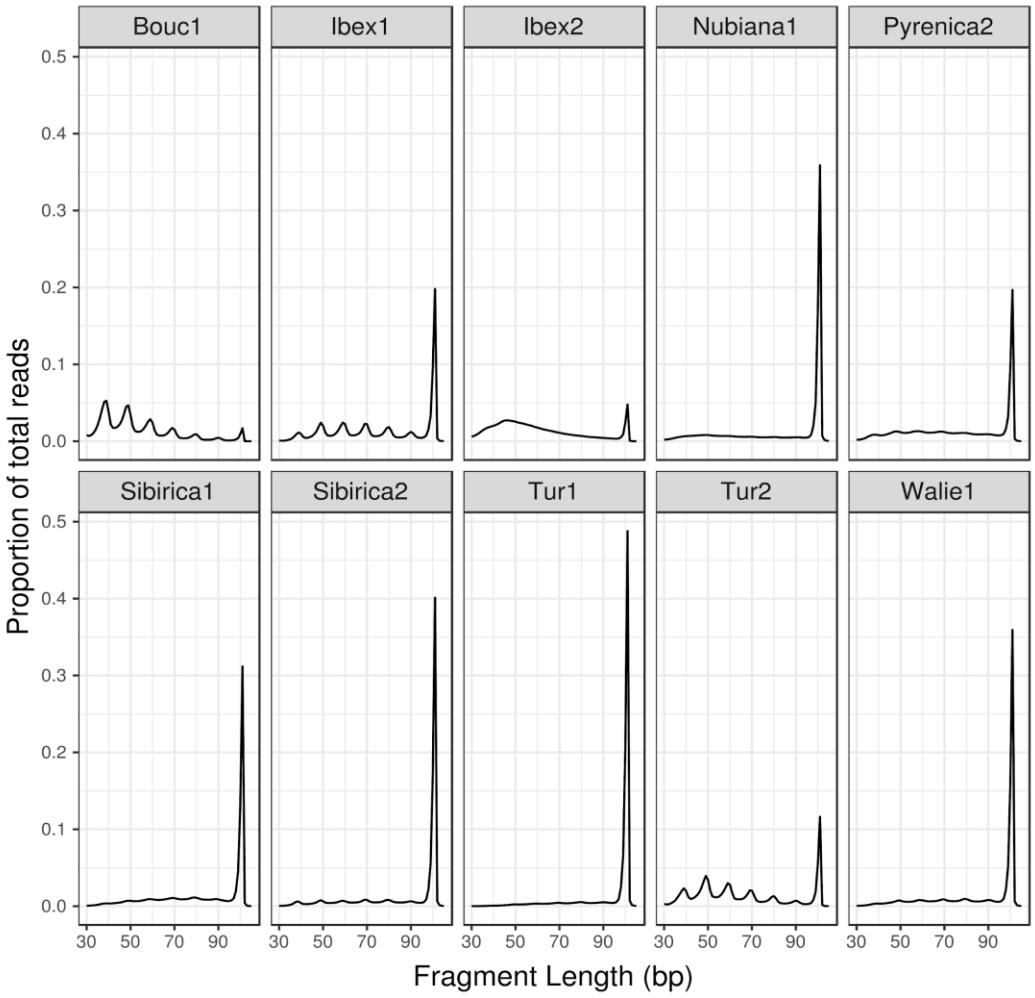


Figure 6.5 - Fragment length distribution of MNHN *Capra* samples. Read count is expressed as a proportion of the total on a per-sample basis. Poor quality genomes (Nubiana2 and Pyrenica1) were excluded.

Mitochondrial alignment and sequence generation

Mitochondrial alignment was performed for all *Capra* samples to the *Capra hircus* mitochondrial reference. Following an initial phylogenetic reconstruction, samples were realigned to a closer reference sequence in order to maximize read alignment. Mitochondrial alignment statistics are displayed in Appendix Table 6.9; sample mtDNA coverage is displayed in Table 6.3.

A maximum likelihood phylogeny was constructed using a set of ancient and modern *Capra* and other *Caprini* mitochondrial sequences and rooted on Yak, and is displayed in Figure 6.6. Bootstrap support for the most of nodes was high, 0.9-1.0 for the majority of nodes. Falconeri1 showed high divergence from other *Capra* mitochondria, and formed a clade with the Barbary Sheep *Ammotragus lervia*, in a sister clade to the Arabian tahr *Arabitragus jayakari*. This clade was sister to all other *Capra* mitochondrion, and also other *Caprini* tribe members: the Bharal *Pseudois nayaur*, and the Himalayan Tahr *Hemitragus jemlahicus*. The later grouped with the reference Siberian Ibex mitochondrion, which did not form a clade with either of the *C. sibirica* mitochondria sequenced here. The other historic *Capra falconeri* sample (Falconeri2) did not group with Falconeri, nor did the markhor reference mitochondrion. The Siberian ibex reference sequence was also divergent from other *Capra* mitochondrion; it did not form a clade with either Sibirica1 or Sibirica2 sequenced here.

Within the *Capra* genus mitochondrial phylogeny, the first node splits the remaining *Capra* mitochondria from those of the Nubian ibex, the Walia ibex, and also the Siberian ibex Sibirica1. The low coverage sample Nubiana2 rooted this clade, but this position likely reflects the low number of called basis for the sample. Otherwise, this clade was poorly differentiated with shallow branching. The next major node splits Nubian ibex (reference mitochondrion, Nubiana1, Nubiana2), the Walia ibex Walie1, and the Siberian ibex Sibirica1 (a likely hybrid, see Table 6.2) from other *Capra*. The next major clade splits European ibex from the remaining *Capra* mitochondria. The two *Capra pyrenaica* samples formed a sister clade to all Alpine ibex sequences, while the two *Capra ibex* historic samples show high similarity with each other, compared to the *Capra ibex* reference mitochondrion.

Three of the four Direkli Cave sequences and all historic Tur form a clade at the next node of the Capra mtDNA phylogeny. Within this clade, Direkli sequences are monophyletic and a sister group to the *Capra caucasica/cylindricornis* mtDNA. The West Caucasian Tur (*Capra*

caucasica) mitochondrial reference groups with Tur1, which was morphologically identified as also being *Capra caucasica*. Tur2, identified as being an East Caucasian Tur (*Capra cylindricornis*), is a sister branch to this clade. All nodes in these clades are high confidence, with bootstrap values of either 99 or 100.

A final node separates *Capra aegagrus* and *Capra hircus* from the remaining mtDNA sequences. This clade consists of a monophyly of two markhor sequences (the markhor reference and Falconeri2), and a *Capra sibirica* historic sample Sibirica2. The final mtDNA sequence reported here, the domestic goat Bouc1 from present-day Levant, falls within the G haplogroup.

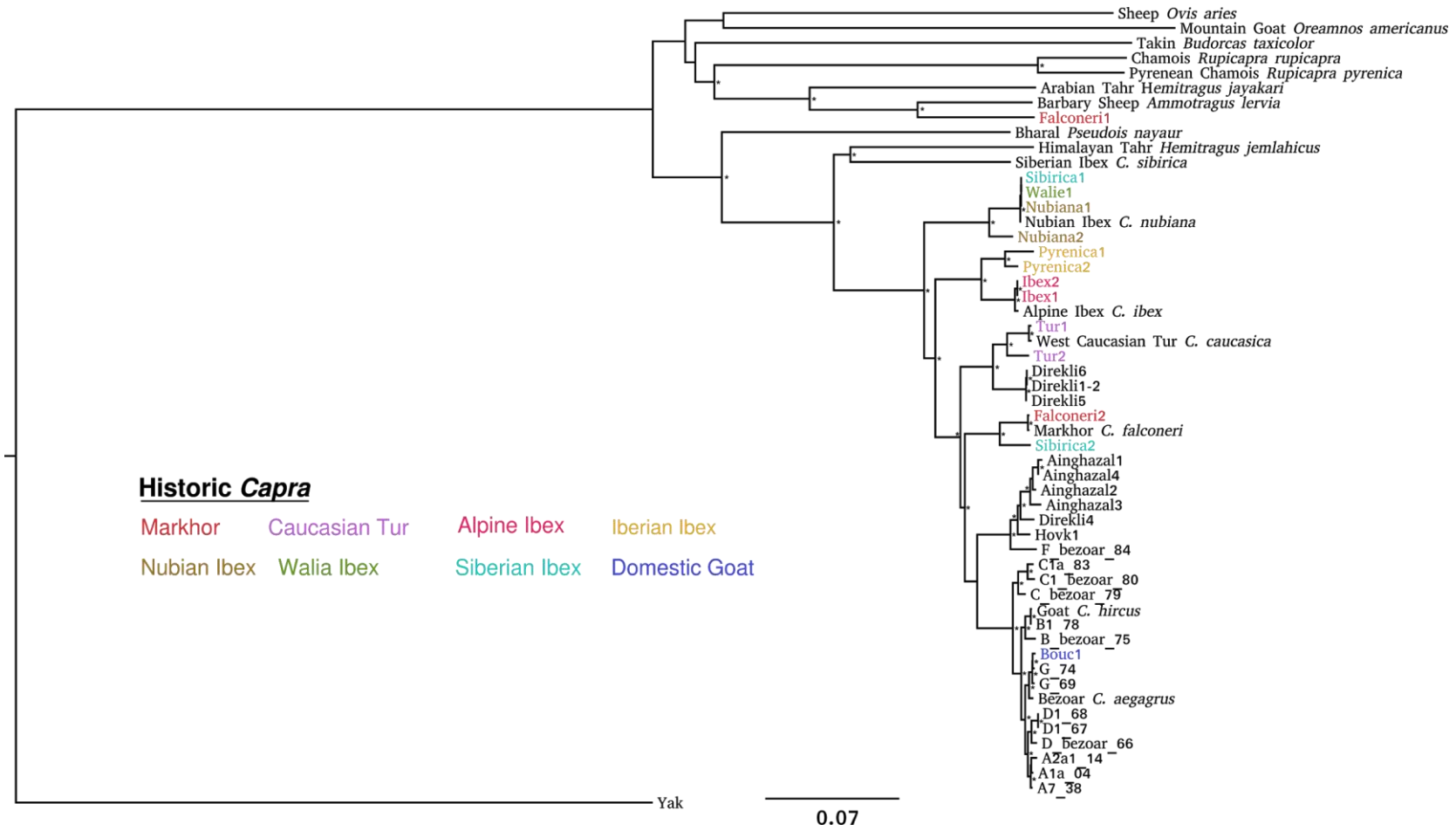


Figure 6.6 - Maximum likelihood tree of historic *Capra* mitochondria, with reference modern and ancient sequences, with 100 bootstraps.

High bootstrap confidence nodes (>0.6) are indicated by asterisks. Reference sequences indicated by binomial name. Rooted on Yak.

Identity-By-State nuclear genome phylogeny

A nuclear genome phylogeny was constructed by applying the neighbour-joining approach to an Identity-By-State matrix, computed by ANGSD and using only transversion. Yak was used to root the tree.

Similar to the mtDNA tree, the markhor Falconeri1 was an outgroup to all other *Capra* genomes analysis (Figure 6.7). A major node divided the remaining genomes into two partitions. The first consisted of ancient and modern domestic goat, wild bezoar, and the remaining markhor Falconeri2. Bouc1, the modern domestic goat from the Levant, showed affinity with modern and ancient Iranian goat genomes, similar to a Bronze Age Israeli goat Yoqneam2.

The second group of samples consisted of "ibex" species: the Nubian, Walia, Alpine, Siberian, Iberian, and Caucasian ibex. For convenience, this clade will be referred to below as the Ibex clade. The Siberian ibex Sibirica2 was the outgroup to all other samples in this clade, despite having a mtDNA genome similar to markhor (Figure 6.6). The next node divided European and Caucasian ibex from the African ibex (Nubiana1 and Walie1) and Sibirica1, which thought to be Siberian ibex/unknown ibex hybrid; this strongly suggests the unknown ibex parent of Sibirica1 is a Nubian ibex. European ibex formed a clade, with the Iberian ibex Pyrenica2 being sister to both alpine ibex genomes. Surprisingly, the Caucasian clade also contained a genome (Direkli4) from the Epipaleolithic site of Direkli Cave in southern Turkey. This sample was a sister to the tur clade, suggesting that it shared an equal degree of alleles with both genomes.

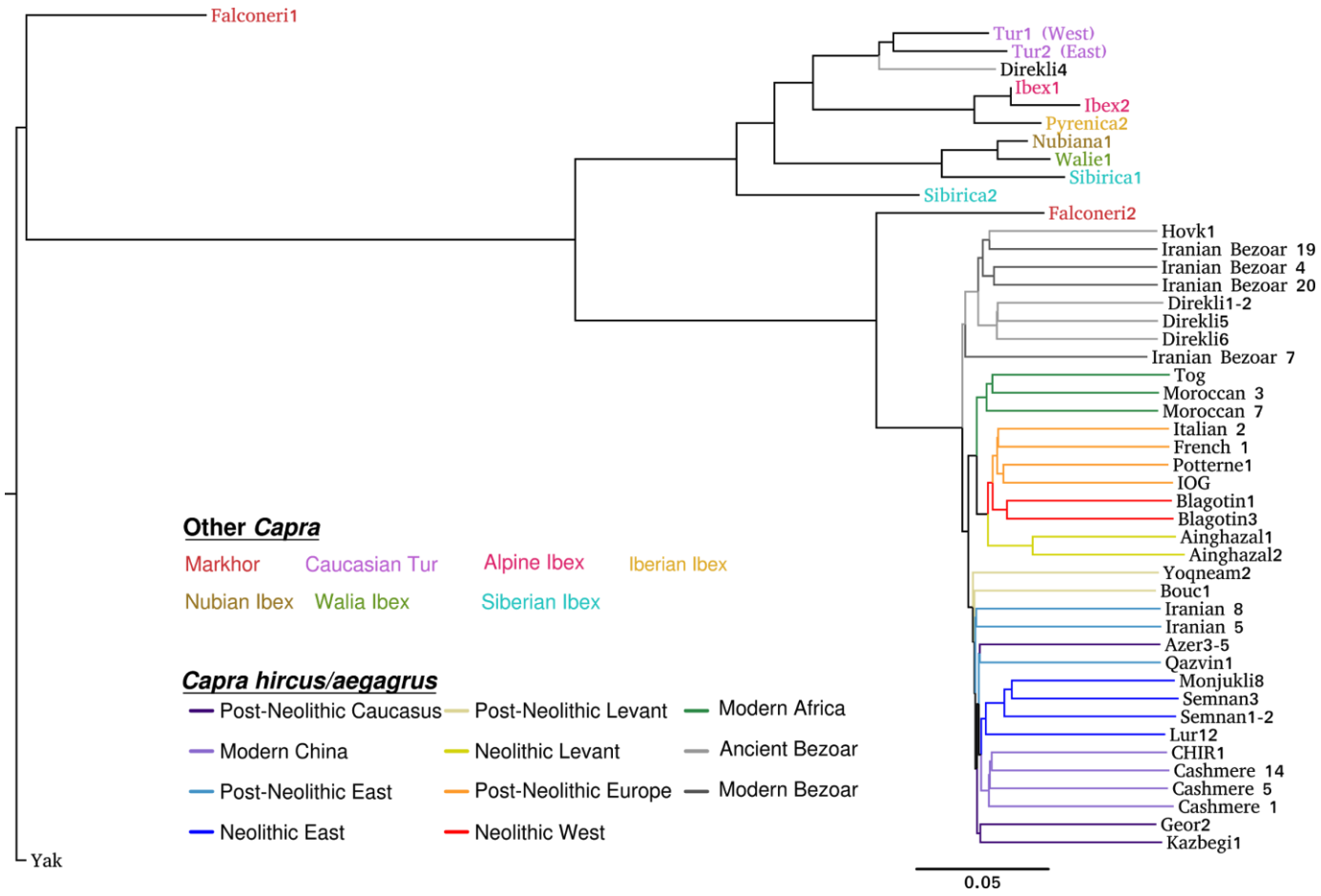


Figure 6.7 - Neighbour-Joining tree of IBS matrix of ancient and modern goat/bezoar, and historic *Capra* samples. Post-Neolithic refers to any sample from the Chalcolithic to present-day. Branches are coloured according to sample context. *Capra* species are differentiated by label colour. Yak is used as an outgroup and to root the tree.

D statistics

D statistic results were computed to address three interconnected questions:

1. The phylogeny of *Capra*.
2. If admixture has occurred from domestic goat to non-bezoar *Capra* species.
3. If admixture has occurred from *Capra* species to domestic goat.

How do different Capra *species relate to each other?*

To investigate the phylogenetic relationship of *Capra*, the statistic *D*(H1, H2, Test, Yak) was computed, to test if the branching pattern observed in the IBS-nj tree (Figure 6.7) was consistent with ANGSD computed derived allele sharing. As Falconeri1 was placed as the outgroup to all other *Capra* the *D*(H1, H2, Falconeri1, Yak) was computed, with the expectation that if Falconeri1 was a true outgroup, no |Z| scores greater than 3 would be obtained. This was not the case; the majority of tests gave significant results (Appendix Figure 6.1). A high degree of derived allele sharing was observed between Falconeri1 and Alpine or Walia ibex, while a low degree was observed with other historic markhor Falconeri2.

Similar logic was used to test the position of Sibirica2 in the IBS-nj tree. Sibirica2 is expected to share a roughly-equal number of derived alleles with pairs of ibex genomes, if it diverged earliest among ibex species. This was not the case, with Sibirica2 consistently sharing less derived alleles with the likely-hybrid Sibirica1 and Pyrenica2, and more with the Alpine and Caucasian ibex genomes (Appendix Figure 6.2). It is expected that greater derived allele sharing would be observed between Sibirica2 and any other ibex genome compared to between Sibirica2 and domestic goat/bezoar, which is indeed observed (Appendix Figure 6.3). Finally, Sibirica2 shared less derived alleles with Falconeri1 than any other genome, supporting the basal position of Falconeri1 in the phylogeny (Appendix Figure 6.4).

The clade of Sibirica1, Nubiana1, and Walie1 was similarly examined, first to address whether each member was equally related to the remaining ibex genomes. All three share less derived alleles with the Iberian ibex Pyrenica2 than with other ibex genomes (Appendix Figure 6.5), suggesting an excess of ancestral or domestic alleles in Pyrenica2. The three appear to share equal amounts of derived alleles with Alpine and Caucasian ibex genomes,

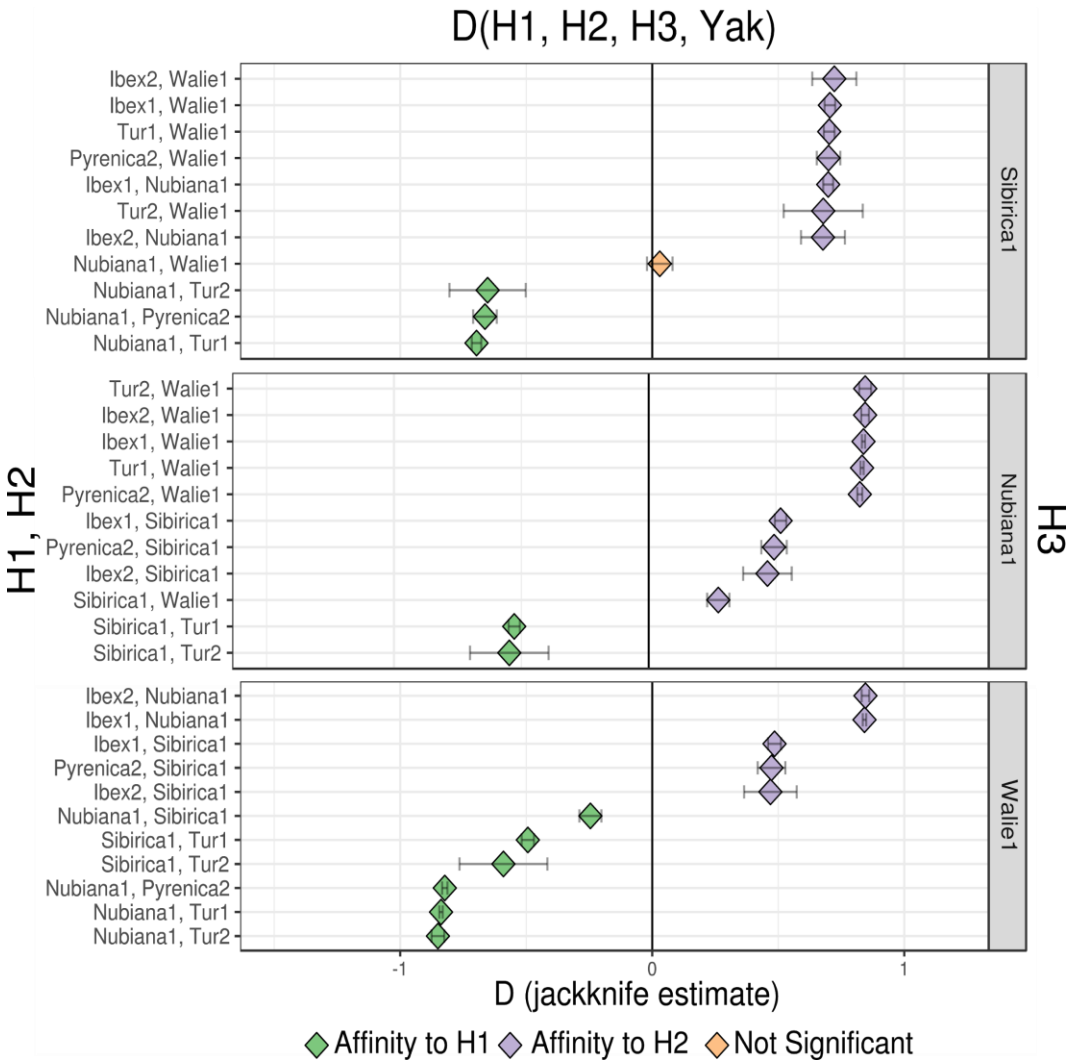


Figure 6.8 - *D* statistic test for relative affinity (derived allele sharing) of the test genomes Sibirica1, Nubiana1, and Walie1 between themselves and other ibex genomes, H1 and H2. Colour indicates test affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

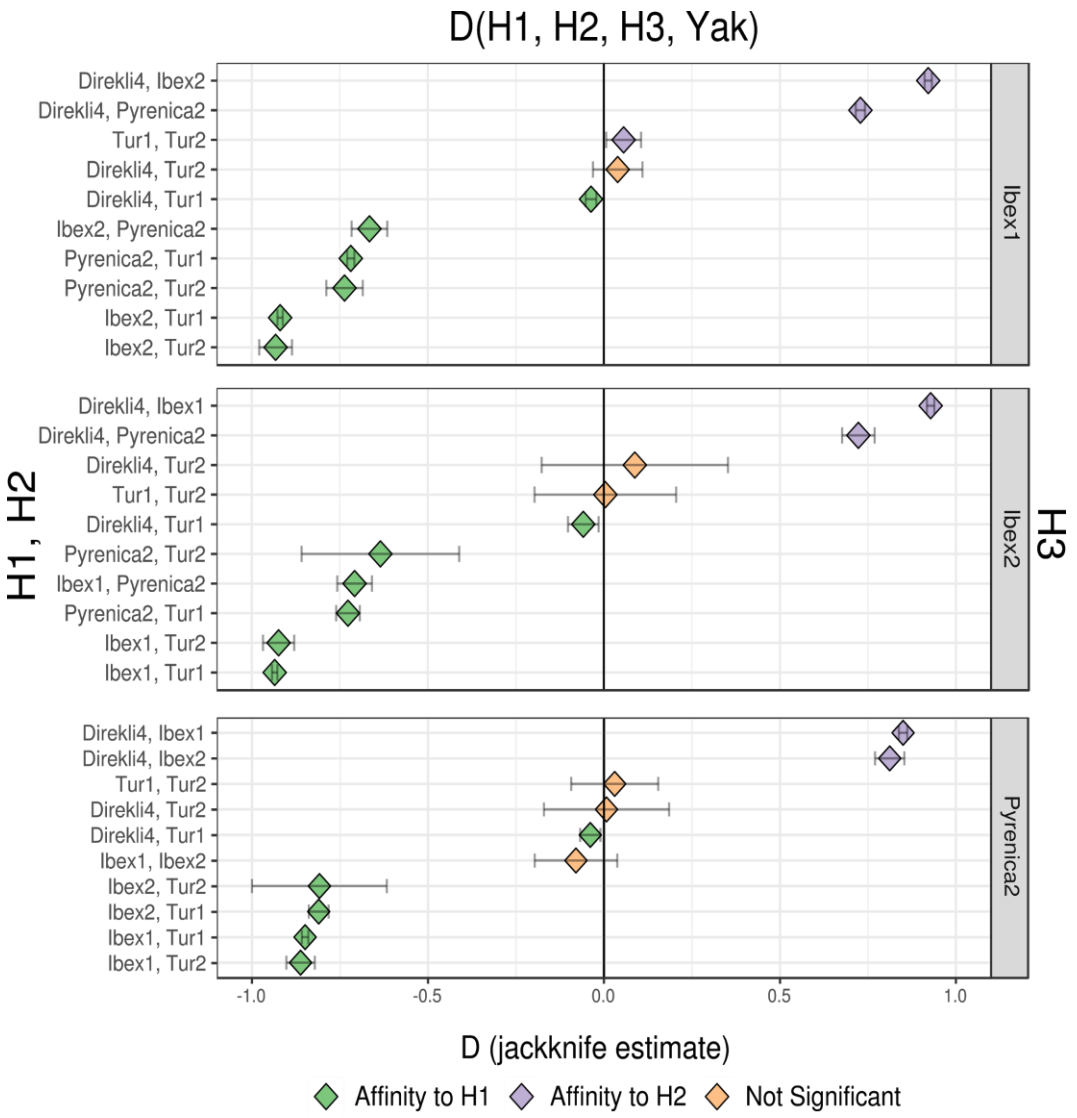


Figure 6.9 - *D* statistic test for relative affinity (derived allele sharing) of the test genomes Ibex1, Ibex2, and Pyrenica2 between themselves and other ibex genomes, H1 and H2. Colour indicates test affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange. Here, Direkli4 is included in tests.

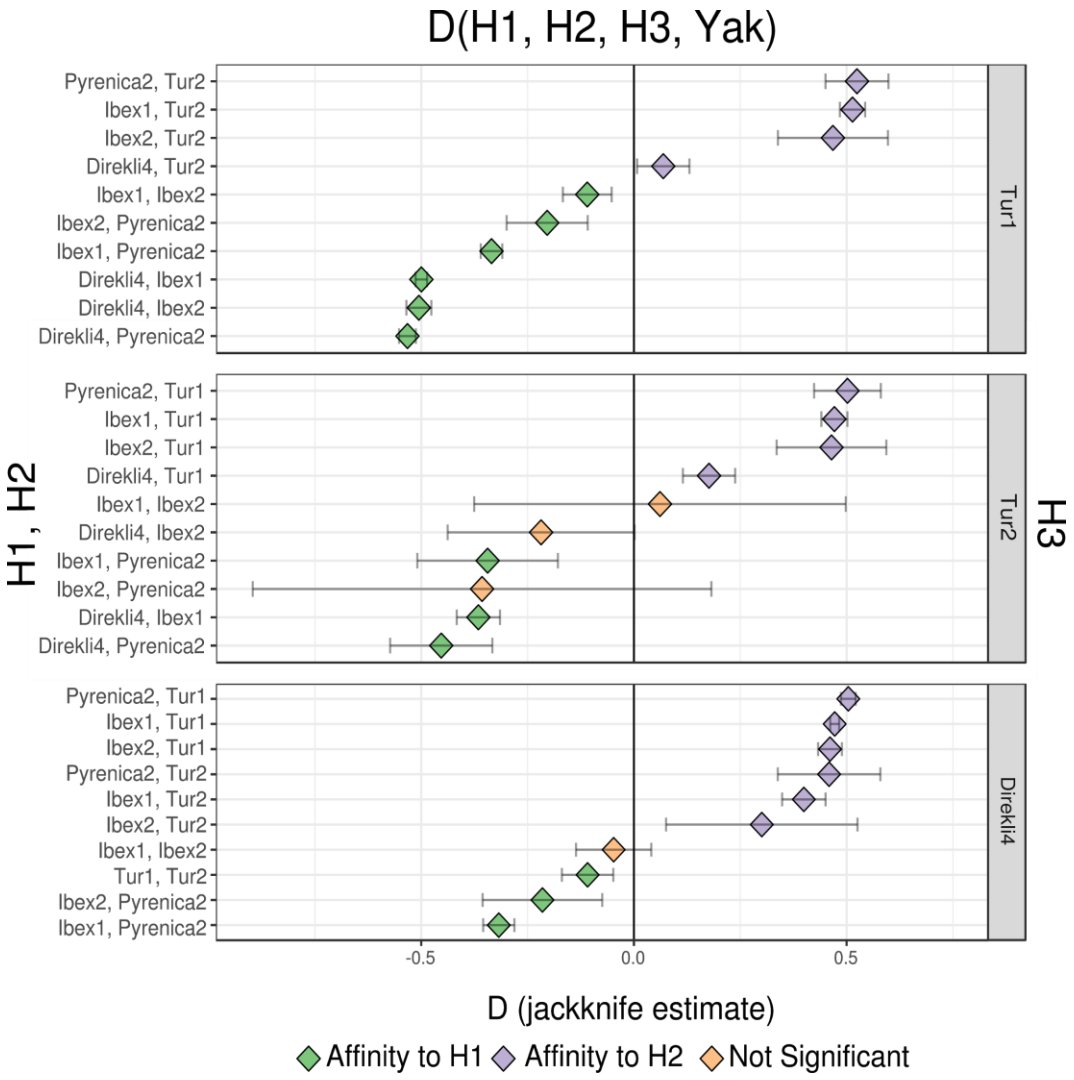


Figure 6.10 - *D* statistic test for relative affinity (derived allele sharing) of the test genomes Tur1, Tur2, and Direkli4 between themselves and other ibex genomes, H1 and H2. Colour indicates test affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange. Here, Direkli4 is included in tests.

although the test *D*(Ibex1, Tur1, Test, Yak) gave significant negative results across all three test genomes, suggesting a slightly greater degree of allele sharing with Ibex1 than with Tur1. Finally, the relationship between these genomes was examined (Figure 6.8), supporting the placement of Sibirica1 as an outgroup to Walie1 and Nubiana1 with the non-significant *D*(Nubiana1, Walie1, Sibirica1, Yak). In all cases, members of this trio showed greater derived allele sharing between each other than with other ibex genomes, supporting a close genetic relationship. The Alpine and Iberian ibex genomes roughly show the same degree of

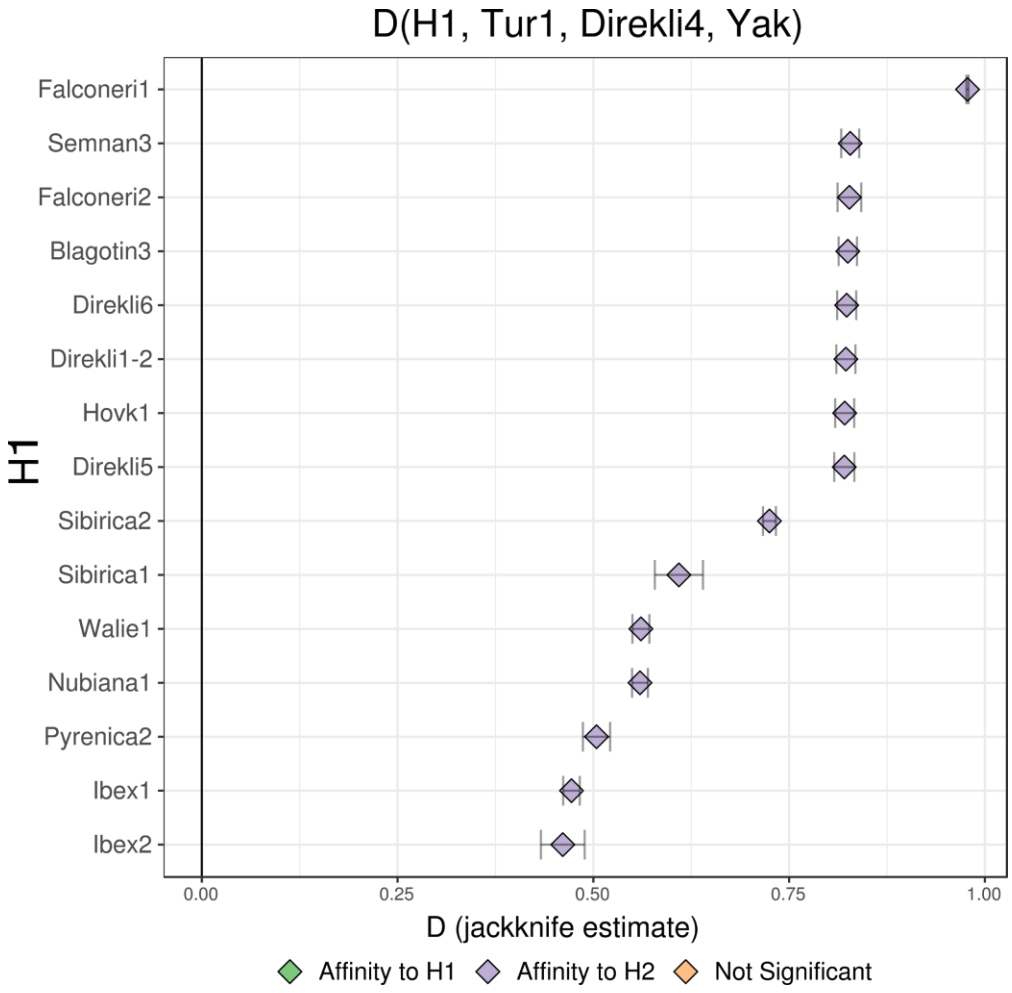


Figure 6.11 - *D* statistic test for relative affinity (derived allele sharing) of Direkli4 with *Capra* genomes relative to Tur1. Colour indicates Direkli4 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

derived allele sharing with the Tur/Direkli4 clade (Figure 6.9). Slightly negative significant results for the test *D*(Direkli4, Tur1, Ibex1/Ibex2/Pyrenica2, Yak) indicates either an excess of derived allele sharing with this clade and Direkli4, or an excess of ancestral or domestic alleles in Tur1. All members of this European ibex clade share more derived alleles among each other than with any of the Tur/Direkli4 clade. The Iberian ibex Pyrenica2 is symmetrically related to both alpine ibex genomes, based on the nonsignificant *D*(Ibex1, Ibex2, Pyrenica2,Yak) result. Consistent with this is greater derived allele between alpine ibex genomes than with the Iberian ibex, based on negative *D*(Ibex2, Pyrenica2, Ibex1, Yak) statistics.

Members of the Tur/Direkli4 clade do not form perfect outgroups with the European ibex clade. All appear to show the hierarchy of derived allele sharing, of the order Ibex1, Ibex2, and Pyrenica2 (Figure 6.10); this observation is limited by the high standard errors of tests involving Tur2, and suggests non-equal amounts of derived or ancestral/domestic alleles in these ibex genomes. The West Caucasian tur Tur1 shares a greater degree of alleles with Tur2 than with Direkli4, and vice versa; *D*(Direkli4, Tur1, Tur2, Yak) produces significant positive results. When Direkli4 is set as the outgroup to both Tur1 and Tur2, non-significant results are obtained, suggesting that the Epipaleolithic sample is indeed equally related to both Caucasian tur genomes. To test if this ancient sample from the Taurus Mountains does share more ancestry with the Caucasian Tur than other *Capra* species, the statistic *D*(H1, Tur1, Direkli4, Yak) was computed, varying H1 between genomes of different *Capra* species. Large positive *D* values were obtained for the range of genomes tested (Figure 6.11), indicating that Direkli4 shares a large degree of derived alleles with the lineage now represented by Caucasian tur species.

The IBS-nj tree placed the markhor Falconeri2 as an outgroup to domestic and wild goat, in a clade sister to other *Capra*. This matches patterns of derived allele sharing, with both test statistics *D*(Falconeri1, H2, Falconeri2, Yak) and *D*(Semnan3, H2, Falconeri2, Yak) indicating greater sharing between Falconeri2 and domestic or wild goat compared to other *Capra* species (Appendix Figures 6.6 and 6.7). The *D*(H1, H2, Falconeri2, Yak), where H1 and H2 are other *Capra* genomes, indicates asymmetric affinity of Falconeri2 with different *Capra* (Appendix Figure 6.8). Although affected by the low coverage of Falconeri2 (0.05×), *D* statistics indicate a consistent lower amount of derived allele sharing with either Siberian ibex genome, and to a lesser extent with the Nubian and Walia ibex.

Has admixture occurred from domestic goat to other Capra *species?*

To investigate if domestic goat populations have interbred with other *Capra* species, sets of *D* statistics were computed. Initially, the test *D*(Sibirica2, Test, Domestic, Yak) was computed, assuming that Sibirica2 was a reasonable outgroup to all other non-goat/bezoar *Capra* genomes. This assumption may is not entirely valid based on other *D* statistics (see above and Appendix Figure 6.1), but these results may be the consequence of domestic admixture creating asymmetric derived allele sharing among other *Capra* genomes. The test *D*(Sibirica2, Test, Domestic/Wild Goat, Yak) gave significant positive scores for all tests, except with Sibirica1 was used as H1 (Figure 6.12). The bulk of these results suggests that all other *Capra* genomes sequenced here (excluding the outgroup Falconeri1, the goat-related Falconeri2, and Sibirica1) have wild bezoar or domestic goat ancestry, in excess of what is observed in Sibirica2. The validity of this results depends on the assignment of Sibirica2 as an outgroup to the "ibex" clade from the IBS-nj phylogeny (Figure 6.7), which is imperfect for the reasons discussed above. The deficit in "domestic" derived alleles in Sibirica1 was also unexpected, and suggests that the ancestry of this hybrid individual is complex.

As an alternative means to investigate admixture from domestic or wild goat into other *Capra* species, the test *D*(H1, H2, Domestic/Wild Goat, Yak) was computed, where H1 and H2 were pairs of genomes from the "ibex" clade of the IBS-nj phylogeny (Figure 6.7). The results for *D*(H1, H2, Semnan3, Yak) are presented in Figure 6.13, and the results for three other wild/domestic goat genomes (Blagotin3, Direkli1-2, and Hovk1) are presented in Appendix Figure 6.9; the pattern of *D* statistic results were qualitatively similar. Two pairs of genomes showed an equal number of "domestic" derived alleles: Tur1/Tur2 and Walie1/Nubiana1, suggesting that for each of these pairs, neither member was admixed with domestic goat to the exclusion of the other. All tests of the form *D*(Tur1/2, Direkli4, Wild/Domestic goat, Yak) produced significant positive results, implying that there was likely gene flow between the ancestors of the "Tur-like" Direkli4 and the ancestors of the remaining "bezoar-like" Direkli wild goat. Greater "domestic" derived alleles were detected in Ibex1 compared with Ibex2, and in Pyrenica2 compared to either Ibex1 or Ibex2, suggesting greater gene flow from domestic goat into the ancestors of Alpine ibex compared with Iberian ibex.

The highest degree of "domestic" derived allele sharing was observed in Tur1, Tur2, Direkli4, and Ibex1 suggesting that if the topology presented in Figure 6.7 is an accurate representation of the species tree, the ancestors of Caucasian tur and Alpine ibex have likely admixed with domestic goat, or vice versa. The lowest amount of shared "domestic" alleles is

observed in Pyrenica2, Nubiana1, and Walie1. If all non-domestic *Capra* genomes are admixed with domestic sources, then these particular lineages have admixed to the smallest extent; a possible alternative is that they have admixed with a source of ancestral alleles.

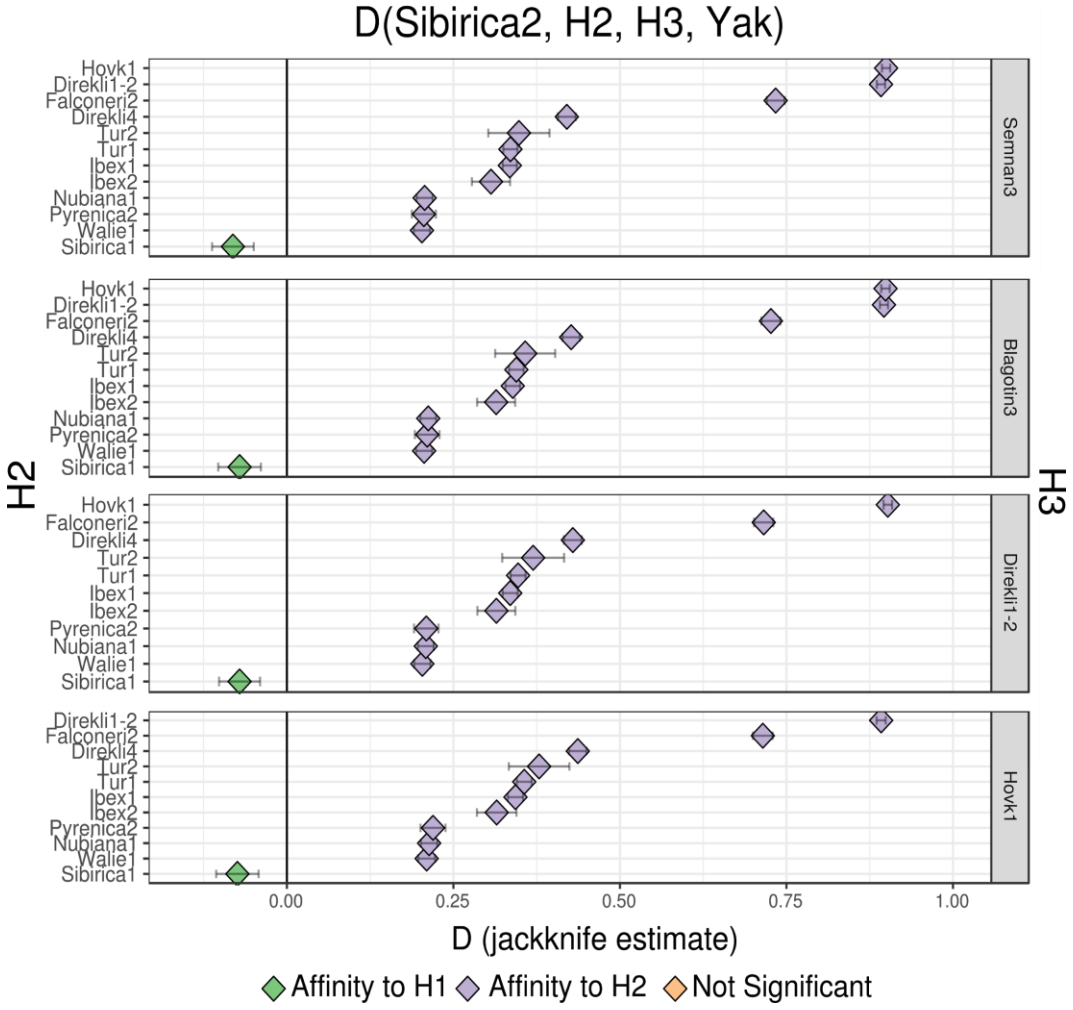


Figure 6.12 - *D* statistic test for admixture from an ancient domestic or wild goat genome (H3) into a *Capra* genome, relative to Sibirica2. Colour indicates H3 affinity (excess of derived alleles shared) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

D statistics were computed to determine if modern domestic goat populations share a greater number of derived alleles with other *Capra* species compared to Neolithic genomes from a similar geographic region. To note is that admixture from domestic goat into other *Capra* could confound the following tests, making some domestic populations share more derived alleles with other *Capra* than others, and consequently to non-zero *D* statistic results.

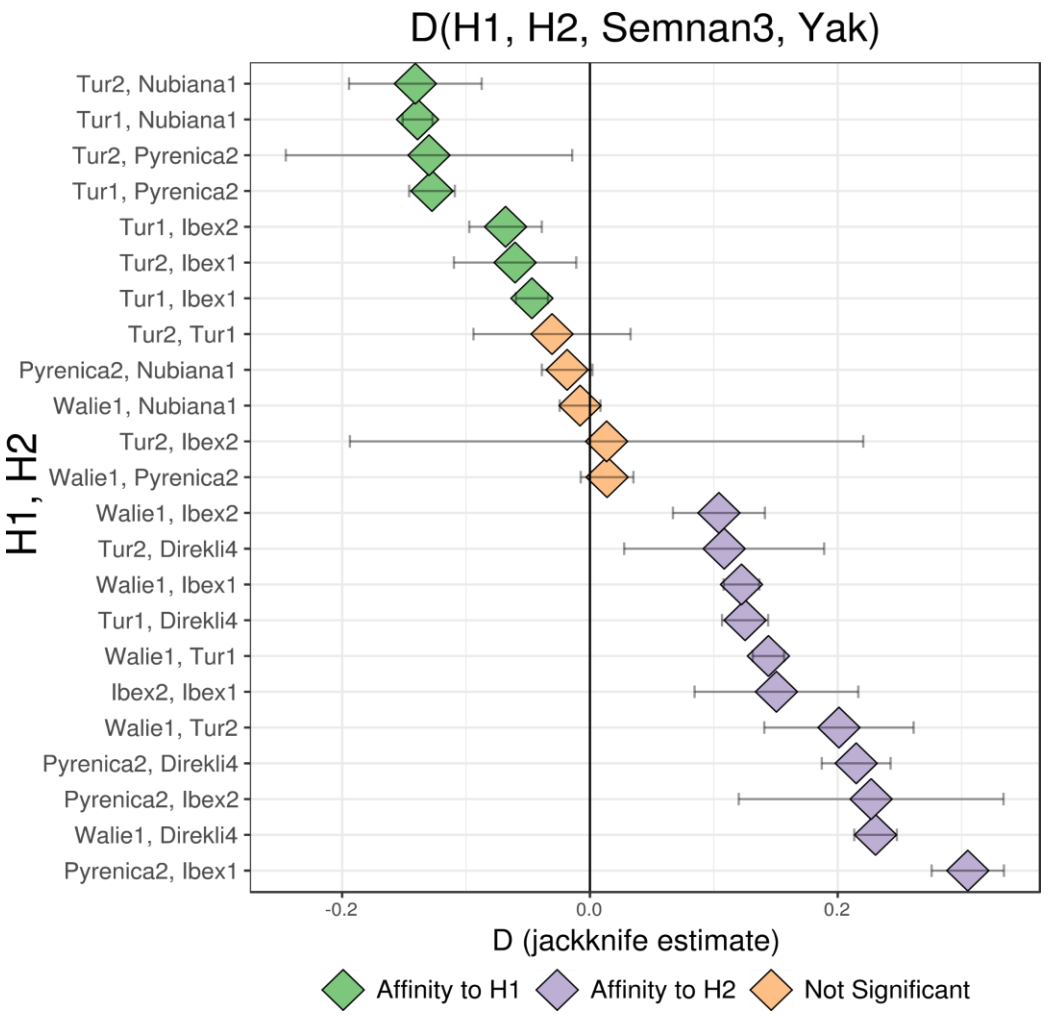


Figure 6.13 - *D* statistic test for relative affinity (derived allele sharing) of the eastern Neolithic domestic Semnan3 (H3) to members of pairs of *Capra* genomes, H1 and H2. Colour indicates H3 affinity (an excess of derived alleles shared) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.

Introgression from European ibex was tested using the *D*(Blagotin3, European goat, European ibex, Yak). European ibex was varied between Ibex1, Ibex2, and Pyrenica2. The Bronze Age goat Potterne1 from Britain was included in this analysis. Absolute *D* values were low, but significant and positive for four modern goat genomes from France and Italy when Ibex1 was used to define the derived allele (Figure 6.14). One of these lost significance when the lower-coverage Pyrenica2 was set to H3 (3.53× vs 0.14×), and no significant results were obtained when Ibex2 was used as H3 (0.05×). *D* statistics suggested that both Potterne1 and the modern Irish goat IOG were not admixed with European ibex, suggesting that the admixture is primarily a geographic phenomenon *i.e.* restricted to European goat populations close to the habits of ibex (the Alps and Iberian Peninsula), but a wider sample of goat genomes is required to definitely conclude this.

Has admixture occurred from non-domestic Capra *species to African domestic goat?*

African goat were then tested in order to determine if they had been affected by past admixture events with African ibex (represented by Nubiana1 and Walie1). As the exact proxy for "unadmixed ancestral African goat" was unclear, three different genomes were used: IOG, the high coverage Irish modern goat that did not appear to be admixed with ibex; Blagotin3, the Neolithic Serbian goat; and Italian3, the modern Italian goat that had positive signals of ibex admixture. Using either Nubiana1 (Figure 6.15) or Walie1 (Appendix Figure 6.10) produced the same general pattern: a subset of Moroccan genomes display greater derived allele sharing with African ibex relative to the "unadmixed" IOG, but not a greater amount compared to Italian3 or Blagotin3. Domestic derived allele sharing is equal between Nubiana1 and Walie1, tested using the statistic *D*(Nubiana1, Walie1, Modern African goat, Yak) (Appendix Figure 6.11). However, greater allele sharing was observed between African goat and Alpine ibex rather than between African goat and African ibex, as indicated by positive results for the test *D*(Nubiana1/Walie1, Ibex1/2, African goat, Yak) (Appendix Figure 6.11). This may be a consequence of greater admixture between the alpine ibex and the ancestors of African goat compared to African ibex and the ancestors of African goat. Alternatively, Alpine ibex may be admixed with domestic goat (see Figure 6.12 above), which would confound *D* statistic results.

Unexpectedly, the Togolese goat Tog showed a significant deficit in African ibex derived alleles relative to any of the European goat. To determine if this was due to admixture from a source of ancestral alleles, the *D*(European goat, African goat, Falconeri1, Yak) was

computed (Appendix Figure 6.12). Significant signals of derived allele sharing between African goat relative to European goat was detected for almost all Moroccan goat, while Tog displayed a deficit in Falconeri1 alleles relative to all modern European genomes. Assuming Falconeri1 does represent an outgroup to all other *Capra* species, a possible explanation for the latter is if the ancestors of Tog admixed with another member of the *Caprini* tribe, resulting in an excess of ancestral alleles. Possible candidates include domestic sheep (*Ovis*

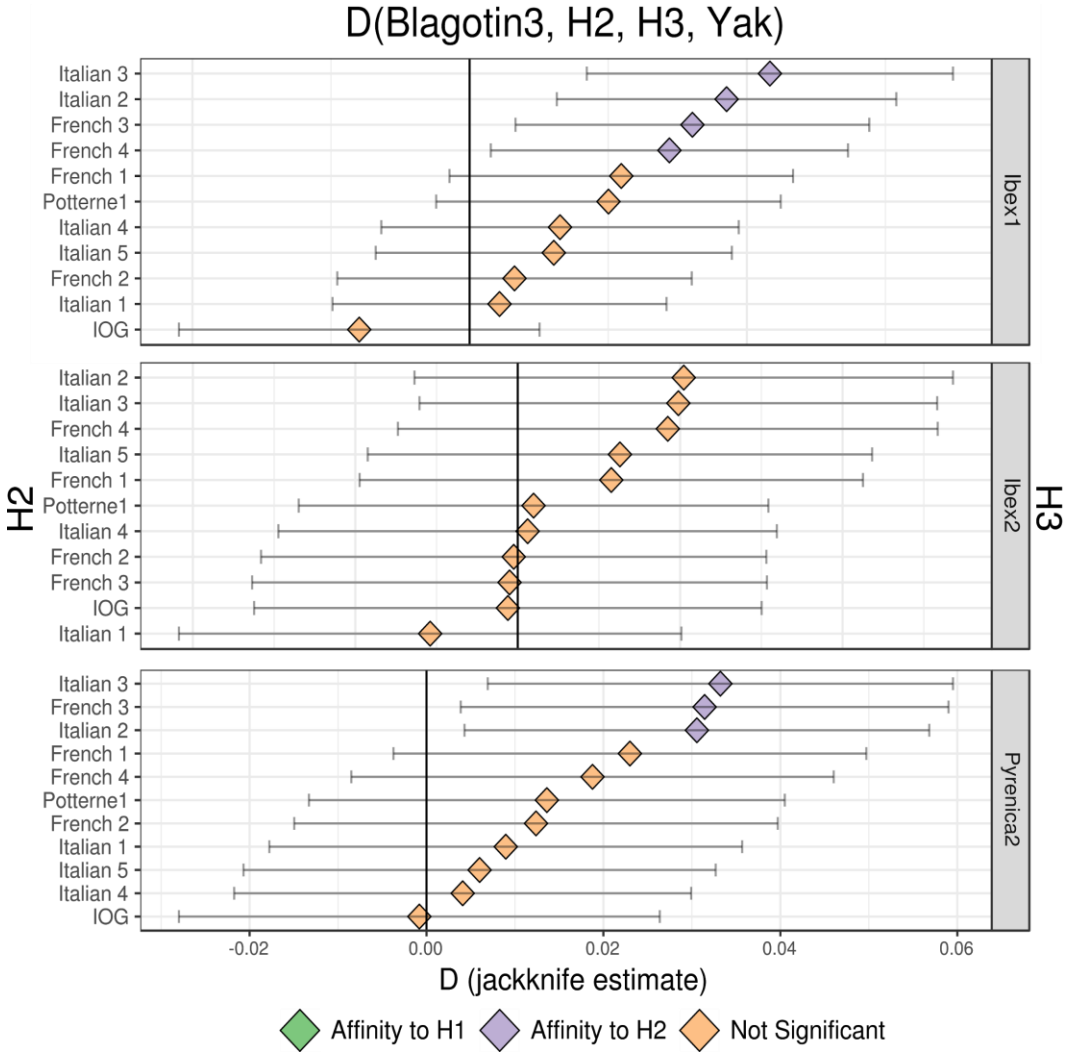


Figure 6.14 - *D* statistic test for admixture from Alpine or Iberian ibex into European goat. The Serbian early Neolithic sample Blagotin3 was used as an unadmixed reference. Colour indicates ibex affinity (an excess of derived alleles shared) to either Blagotin3 or the modern European goat, with non-significance (|Z| < 3) indicated by the colour orange.

aries), the Barbary sheep (*Ammotragus lervia*) from North Africa, or the Arabian Tahr (*Arabitragus jayakari*) from the same peninsula. However, none of these currently are found in Togo. The excess of markhor derived alleles between Moroccan goat compared to European could be due to admixture between a Moroccan-like source and the ancestors of Falconeri1, in which case it is not suitable as an outgroup.

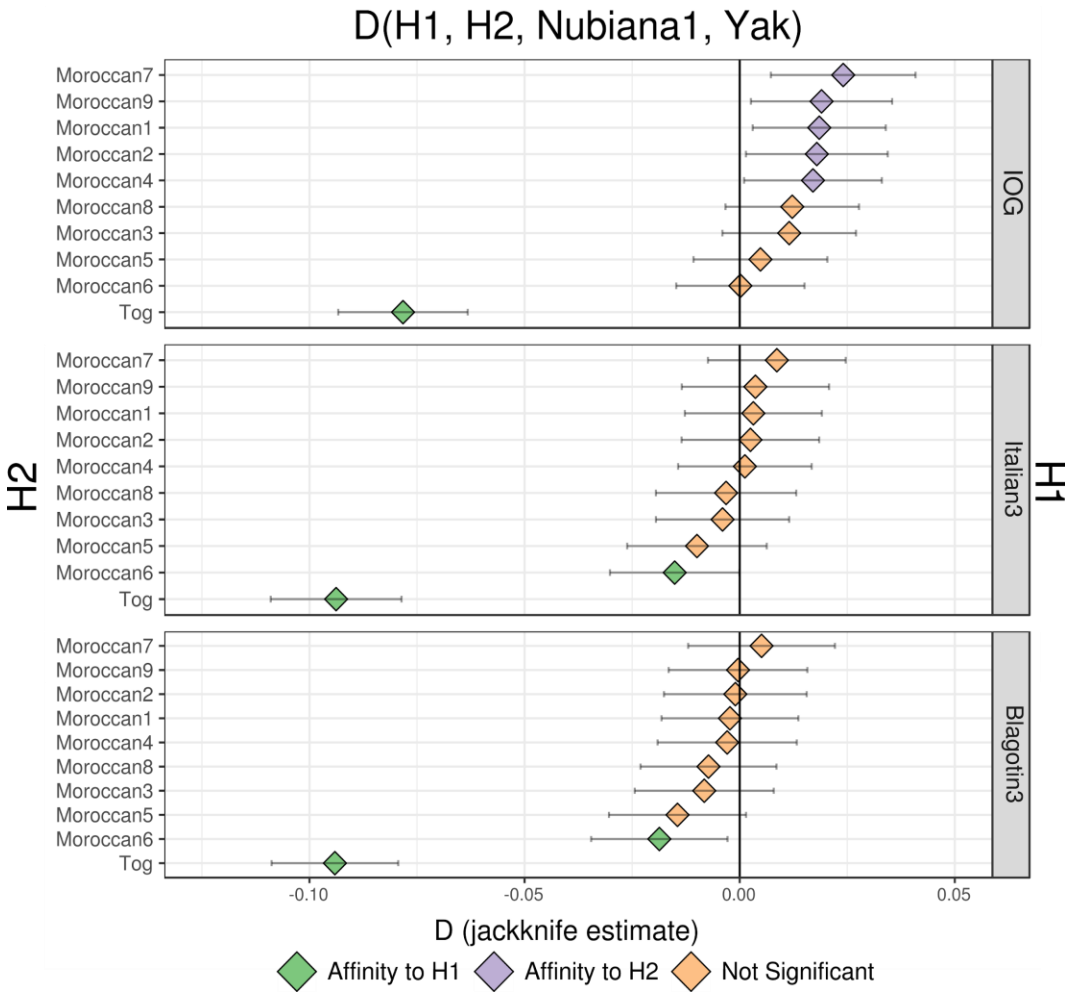


Figure 6.15 - *D* statistic test for admixture from the Nubian ibex Nubiana1 (H3) to modern African goat (H2) relative to several European genomes (H1). Colour indicates Nubiana1 affinity (degree of derived allele sharing) to either the European genome H1 or the modern African goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.

Has admixture occurred from non-domestic Capra *species to Asian domestic goat?*

To test if admixture has occurred from other *Capra* species to the ancestors of modern Asian goat, the statistic *D*(Semnan3, Modern Asian, *Capra* species, Yak) was computed, using the early Neolithic Semnan3 from eastern Iran as an unadmixed reference, and *Capra* species present in the Asian continent. Using Falconeri1 to define the derived allele resulted in positive scores for all modern Iranian goat and the large majority of Chinese (Figure 6.16), indicating an excess of markhor-like derived alleles in modern Asian genomes compared to Neolithic Iranian. In contrast, Falconeri2 as H3 produced only non-significant *D* statistics, which would be consistent with it being a markhor with substantial domestic goat ancestry. Setting either Sibirica2 or Tur1 produces similar patterns of allele sharing (Appendix Figures 6.13 and 6.14). The relatively-low coverage of Tur2 limits any confident assertions on less admixture from *Capra cylindricornis* to domestic goat compared to *Capra caucasica*, while the hybrid ancestry of Sibirica1 confounds such *D* statistic tests.

Admixture from the possible genus outgroup Falconeri1 (Figure 6.16) is not expected to be confounded by admixture from other *Capra* species, ignoring the possibility of gene flow between *Capra* species. However, admixture from members of the "ibex" clade may create false signals of admixture from other members, due to shared "ibex" derived alleles. Testing for admixture from European ibex gives positive results, demonstrating this potential issue. To attempt to distinguish admixture from potential "ibex" clade species, the statistic *D*(H1, H2, Modern Asian, Yak) was computed, where H1 and H2 are different *Capra* genomes. This test should indicate from which *Capra* genome the majority "ibex" derived alleles are likely to originate from, but does not account for admixture from multiple sources. Additionally, admixture from domestic goat to one of H1 or H2 will result in an excess of derived allele sharing, making the test unideal for the purpose of determining likely admixture sources. The markhor genomes were not included due to their differing affinities to domestic goat and other *Capra* (Figure 6.7).

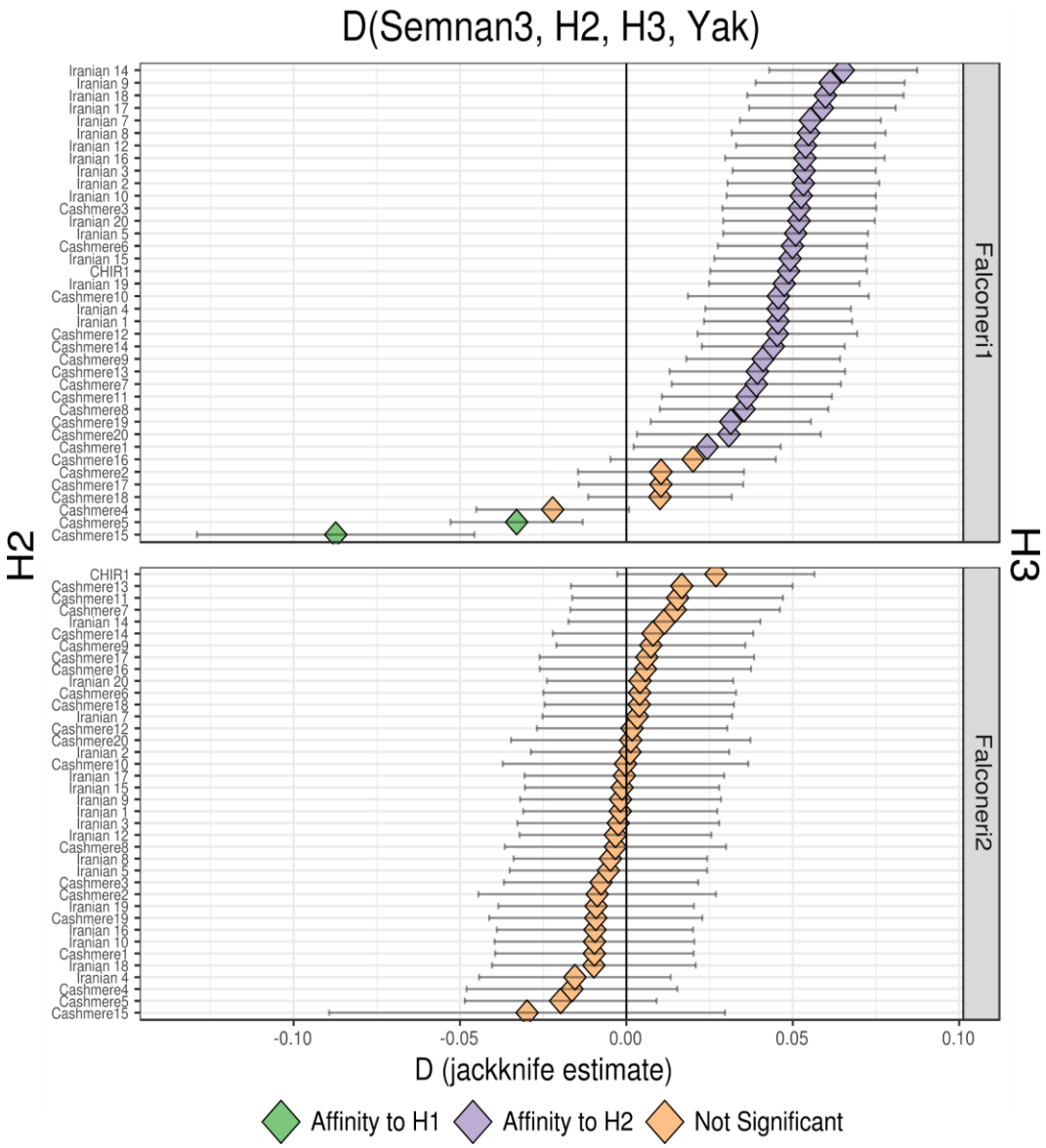


Figure 6.16 - *D* statistic test for admixture from samples identified as *Capra falconeri* (H3) into modern Asian goat (H2). Colour indicates markhor affinity (degree of derived allele sharing) to either Semnan3 or the modern Asian goat H2, with non-significance (|Z| < 3) indicated by the colour orange.

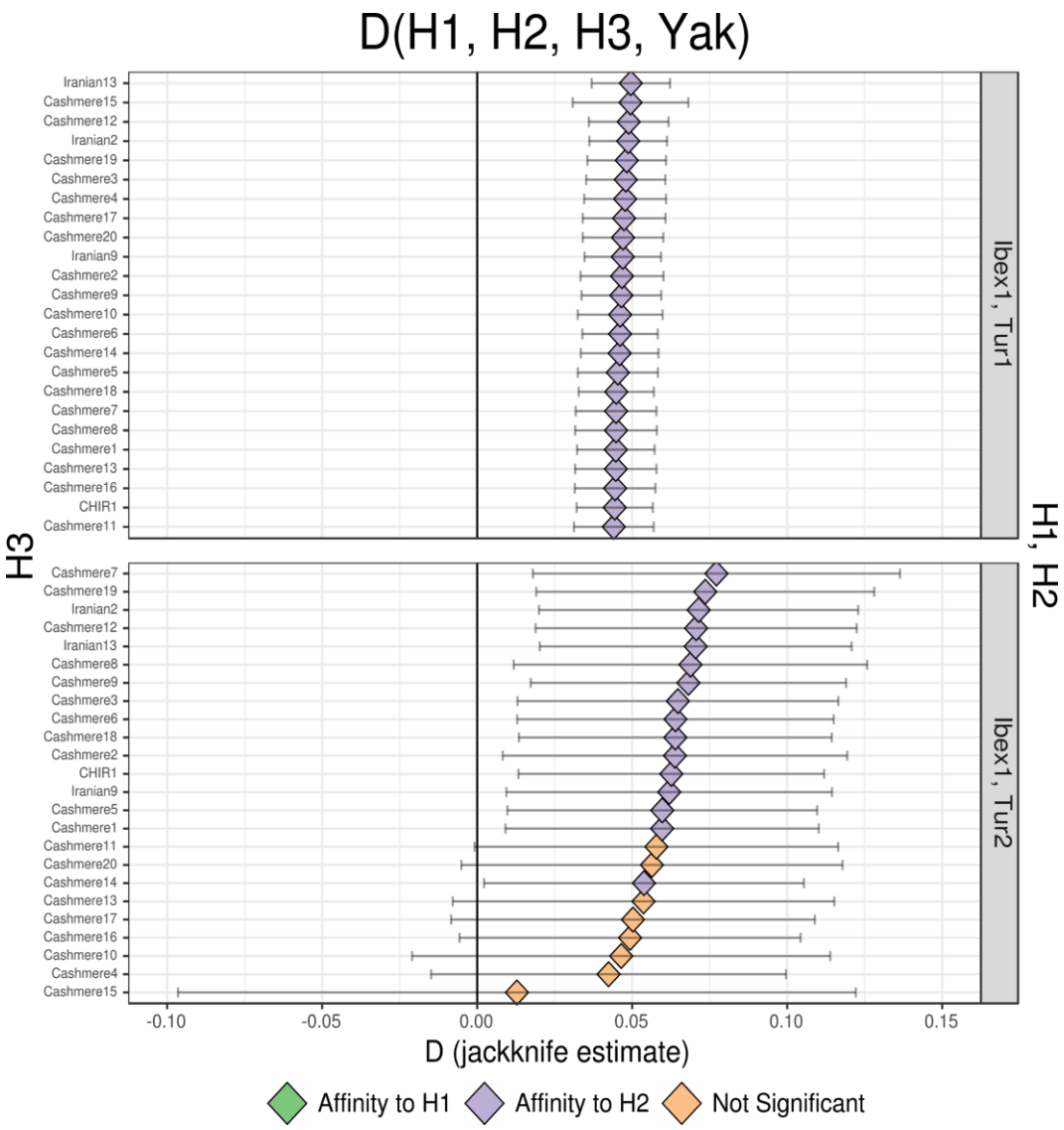


Figure 6.17 - *D* statistic for derived allele sharing of modern Asian goat genomes (H3) with the Alpine ibex Ibex1 or Caucasian ibex (H1, H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

Sibirica1 was excluded due to its hybrid status. Both Alpine ibex genomes were included as representatives of *Capra* unlikely to have admixed with domestic goat in Asia. There is a bias towards allele sharing between modern Asian goat and Caucasian ibex, compared with Asian goat and Alpine ibex (Figure 6.17), regardless of which alpine ibex genome is used (Appendix Figure 6.15). Although this is consistent with admixture from Caucasian ibex, these genomes also show greater derived allele sharing with the Late Pleistocene bezoar sample Hovk1 than Alpine ibex do (Appendix Figure 6.9), implying a larger degree of admixture with the *Capra aegagrus*/*hircus* clade.

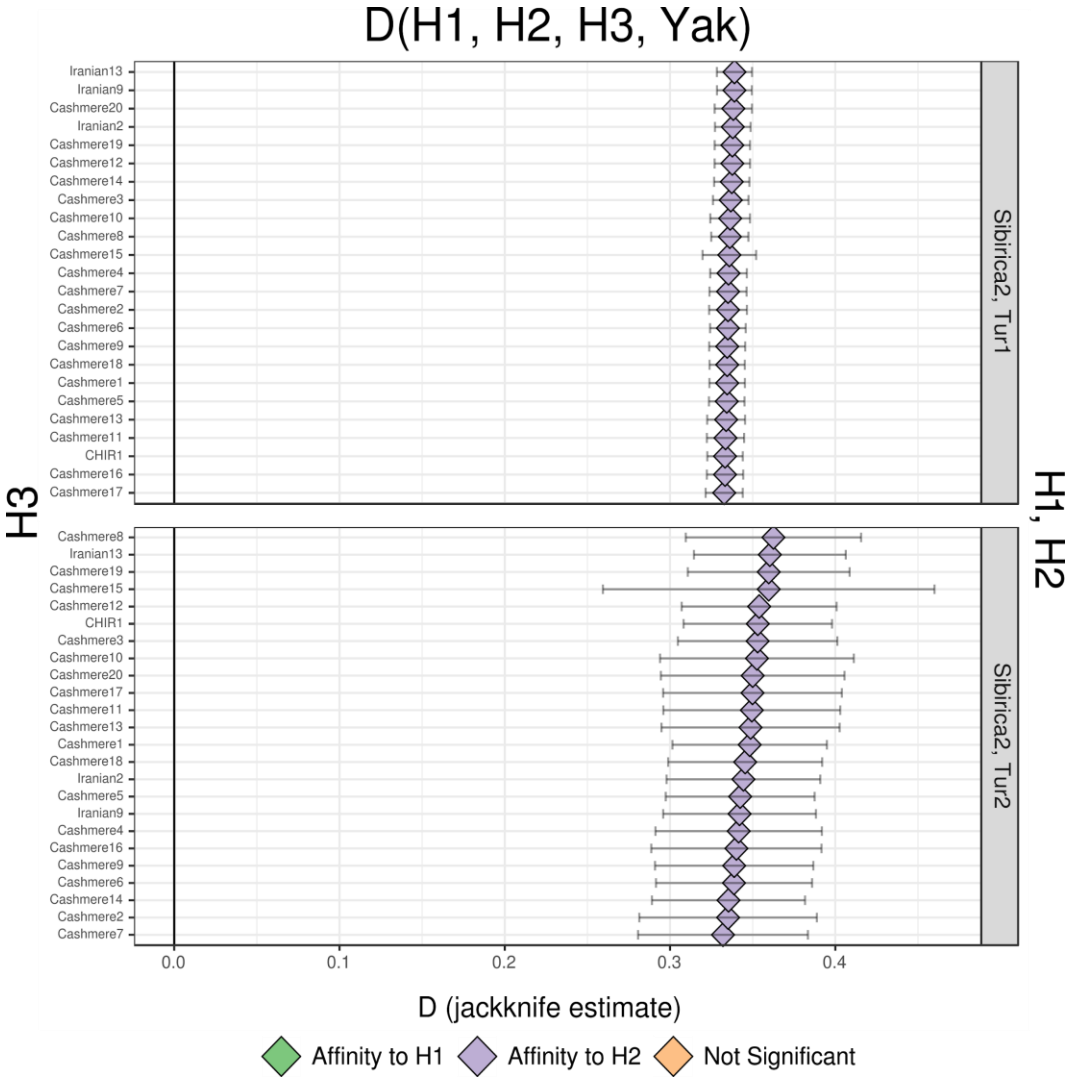


Figure 6.18 - *D* statistic for derived allele sharing of modern Asian goat genomes (H3) with the Siberian ibex Sibirica2 or Caucasian ibex (H1, H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

Non-significant results are obtained for the vast majority of tests in the form *D*(Tur1, Tur2, Modern Asian, Yak) (Appendix Figure 6.16). This suggests that the "ibex" alleles present in modern Asian goat originate from a population equally related to both *Capra caucasica* or *Capra cylindricornis*, or there is not a sufficient number of sites shared across all samples to detect a difference in derived allele sharing between either tur genome. Alternatively, both Caucasian species may be admixed with wild/domestic goat to the same extent. In addition, there is a strong tendency for modern Asian goat to share more derived alleles with the Alpine ibex Ibex1 rather than Ibex2 (Appendix Figure 6.16), consistent with the excess of domestic alleles in Ibex1 detected in Figure 6.12.

Surprisingly, there are clear indications of greater allele sharing between modern Asian genomes and Alpine ibex, compared to between modern Asians and Siberian ibex (Appendix Figure 6.17). Again, there are many possible confounders. Admixture of Siberian ibex with a source of ancestral alleles (*i.e.* markhor, if the phylogeny presented in Figure 6.7 is accurate) is one possibility. Another is if Alpine ibex have experienced extensive gene flow with domestic goat, which is not implausible. Gene flow from Siberian ibex may have occurred, but this is difficult to disentangle from the apparent gene flow with Caucasian tur.

Has admixture occurred from non-domestic Capra *species to ancient eastern goat?*

Ancient goat populations in eastern regions (Iran, Turkmenistan, Uzbekistan) may have also admixed with other *Capra* species. The test *D*(Semnan3, Post-Neolithic eastern goat, Tur1, Yak) produced significant positive results for three genomes: the Bronze Age samples Qazvin1 and Bulak2, and the medieval Darre2 (Figure 6.19). Negative results (*i.e.* significantly less derived allele sharing) are obtained for several earlier (Chalcolithic) genomes. As this "Tur-like" ancestry could have several possible sources in Asia, the statistic *D*(Caucasian ibex, Siberian ibex, Post-Neolithic eastern goat) was calculated. Similar to modern Asian genomes, there is greater derived allele sharing observed with Caucasian ibex and the Post-Neolithic eastern goat (Appendix Figure 6.19). A hierarchy of allele sharing with ancient eastern goat is observed: Tur>Ibex>Nubiana>Sibirica (Appendix Figures 6.19 and 6.20). Again, these observations of greater derived allele sharing may be confounded by gene flow from the *Capra aegagrus*/*hircus* clade to the ancestors these *Capra* species.

Has admixture occurred from non-domestic Capra *species to ancient Levantine goat?*

As wild and domestic goat herds in Neolithic Levant likely bordered or overlapped with the wild Nubian ibex distribution, it is possible that admixture between the populations occurred. Using Ainghazal1 as a reference, admixture was tested using the statistic *D*(Ainghazal1, Post-Neolithic Levantine/African goat, African Ibex, Yak). Positive results were obtained for the Bronze Age goat

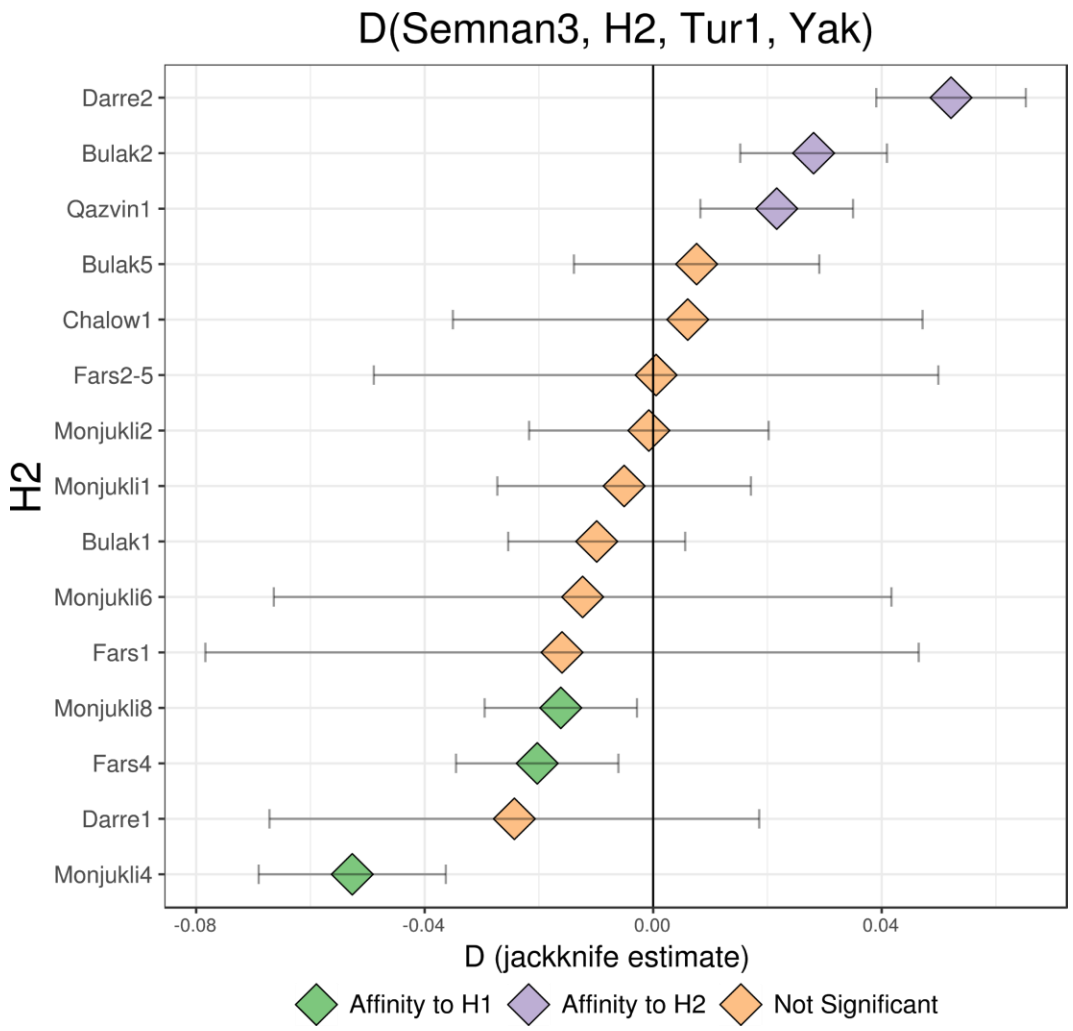


Figure 6.19 - *D* statistic test for admixture from the West Caucasian ibex Tur1 (H3) to ancient post-Neolithic Eastern (Iran, Turkmenistan, Uzbekistan) goat (H2) relative to the Neolithic Iranian goat Semnan3 (H1). Colour indicates Tur1 affinity (degree of derived allele sharing) to either Semnan3 H1 or the post-Neolithic Eastern goat H2, with non-significance (|Z| < 3) indicated by the colour orange.

Yoqneam2, and also the Neolithic Serbian Blagotin3, and modern African genomes (Figure 6.20). A single negative result was obtained for the Chalcolithic sample Shiqmim1. The

remaining samples produced non-significant results or were of too-low a coverage to include. This result could be confounded by the presence of Direkli-like ancestry in later genomes. Consistent with this is the observation that the test *D*(Blagotin3, Post-Neolithic Levantine/African goat, African Ibex, Yak) produces several negative results, including for

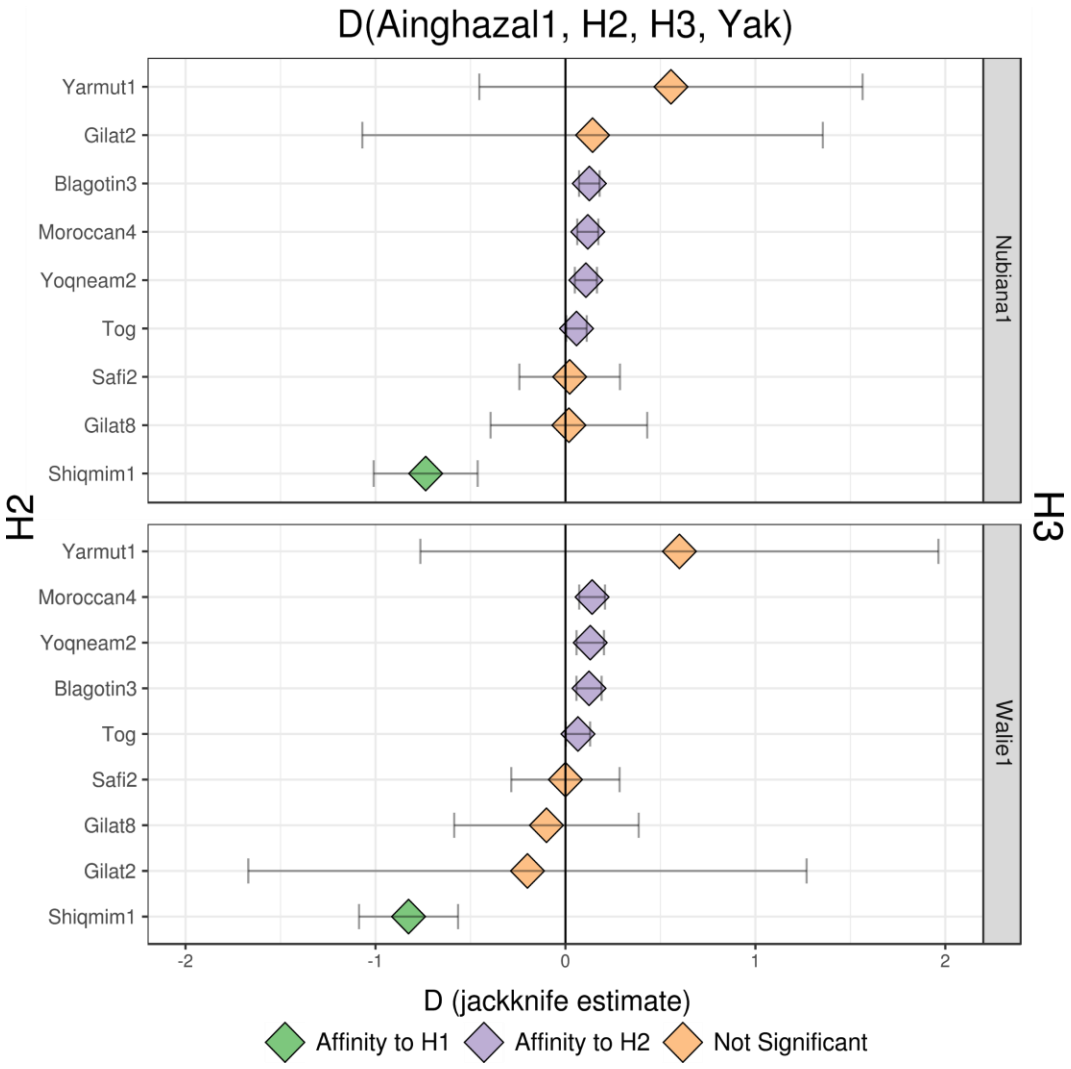


Figure 6.20 - *D* statistic test for admixture from the African ibex Nubiana1 or Walie1 (H3) to ancient post-Neolithic Levantine goat (H2) relative to the Neolithic Levant goat Ainghazal1 (H1). Colour indicates African ibex affinity (degree of derived allele sharing) to either Ainghazal1 H1 or the post-Neolithic Levantine goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.

Ainghazal1 and Yoqneam2 (Appendix Figure 6.21). Direkli-like ancestry may have mediated Tur/ibex-like derived alleles in later populations, as typified by the Neolithic Serbian genome which has a particularly large amount of Direkli-like ancestry (Chapter 4). The deficit in "ibex-like" alleles in Shiqmim1 is more pronounced when relative to Blagotin3.

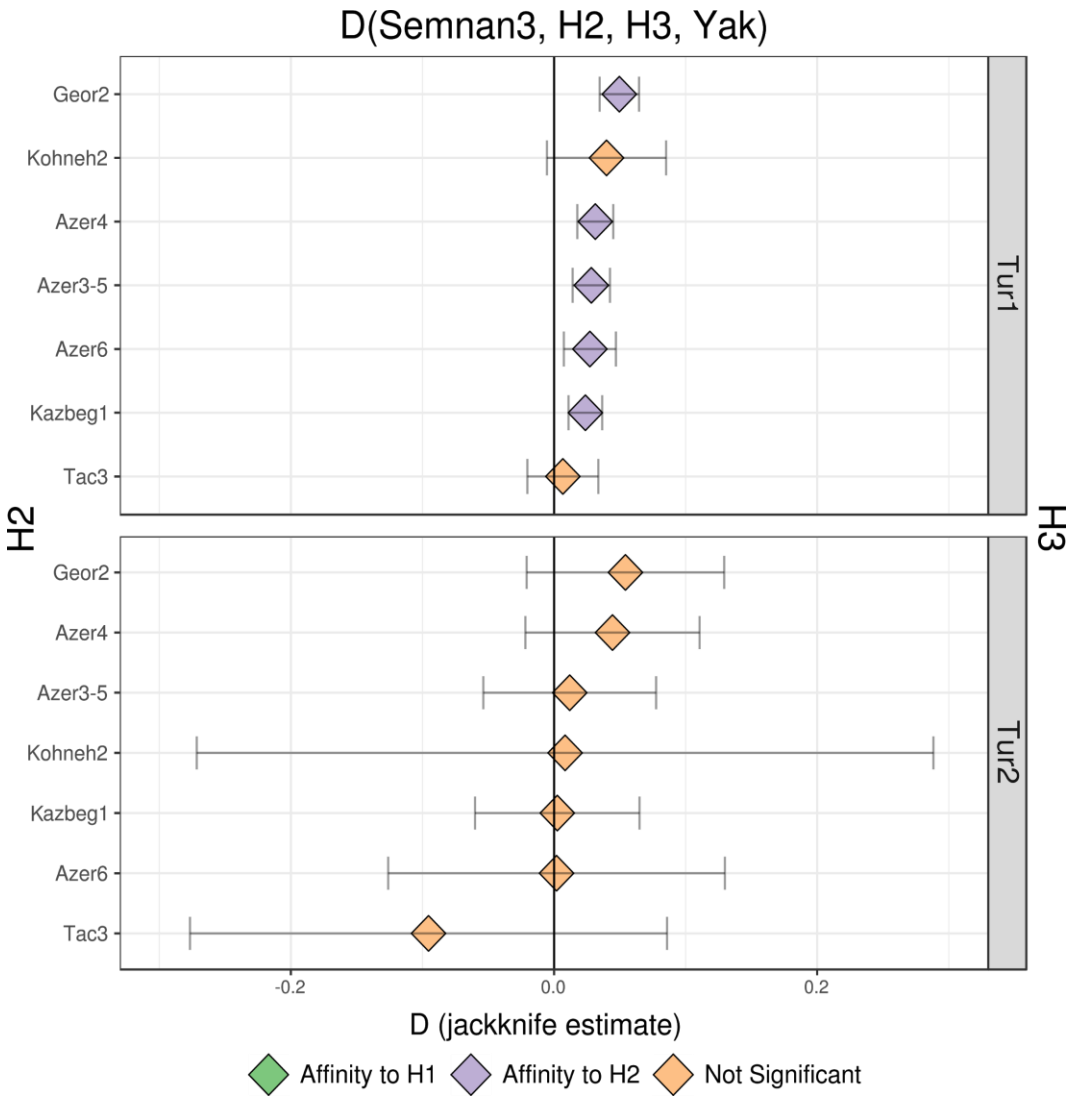


Figure 6.21 - *D* statistic test for admixture from Caucasian ibex/tur genomes (H3) into ancient Caucasian goat (H2) relative to the Iranian Neolithic Semnan3 (H1). Colour indicates tur affinity (degree of derived allele sharing) to either Semnan3 or the ancient Caucasian goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.

Has admixture occurred from non-domestic Capra *species to ancient Caucasus goat?*

Domestic goat populations from the Caucasus region may have also experienced admixture with the derived allele sharing. The statistic *D*(Semnan3, Ancient Caucasian Goat, Tur1, Yak) was significant and positive for five out of seven tests, including the Chalcolithic Azerbaijan samples Azer6 (Figure 6.21). Tur2 did not produced significant test statistics. When Blagotin3 was set as the baseline, ancient Caucasian goat shared less derived alleles with the Caucasian tur, relative to Blagotin3 (Appendix Figure 6.22). This statistic was likely confounded by the Direkli-like ancestry present in western Neolithic genomes (Chapter 4), which may have interacted with the Tur-like population represented by Direkli4. Direkli4 would represent a source of "Tur" alleles that could have entered the ancestors of western Neolithic goat via the "bezoar" Direkli goat, thus increasing the number of "Tur" derived alleles in their collective gene pool. However, the bezoar population which actually admixed with the ancestors of western goat may themselves not have interacted with this Tur-like population. Alternatively, Caucasian Tur may be admixed with a domestic population closer related to Blagotin3 than Semnan3, thus inflating the Blagotin3/Caucasian tur shared derived alleles. The positive result of the test *D*(Blagotin3, Direkli1-2, Tur1, Yak) (Z=8.48) suggests the former of these explanations to be more accurate, but admixture with local Caucasian bezoar populations may also complicate this test.

Has admixture occurred from non-domestic Capra *species to ancient Turkish goat?*

The ancestors Bronze Age Turkish samples Acem1 and Acem2 may have admixed with sources of other *Capra* alleles, compared to the Neolithic Serbian genome Blagotin3. The statistic *D*(H1, H2, H3, Yak) was computed, where H1 and H2 were two of Acem1, Acem2, or Blagotin3, and H3 was a *Capra* genome or Dirkeli1-2. In all tests of the form *D*(Acem1, Acem2, H3, Yak) positive results were obtained (Figure 6.22), suggesting an excess of ancestral alleles in Acem1. Consistent with this, all tests of the form *D*(Blagotin3, Acem1, H3, Yak) gave negative results. Only the test statistic *D*(Blagotin3, Acem2, H3, Yak) gave results which varied with the identity of H3. All members of the "ibex" clade gave non-significant results when set to H3. However, the markhor outgroup Falconeri1 showed greater allele sharing with Acem2 than Blagotin3. When the Epipaleolithic bezoar Direkli1-2 was used as H3 negative *D* scores were obtained, implying greater derived allele sharing between Direkli1-2 and Blagotin3 than between Direkli1-2 and Acem2.

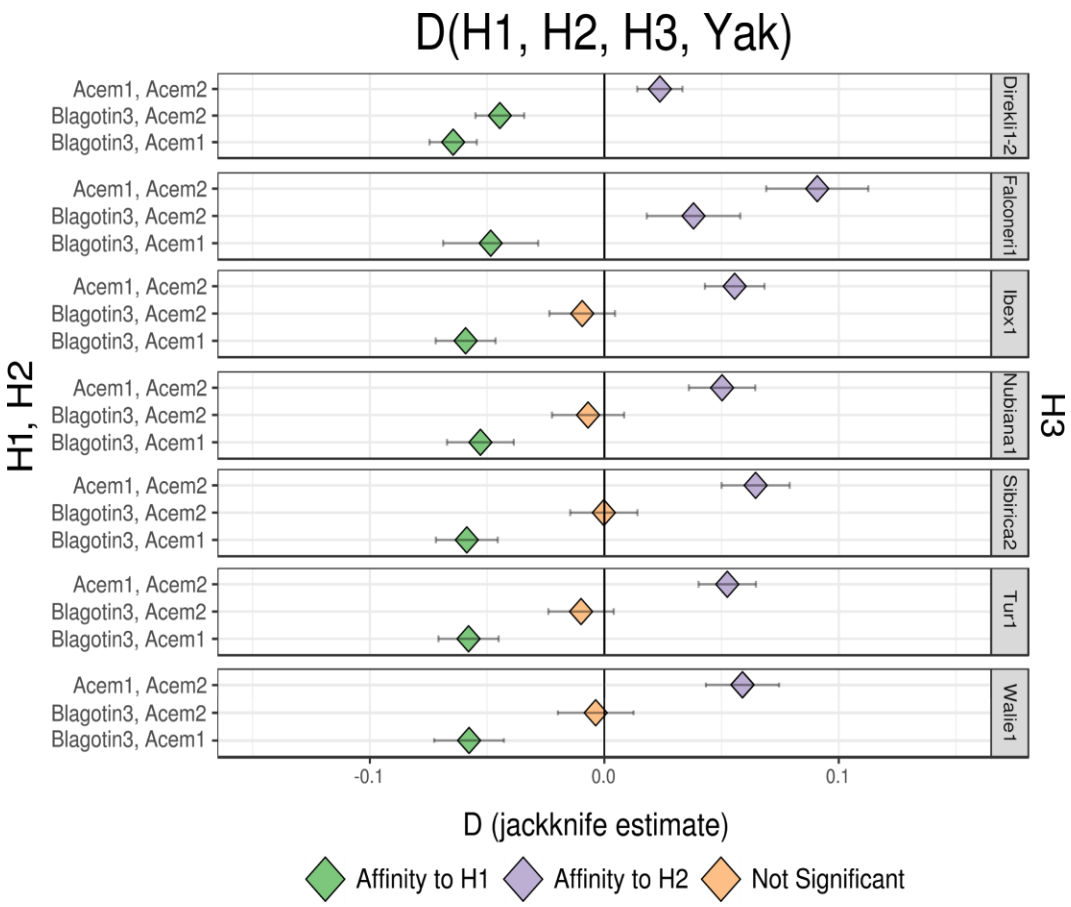


Figure 6.22 - *D* statistic for derived allele sharing of various *Capra* genomes (H3) with pairs of western ancient goat Blagotin3, Acem1, and Acem2 (H1, H2). Colour indicates *Capra* species affinity (degree of derived allele sharing) to either H1 or H2, with non-significance (|Z| < 3) indicated by the colour orange.

Discussion

The mitochondrial phylogeny of *Capra* and *Caprini*

The mtDNA phylogeny of the tribe *Caprini* and genus *Capra* generated here (Figure 6.6) is broadly in line with previous work. A notable difference is that sheep is not the outgroup to all other *Caprini* species in the phylogeny shown in Figure 6.6, while it is in the tree presented in (Hassanin et al. 2012); the ordering of nodes also differs between the two analyses. Low (≤0.6) bootstrap values are obtained here for these nodes, so there is limited confidence placed in the gross details of the phylogeny. The relatively close relationship between *Capra* mtDNA and mtDNA of the Bharal and Himalayan Tahr is consistent, however, and the monophyly of the Siberian ibex reference and Himalayan Tahr is also replicated. Additionally, a mtDNA sequence recovered from the historic markhor Falconeri1 shows affinity with the Barbary sheep (*Ammotragus lervia*), a *Caprini* species native to Northern Africa (Geist 1990). Hybridization between species within the tribe *Caprini* has been proposed previously, but incomplete lineage sorting is also a possible explanation (Ropiquet & Hassanin 2006). An alternative, misspecification of species identity, is discussed below.

The composition of clades within the *Capra* genus was mostly as expected. Nubian and Walia ibex show high mtDNA affinity, as previously reported (Gebremedhin et al. 2009). The low variation as expressed by shallow branching between the Nubian ibex reference, Nubiana1, and Walie1 suggests a relatively restricted maternal lineage; the poor preservation of Nubiana2 prevents fair comparison. The presence of the hybrid sample Sibirica1 in this African clade resolved the question of its parentage (Table 6.2), suggesting its dam was a Nubian ibex. The European ibex clade is also straightforward, with both species forming distinct sister clades, consistent with a relatively recent shared mtDNA ancestor and the "single migration" model of European ibex origins.

Caucasian ibex form a clade that is sister to the T haplogroup observed in Direkli bezoar. The West Caucasian Tur reference sequence forms a branch with Tur1, which was identified as *Capra caucasica*. The *Capra cylindricornis* sample Tur2 is sister to this *C. caucasica* clade. The Direkli T haplogroup is equally related to both Western and Eastern Caucasian ibex mtDNA, suggesting that the T maternal ancestor diverged from the common ancestor of both

Tur prior to their speciation. The entire "Tur-like" clade is robustly supported by bootstrap values >0.9.

The remaining markhor sequence generated here, Falconeri2, shows high similarity with the markhor reference sequence. Surprisingly, mtDNA from the Siberian ibex sample Sibirica2 is sister to both of these *C. falconeri* sequences, and does not group with the relatively-divergent Sibirica1 sequence. Incomplete lineage sorting or gene flow between the two Central Asian species are both possible explanations for this observation.

Resolving the *Capra* phylogeny using nuclear genomes

The approach taken here did not fully resolve the nuclear genome phylogeny of the *Capra* genus. Placing two species in particular did not prove tractable with the current data: the markhor *C. falconeri* and the Siberian ibex *C. sibirica*. The remaining groups formed clades in line with expectation. The two Alpine ibex genomes formed a clade that was sister to the higher-quality Iberian ibex sample Pyrenica2 (Figure 6.7), consistent the mtDNA phylogeny. This whole genome result fit a single migration model for European ibex dispersion, in line with mtDNA. This European ibex clade shares a recent node with the Caucasian Tur clade, a result that echoes the suggestion from palaeontological link between European and Caucasian ibex (Crégut-Bonnoure 1991; Rivals 2006). Neither Tur genome appears to share more bezoar or domestic-like alleles (Appendix Figure 6.9), which would argue against the hybridization model of *Capra caucasica* speciation (Groves & Grubb 2011). In the case both European and Caucasian ibex, the data support simple speciation events from a shared common ancestor, likely a consequence of climate-induced vicariance.

The close relationship between the African ibex *Capra nubiana* and *Capra walie* suggested by previous data (Gebremedhin et al. 2009) is supported here also. Although nuclear genome data was available for only one Nubian ibex, it formed a clade with Walie1, and is consistent with the origin of the Walia ibex from a geographically-isolated Nubian ibex subpopulation (Groves & Grubb 2011). Sibirica1, the Siberian- ibex hybrid specimen, is placed as a sister branch to both of these African ibex, which would be expected based on its partially-divergent ancestry.

The phylogenetic position of Sibrica2 and Siberian ibex is less clear. The species may truly be an outgroup to all other "ibex" species, as suggested by its position in the IBS-nj tree

(Figure 6.7). However, the absence of other non-hybridized *C. sibirica* genomes limits any confident statement regarding its phylogenetic position. For example, Siberian ibex may actually form a clade with Nubian ibex, but Sibirica2 may happen to have divergent ancestry, resulting in its observed location. The mitochondria of this sample forms a clade with markhor mtDNA, not inconsistent with this hypothesis, but incomplete lineage sorting remains a possibility.

The markhor genomes presented here provide a similar phylogenetic challenge. Falconeri1 was demonstrated to have a relatively high proportion of ancestral alleles (Appendix Figure 6.4), which would be expected if the species was basal within the *Capra* genus. The mitochondria of this sample is also divergent relative to previously-sequenced *Capra* mtDNA, grouping with the Barbary Sheep mtDNA (Figure 6.6). Although Siberian ibex mtDNA has been shown to group with that of the Himalayan Tahr (Hassanin et al. 1998; Hassanin et al. 2012), this would be an even greater case of incomplete lineage sorting or admixture during divergence (Ropiquet & Hassanin 2006). A simple explanation for the position of Falconeri1 would be if it were not actually a representative of the species *Capra falconeri*, and was instead a mislabeled divergent *Caprini* species (see discussion below). Alternatively, the sample may be admixed with another *Caprini* species, thus inflating the number of ancestral alleles present in its genome.

Falconeri2 does not resolve this phylogenetic uncertainty. The sample is clearly related to but diverged from the entire *Capra hircus*/*aegagrus* clade in the IBS-tree (Figure 6.7). If the position of Falconeri1 in the phylogeny was correct, then it would be plausible that Falconeri2 was substantially admixed with domestic goat, increasing its allele sharing with domestic goat compared to other markhor. However, Falconeri1 shares a deficit of derived alleles with Falconeri2 (Appendix Figure 6.1), which is not consistent with this explanation. Ancestral structuring within the markhor species would not explain such a strong difference in derived allele sharing. If Falconeri2 was indeed a recent hybrid, reference genome bias may inflate the number of ancestral alleles aligned, thus decreasing the proportion of shared derived alleles between Falconeri1 and Falconeri2. In the absence of genomic data from wild *C. falconeri* or other *Caprini* members, the phylogenetic position of the markhor cannot be resolved here, but it should be noted the Y chromosome data currently supports a close relationship of markhor and goat/bezoar (Pidancier et al. 2006), in a phylogeny close to the one presented in Figure 6.7. This would suggest that Falconeri1 is either admixed or misidentified, which must be addressed to properly contextualize these results.

Direkli4

The discovery of a Tur-like nuclear genome from Direkli Cave - labelled Direkli4 - was unexpected. In Chapters 4, three other Direkli genomes were presented, all showing high affinity with bezoar ibex. Each of these samples had mtDNA related to the Caucasian tur, which could be explained by incomplete lineage sorting or possibly gene flow from bezoar to tur. In contrast, Direkli4 had a mtDNA sequence similar to the F bezoar haplogroup, but shows nuclear genome affinity with both species of Caucasian tur. The sample is securely dated to the late Epipaleolithic, *c.* 12,000 cal BC (Table 2.1). This result is somewhat unprecedented in the *Capra* genus; outside of paleontological evidence for tur in the French Pyrenees (Crégut-Bonnoure 1991; Rivals 2006), there have been no ancient remains for the species beyond the Caucasus Mountains (Weinberg 2002). Less surprising but still of interest is the apparent co-existence of divergent *Capra* populations in the locale of Direkli Cave in the Central Taurus Mountains: a bezoar population represented by Direkli1-2, Direkli5, and Direkli6, and a "Tur-like" population represented by Direkli4. The two groups are genetically distinct, as evidenced by the IBS-nj phylogeny (Figure 6.7), and patterns of derived allele sharing with other bezoar (Figure 6.12). It is possible that the Tur-like population inhabited higher-altitude regions, as is observed in the Caucasus today where the two species co-occur (Heptner et al. 1961).

To my knowledge, distinct populations or subspecies of *Capra* have not been reported in the Taurus Mountains. It is unknown if the population/s are extant, as there is no nuclear genome data from wild goat in the region. This results again highlights our current inadequate understanding of *Capra* genomics.

Capra admixture with domestic goat

Admixture between domestic goat and *Capra* species proved difficult to disentangle. As mentioned above, admixture from domestic goat into a *Capra* species will confound attempts to detect admixture from that species to domestic goat, when using the ABBA/BABA test. Populations related to the domestic population which introgressed into the *Capra* species will share more derived alleles with that admixed *Capra*, and appear admixed when using the *D* statistic. The same applies in cases when gene flow has occurred from a *Capra* species to certain domestic goat population; that goat population will appear to share more derived alleles with all related *Capra* species. This can be somewhat addressed by the appropriate *D* test, but will also be confounded by other episodes of gene flow, and will not reveal multiple

episodes of gene flow between different *Capra* species, unless conditioned on species-specific derived alleles.

The use of bezoar genomes prior to the spread of domestic goat give indications of the extent of gene flow into *Capra* species (Appendix Figure 6.9). Ideally, ancient *Capra* genomes of each species prior to the introduction of the domestication of goat would be used to directly test for admixture without them, only relative measures of derived allele sharing are presented. The statistic *D*(H1, H2, Hovk1, Yak), where H1 and H2 are *Capra* species, gives a general pattern of domestic affinity within the "ibex", which increases from Nubian, Iberian, and Walia ibex genomes, to the Alpine ibex and both Tur species. The apparent domestic admixture into Alpine ibex is not surprising, given their recent population history and other research (Grossen et al. 2014). The finding for Tur is also not unexpected, as *Capra aegagrus* are also found on the slopes of the Caucasus (Schaller 1977; Shackleton 1997). The lack of *C. aegagrus* alleles in Nubian ibex implies their interaction in early Holocene Levant was limited (Uerpmann 1987), but this inference is restricted to a single genome.

Although gene flow in both directions likely confounded *D* statistic tests here, some instances stand and require discussion. The Togolese modern goat Tog shows a large deficit of African ibex derived alleles relative to Moroccan goat (Figure 6.15 and Appendix Figure 6.10). Gene flow from a population related to Moroccan goat to Nubian and Walia ibex could explain this derived allele imbalance, but the African ibex appears to contain the smallest amount of domestic derived alleles of all ibex tested here (Figure 6.13 and Appendix Figure 6.9). These two observations are not mutually exclusive, but can be distinguished by an unadmixed Nubian ibex genome *i.e.* from Northern Africa prior to the introduction of domestic goat. Alternatively, the ancestors of Tog may have admixed with a source of ancestral alleles, but there are no likely candidates in western Africa. The history of African goat breeds is largely unclear, so it is unknown from where modern Togolese village goat were introduced from, or how many distinct populations may have contributed to their gene pool.

Other domestic goat populations show possible signs of admixture. A subset of European genomes give positive signals of admixture from Ibex1 and Pyrenica2 (Figure 6.14). This is unlikely to be confounded entirely by admixture into European ibex, as the signal is not consistent across all European goat. Moroccan goat appear to show an excess of African ibex ancestry relative to European goat (Figure 6.15), but this cannot be disentangled from gene flow with alpine ibex (Appendix Figure 6.11). A large number of modern Asian goat appear

to have an excess of Falconeri1-derived alleles, relative to the Neolithic goat Semnan3 (Figure 6.16); the non-significance for tests using Falconeri2 may be a product of its relatively low genomic coverage. Asian goat likely have admixed with Caucasian Tur (Figure 6.16-6.18, Appendix Figure 6.18), but directionality is difficult to infer. The data here do not indicate that Asian goat are admixed with Siberian ibex (Appendix Figure 6.17-6.19). Ancient goat in the Caucasus region may also be admixed with Tur (Figure 6.21 and Appendix Figure 6.22), but again this may be due to gene flow in the other direction. Western goat admixture with Direkli bezoar, which may have contributed a degree of the Tur-like ancestry (represented by Direkli4) to the genomes of western goat; a hint of this is observed in the *D* statistic result *D*(Blagotin3, Semnan3, Direkli4, Yak) of Z = -11.01, indicating a larger amount of Direkli4 derived alleles in the Neolithic Serbian Blagotin3 than the eastern Iranian Semnan3, but this may reflect "Direkli" rather than "Tur-like" ancestry. The current data does not indicate a substantial interaction between Levantine goat and Nubian ibex (Figure 6.20 and Appendix Figure 6.21), but population turnover and domestic gene flow to Nubian ibex may confound this.

The goat genomes from Bronze Age Acemhöyük, Turkey, show an unusual pattern of allele sharing with other *Capra*. Acem2 consistently shows a greater number of derived alleles than Acem1, in all cases of the test *D*(Acem1, Acem2, *Capra* genome, Yak), while Blagotin3 more derived alleles than Acem1 in the *D*(Blagotin3, Acem1, *Capra* genome, Yak) (Figure 6.22). For most ibex, the *D* statistics indicate that Acem2 and Blagotin3 have a roughly equal number of derived alleles. However, two cases interrupt this pattern: Blagotin3 appears to share more derived alleles with Direkli1-2 than Acem2 does, while Acem2 shares more derived alleles with the apparent-outgroup Falconeri1 than Blagotin3 does. Acem2 may be admixed with a population related to Falconeri1, which would explain the later statistic. This does not explain the first observation; if the Acem2 has an excess of ancestral alleles via markhor-like admixture, all other statistics in the form *D*(Blagotin3, Acem2, ibex clade genome, Yak) should be similarly biased towards BABA sites *i.e.* greater derived allele sharing between Blagotin3 and the *Capra* genome. As already discussed, these test results are actually non-significant in the cases examined (Figure 6.22). The explanation that Acem2 happens to descend from a population of relatively low Direkli ancestry is unlikely, given positive results of the *D*(Acem1, Acem2, Direkli1-2, Yak). The Falconeri1-like ancestry may have mediated the flow ibex-like ancestry into Acem2, resulting in the positive Z scores for the test (Acem1, Acem2, Ibex clade genome, Yak), but the evidence for this is weak (Appendix Figure 6.1). It may be that Direkli1-2 has a strong deficit in Falconeri1-like alleles

(or vice-versa); admixture into the ancestors of Blagotin3 from Direkli1-2 would deflate the apparent Falconeri1-derived content of that lineage. However, that should also affect tests involving Blagotin3 and Acem1, rather than just Blagotin and Acem2. Additionally, the statistic *D*(Blagotin3, Direkli1-2, Falconeri1, Yak) is nonsignificant and actually mildly positive (Z=2.8), indicating that this explanation is not likely. Evidently, just two genomes from this time period are sufficient to indicate surprising ancestry diversity in Middle Bronze Age Anatolian goat.

Limitations

There were considerable limitations to the analyses presented in this chapter. A major element of this was nature of the sampling and data: relatively low genomic coverage (< 1×), obtained from material not directly sampled from the species in their current habitat. This added considerable uncertainty to analyses; a given test statistic result could be the product of a low-coverage effect (*e.g.* non-significance), or due to undocumented events in the history of the museum samples (*e.g.* hybridization). The museum samples themselves were mostly obtained from zoo populations (Appendix Table 6.10), which are known to show domestic goat admixture in the case of markhor populations (Hammer et al. 2008). The reasoning used here often relied on prior expectations based on previous mitochondrial DNA, creating a reliance on single-locus studies of limited scope.

As the biological specimen used to generate genomes here were sampled several decades prior to sequencing, there were more opportunities for sample mix-up/mislabeling to occur, compared to if the specimen were obtained from extant wild populations. Mitochondrial identity was used to increase species identification certainty; for example, the mitochondria of Nubian and Walia ibex formed a clade with the Nubian ibex reference, which was taken as a sign of their veracity. Again, this creates a dependency on single locus data and past publications. For instance, the grouping of the Sibirica2 mtDNA with markhor sequences, and high divergence of the Falconeri1 mtDNA not unlike that seen in the Siberian ibex reference, could suggest that the sample labels had been erroneously interchanged at some point. Sibrica2 would therefore be a markhor with a markhor-like mtDNA, less divergent from Falconeri2, while Falconeri1 would actually be a Siberian ibex rooting the clade. However, molecular sex assignment excludes this scenario: Falconeri1 was known to derive from a male specimen (Appendix Table 6.10), which matches the assigned sex, but Sibirica2 was determined to be female (Table 6.3). Excluding multiple mis-assignments occurring, the

Falconeri1-Sibirica2 interchange likely did not happen. This does not exclude other mislabeling having occurred; additionally, the reference mtDNA species assignment may be erroneous. For instance, the majority of *Capra* mitochondrial reference specimens derive from the same museum and zoo as sampled here (the MNHN) (Ropiquet & Hassanin 2006; Hassanin et al. 2009; Hassanin et al. 2012), but according to the later reference these originate from wild animals. For the vast majority of these species, no other complete mitochondrial sequences are available, preventing external verification of the reference sequences themselves. The uncertainty in the origin of either the reference mitochondria or the museum samples sequenced here persists, and hinders drawing any strong conclusions of the phylogeny of *Capra*, or its apparent complexity.

The IBS approach also represents a limitation in attempting to produce a phylogeny of *Capra*. Calculating identity-by-state values using ANGSD does provide the advantage of not having to call genotypes directly, or to provide pre-ascertained polymorphic sites; the later point is important when exploring variation of an understudied species. However, the sites used are likely affected by the large number of domestic goats included in IBS calculation, the minimum covered individual filter (half the number of analyzed samples), and also by reference alignment bias. Ideally, the only sites included in IBS determination would be covered in all individuals, but this is not practical given the low coverage of many of the genomes. The per-site filter of 50% of individuals will be biased towards those sites covered in the higher-coverage domestic goat genomes, leading to non-*Capra hircus* variation being relatively poorly represented. Reference bias is also a possible issue, given that the reference genome used was not equally related to all aligned species. Genome alignment using permissive parameters and reduced stringency of read mapping quality should reduce this effect, allowing more variation in the *Capra* genomes to be captured and providing a more accurate measures of allele sharing. Additionally, random read sampling was implemented during IBS estimation, reducing the potential effect of reference genome bias. However, some reference bias likely persists, favouring the alignment of reads similar to the domestic goat reference genome.

Conceivably, *Capra* genomes that have admixture with domestic goat will thus be overestimated in their affinity with goat. The net effect is that the sites used to calculate IBS may overly-represent both domestic goat variation and the variation that distinguishes domestic goat from other *Capra*. The consequence of this may be the main cluster of the IBS-nj phylogeny *i.e. Capra aegagrus/hircus* (possibly together with the markhor) are not a sister

clade to all other ibex, and this actually represent an ascertainment bias. A preferable approach may have been to calculate pairwise sample divergence without dependence on sites being present in multiple individuals, as in the reconstruction of the *Elephantidae* family (Palkopoulou et al. 2018). This method also allows bootstrapping - subsampling of DNA sequence with replacement to estimate confidence in tree nodes (Efron et al. 1996) - which is lacking in the implementation of ANGSD.

To address the possibility that the over-representation of domestic goat or bezoar may have altered the estimated phylogeny, IBS calculation was repeated using just four goat/bezoar genomes, using the same per-site filter (50% of individuals covered). The recovered tree (Appendix Figure 6.23) was identical to that presented in Figure 6.7 with two distinctions: the positions of both *Capra sibirica* samples. The suspected-hybrid individual Sibirica1 was placed as an outgroup to all samples other than Falconeri1, while the remaining *C. sibirica* genome Sibirica2 falls within the "ibex" apart from other *Capra* species. This differences in the computed phylogenies suggests that *Capra sibirica* variation, and non-domestic *Capra* variation in general, was poorly captured in the initial IBS computation. However, the general structure of the neighbour-joining phylogeny held between trees: the primary division of non-*Capra falconeri* samples occurring between domestic goat/bezoar and the "ibex" species.

Other technical issues may have caused a bias in results. The genomes here were generated using different sequencing approaches: Illumina pair-end and Illumina single-end sequencing. Additionally, DNA libraries were prepared in different laboratories, using different library protocols: USER- and non USER-treated double-strand libraries for the ancient and historic samples (Meyer & Kircher 2010; Rohland et al. 2015), and the longer-read libraries of the modern tissue-extracted DNA. *D* statistics using data from different technology can produce false positive signals of affinity (Prüfer et al. 2012); although all genomes here were produced using Illumina HiSeq technology, it is possible that batch effects between laboratories affected the test statistic. Reference bias may have again contributed to error in *D* statistic estimation, reducing the frequency at which non-*Capra aegagrus*/*hircus* reads containing derived alleles align to the reference genome. This would break the null hypothesis that when the phylogeny is true, ABAB and BABA sites will be sampled at a roughly-equal frequency. Back mutation may also affect patterns of allele sharing between different species and contribute to error in *D* statistic estimation.

The approach taken here did not exploit recent methods to estimate the directionality of gene flow. The D_{FOIL} statistic is one such example, which exploits five-taxon phylogenies in order to polarize admixture (Pease & Hahn 2015). This method utilizes the change in relationship between members of the tree not directly involved in admixture *i.e.* the change in affinity to an admixed population a second related population will exhibit. The difference in the magnitude of the change in allele sharing between two possible introgressors can indicate the population most likely to have admixed directly. Another developed method uses expected coalescence times to estimate when admixture occurred and also the direction (D_1 and D_2) (Hibbins & Hahn 2018). Both the D_{FOIL} and D_2 statistics have limitations regarding the topologies of tests, which would present an issue here as uncertainty remains as to the exact *Capra* phylogeny. Finally, the detection of an undescribed Tur-like population in Epipaleolithic Turkey suggests that other unsampled ghost populations may exist, and may have admixed in an asymmetric fashion with different *Capra* and domestic goat group. Such gene flow could interfere with the standard *D* test and alternative statistics.

Conclusion

The phylogeny of the genus *Capra* has proved difficult in the past to resolve with morphological cladistic approaches, and here the problem is found to be similarly challenging with both mitochondrial and nuclear genome data. The mtDNA phylogeny is in broad agreement with previously published work, and indicates substantial interspecies gene flow during divergence or incomplete lineage sorting. A tree constructed from nuclear genome identity-by-state suggests a major division between the monophyletic ibex species and bezoar/domestic goat. The position of the markhor *Capra falconeri* was not resolved, which may be the consequence of admixture with domestic species, admixture with a source of ancestral alleles, or sample misidentification. African, European, and Caucasian ibex form clades, as expected under simple models of speciation by vicariance. Unexpectedly, a genome from the Epipaleolithic site of Direkli Cave in the Taurus Mountains shows high affinity with the Caucasian ibex clade, despite having a bezoar-type mitochondrion. These results demonstrate the current lack of understanding of the phylogenomic history of the genus *Capra*, which could be improved with a greater survey of wild goat species genetic diversity.

Chapter 6 References

Alberto, F.J. et al., 2018. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

Bickhart, D.M. et al., 2017. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature genetics*, 49(4), pp.643–650.

Biebach, I. & Keller, L.F., 2009. A strong genetic footprint of the re-introduction history of Alpine ibex (Capra ibex ibex). *Molecular ecology*, 18(24), pp.5046–5058.

The Broad Institute, 2018. *Picard Tools*, The Broad Institute. Available at: https://broadinstitute.github.io/picard/.

Crégut-Bonnoure, E., 1991. Intérêt biostratigraphique de la morphologie dentaire de Capra (Mammalia, Bovidae). *Annales zoologici Fennici*, 28(3/4), pp.273–290.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), pp.1792–1797.

Efron, B., Halloran, E. & Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), pp.13429–13434.

Ejigu, D., Bekele, A. & Powell, L., 2017. Walia ibex have increased in number and shifted their habitat range within Simien Mountains National Park, Ethiopia. *Journal of Mountain Ecology*, 10(0). Available at: http://www.mountainecology.org/index.php/me/article/view/212.

Engländer, H., 1986. Capra pyrenaica Schinz, 1838--Spanischer Steinbock, Iberiensteinbock. In F. Niethammer & F. Krapps, eds. *Handbuch der Säugetiere Europeas*. Wiesbaden: AULA-Verlag, pp. 405–420.

Fedosenko, A.K. & Blank, D.A., 2001. Capra sibirica. *Mammalian Species*, pp.1–13.

Folch, J. et al., 2009. First birth of an animal from an extinct subspecies (Capra pyrenaica pyrenaica) by cloning. *Theriogenology*, 71(6), pp.1026–1034.

Geberemedhin, B. & Grubb, P., 2008. Capra walie. The IUCN Red List of Threatened Species 2008. *IUCN*. Available at: http://www.iucnredlist.org/details/3797/0.

Gebremedhin, B. et al., 2009. Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian walia ibex. *Animal conservation*, 12(2), pp.89–100.

Geist, V., 1990. Goat Antelopes. In D. Macdonald, ed. *Hoofed Mammals (All the World's Animals)*. New York: Torstar Books, pp. 144–149.

Gray, A.P., 1954. *Mammalian Hybrids: A Check-list with Bibliography*, Commonwealth Agricultural Bureaux.

Grossen, C. et al., 2014. Introgression from Domestic Goat Generated Variation at the

Major Histocompatibility Complex of Alpine Ibex. *PLoS genetics*, 10. Available at: http://dx.doi.org/10.1371/journal.pgen.1004438.

Grossen, C. et al., 2018. Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. *Evolutionary applications*, 11(2), pp.123–139.

Groves, C. & Grubb, P., 2011. *Ungulate Taxonomy*, Baltimore: JHU Press.

Guindon, S. et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), pp.307–321.

Hammer, S.E., Schwammer, H.M. & Suchentrunk, F., 2008. Evidence for introgressive hybridization of captive markhor (Capra falconeri) with domestic goat: cautions for reintroduction. *Biochemical genetics*, 46(3-4), pp.216–226.

Hassanin, A. et al., 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *Journal of molecular evolution*, 68(4), pp.293–310.

Hassanin, A. et al., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes rendus biologies*, 335(1), pp.32–50.

Hassanin, A., Pasquet, E. & Vigne, J.-D., 1998. Molecular Systematics of the Subfamily Caprinae (Artiodactyla, Bovidae) as Determined from Cytochrome b Sequences. *Journal of mammalian evolution*, 5(3), pp.217–236.

Hecker, H.M., 1976. *The Faunal Analysis of the Primary Food Animals from Pre-pottery Neolithic Beidha (Jordan)*. Columbia University. Available at: https://elibrary.ru/item.asp?id=7162726.

Heptner, V.G., Nasimovich, A.A. & Bannikov, A.G., 1961. *Mammals of the Soviet Union* V. G. Heptner & N. P. Naumov, eds., Moscow: Vysshaya Shkola.

Hibbins, M.S. & Hahn, M.W., 2018. Population genetic tests for the direction and relative timing of introgression. *bioRxiv*, p.328575. Available at: https://www.biorxiv.org/content/early/2018/05/23/328575.

Huffman, B., 2004. Capra falconeri. *Ultimate Ungulate*. Available at: http://www.ultimateungulate.com/Artiodactyla/Capra_falconeri.html.

Jónsson, H. et al., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* , 29(13), pp.1682–1684.

Kazanskaia, E.I., Kuznetsova, M.V. & Danilkin, A.A., 2007. [Phylogenetic reconstructions in the genus Capra (Bovidae, Artiodactyla) based on the mitochondrial DNA analysis]. *Genetika*, 43(2), pp.245–253.

Keane, T.M. et al., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology*, 6, p.29.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R., 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics*, 15, p.356.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Li, X. et al., 2017. Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. *Scientific reports*, 7(1), p.15142.

Manceau, V. et al., 1999. Systematics of the genus Capra inferred from mitochondrial DNA sequence data. *Molecular phylogenetics and evolution*, 13, pp.504–510.

Mason, I.L., 1986. *Evolution of Domesticated Animals*, John Wiley & Sons, Incorporated.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297–1303.

Meyer, M. & Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols*, 2010(6), p.db.prot5448.

Nievergelt, B., 1981. *Ibexes in an African Environment: Ecology and Social Systems of the Walia Ibex in the Simen Mountains, Ethiopia*, Berlin: Springer-Verlag.

Nievergelt, B., 1998. Observations on the walia ibex, the klips pringer and the Ethiopian wolf. *Walia: J. Ethiop. Wildl. Nat. Hist. Soc*, pp.44–51.

Palkopoulou, E. et al., 2018. A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences of the United States of America*. Available at: http://dx.doi.org/10.1073/pnas.1720554115.

Pease, J.B. & Hahn, M.W., 2015. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Systematic biology*, 64(4), pp.651–662.

Perez, J.M. et al., 2002. Distribution, status and conservation problems of the Spanish Ibex, Capra pyrenaica (Mammalia: Artiodactyla)+. *Mammal review*, 32(1), pp.26–39.

Pidancier, N. et al., 2006. Evolutionary history of the genus Capra (Mammalia, Artiodactyla): discordance between mitochondrial DNA and Y-chromosome phylogenies. *Molecular phylogenetics and evolution*, 40(3), pp.739–749.

Pilgrim, G.E., 1947. The Evolution of the Buffaloes, Oxen, Sheep and Goats. *Zoological journal of the Linnean Society*, 41(279), pp.272–286.

Prüfer, K. et al., 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature*, 486(7404), pp.527–531.

Rivals, F., 2006. Discovery of Capra caucasica and Hemitragus cedrensis (Mammalia, Bovidae) in the Late Pleistocene levels of the Caune de l'Arago (Tautavel, France): Biochronological implication in the Mediterranean Basin context. *Geobios . Memoire special*, 39, pp.85–102.

Rohland, N. et al., 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 370(1660), p.20130624.

Ropiquet, A. & Hassanin, A., 2006. Hybrid origin of the Pliocene ancestor of wild goats. *Molecular phylogenetics and evolution*, 41(2), pp.395–404.

Schaller, G.B., 1977. *Mountain Monarchs: Wild Sheep and Goats of the Himalaya*, London: University of Chicago Press.

Schmitt, J. & Ulbrich, F., 1968. Die Chromosomen verschiedener Caprini Simpson, 1945. *Zeitschrift für Säugetierkunde*, 33, pp.180–186.

Schubert, M. et al., 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics*, 13, p.178.

Shackleton, D.M., 1997. *Wild sheep and goats and their relatives: status survey and conservation action plan for Caprinae*, IUCN. Available at: https://portals.iucn.org/library/sites/library/files/documents/1997-006.pdf.

de Smet, K. et al., 2008. Capra nubiana. The IUCN Red List of Threatened Species 2008. *IUCN*. Available at: http://www.iucnredlist.org/details/3796/0.

Stüwe, M. & Nievergelt, B., 1991. Recovery of alpine ibex from near extinction: the result of effective protection, captive breeding, and reintroductions. *Applied animal behaviour science*, 29(1), pp.379–387.

Uerpmann, H.-P., 1987. *The ancient distribution of ungulate mammals in the Middle East: fauna and archaeological sites in Southwest Asia and Northeast Africa*, Wiesbaden: L. Reichert Verlag.

Weinberg, P., 2008. Capra caucasica. The IUCN Red List of Threatened Species 2008. *IUCN*. Available at: http://www.iucnredlist.org/details/3794/0.

Weinberg, P.J., 2002. Capra cylindricornis. *Mammalian Species*, pp.1–9.

Weinberg, P.J., Akkiev, M.I. & Buchukuri, R.G., 2010. Clineal variation in Caucasian tur and its taxonomic relevance. *Galemys*, 22(1), pp.375–394.

Zalikhanov, M.C., 1967. *Tur in Kabardin-Balkaria*, Nal'chik: Kabardin-Balkarian Publishers.

Chapter 7: Conclusion and future directions

Overview of thesis findings

Genetic structure in Neolithic goat

Multiple lines of genetic evidence suggest that goat populations in the Neolithic were highly structured. Neolithic goat from western Turkey and southeast Europe overwhelmingly carry the A mitochondrial haplogroup, and show high affinity with modern goat from Europe. Principal components analysis, model-based ancestry estimation, and allele-sharing patterns indicate the Neolithic Levantine goat share ancestry with this western group, but are also differentiated. These Levantine samples also all have the F haplogroup, one not typically associated with domestic goat populations. Further discontinuity is seen between these genomes and those of Neolithic Iran, which show greater affinity with modern Asian (Iranian and Chinese) goat. Their mitochondrial lineages are highly diverse, showing haplogroups now infrequent in domestic goat (B, D, and G).

The observed differentiation between these regional populations may be a consequence of drift and isolation by distance as the practice of goat herding diffused from a specific locale. However, this explanation does not fully agree with the current zooarchaeological evidence, which suggests a more dispersed emergence of goat herding in southwest Asia. Direct detection of differential wild influx to Neolithic goat populations (see below) also supports this model.

Localized wild recruitment of bezoar to domestic stock

The discovery of well-preserved samples from the Epipaleolithic hunting site of Direkli Cave, southern Turkey, was highly fortuitous; their genomes permitted the direct observation of gene flow from wild bezoar populations to specific Neolithic goat populations. Treemix, qpGraph, and *D* statistics all indicate that western Neolithic populations received ancestry from a population related to these wild goats, unlike those of Iran. The legacy of this admixture has consequences even today, with European goat exhibiting the Direkli-like ancestry - although not to the same extent as goat from Neolithic Serbia, implying some degree of population turnover or structure within European Neolithic goat herds. qpGraph modelling also suggests that Neolithic Levantine goat also bear this ancestry, and that eastern

Neolithic goat may have had input from a distinct wild population (or populations). The differential affinity of the >47,000 BP Hovk1 sample to Neolithic goat demonstrates that bezoar genetic structure was pre-LGM in origin, the traces of which linger in the genomes of present-day domestic goat.

These two findings hint of a goat domestication process that was diffuse and dynamic, with a variegated domestic genome emerging from local traditions and livestock management strategies.

A reduction in goat population structure following the Neolithic

Following the Neolithic, a shift in the genetic makeup of goat populations in southwest Asia appears to have occurred. This is strikingly evident in the mitochondrial genome: haplogroup A now appears at a high frequency in the Levant and eastern regions, with some persistence of other mtDNA lineages. In eastern regions there is evidence that this turnover coincided with an increase in western-like ancestry in nuclear genomes. The general persistence of eastern Neolithic-like ancestry despite the change in the mtDNA pool implies that continuity was maintained by local males. In the Levant, the previous haplogroup F sequences were wholly replaced by haplogroup A and also those found in Neolithic Iran. A range of measures of genetic affinity (outgroup f_3, NGSadmix, PCA) indicate that Chalcolithic, Bronze Age, and Iron Age Levantine goat are distinct from those of the Neolithic, showing more shared ancestry with eastern Neolithic populations. Bronze Age genomes from central Turkey also appear to contain a greater proportion of eastern Neolithic-like ancestry compared with Neolithic genomes from western Turkey/southeast Europe, inferred from *D* statistics, f_3 outgroup values, qpGraph modelling, and their positions in PCA space.

These findings reflect those reported in human studies, where ancient genomes from southwest Asia have revealed Neolithic population structure followed by genetic homogenization in the succeeding eras (Broushaki et al. 2016; Lazaridis et al. 2016; Harney et al. 2018). Nuclear genome F_{ST} between goat populations also decrease substantially following the Neolithic (Appendix Table 4.10); the parallels between the partner species' genomes during these periods suggest that the underlying demographic events were correlated. The establishment of trade networks may have promoted the movement of both people and livestock across large distances, facilitating admixture between previously-disparate populations. Urbanization may have also played a role in this process of greater

connectivity. The absence of genomes of either species from Mesopotamia, the birthplace of city-states in west Eurasia, currently prevents assessment of the role played by that region.

Early signatures of selection in goat

The analyses presented in Chapter 5 represent the earliest reported potential signatures for selection in livestock species by early farmers. The F_{ST} outlier approach, conditioned on reduced genetic diversity compared to modern bezoar, identified a total of nineteen distinct regions, two of which were shared across western and eastern Neolithic genomes. Both loci overlapped with or fell nearby genes well-established to affect pigmentation in domestic animals: *KIT* and *KITLG*. Although it is possible that the detected F_{ST} patterns may have arisen from another process *e.g.* large diversity at these loci in modern bezoar, or a change in recombination rates along the genome, these genes are plausible targets of selection by early farmers, due to the inherently visible phenotypes both genes influence. Other genes overlapping outlier regions represent potential adaptation to the anthropogenic domestic herd environment, including fertility traits, toxin and hormone metabolism. This analysis was limited by the fact of that the Neolithic genomes derived from a limited number of archaeological sites, lacked an optimal wild goat reference population, and did not account for factors such as demographic history. Thus, the findings presented here should be treated as preliminary (or not generalizable) and addressed with LD- and demography-aware methods using a more diverse panel of ancient genomes.

The *Capra* phylogeny: unresolved questions

Genetic exchange between domestic goat populations and other *Capra* species likely occurred, based on differential allele sharing between those species and Neolithic goat populations. There is some evidence of gene flow between European (particularly Alpine) ibex populations and European goat, and between African ibex and Moroccan goat. The Caucasian Tur appear to have interacted with domestic populations to a greater extent than other *Capra* species. This may be confounded by admixture with a population related to the Direkli bezoar, which plausibly experienced gene flow with a "Tur-like" population represented by Direkli4. Differentiating admixture from different *Capra* species proved difficult with the approach taken here, as did discerning the directionality of admixture. Some puzzling results remain unexplained, including the excess of ancestral alleles of the Togolese goat (Tog) and the difference in *Capra* derived allele sharing in Bronze Age Turkish goat.

The phylogeny of *Capra* was only partially resolved, but appears to be consistent with that observed in Y chromosome trees: a primary divide between "ibex" species and a clade consisting of bezoar and goat, and possibly markhor. The markhor and also the Siberian ibex proved to be problematic in placing. Markhor may represent an outgroup to the genus or a sister species to goat/bezoar. Despite their mtDNA clustering with the Himalayan Tahr, Siberian ibex fall within the "ibex" clade, either as an outgroup to the entire clade or to European and Caucasian Tur. It is evident that a larger sample of each species, preferably from their natural habitat, is required to achieve an accurate estimate of the relationships between *Capra* species; ancient genomes of each would also allow admixture with domestic goats to be quantified.

The *Capra* genomes presented here give further indications of our collective ignorance of the genus and its history. The discovery of a "Tur-like" nuclear genome in Direkli Cave, securely dated to a late Epipaleolithic context and accompanied by "Tur-like" mtDNA sequences from other Direkli samples, is evidence of previously-unknown complexity within the genus. This discovery may represent a now-extinct remnant of the Caucasian Tur, or the lineage may actually persist to this day; at the risk of repetition, more extensive sampling of wild *Capra* populations across their fragmented distributions is required to better understand the evolutionary history of the genus.

Outstanding issues and future directions

Prospective datasets

A major limitation in the analyses of the latter chapters of this thesis was the genomic resources available for goat. Prior to 2017, whole genome data was published from goat populations from just China, Morocco, and Australia (Dong et al. 2013; Dong et al. 2015; Benjelloun et al. 2015). The publications of (Li et al. 2017) and (Alberto et al. 2018) expanded sequenced populations to include France, Italy, Iran, China, and also Iranian bezoar. This leaves goat populations from large swathes of the Eurasia continent unsequenced, restricting the breadth of analyses that can be performed. The problem is more pronounced in bezoar ibex and other *Capra* species; wild goat genomes are currently restricted to certain regions of Iran, while no whole genome sequence data is available across the rest of the genus. This will likely change in the near future, and greatly expand the scope of inference possible using ancient genomes. For example, modern genomic data from Central Asia (*i.e.* Pakistan, Uzbekistan), Turkey, and other parts of Europe will allow a more complete model of goat domestication history to be constructed, using ancient genomes to aid contextualization.

A 1,000 genome project for goat was established recently (Amills et al. 2017), and will dramatically improve research possibility in goat genomics. In the shorter term, there is the prospect of a worldwide 50K SNP dataset becoming available to the goat genomics community. This will be composed of genome-wide data from approximately 4,000 individuals across 144 breeds, and will hopefully increase our understanding of global goat genetic diversity. Both datasets also imply that imputation of ancient genomes will become a possibility in the future, allowing the many low coverage genomes presented here to be fully leveraged, and haplotype-aware methodologies employed to further dissect the goat domestication process.

Alternative approaches

The methods taken in this thesis to investigate the genetic history of goat and the *Capra* genus were by no means exhaustive; several avenues of analyses using this genomic data remain unexplored. For example, the demographic history of goat and bezoar was not examined with the current data. Modern genomes of domestic goat and bezoar suggest that the wild progenitor shows less diversity than domestic populations (Alberto et al. 2018). It is

unknown whether this a consequence of recently population decline in bezoar, recent population expansion in goat, or due to gene flow from divergent sources to domestic herds. The samples presented here are ideally placed to investigate these possibilities. Measures of genomic diversity such as runs of homozygosity in high coverage individuals or pairwise differences within populations are possible means of addressing this. If deeper sequencing of these ancient genomes was performed, powerful demographic techniques such as G-PhoCS or MSMC would also be viable (Gronau et al. 2011; Schiffels & Durbin 2014).

The genomic sculpting by the forces of selection during domestication also remain under-analyzed. The method described in Chapter 5 focused entirely on Neolithic populations, with the intent of identifying the strongest and earliest genetic consequences of goat herding. As such, contiguous regions of high divergence and a decline in genetic diversity were criteria to identify possible targets of selection. This leaves many unaddressed processes unexplored. For example, balancing selection may have promoted genetic diversity at certain loci, perhaps mediated by gene flow from other bezoar or *Capra* populations, as reportedly occurred in the Alpine ibex (Grossen et al. 2014). Domestic goat populations may have also accumulated deleterious mutations due to a relaxation of selection pressures in the herd environment. This is thought to have occurred in horses (Schubert et al. 2014). Other methods of investigating the genetic changes throughout the domestication process include the population branch statistic, to identify population-specific differentiated regions relative to a sister population and an outgroup, to identify population-specific differentiated regions, and the LSD ("Levels of exclusively Shared Differences") statistic, which can incorporate multiple populations and time-series data (Librado et al. 2017).

Finally, the Y chromosome was not exploited as a marker of male-specific processes. East Asian breeds carry a unique haplogroup Y2B not found in western Eurasia (Waki et al. 2015); Turkish breeds have a unique Y2C haplogroup, while Y1B is primarily found in North Europe. Based on the nuclear genome data presented in this thesis, it would be expected that the samples from Neolithic Serbia bear the Y1B haplogroup, while those from eastern Iran have Y2C. This genomic time series could indicate if regional male-line stasis was the norm or if they showed the same dramatic changes as the mitochondria.

Old questions

Several lines of inquiry are unanswered or were raised by the results of presented in this thesis. The light shed on the goat domestication process by the genomes reported here also highlight the areas still in shadow. The nature of ancient DNA preservation results in some regions being under-sampled or not sampled at all. In this thesis, several regions stand out as having not yielded goat DNA, which are pertinent to understanding the early history of goat management. These include eastern and central Turkey, northern Levant, the Caucasus region, and central/northern Iran. Genomes from these regions would allow the genetic consequences of the spread of goat herding to be quantified *i.e.* if it was associated with a substantial loss of diversity. In addition, the timing and nature of the interaction of goat herds with the wild could be properly dissected; for example, exactly how the earliest goat herds in the Levant related to those of eastern Turkey and the Zagros.

Ancient bezoar genomes are also under-sampled. Pre-domestic (*i.e.* prior to 8,000-8,500 BC) genomes from three regions in particular would be highly informative: the Levant (preferably southern regions), eastern Turkey, and central/eastern Iran. In the case of the Levant and Iran, such wild genomes would contextualize the Neolithic samples reported from those regions, and more-precisely quantify local recruitment of bezoar. Eastern Turkish bezoar genomes would put those wild samples and also those from Direkli Cave themselves in context, as the geographic structure of *Capra aegagrus* prior to the domestication process is still very-much unclear.

The nature of the spread of goat herding in Europe also presents outstanding questions. The Neolithic Serbian genomes appear to show greater Direkli ancestry than a Bronze Age genome from Britain or modern French genomes, as assessed by Treemix and f_4 ratios. Admixture from European ibex was detected in only a subset of these genomes, eliminating this as an explanation for the difference in derived alleles shared with Direkli1-2. An alternative explanation, as discussed in Chapter 4, would be if the Neolithic Serbian goat are not the direct ancestors of the other European goat genomes sampled here. Instead, a different population with a smaller proportion of Direkli-like ancestry may have primarily contributed to those other European goats. These populations may coincide with the two main routes of the spread of the Neolithic in Europe: the Danubian and the Mediterranean routes. This has already been suggested based on goat Y chromosome frequencies (Vidal et al. 2017). To address this, genomes from both central and southern Europe would be required.

Additionally, it is unknown if the dramatic change in European human populations during the Bronze Age (Gamba et al. 2014; Allentoft et al. 2015) had an analogous event in livestock species; a time series of goat genomes from the same region would allow this possibility to be investigated.

New data

To address these outstanding questions, an additional 52 bone elements were screened (method as described in Chapter 2), from a range of regions and time periods including *c.* 10,000 BC Körtik Tepe in southeast Turkey, Neolithic sites in central Turkey, and several pre-Holocene Iranian caves (Appendix Table 7.1). 19 goat or *Capra* were identified, which were supplemented with 7 samples screened and determined to be goat by Andrew Hare and Marta Verdugo. Unfortunately, no goat samples with preserved endogenous DNA were obtained from eastern Turkey or pre-domestic Iranian contexts. A subset of the remainder were selected for Illumina HiSeq sequencing or mtDNA enrichment as described in Chapter 2, aligning to ARS1 according to the method described in Chapter 6. A summary of these 26 samples and sequencing results are presented in Table 7.1. Nuclear genome data was generated for 14 samples, giving a median ×-fold coverage of 0.295 (Appendix Table 7.3). Mitochondrial sequences were constructed for 24 samples, having a median ×-fold coverage of 75.71 (Appendix Table 7.4).

Mitochondrial genomes from Central and western Turkey confirm the presence of the A haplogroup in the region during the Neolithic (Figure 7.1). A surprising discovery is the haplogroup in the site of Tepe Abdul Hosein in the Zagros Mountains, one of the earliest Neolithic sites in the region dating to *c.* 8,000 cal BC (Pullar et al. 1990; Broushaki et al. 2016). This sample, Abdul6, dates to 8,166-7,606 cal BC (two sigma confidence) and is not an outgroup to previously-reported A haplogroup sequences (Appendix Figure 7.1). A plurality of mtDNA haplogroups are found among the other six sequences recovered from the site including G, G' and B, in line with those previously found in Neolithic Iran (Chapter 3). Two sequences from Montou, a Bronze Age site in the French Pyrenees, form a clade with other domestic C haplogroup sequences. All Israeli goat mtDNA sequences from Iron Age and later periods are of the A clade, similar to the frequency observed in modern Levantine goat (Naderi et al. 2008). A single mtDNA from the Sasanian (200-600 AD) levels of the Gorgan Wall, eastern Iran, is also of the A haplogroup. Of the three mtDNA from medieval Tamara Fort, Georgia, two form a clade with the East Caucasian Tur historic sample Tur2

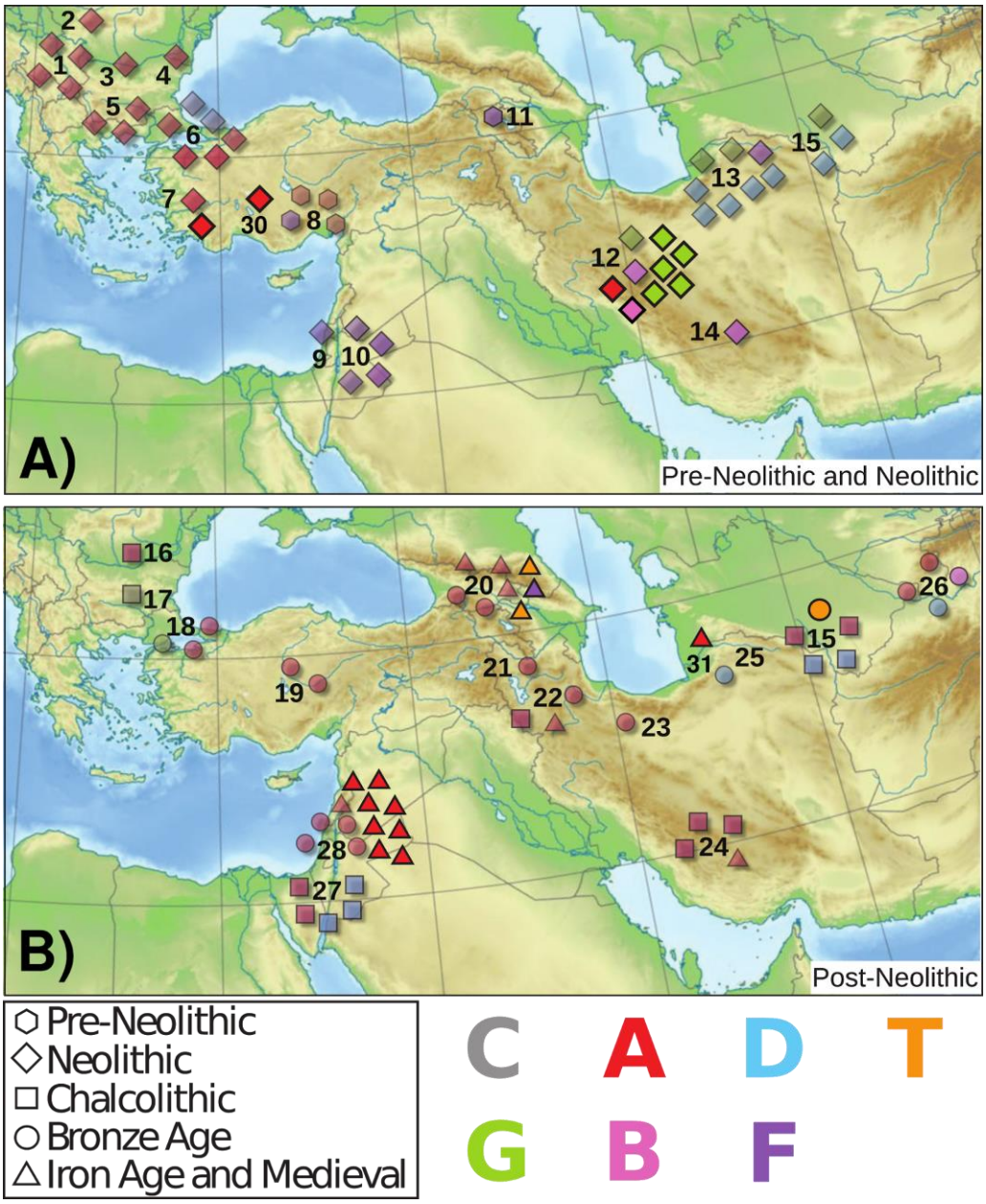


Figure 7.1 - Chapter 7 mtDNA, added to previously sequenced ancient *Capra* mtDNA. Samples presented in Figure 3.4 are indicated here using faded symbols; new samples are marked by thick borders. Site identifiers are the same as in Figure 2.5 and Table 2.2, with the addition of those listed in Appendix Table 7.5. 7. Ulucak Höyük, 12B. Tepe Abdul Hosein, 15B. Ulugh Depe, 20B. Dariali Tamara Fort (Kazbegi), 30. Erbaba Höyük, 31. Gorgan Wall, 28E. Tel'Aroer, 28F. Banias, 28G. Jerusalem, 28H. Nizzana, 28I Sepphoris, 28J. Tel Zahara. Not shown: two haplogroup C sequences from Bronze Age Montou, southern France (32). Map adapted from (Виктор 2010).

(Caucasus1 and 2), while the remaining (Caucasus3) falls within the primarily-wild clade F. Finally, a sequence (Ulugh1) from Bronze Age Ulugh Depe, Turkmenistan, is an outgroup to all Tur-like mtDNA.

To perform a preliminary investigation of the nuclear genome affinities of these sequenced *Capra*, an IBS matrix was constructed for all genomes of at least 0.01× coverage (according to the method described in Chapter 6). A neighbour- joining tree of this matrix indicated that Caucasus1, the only nuclear genome generated from the Tamara Fort samples presented here, formed a clade with Tur2, indicating the sample was likely an East Caucasian Tur (Appendix Figure 7.2). Samples outside of the *Capra aegagrus/hircus* clade were removed and multidimensional scaling performed on IBS matrix (Appendix Text 7.1), and the resulting dimensions plotted (Figure 7.2). Two genomes showed greater affinity with wild bezoar than with domestic goat: the Bronze Age Turkmen sample Ulugh1 and a single sample from Tepe Abdul Hosein, Abdul4, and likely reflect hunted bezoar by the inhabitants of these settlements. The remaining genomes from Tepe Abdul Hosein form a cluster with the Lur12, the previously-reported sample from the same site. This group is distinct from Fars2-5, the Neolithic genome from *c.* 7,000 BC Rahmat Abad in southwest Iran, and from the Neolithic sites of Sang-e Chakmaq and Monjukli Depe in northeast Iran and Turkmenistan respectively. Also deriving from eastern Iran but the later Sasanian time period (200-600 AD), Gorgan2 groups with a later medieval goat Darre2 from southeast Iran, the Bronze Age Qazvin1 from north central Iran, and the Bronze Age Azer3-5 from the Iranian Azerbaijan site of Tepe Hasanlu. This again illustrates the difference in genetic makeup of Neolithic compared with later goat genomes from Iran.

Ulucak3, a 0.34× genome from the western Turkish site of Ulucak Höyük and dating to 6,366-6,094 cal BC, clusters with slightly younger genomes from southeast Europe, consistent with spread from farming to Europe from southwest Asia (Lazaridis et al. 2016). Two Bronze Age genomes from southern France show high affinity with modern goat from France and Italy, and fall apart from modern and ancient samples from Britain and Ireland, Potterne1 and IOG. Genomes from the Levant show a similar change in genetic makeup as described in Chapter 4. Goat from ‘Ain Ghazal show a much greater affinity with those from Neolithic Turkey and Europe (and here show less affinity with modern African genomes than was observed in Figure 4.6, possibly a consequence in the different variant sites between ANGSD IBS and those defined in Chapter 4 using primarily modern genomes). Goat from Chalcolithic Levant are significantly shifted towards eastern genomes, which is seen to an

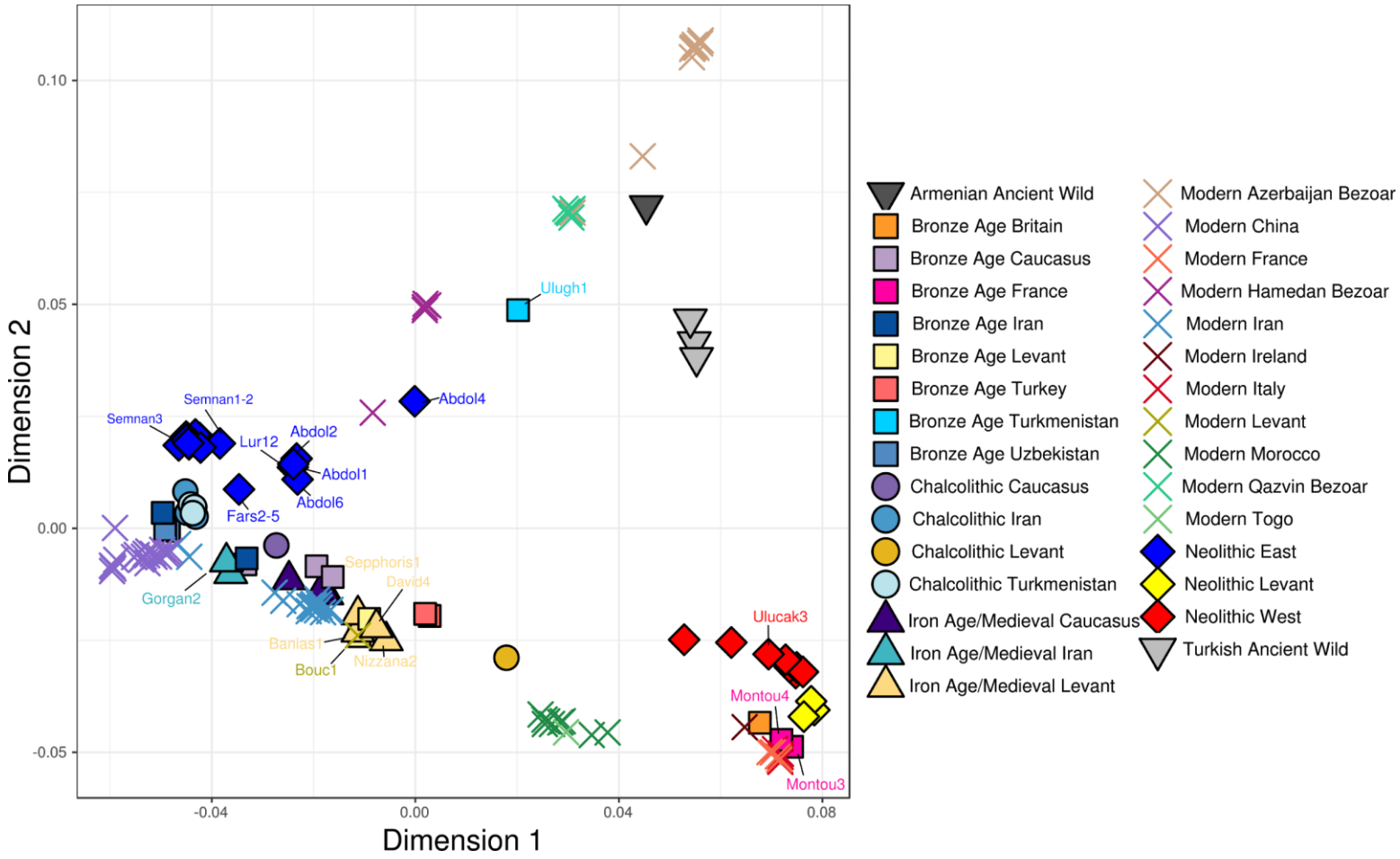


Figure 7.2 - IBS-MDS plot of ancient and modern goat/bezoar. Newly sequenced ancient goat/bezoar and key previously sequenced ancient goat are indicated on the plot. A plot with all samples labelled is presented in Appendix Figure 7.3.

even greater extent in the Bronze Age onwards. Four Iron Age or Classical period goat from Israel fall in a group with two Israeli Bronze Age genomes and also the modern Bouc1; these genomes are relatively close to those from modern Iran, Caucasus Bronze Age and later context, and also Bronze Age Turkey. These samples support apparent continuity from the Bronze Age onwards, following a dramatic turnover in goat populations at some point in the late Neolithic or Chalcolithic eras.

Together, these data indicate that despite some predictions being correct - *i.e.* the occurrence of haplogroup A in central Turkey - there are still undiscovered complexities in the populations involved in early goat management. The observation that haplogroup A (Abdul6) was found in the Zagros Mountains *c.* 8,000 cal BC contrasts its absence in modern day Zagros bezoar (Naderi et al. 2008), and demonstrates the potential issues of extrapolating present-day distributions of genetic diversity into the past. It is still to be determined the contribution, if any, this Tepe Abdul Hosein population played in domestic goat variation. Abdul6 dates to a few centuries prior to the earliest traces of wild goat management at Ganj Dareh (Zeder 2005). The site of Tepe Abdul Hosein itself is geographically close to Ganj Dareh, so it is possible that the Abdul Hosein caprines were also subject to a degree of human control. This or a related population may be the source of the domestic haplogroup A; however, Abdul6 is not an outgroup to all other haplogroup A sequences so this may not be the case (Appendix Figure 7.1).

More intriguing is the position of the remaining Tepe Abdul Hosein goats as the sister clade to all domestic goat in the IBS-nj phylogeny (Appendix Figure 7.2), suggesting that these represent the closest yet-sequenced genomes to the earliest managed goat. Even if the Tepe Abdul Hosein samples played no part in the composition of early herds, they are further evidence of the previously- undocumented complexity in the genetic diversity of goat in prehistoric southwest Asia. Ancient DNA analysis permits forgotten populations such as these and the Direkli Tur to be recovered, their genetic history reconstructed, and the deepening of our understanding of one of humanity’s oldest partner animals.

Table 7.1 - Summary of goat genomes presented in Chapter 7. Samples ages marked with * were directly radiocarbon dated; raw results are presented in Appendix Table 7.2. Endogenous percentage and autosomal coverage calculated after mapping quality 30 filter. Endog. = Endogenous. Cov. = Coverage. Hap. = Haplotype. C.D. = Cannot Determine.

Sample	Site	Region	Context	Age	Sex	Endog. %age	Autos. Cov.	mtDNA Cov.	% bases called	mtDNA hap.
Abdul1	Tepe Abdul Hosein	Luristan, Iran	Neolithic	~8,000 cal BC	M	20.49	0.14	47.91	99.90	G'
Abdul2	Tepe Abdul Hosein	Luristan, Iran	Neolithic	~8,000 cal BC	M	24.03	0.21	86.78	100.00	G
Abdul4	Tepe Abdul Hosein	Luristan, Iran	Neolithic	~8,000 cal BC	M	18.30	0.13	63.52	100.00	G
Abdul5	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	~8,000 cal BC	M	3.09	-	3.36	64.66	G
Abdul6	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	*8,166-7,606 cal BC	M	9.02	0.07	6.29	100.00	A
Abdul7	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	~8,000 cal BC	M	0.78	-	4.00	74.30	G'
Abdul8	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	~8,000 cal BC	M	0.62	-	-	-	-
Abdul9	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	~8,000 cal BC	F	2.62	-	6.84	97.09	B
Aroer1	Tel'Aroer	Negev Desert, Israel	Iron Age - Hellenistic	~1,000 - 31 BC	C.D.	0.45	-	-	-	-
Aroer2	Tel'Aroer	Negev Desert, Israel	Iron Age - Hellenistic	~1,000 - 31 BC	M	3.89	-	64.31	64.31	A
Banias1	Banias	Israel	Roman - Islamic	~63 BC - 634 AD	F	14.18	0.02	33.25	100.00	A

Banias2	Banias	Israel	Roman - Islamic	~63 BC - 634 AD	M	10.15	-	118.04	118.04	A
Caucasus1	Tamara Fort	Kazbegi, Georgia	Medieval	15th-21st cent. AD	M	27.87	0.49	64.63	98.64	T
Caucasus2	Tamara Fort	Kazbegi, Georgia	Medieval	15th-21st cent. AD	M	0.87	-	379.71	379.71	T
Caucasus3	Tamara Fort	Kazbegi, Georgia	Medieval	15th-21st cent. AD	M	1.13	-	464.51	464.51	F
David4	Jerusalem	Israel	Iron Age	8th cent. BC	M	18.15	1.02	245.69	99.51	A
Erbaba2	Erbaba Höyük	Turkey	Neolithic	*6,567-6,379 cal BC	M	0.66	-	394.13	394.13	A
Gorgan2	Gorgan Wall	Golestān, Iran	Sasanian	~400 AD	M	34.25	0.16	23.32	100.00	A
Montou3	Montou	France	Early Bronze Age	1900-1700 cal BC	M	30.83	0.88	147.96	100.00	C
Montou4	Montou	France	Late Bronze Age	1100-900 cal BC	F	47.39	1.91	155.34	100.00	C
Nizzana2	Nizzana	Israel	Byzantine - Islamic	~63 BC - 634 AD	M	42.31	0.46	169.55	100.00	A
Nizzana3	Nizzana	Israel	Byzantine - Islamic	~63 BC - 634 AD	C.D.	1.09	-	46.10	46.10	A
Sepphoris1	Sepphoris	Israel	Roman - Byzantine	~~63 BC - 400 AD	F	65.51	1.47	154.39	100.00	A
Ulucak3	Ulucak Höyük	Turkey	Neolithic	*6,366-6,094 cal BC	F	12.63	0.34	125.03	100.00	A
Ulugh1	Ulugh Depe	Turkmenistan	Bronze Age	~ 3,000 BC	M	21.25	0.25	14.08	97.45	T
Zahara1	Tel Zahara	Israel	Roman - Islamic	1,291-1,917 AD	C.D.	0.97	-	377.66	377.66	A

Chapter 7 References

Alberto, F.J. et al., 2018. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

Allentoft, M.E. et al., 2015. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555), pp.167–172.

Amills, M., Capote, J. & Tosser-Klopp, G., 2017. Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. *Animal genetics*, 48(6), pp.631–644.

Benjelloun, B. et al., 2015. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Frontiers in genetics*, 6, p.107.

Broushaki, F. et al., 2016. Early Neolithic genomes from the eastern Fertile Crescent. *Science*. Available at: http://dx.doi.org/10.1126/science.aaf7943.

Dong, Y. et al., 2015. Reference genome of wild goat (capra aegagrus) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC genomics*, 16, p.431.

Dong, Y. et al., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nature biotechnology*, 31(2), pp.135–141.

Gamba, C. et al., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications*, 5, p.5257.

Gronau, I. et al., 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics*, 43(10), pp.1031–1034.

Grossen, C. et al., 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. *PLoS genetics*, 10. Available at: http://dx.doi.org/10.1371/journal.pgen.1004438.

Harney, É. et al., 2018. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nature communications*, 9(1), p.3336.

Lazaridis, I. et al., 2016. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617), pp.419–424.

Librado, P. et al., 2017. Ancient genomic changes associated with domestication of the horse. *Science*, 356(6336), pp.442–445.

Li, X. et al., 2017. Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. *Scientific reports*, 7(1), p.15142.

Naderi, S. et al., 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America*, 105, pp.17659–17664.

Pullar, J. et al., 1990. *Tepe Abdul Hosein: a Neolithic site in Western Iran : excavations 1978*, Oxford: B.A.R.

Schiffels, S. & Durbin, R., 2014. Inferring human population size and separation history from multiple genome sequences. *Nature genetics*, 46(8), pp.919–925.

Schubert, M. et al., 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(52), pp.E5661–9.

Vidal, O. et al., 2017. Differential distribution of Y-chromosome haplotypes in Swiss and Southern European goat breeds. *Scientific reports*, 7(1), p.16161.

Waki, A. et al., 2015. Paternal phylogeography and genetic diversity of East Asian goats. *Animal genetics*, 46(3), pp.337–339.

Zeder, M.A., 2005. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. In J. Peters, A. V. Den Driesch, & D. Helmer, eds. *The First Steps of Animal Domestication: New Archaeozoological Approaches*. Oxford: Oxbow Books, pp. 125–146.

Виктор, В., 2010. *Relief Map of Middle East*, Available at: https://commons.wikimedia.org/wiki/File:Relief_Map_of_Middle_East.jpg.

Appendix

Appendix Material for Chapter 1

Appendix Table 1.1 - Attributions of photos used in Figure 1.1.

Photo	Species	URL	Creative Commons License	Author
Top Left	*Capra aegagrus*	https://commons.wikimedia.org/wiki/File:Athina_Ogrod_Narodowy_kozy.jpg	Creative Commons Attribution-Share Alike 3.0 Unported	Andrzej Otrębski
Bottom Left	*Capra hircus*	https://commons.wikimedia.org/wiki/File:Bouc_commun_proven%C3%A7al.JPG	Public Domain	Mimicki
Top Centre	*Capra hircus*	https://commons.wikimedia.org/wiki/File:Capra-aegagrus-hircus-6.jpg	Creative Commons CC0 1.0 Universal Public Domain Dedication	MartinThoma
Bottom Centre	*Capra hircus*	https://commons.wikimedia.org/wiki/File:Chevres_Rove_BdRhone.jpg	Creative Commons Attribution-Share Alike 3.0 Unported	Damien Hardy
Top Right	*Capra hircus*	https://commons.wikimedia.org/wiki/File:Goats_on_Road.jpg	Creative Commons Attribution-Share Alike 3.0 Unported	Syed Ahsan Raza Kazmi
Bottom Right	*Capra hircus*	https://commons.wikimedia.org/wiki/File:Mr._Goat.jpg	Creative Commons Attribution-Share Alike 3.0 Unported	HabibullahQureshi

Appendix Material for Chapter 2

Appendix Table 2.1 - All screened bone elements presented in Chapter 2. Samples are ordered by Internal Name. Genetic species was identified using Fastq Screen. Endogenous percentage was calculated as the number of reads aligning to the appropriate reference genome following MapQ 30 filtering, divided by the total number of reads following adaptor trimming. Samples with internal names beginning in AH or VEM were screened by Andrew Hare or Victoria Mullin respectively. End. = Endogenous.

Sample Name	Internal Name	Site	Region	Context	Genetic Species	End. %age	Skeletal Element	Collaborator
Semnan8	AH004	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	Goat	9.91	Petrous	Marjan Mashkour
Semnan9	AH005-1	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	Goat	53.27	Petrous	Marjan Mashkour
Semnan10	AH005-2	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	Goat	19.17	Petrous	Marjan Mashkour
Semnan13	AH005-3	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	Goat	34.7	Partial Petrous	Marjan Mashkour
Semnan17	AH006-4	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic	Goat	7.37	Partial Petrous	Marjan Mashkour
Ghosh5	AH032	Abu Ghosh	Judean Hills, Jordan	Mid PPNB	Goat	0.04	Petrous	Liora Horwitz
Kortik1	AH038_1	Körtik Tepe	Turkey	PPNA	Sheep	0.23	Petrous	Levant Attici
Kortik2	AH038_2	Körtik Tepe	Turkey	PPNA	-	-	Petrous	Levant Attici
Kortik3	AH039	Körtik Tepe	Turkey	PPNA	Sheep	0.38	Petrous	Levant Attici
Kortik4	AH040	Körtik Tepe	Turkey	PPNA	Sheep	2.25	Petrous	Levant Attici
Kortik5	AH041_1	Körtik Tepe	Turkey	PPNA	-	-	Petrous	Levant Attici

Kortik6	AH041_2	Körtik Tepe	Turkey	PPNA	-	-	Petrous	Levant Attici
Geor2	AH066	Tamara Fort	Kazbegi, Georgia	11th-15th Century AD	Goat	68.31	Petrous	Marjan Mashkour
AP38	AP38	Aşağı Pınar	Kırklareli, Turkey	Middle/Late Neolithic	Goat	0.5	Non Petrous	Norbert Benecke
AP44	AP44	Aşağı Pınar	Kırklareli, Turkey	Mid/Late Neolithic (5500-5000 BC)	Goat	0.8	Non Petrous	Norbert Benecke
AP45	AP45	Aşağı Pınar	Kırklareli, Turkey	Mid/Late Neolithic (5300-5000 BC)	Goat	1.48	Non Petrous	Norbert Benecke
AP46	AP46	Aşağı Pınar	Kırklareli, Turkey	Mid/Late Neolithic	Goat	0.28	Non Petrous	Norbert Benecke
AP49	AP49	Aşağı Pınar	Kırklareli, Turkey	Mid/Late Neolithic	Goat	2.02	Non Petrous	Norbert Benecke
AP50	AP50	Aşağı Pınar	Kırklareli, Turkey	Neolithic (5300-5000 BC)	Goat	0.47	Non Petrous	Norbert Benecke
Cav8	Cav8	Čavdar	Sofia District, Bulgaria	Neolithic (6000-5500 BC)	Goat	0.03	Non Petrous	Norbert Benecke
Dra34	Dra34	Merdžumekja	Drama, Bulgaria	Chalcolithic	Goat	0.11	Non Petrous	Norbert Benecke
Kan19	Kan19	Kanlıgeçit	Kırklareli, Turkey	Bronze Age (2700-2200 BC)	Goat	0.06	Non Petrous	Norbert Benecke
Kan23	Kan23	Kanlıgeçit	Kırklareli, Turkey	Bronze Age	Goat	0.5	Non Petrous	Norbert Benecke
Kan25	Kan25	Kanlıgeçit	Kırklareli, Turkey	Bronze Age (2700-2200 BC)	Goat	0.04	Non Petrous	Norbert Benecke
Acem1	KD001	Acemhöyük	Aksaray Plain, Turkey	Bronze Age (~2500 BC)	Goat	56.2	Petrous	Benjamin Arbuckle
Mentese1	KD002	Menteşe	Turkey	Neolithic	Sheep	7.23	Petrous (L)	Lionel Gourichon
Mentese2	KD003	Menteşe	Turkey	Neolithic	Sheep	4.22	Petrous (L)	Lionel Gourichon
Mentese3	KD004	Menteşe	Turkey	Neolithic	Sheep	0.29	Petrous (R)	Lionel Gourichon
Mentese4	KD005	Menteşe	Turkey	Neolithic	Sheep	4.84	Petrous (R)	Lionel Gourichon

Mentese5	KD006	Menteşe	Turkey	Neolithic	Sheep	10.15	Petrous (L)	Lionel Gourichon
Cukuirci1	KD007	Çukuriçi Höyük	Turkey	Neolithic	Sheep	2.96	Petrous	Alfred Galik
Shill1	KD008	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	2nd Phalange	Jean-Dennis Vigne
Shill2	KD009	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	Petrous	Jean-Dennis Vigne
Shill3	KD010	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	Petrous	Jean-Dennis Vigne
Shill4	KD011	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	Petrous	Jean-Dennis Vigne
Shill5	KD012	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	Petrous	Jean-Dennis Vigne
Shill6	KD013	Shillourokambos	Parekklisia, Cyprus	Neolithic	-	-	Petrous	Jean-Dennis Vigne
Hallan1	KD014	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan2	KD015	Hallan Çemi	Turkey	Epipaleolithic	Sheep	0.55	Petrous	Melinda Zeder
Hallan3	KD016	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan4	KD017	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan5	KD018	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan6	KD019	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan7	KD020	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Hallan8	KD021	Hallan Çemi	Turkey	Epipaleolithic	-	-	Petrous	Melinda Zeder
Zuid1	KD022	Voorschoten	Zuid Holland, Netherlands	Neolithic	-	-	Petrous	Youri van den Hurk
Zuid2	KD023	Stevenshofjespolder	Zuid Holland, Netherlands	Iron Age - Roman	Cow	5.25	Petrous	Youri van den Hurk
Utrecht1	KD024	Houten-Hoogdijk	Utrecht, Netherlands	Neolithic	Pig	39.38	Petrous	Youri van den Hurk
Utrecht2	KD025	Bunnik Odij	Utrecht, Netherlands	Mid Iron Age/Late Roman	Pig	50.48	Petrous	Youri van den Hurk

Mainz1	KD026	Heim, Mainz	Neubaugelande, Germany	1st/2nd century AD	Sheep	44.2	Petrous	Helmut Hemmer
Mainz2	KD027	Heim	Mainz, Germany	Mid-1st century AD	Sheep	51.24	Petrous	Helmut Hemmer
Frankfurt1	KD028	Frankfurt-Heddern	Germany	2/3rd century AD	Sheep	43.82	Petrous	Helmut Hemmer
Alzey1	KD029	Alzey	Germany	2nd century AD	Sheep	54.44	Petrous	Helmut Hemmer
Mainz3	KD030	Mainz	Germany	Roman	Sheep	31.82	Petrous	Helmut Hemmer
Mainz4	KD031	Mainz	Germany	Roman	Sheep	41.35	Petrous	Helmut Hemmer
Herxheim1	KD032	Herxheim	Germany	Neolithic	Pig	47.13	Petrous	Andrea Zeeb-Lanz
Herxheim2	KD033	Herxheim	Germany	Neolithic	Pig	40.36	Petrous	Andrea Zeeb-Lanz
Herxheim3	KD034	Herxheim	Germany	Neolithic	Sheep	54.33	Petrous	Andrea Zeeb-Lanz
Herxheim4	KD035	Herxheim	Germany	Neolithic	Dog	47.68	Petrous	Andrea Zeeb-Lanz
Herxheim5	KD036	Herxheim	Germany	Neolithic	-	-	Petrous	Andrea Zeeb-Lanz
Herxheim6	KD037	Herxheim	Germany	Neolithic	Pig	58.52	Petrous	Andrea Zeeb-Lanz
Herxheim7	KD038	Herxheim	Germany	Neolithic	Sheep	55.33	Petrous	Andrea Zeeb-Lanz
Herxheim8	KD039	Herxheim	Germany	Neolithic	Sheep	29.11	Petrous	Andrea Zeeb-Lanz
Blagotin1	KD040	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Goat	50.79	Petrous (R)	David Orton
Blagotin2	KD041	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Goat	38.81	Petrous (L)	David Orton
Blagotin3	KD042	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Goat	66.45	Petrous (L)	David Orton
Blagotin4	KD043	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	6.33	Petrous (L)	David Orton
Blagotin5	KD044	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	-	-	Petrous (R)	David Orton

Blagotin6	KD045	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	18.36	Petrous	David Orton
Blagotin7	KD046	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	17.7	Petrous (R)	David Orton
Blagotin8	KD047	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	2.58	Petrous (L)	David Orton
Blagotin9	KD048	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	7.91	Petrous (R)	David Orton
Blagotin10	KD049	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	21.06	Petrous (R)	David Orton
Blagotin11	KD050	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	12.55	Petrous (L)	David Orton
Blagotin12	KD051	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	-	-	Petrous (L)	David Orton
Blagotin13	KD052	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	9.4	Petrous (R)	David Orton
Blagotin14	KD053	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Deer	0.35	Petrous (R)	David Orton
Blagotin15	KD054	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	4.23	Petrous (R)	David Orton
Blagotin16	KD055	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Goat	46.04	Petrous (L)	David Orton
Blagotin17	KD056	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	8.71	Petrous (L)	David Orton
Bubanj1	KD057	Bubanj-Nevo	Selo, Serbia	Chalcolithic	Sheep	49.06	Petrous	David Orton
Blagotin18	KD058	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	31.62	Petrous	David Orton
Blagotin19	KD059	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	55.05	Petrous	David Orton
Blagotin20	KD060	Blagotin-Poljna	Trstenik, Serbia	Early Neolithic (~6200 cal BC)	Sheep	35.74	Petrous	David Orton

Mongolia1	KD061	?	Mongolia	Bronze/Iron Age	Sheep	76.31	Petrous	Ulaanbaatar University
Tarxein3	KD064	Tarxein	Malta	?	Sheep	6.28	Petrous	Finbar McCormick
Tarxein4	KD065	Tarxein	Malta	?	-	-	Petrous	Finbar McCormick
Tarxein5	KD066	Tarxein	Malta	?	-	-	Petrous	Finbar McCormick
Marmara1	KD067	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	38.5	Petrous	Canan Çakırlar
Marmara2	KD068	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	-	-	Petrous	Canan Çakırlar
Marmara3	KD069	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	46.33	Petrous	Canan Çakırlar
Marmara4	KD070	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	25.86	Petrous	Canan Çakırlar
Marmara5	KD071	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	21.95	Petrous	Canan Çakırlar
Marmara6	KD072	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	12.96	Petrous	Canan Çakırlar
Marmara7	KD073	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	14	Petrous	Canan Çakırlar
Marmara8	KD074	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	10.43	Petrous	Canan Çakırlar
Marmara9	KD075	Yenikapı	Istanbul, Marmara, Turkey	Neolithic	Sheep	4.77	Petrous	Canan Çakırlar
Direkli1	KD076	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	Goat	17.59	Petrous	Benjamin Arbuckle
Direkli2	KD077	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	Goat	60.32	Petrous	Benjamin Arbuckle
Direkli3	KD078	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	-	-	Petrous	Benjamin Arbuckle

Direkli4	KD079	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	Goat	25.89	Petrous	Benjamin Arbuckle
Direkli5	KD080	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	Goat	7.08	Petrous	Benjamin Arbuckle
Direkli6	KD081	Direkli Cave	Taurus Mountains, Turkey	Late Epipaleolithic (~9500 cal BC)	Goat	16.05	Petrous	Benjamin Arbuckle
Dover1	KD082	Tel Dover	Israel	Pottery Neolithic	-	-	Petrous	Liora Horwitz
Ashqel1	KD083	Ashkelon	Israel	PPNC	-	-	Petrous	Liora Horwitz
Plocnik1	KD084	Plocnik	Serbia	Neolithic	Sheep	1.75	Petrous	David Orton
Ashqel2	KD085	Ashkelon	Israel	PPNC	-	-	Petrous	Liora Horwitz
Ashqel3	KD086	Ashkelon	Israel	PPNC	-	-	Petrous	Liora Horwitz
Yoqneam1	KD087	Tel Yoqne'am	Haifa, Israel	Iron Age/Later	Sheep	1.76	Petrous	Liora Horwitz
Ashqel4	KD088	Ashkelon	Israel	PPNC	-	-	Petrous	Liora Horwitz
Yoqneam2	KD089	Tel Yoqne'am	Haifa, Israel	Mid Bronze Age (~1650-1550/1540 BC)	Goat	45.66	Petrous	Liora Horwitz
Ashqel5	KD090	Ashkelon	Israel	PPNC	-	-	Petrous	Liora Horwitz
Yoqneam3	KD092	Tel Yoqne'am	Haifa, Israel	Iron Age/Later	Sheep	29.77	Petrous	Liora Horwitz
Yoqneam4	KD094	Tel Yoqne'am	Haifa, Israel	Iron Age/Later	Sheep	5.03	Petrous	Liora Horwitz
Atlityam1	KD095	Atlit Yam	Israel	Late PPNB/PPNC	-	-	Petrous	Liora Horwitz
Atlityam2	KD096	Atlit Yam	Israel	Late PPNB/PPNC	-	-	Petrous	Liora Horwitz
Safi1	KD097	Tel es-Safi	Israel	Iron Age	Cow	0.19	Petrous	Liora Horwitz
Yarmut1	KD098	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age (~2700-2500 BC)	Goat	0.23	Petrous	Liora Horwitz
Yarmut2	KD099	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz

Yarmut3	KD100	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Yarmut4	KD101	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Yarmut5	KD102	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Yarmut6	KD103	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Yarmut7	KD104	Tel Yarmuth	Bet Shemesh, Israel	Early Bronze Age (~2650-2200 BC)	Goat	0.14	Petrous	Liora Horwitz
Dzhulyunitsa1	KD105	Dzhulyunitsa	Bulgaria	Neolithic	Sheep	16.81	Petrous	Canan Çakırlar
Dzhulyunitsa2	KD106	Dzhulyunitsa	Bulgaria	Neolithic	Sheep	9.88	Petrous	Canan Çakırlar
Fars1	KD107	Rahmat Abad	Fars, Iran	Early Chalcolithic / Middle Bakun phase (~4600 BC)	Goat	1.13	Petrous	Marjan Mashkour
Khouz1	KD108	Abu Fandoweh	Khouzestan, Iran	Chalcolithic	-	-	Petrous	Marjan Mashkour
Khouz2	KD109	Tepe Sohz	Khouzestan, Iran	Chalcolithic	-	-	Petrous	Marjan Mashkour
Khouz3	KD110	Tepe Sohz	Khouzestan, Iran	Chalcolithic	-	-	Petrous	Marjan Mashkour
Semnan1	KD111	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic (~8000 BP)	Goat	37.82	Petrous	Marjan Mashkour
Semnan2	KD112	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic (~8000 BP)	Goat	40.16	Petrous	Marjan Mashkour
Semnan3	KD113	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic (~7000 BP)	Goat	66.23	Petrous	Marjan Mashkour
Kazbeg1	KD114	Tamara Fort	Kazbegi, Georgia	Medieval 10th AD	Goat	73.94	Petrous	Marjan Mashkour
Lur1	KD115	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur2	KD116	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur3	KD117	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Quwain1	KD118	Umm al Quwain	United Arab Emirates	Mid Neolithic	-	-	Petrous	Marjan Mashkour

Quwain2	KD119	Umm al Quwain	United Arab Emirates	Mid Neolithic	-	-	Petrous	Marjan Mashkour
Quwain3	KD120	Umm al Quwain	United Arab Emirates	Mid Neolithic	-	-	Petrous	Marjan Mashkour
Khouz4	KD121	Tepe Sohz	Khouzestan, Iran	Chalcolithic	-	-	Petrous	Marjan Mashkour
Semnan4	KD122	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic (~7000BP)	Sheep	17.72	Petrous	Marjan Mashkour
Semnan5	KD123	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic (~8000BP)	Sheep	51.53	Petrous	Marjan Mashkour
Lur4	KD124	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur5	KD125	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur6	KD126	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur7	KD127	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Miqne1	KD128	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Sheep	0.45	Petrous	Liora Horwitz
Miqne2	KD129	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Sheep	0.51	Petrous	Liora Horwitz
Miqne3	KD130	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Sheep	0.66	Petrous	Liora Horwitz
Miqne4	KD131	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Cow	0.2	Petrous	Liora Horwitz
Miqne5	KD132	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Goat	0.15	Petrous	Liora Horwitz
Miqne6	KD133	Tel Miqne-Ekron	Shephelah, Israel	Late Bronze/Iron Age	Sheep	1.11	Petrous	Liora Horwitz
David1	KD134	City of David	Jerusalem, Israel	Iron Age	Sheep	29.6	Petrous	Liora Horwitz
David2	KD135	City of David	Jerusalem, Israel	Iron Age	Cow	12.25	Petrous	Liora Horwitz
David3	KD136	City of David	Jerusalem, Israel	Iron Age	Cow	23.95	Petrous	Liora Horwitz
David5	KD138	City of David	Jerusalem, Israel	Iron Age	Horse	0.29	Petrous	Liora Horwitz
Yoqneam5	KD139	Tel Yoqne'am	Haifa, Israel	Iron Age/Later	Cow	32.83	Petrous	Liora Horwitz

Dalit1	KD140	Tel Dalit	Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Dalit2	KD141	Tel Dalit	Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Dalit3	KD142	Tel Dalit	Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Ashqel8	KD143	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Ashqel9	KD144	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Ashqel10	KD145	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Ashqel11	KD146	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Ashqel12	KD147	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Ashqel13	KD148	Ashkelon	Israel	PPNC	-	-	Partial Petrous	Liora Horwitz
Lur8	KD149	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	-	-	Petrous	Marjan Mashkour
Lur9	KD150	Kelek Asad Morad	Luristan, Iran	Pre-Pottery Neolithic	Goat	0.03	Partial Petrous	Marjan Mashkour
Semnan6	KD151	Sang-e Chakmaq	Semnan, Iran	Pre-Pottery Neolithic (~8000BP)	Gazelle	27.54	Partial Petrous	Marjan Mashkour
Semnan7	KD152	Sang-e Chakmaq	Semnan, Iran	Pottery Neolithic (~7000BP)	Goat	85.89	Partial Petrous	Marjan Mashkour
Azer1	KD153	Kul Tepe	Azerbaijan, Iran	Chalcolithic/Bronze/Iron Age	Cow	26.95	Petrous	Marjan Mashkour
Quwain4	KD154	Umm al Quwain	United Arab Emirates	Mid Neolithic	-	-	Petrous	Marjan Mashkour
Quwain5	KD155	Umm al Quwain	United Arab Emirates	Mid Neolithic	-	-	Petrous	Marjan Mashkour

Quwain6	KD156	Umm al Quwain	United Arab Emirates	Mid Neolithic	Cow	0.15	Petrous	Marjan Mashkour
Acem2	KD157	Acemhöyük	Aksaray Plain, Turkey	Mid Bronze Age (1700BC)	Goat	73.76	Petrous	Benjamin Arbuckle
Cukuirci2	KD158	Çukuriçi Höyük	Turkey	Neolithic	Sheep	1.49	Petrous	Alfred Galik
Qassis1	KD159	Qassis/Kassis/Tel es-Kassis	Israel	Early Bronze Age	Cow	4.17	Non Petrous	Liora Horwitz
Musos1	KD160	Musos	Israel	Late Bronze - Iron Age	Cow	2.37	Petrous	Liora Horwitz
Afridar1	KD161	Ashqelon Ashridar	Israel	Chalcolithic	-	-	Petrous	Liora Horwitz
Kitan1	KD162	Tel Kitan	Israel	Early Bronze Age	-	-	Petrous	Liora Horwitz
Shiqmim1	KD163	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	Goat	0.17	Petrous	Liora Horwitz
Shiqmim2	KD164	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	Sheep	0.63	Petrous	Liora Horwitz
Gilat1	KD165	Gilat	Northern Negev, Israel	Chalcolithic (4500-3500 BC)	-	-	Petrous	Liora Horwitz
Gilat2	KD166	Gilat	Northern Negev, Israel	Chalcolithic (4500-3500 BC)	Goat	0.34	Petrous	Liora Horwitz
Ghosh1	KD167	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh2	KD168	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh3	KD169	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh4	KD170	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Karainb1	KD171	Karain Cave B	Turkey	?	-	-	Carpal	Levant Attici
Karainb2	KD172	Karain Cave B	Turkey	?	-	-	Carpal	Levant Attici
Karainb3	KD173	Karain Cave B	Turkey	?	-	-	Carpal	Levant Attici

Kazbeg2	KD174	Tamara Fort	Kazbegi, Georgia	Medieval	Sheep	53.07	Petrous	Marjan Mashkour
Azer2	KD175	Tepe Hasanlu	Azerbaijan, Iran	Iron Age III	Sheep	32.56	Petrous	Marjan Mashkour
Azer3	KD176	Tepe Hasanlu	Azerbaijan, Iran	Early Bronze (3000-2100 BC)	Goat	47.32	Petrous	Marjan Mashkour
Azer4	KD177	Tepe Hasanlu	Azerbaijan, Iran	Achaemenid / Iron Age (550-330 BC)	Goat	63.64	Petrous	Marjan Mashkour
Azer5	KD178	Tepe Hasanlu	Azerbaijan, Iran	Early Bronze (3000-2100 BC)	Goat	66.99	Petrous	Marjan Mashkour
Qazvin1	KD179	Tepe Chizar	Qazvin, Iran	Middle Bronze Age (2400-1900 BC)	Goat	54.71	Petrous	Marjan Mashkour
Qazvin3	KD181	Tepe Chizar	Qazvin, Iran	Neolithic/Chalcolithic	Sheep	1.76	Petrous	Marjan Mashkour
Qom1	KD182	Qoli Darvish	Qom, Iran	Chalcolithic/Iron Age	Sheep	15.97	Petrous	Marjan Mashkour
Fars2	KD183	Rahmat Abad	Fars, Iran	Pottery Neolithic / Formative Mushki (6700-6471 BC)	Goat	1.45	Petrous	Marjan Mashkour
Fars3	KD184	Tepe Qasr Ahmad	Fars, Iran	Early Neolithic (7,000-6,000 BC)	Sheep	25.16	Petrous	Marjan Mashkour
Lur10	KD185	Tepe Ghela Gap	Luristan, Iran	Chalcolithic/Iron Age	Sheep	29.58	Petrous	Marjan Mashkour
Lur11	KD186	Tepe Ghela Gap	Luristan, Iran	Chalcolithic/Iron Age	Sheep	2.24	Petrous	Marjan Mashkour
Lur12	KD187	Tepe Abdul Hosein	Luristan, Iran	Pre-Pottery Neolithic (~8000 cal BC)	Goat	8.42	Petrous	Marjan Mashkour
Khor1	KD188	Chalow	Khorasan, Iran	Bronze Age	Sheep	3.24	Petrous	Marjan Mashkour
Khor2	KD189	Chalow	Khorasan, Iran	Bronze Age	Sheep	0.69	Petrous	Marjan Mashkour
Fars4	KD190	Mianrud	Fars, Iran	Early / Mid Chalcolithic (5550 - 4200 BC)	Goat	16.19	Partial Petrous	Marjan Mashkour

Azer6	KD191	Soha Chay Tepe	Azerbaijan, Iran	Bronze Age (2400-1900 BC) / Chalcolithic (~4200 BC)	Goat	28.34	Petrous	Marjan Mashkour
Azer7	KD192	Soha Chay Tepe	Azerbaijan, Iran	Bronze Age	Gazelle	2.59	Petrous	Marjan Mashkour
Fars5	KD193	Rahmat Abad	Fars, Iran	Pottery Neolithic / Formative Mushki (6700-6471 BC)	Goat	2.59	Petrous	Marjan Mashkour
Dzuduana1	KD194	Dzudzuana Cave	Georgia	Paleolithic	Cow	4.84	Petrous	Ron Pinhasi
Aliabad3	KD195	Ali Abad	Azerbaijan, Iran	Mid-Late Bronze Age (2200-1200 BC)	Sheep	4.73	Petrous	Marjan Mashkour
Kohneh2	KD197	Kohneh Tepesi	Azerbaijan, Iran	Early Bronze Age (3300-3000 BC)	Goat	6.81	Petrous	Marjan Mashkour
Chalow1	KD198	Chalow	Khorasan, Iran	Bronze (2300-2000 BC)	Goat	6.66	Petrous	Marjan Mashkour
Nahal1	KD200	Nahal Tillah	Israel	Chalcolithic	Dog	0.01	Petrous	Liora Horwitz
Gilat6	KD201	Gilat	Northern Negev, Israel	Chalcolithic (Early Ghassulian) (4500-4200 BC)	-	-	Petrous	Liora Horwitz
Gilat7	KD202	Gilat	Northern Negev, Israel	Chalcolithic (Early Ghassulian) (4500-4200 BC)	-	-	Partial Petrous	Liora Horwitz
Gilat8	KD203	Gilat	Northern Negev, Israel	Chalcolithic (Early Ghassulian) (4500-4200 BC)	Goat	0.83	Petrous	Liora Horwitz
Shiqmim6	KD204	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	-	-	Petrous	Liora Horwitz
Shiqmim7	KD205	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	-	-	Petrous	Liora Horwitz
Ghosh8	KD206	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz

Ghosh9	KD207	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh10	KD208	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh11	KD209	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh12	KD210	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh13	KD211	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Ghosh14	KD212	Abu Ghosh	Judean Hills, Israel	PPNB	-	-	Petrous	Liora Horwitz
Bir1	KD213	Bir es-Safadi	Israel	Chalcolithic	-	-	Petrous	Liora Horwitz
Bir2	KD214	Bir es-Safadi	Israel	Chalcolithic	-	-	Petrous	Liora Horwitz
Safi2	KD215	Tel es-Safi	Ashkelon, Israel	Early Bronze Age III (2900-2570 BC)	Goat	3.28	Petrous	Liora Horwitz
Cardo1	KD216	Cardo	Northern Negev, Israel	Roman	-	-	Petrous	Liora Horwitz
Masos3	KD217	Tel Masos	Northern Negev, Israel	Iron Age	-	-	Petrous	Liora Horwitz
Gilat9	KD218	Gilat	Northern Negev, Israel	Chalcolithic (Early Ghassulian) (4500-4200 BC)	-	-	Petrous	Liora Horwitz
Gilat10	KD219	Gilat	Northern Negev, Israel	Chalcolithic (Early Ghassulian) (4500-4200 BC)	Goat	0.15	Petrous	Liora Horwitz
Shiqmim8	KD220	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	Sheep	0.12	Petrous	Liora Horwitz
Shiqmim9	KD221	Shiqmim	Northern Negev, Israel	Chalcolithic (Late Ghassulian) (4300-3700 BC)	Goat	0.1	Petrous	Liora Horwitz
Michal1	KD222	Tel Michal	Israel	Mid Bronze - Iron Age	Sheep	0.93	Petrous	Liora Horwitz
Lux1	KD223	Luxmanda	Tanzania	Neolithic	-	-	Petrous	Mary Prendergast

Lux2	KD224	Luxmanda	Tanzania	Neolithic	-	-	Petrous	Mary Prendergast
Lux3	KD225	Luxmanda	Tanzania	Neolithic	-	-	Petrous	Mary Prendergast
Lux4	KD226	Luxmanda	Tanzania	Neolithic	-	-	Petrous	Mary Prendergast
Lux5	KD227	Luxmanda	Tanzania	Neolithic	-	-	Petrous	Mary Prendergast
Bulak1	KD228	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age (~2000-1700 BC)	Goat	51.16	Petrous	Norbert Benecke
Bulak2	KD229	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age (~2000-1700 BC)	Goat	49.47	Petrous	Norbert Benecke
Bulak3	KD230	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age (~2000-1700 BC)	Goat	39.83	Petrous	Norbert Benecke
Bulak4	KD231	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age (~2000-1700 BC)	Goat	5.6	Petrous	Norbert Benecke
Bulak5	KD232	Tilla Bulak	Surkhandarja, Uzbekistan	Bronze Age (~2000-1700 BC)	Goat	18.13	Petrous	Norbert Benecke
Darre1	KD233	Darre-ye Bolāghi	Fars, Iran	Chalcolithic (5th mil. BC)	Goat	7.1	Petrous	Norbert Benecke
Darre2	KD234	Darre-ye Bolāghi	Fars, Iran	Chalcolithic (5th mil. BC)	Goat	70.31	Petrous	Norbert Benecke
Monjukli1	KD235	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Goat	7.87	Petrous	Norbert Benecke
Monjukli2	KD236	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Goat	8.18	Petrous	Norbert Benecke
Monjukli3	KD237	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Sheep	11.44	Petrous	Norbert Benecke
Monjukli4	KD238	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Goat	20.08	Petrous	Norbert Benecke
Monjukli5	KD239	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Sheep	13.23	Petrous	Norbert Benecke

Monjukli6	KD240	Monjukli Depe	Meana-Čaača, Turkmenistan	Early Chalcolithic (~5100-4500 cal BC)	Goat	5.84	Petrous	Norbert Benecke
Monjukli7	KD241	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic (~6400-5900 cal BC)	Goat	3.18	Petrous	Norbert Benecke
Monjukli8	KD242	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic (~6400-5900 cal BC)	Goat	17.43	Petrous	Norbert Benecke
Monjukli9	KD243	Monjukli Depe	Meana-Čaača, Turkmenistan	Pottery Neolithic (~6400-5900 cal BC)	Goat	1.99	Petrous	Norbert Benecke
Georgia1	KD244	Tamara Fort	Kazbegi, Georgia	Medieval (11th-15th century AD)	Sheep	74.85	Petrous	Marjan Mashkour
Georgia2	KD245	Tamara Fort	Kazbegi, Georgia	Medieval (7th-11th century AD)	Sheep	75.09	Petrous	Marjan Mashkour
Georgia3	KD246	Tamara Fort	Kazbegi, Georgia	Medieval (7th-11th century AD)	Sheep	58.34	Petrous	Marjan Mashkour
Georgia4	KD247	Tamara Fort	Kazbegi, Georgia	Medieval (7th-11th century AD)	Sheep	40.45	Petrous	Marjan Mashkour
Dzharkutan1	KD248	Dzharkutan	Uzbekistan	Bronze/Iron Age (3rd/2nd Mil. BC)	-	-	Petrous	Marjan Mashkour
Dzharkutan2	KD249	Dzharkutan	Uzbekistan	Bronze/Iron Age (3rd/2nd Mil. BC)	Sheep	0.4	Petrous	Marjan Mashkour
Dzharkutan3	KD250	Dzharkutan	Uzbekistan	Bronze/Iron Age (3rd/2nd Mil. BC)	Sheep	1.56	Petrous	Marjan Mashkour
Dzharkutan4	KD251	Dzharkutan	Uzbekistan	Bronze/Iron Age (3rd/2nd Mil. BC)	Sheep	0.57	Petrous	Marjan Mashkour
Koktepe1	KD252	Kok Tepe	Uzbekistan	Iron Age (2nd/1st Mil. BC)	-	-	Petrous	Marjan Mashkour
Padayatak1	KD253	Padayatak	Uzbekistan	Iron Age (2nd/1st Mil. BC)	Sheep	59.85	Petrous	Marjan Mashkour
Sangir1	KD254	Sangir Tepe	Uzbekistan	Achaemenid / Iron Age (550-330 BC)	Sheep	52.71	Petrous	Marjan Mashkour

Aliabad1	KD255	Ali Abad	Azerbaijan, Iran	Mid-Late Bronze Age (2200-1200 BC)	-	-	Long Bone	Marjan Mashkour
Aliabad2	KD256	Ali Abad	Azerbaijan, Iran	Mid-Late Bronze Age (2200-1200 BC)	-	-	Long Bone	Marjan Mashkour
Kov27	Kov27	Kovačevo	Blagoevgrad, Bulgaria	Neolithic (6200-5600 BC)	Goat	0.02	Non Petrous	Norbert Benecke
Kov57	Kov57	Kovačevo	Blagoevgrad, Bulgaria	Neolithic (6200-5600 BC)	Goat	5.74	Non Petrous	Norbert Benecke
Kov60	Kov60	Kovačevo	Blagoevgrad, Bulgaria	Neolithic (6200-5600 BC)	Goat	0.52	Non Petrous	Norbert Benecke
Ovc11	Ovc11	Ovčarovo-gorata	Tărgovište, Bulgaria	Early Neolithic (5,700-5,500 BC)	Goat	0.09	Non Petrous	Norbert Benecke
Pie17	Pie17	Pietrele	Giurgiu, Romania	Chalcolithic (4450-4250 BC)	Goat	0.18	Non Petrous	Norbert Benecke
Tac1	Tac1	Tachti Perda	Kakheti, Georgia	Late Bronze (1400-1000 BC)	Goat	0.09	Non Petrous	Norbert Benecke
Tac2	Tac2	Tachti Perda	Kakheti, Georgia	Iron Age (1000-700 BC)	Goat	0.24	Non Petrous	Norbert Benecke
Tac3	Tac3	Tachti Perda	Kakheti, Georgia	Late Bronze (1400-1000 BC)	Goat	13.58	Non Petrous	Norbert Benecke
Uiv17	Uiv17	Uivar	Timişoara, Romania	Neolithic (5250-5050 BC)	Goat	0.1	Non Petrous	Wolfram Schier
Ulu38	Ulu38	Ulucak Höyük	Turkey	Early Neolithic (6400-6100 BC)	Goat	0.03	Non Petrous	Canan Çakırlar
Hovk1	VEM119	Hovk-1 Cave	Tavush, Armenia	Paleolithic	Goat	29.72	Petrous	Ron Pinhasi
Ainghazal1	VEM140	'Ain Ghazal	Amman, Jordan	Middle PPNB	Goat	2.22	Petrous	Louise Martin
Ainghazal3	VEM142	'Ain Ghazal	Amman, Jordan	Middle PPNB	Goat	0.21	Petrous	Louise Martin
Ainghazal2	VEM157	'Ain Ghazal	Amman, Jordan	Middle PPNB	Goat	2.58	Petrous	Louise Martin
Ainghazal4	VEM159	'Ain Ghazal	Amman, Jordan	Middle PPNB	Goat	0.65	Petrous	Louise Martin
Potterne1	VEM199	Potterne	Wiltshire, UK	Bronze Age (2040-990 BC)	Goat	71.12	Petrous	Lisa Brown/Wiltshire Museum

Appendix Table 2.2 - Reference genomes used in Fastq Screen.

Species	Genbank Accession	Assembly Name
Goat	GCA_000317765.1	CHIR_1.0
Sheep	GCA_000298735.1	Oar_v3.1
Human	GCA_000001405.16	GRCh38.p1
Cow	GCA_000003055.5	Bos_taurus_UMD_3.1.1
Dog	GCA_000002285.2	CanFam3.1
Cat	GCA_000181335.2	Felis_catus-6.2
Pig	GCA_000003025.4	Sscrofa10.2
Deer	GCA_000751575.1	kmer631

Appendix Table 2.3 - Mitochondrial sequences used in Maximum Likelihood phylogeny. Genomes marked in bold were used in a circularized form for sample realignment, and as reference individuals in phylogenies.

Name used	Accession No.	Haplogroup	Name used	Accession No.	Haplogroup
A1_01	KR059146.1	A1	A_47	KR059191.1	A
A1a_02	KR059147.1	A1a	A_48	KR059192.1	A
A1a_03	KR059148.1	A1a	A_49	KR059193.1	A
A1a_04	KR059149.1	A1a	A_50	KR059194.1	A
A1a_05	KR059150.1	A1a	A_51	KR059195.1	A
A1a_06	KR059151.1	A1a	A_52	KR059196.1	A
A2_07	KR059152.1	A2	A_53	KR059197.1	A
A2_08	KR059153.1	A2	A_54	KR059198.1	A
A2_09	KR059154.1	A2	A_55	KR059199.1	A
A2a_10	KR059155.1	A2a	A_56	KR059200.1	A
A2a_11	KR059156.1	A2a	A_57	KR059201.1	A
A2a1_12	KR059157.1	A2a1	A_58	KR059202.1	A
A2a1_13	KR059158.1	A2a1	A_59	KR059203.1	A
A2a1_14	KR059159.1	A2a1	A_60	KR059204.1	A
A2a1_15	KR059160.1	A2a1	A_61	KR059205.1	A
A2a1_16	KR059161.1	A2a1	A_62	KR059206.1	A
A2a1_17	KR059162.1	A2a1	A_63	KR059207.1	A
A2a1_18	KR059163.1	A2a1	A_64	KR059208.1	A
A2a1_19	KR059164.1	A2a1	A_65	KR059209.1	A
A3_20	KR059165.1	A3	D_bezoar_66	KR059210.1	D
A3_21	KR059166.1	A3	**D1_67**	KR059211.1	D1
A3_22	KR059167.1	A3	D1_68	KR059212.1	D1
A3_23	KR059168.1	A3	**G_69**	KR059213.1	G
A3_24	KR059169.1	A3	G_70	KR059214.1	G
A4_25	KR059170.1	A4	G_71	KR059215.1	G
A4_26	KR059171.1	A4	G_72	KR059216.1	G
A4_27	KR059172.1	A4	G_73	KR059217.1	G
A4_28	KR059173.1	A4	G_74	KR059218.1	G

A5_29	KR059174.1	A5	B_bezoar_75	KR059219.1	B
A5_30	KR059175.1	A5	B1_78	KR059220.1	B1
A5_31	KR059176.1	A5	C_bezoar_79	KR059221.1	C
A5_32	KR059177.1	A5	C1_bezoar_80	KR059222.1	C1
A6_33	KR059178.1	A6	**C1a_81**	KR059223.1	C1a
A6_34	KR059179.1	A6	C1a_82	KR059224.1	C1a
A7_36	KR059180.1	A7	C1a_83	KR059225.1	C1a
A7_37	KR059181.1	A7	**F_bezoar_84**	KR059226.1	F
A7_38	KR059182.1	A7	**Goat Reference**	NC_005044.2	B
A_39	KR059183.1	A	Bezoar Reference	NC_028161.1	G
A_40	KR059184.1	A	**West Caucasian Tur**	NC_020683.1	Outgroup
A_41	KR059185.1	A	Nubian Ibex	NC_020624.1	Outgroup
A_42	KR059186.1	A	Markhor	NC_020622.1	Outgroup
A_43	KR059187.1	A	Alpine Ibex	NC_020623.1	Outgroup
A_44	KR059188.1	A	Siberian Ibex	NC_020626.1	Outgroup
A_45	KR059189.1	A	Sheep	NC_001941.1	Outgroup
A_46	KR059190.1	A			

Appendix Table 2.4 - Chapter 2 Illumina HiSeq sequencing and whole genome alignment results. Filtered reads are following cutadapt trimming and minimum read length 30bp filter. Endogenous was calculated as reads aligned following mapQ 30 filter divided by the filtered read count. Samples were aligned to CHIR_1.0. Pairs of samples which were later combined into single samples are present both prior to and following merging. Coverage statistics calculated after mapping quality 30 filter.

Sample	Raw Reads	Filtered Reads	Aligned Reads	Rmdup Total Reads	Rmdup Aligned Reads	>q30 Aligned Reads	Endog. %age	Coverage (q30)
Acem1	535429853	479546615	390919313	398281063	309653761	223852082	56.20	4.76
Acem2	619205809	593031928	550580477	524275148	483899634	386712744	73.76	8.67
Ainghazal1	83187399	75269854	3198903	74662520	2591569	1655535	2.22	0.03
Ainghazal2	128756310	119316119	5189545	118358959	4232385	3050323	2.58	0.06
Ainghazal3	74822625	72860762	258705	72839487	237430	155923	0.21	0.003
Ainghazal4	63597170	58367881	581645	58321821	535585	380297	0.65	0.01
AP45	57607602	48701302	2724863	47069521	1093082	695529	1.48	0.02
AP49	55105031	48449690	2170311	47692824	1413445	965079	2.02	0.02
Azer3	177036801	167817722	99622554	144764561	76569393	55887363	38.61	1.12
Azer4	232495520	224771105	193580054	173573899	142382848	110460711	63.64	2.57
Azer5	266087132	255480017	227463337	228668478	200812484	159978268	69.96	3.54
Azer3-5	443123933	423297739	327085891	373433039	277381877	215865631	57.81	4.66
Azer6	76508162	74024096	24448642	67035604	17460150	12825718	19.13	0.28
Blagotin1	1374961979	1216553609	1027304768	725051664	535802823	368273417	50.79	6.99

Blagotin2	1015818158	912908506	685826030	552273147	325190671	214314969	38.81	4.02
Blagotin3	785079584	1154899365	5216937865	930254368	820784420	618138762	66.45	11.47
Blagotin16	731037823	629697308	482748215	439295991	292346898	202242150	46.04	3.51
Bulak1	85489637	82832962	60012053	73493716	50672807	37602935	51.16	0.87
Bulak2	341223653	324211812	233519547	268169941	178162940	132665825	49.47	2.67
Bulak3	85110322	82036020	47896157	73277455	39137592	29187922	39.83	0.61
Bulak5	84666432	82853966	26122691	75571511	18840236	13700560	18.13	0.27
Chalow1	31355259	30944769	3284797	30548747	2888775	2033240	6.66	0.05
Darre1	29565706	28494367	3162626	28125220	2793479	1997846	7.10	0.04
Darre2	261028786	250473871	219847784	230083532	199457445	161767556	70.31	3.93
Direkli1	363741090	342965451	128733852	327773733	113542134	89340268	27.26	2.00
Direkli2	1131743295	1009580745	812303750	835625509	638348514	485788542	58.13	9.55
Direkli1-2	1495484385	1352546196	941037602	1163399242	751890648	575128810	49.44	11.55
Direkli5	235416747	189541015	23245207	184814055	18518247	13083848	7.08	0.27
Direkli6	800438918	702517029	273524868	606627244	177635083	97371651	16.05	1.93
Fars1	99020181	91837535	1548425	91768937	1479827	1034923	1.13	0.02
Fars2	32552505	31943251	463729	31916704	437182	327887	1.03	0.01
Fars4	297308440	292702230	67002909	285797590	60098269	46280618	16.19	1.05
Fars5	58006070	56568414	1827163	56435501	1694250	1266654	2.24	0.02
Fars2-5	90558575	88511665	2290892	88352205	2131432	1594541	1.80	0.03

Geor2	95217083	91863838	77607028	80692443	67514099	55121287	68.31	1.5
Ghosh5	20088538	19681029	12497	19680372	11840	7964	0.04	<0.01
Gilat2	24581161	21607163	121796	21600257	114890	74236	0.34	<0.01
Gilat8	102535491	97354551	1606816	97093658	1345923	805562	0.83	0.02
Gilat10	27002137	25701503	78520	25688876	65893	38180	0.15	<0.01
Hovk1	493158399	479794460	216201216	450104781	186511537	133767664	29.72	3.08
Kazbeg1	238441921	229262502	212461394	208682478	191881370	154299869	73.94	3.84
Kohneh2	36329162	33932789	3527754	33598312	3193277	2288871	6.81	0.04
Kov57	62554831	58600604	5833945	57233074	4466415	3284888	5.74	0.07
Lur9	28327887	27426734	14520	27424848	12634	6863	0.03	<0.01
Lur12	622799577	606327375	90171769	589678393	73522787	49649146	8.42	1.05
Miqne5	22442210	20880056	45657	20878262	43863	30629	0.15	<0.01
Monjukli1	136651989	134153896	17216066	131030587	14092757	10310059	7.87	0.24
Monjukli2	104714174	103396156	13168495	101263399	11035738	8285525	8.18	0.21
Monjukli4	152229139	149118748	61381879	130057199	42320330	26112882	20.08	0.6
Monjukli6	23893097	23044817	2103345	22801351	1859879	1331463	5.84	0.03
Monjukli8	703151022	693598812	218375612	621245514	145172327	108285172	17.43	2.57
Potterne1	235320071	230025893	202851037	211757594	184582738	150603738	71.12	3.67
Qazvin1	348762506	336691099	169724178	257342838	184052376	140793855	54.71	3.16
Safi2	60911590	59866759	3305158	59491905	2930304	1949107	3.28	0.04

Semnan1	318847280	303872175	150660326	273203164	118561944	91460919	33.48	2.05
Semnan2	1253857214	1196976627	595134680	904437365	323084992	230196578	25.45	4.80
Semnan1-2	1572704494	1500848802	745795006	1177640529	441646936	321657497	27.31	6.85
Semnan3	1290320470	1200758638	1053612961	943384035	792665122	624844041	66.23	14.89
Semnan7	264055600	235981957	70620635	211030082	128698407	181257506	85.89	3.28
Semnan8	145452599	131528398	27853991	121541550	17901036	12044508	9.91	0.21
Semnan9	363158752	344779210	271580143	252094796	178931891	134297213	53.27	3.05
Semnan10	526572133	482652509	195110498	398958390	111421418	76498235	19.17	1.43
Semnan13	385192146	370035589	215225143	290004604	135194158	100635730	34.70	2.54
Semnan17	85716449	82439264	15423383	75075910	8060029	5536825	7.37	0.12
Shiqmim1	77911468	75040580	261432	75002748	223600	127811	0.17	<0.01
Shiqmim9	34543992	32852520	63403	32843295	54178	34456	0.10	<0.01
Tac3	57168101	49613035	17565101	40038836	7990902	5436690	13.58	0.13
Yarmut1	39436031	38023365	134278	38015875	126788	88570	0.23	<0.01
Yarmut7	39308249	36288841	78084	36285009	74252	49504	0.14	<0.01
Yoqneam2	246811430	240328651	151357469	220728269	131757087	100785472	45.66	2.2

Appendix Table 2.5 - Chapter 2 mitochondrial capture sequencing and mtDNA alignment results. Alignment statistics presented following realigned to a closer reference sequence as defined in Appendix Table 2.3. Samples are presented following merging of likely identical individuals.

Sample	Raw Reads	Filtered Reads	>q30 Aligned Reads	Coverage	Called Sites	%age Called	mtDNA Hap
Acem1	-	-	121434	411.03	16642	100.00	A
Acem2	-	-	214607	849.8	16642	100.00	A
Ainghazal1	14251531	13532330	1930	5.46	12484	75.01	F
Ainghazal2	23443628	22984015	4464	12.78	15492	93.08	F
Ainghazal3	19347204	18976966	1363	4.25	10518	63.20	F
Ainghazal4	4413908	3983747	20575	65.44	16266	97.73	F
AP38	3534095	3019548	2258	7.61	15606	93.77	C
AP44	4165328	3806850	1078	3.57	10353	62.21	A
AP45	2289122	2033528	1874	6.47	14762	88.70	A
AP46	4532848	4175633	2834	9.77	16264	97.72	C
AP49	5315007	4799636	3744	12.69	16321	98.07	A
AP50	1849766	1709269	4262	14.31	16437	98.76	A
Azer3-5	-	-	136757	487.2	16642	100.00	A
Azer4	-	-	80707	309.59	16642	100.00	A
Azer6	94508	93159	9876	44	16632	99.93	A
Blagotin1	-	-	156487	544.63	16642	100.00	A

Blagotin16	-	-	101595	296.4	16642	100.00	A
Blagotin2	-	-	109741	253.67	16642	100.00	A
Blagotin3	-	-	280193	885.45	16642	100.00	A
Bulak1	-	-	56375	220.47	16642	100.00	A
Bulak2	-	-	67627	284.97	16642	100.00	A
Bulak3			50383	180.39	16642	100.00	A
Bulak4	456947	447106	17250	60.58	16591	99.69	B
Bulak5	-	-	31277	113.99	16642	100.00	D
Cav8	1950148	1805472	146	0.58	7668	46.1	A
Chalow1	2794368	2771283	69257	341.05	16642	100.00	D
Darre1	801170	777629	26121	42.51	16332	98.13	A
Darre2	-	-	75967	365.56	16642	100.00	A
Direkli1-2	-	-	263108	998.17	16436	98.76	T
Direkli4	213614388	187947946	42267	142.69	16481	99.03	F
Direkli5	-	-	38545	120.94	16327	98.10	T
Direkli6	-	-	147902	523.8	16405	98.57	T
Dra34	3962976	3762093	1594	5.44	13942	83.77	G
Fars1	787006	724742	18142	40.38	16334	98.14	A
Fars2-5	7737720	7595927	190296	284.47	16642	100.00	B
Fars4	-	-	30679	109.24	16643	100.00	A

Geor2	-	-	35923	108	16535	99.35	A
Ghosh5	71302178	70329490	1970	6.25	13020	78.23	F
Gilat10	5784648	5582429	639	2.05	5636	33.86	A
Gilat2	1779611	1617254	7066	16.58	15564	93.52	A
Gilat8	6635559	6474630	34690	133.77	16557	99.48	D
Hovk1	-	-	106522	519.39	16564	99.53	F
Kan19	5512329	5324761	2524	8.55	15133	90.93	A
Kan23	5371322	4640116	3297	10.56	16132	96.93	G
Kan25	5661187	5395828	2737	8.97	15222	91.46	A
Kazbeg1	-	-	48701	256.38	16641	99.99	A
Kohneh2	1347714	1266365	9516	29.56	16228	97.51	A
Kov27	2677699	2509963	802	3.12	9840	59.12	A
Kov57	3343260	3169499	1695	5.86	14259	85.68	A
Kov60	3241805	3013787	1884	6.97	15335	92.14	A
Lur12	-	-	127542	480	16641	100.00	G
Lur9	4076792	3841739	653	1.69	16643	100.00	B
Miqne5	2352930	2165308	5389	12.22	15322	92.06	A
Monjukli1	-	-	11337	96.78	16634	99.95	A
Monjukli2	-	-	22198	90.55	16641	100.00	D
Monjukli4	-	-	24260	123.04	16642	100.00	A

Monjukli6	494880	484544	31473	123.04	16641	100.00	D
Monjukli7	1436980	1372200	8835	31.87	16596	99.72	D
Monjukli8	-	-	97168	411.74	16641	100.00	D
Monjukli9	1276894	1232219	9637	33.02	16489	99.07	G
Ovc11	2146471	1996697	2521	8.63	15392	92.48	A
Pie17	3609599	3315507	2175	6.98	14815	89.02	A
Potterne1	-	-	36623	217.21	16642	100.00	A
Qazvin1	-	-	89984	448.01	16642	100.00	A
Safi2	1886858	1858363	36440	134.14	16628	99.91	A
Semnan1-2	-	-	211679	533.55	16595	99.71	B
Semnan10	-	-	122939	369.08	16641	100.00	G
Semnan13	-	-	27173	109.05	16641	100.00	D
Semnan17	-	-	35406	149.93	16641	100.00	D
Semnan3	-	-	328688	887.12	16641	100.00	D
Semnan7	-	-	103257	204.62	16641	100.00	D
Semnan8	-	-	26486	76.45	16641	100.00	D
Semnan9	-	-	103357	394.12	16641	100.00	G
Shiqmim1	44001686	43434209	4151	7	13292	79.87	D
Shiqmim9	1920566	1839287	247	0.85	1139	6.84	D
Tac1	4805905	4499253	3389	11.45	16150	97.04	A

Tac2	4353357	4024959	2713	12.39	16641	99.99	A
Tac3	1791228	1668795	6193	25.27	16564	99.53	A
Uiv17	2322143	2160471	268	0.95	1319	7.93	A
Ulu38	1743758	1612570	139	0.46	5849	35.1	A
Yarmut1	3524909	3392110	13436	40.41	16236	97.55	A
Yarmut7	1591537	1456896	3976	11.1	15282	91.82	A
Yoqneam2	-	-	92996	472.98	16642	100.00	A

Appendix Table 2.6 - Radiocarbon dating information for goat identified in Chapter 2. Two sigma calibration was performed using OxCal 4.3 (Bronk Ramsey 1994; Ramsey & Lee 2013) and IntCal 13 (Niu et al. 2013).

Sample Name	C14 Code	Context	Conventional Age (BP)	Calibrated C14 date (95.4% Probability)
Acem1	UBA-30288	Bronze Age	3782 +/- 41	2346-2040 cal BC
AP38	KIA-42163	Neolithic	6390 +/- 30	5468-5316 cal BC
AP46	KIA-42164	Neolithic	6210 +/- 30	5293-5057 cal BC
Blagotin1	UBA-30289	Neolithic	7391 +/- 56	6398-6098 cal BC
Blagotin2	UBA-30290	Neolithic	7361 +/- 62	6379-6078 cal BC
Blagotin3	UBA-30292	Neolithic	7135 +/- 53	6096-5892 cal BC
Darre2	UBA-34977	Chalcolithic	337 +/- 30	1473-1641 cal AD
Direkli1-2	Beta-425280	Late Epipaleolithic	11370 +/- 40	11351-11166 cal BC
Direkli4	Beta-432464	Late Epipaleolithic	12130 +/- 40	12191-11882 cal BC
Dra34	ERL-12297	Chalcolithic	5636 +/- 49	4580-4354 cal BC
Fars2-5	Beta-470335	Neolithic	7980 +/- 30	7047-6772 cal BC
Fars4	UBA-34976	Chalcolithic	6311 +/- 42	5460-5211 cal BC
Hovk1	UBA-31978	Paleolithic	>47074	NA
Kan23	KIA-42159	Bronze Age	4020 +/- 40	2833-2465 cal BC
Lur12	Beta-470334	Neolithic	8810 +/- 30	8171-7745 cal BC
Semnan1-2	UBA-33144	Neolithic	8157 +/- 74	7454-6850 cal BC
Semnan3	UBA-33145	Neolithic	7214 +/- 53	6214-6004 cal BC

Appendix Material for Chapter 3

Appendix Table 3.1 - Partitions used for BEAST mtDNA analysis.

Partition	Region
C1-2	2743-3697\3 2744-3698\3 3908-4949\3 3909-4947\3 5331-6873\3 5332-6874\3 7018-7699\3 7019-7700\3 7773-7929\3 7774-7930\3 7931-7969 7970-8609\3 7971-8610\3 8611-9394\3 8612-9392\3 9464-9809\3 9465-9807\3 9880-10168\3 9881-10169\3 10170-11547\3 10171-11545\3 11750-13553\3 11751-13554\3 13555-13570 13571-14081\3 13573-14080\3 14155-15292\3 14156-15293\3
C3	2745-3696\3 3910-4948\3 5333-6875\3 7020-7701\3 7775-7928\3 7972-8608\3 8613-9393\3 9466-9808\3 9882-10167\3 10172-11546\3 11752-13552\3 13572-14079\3 14157-15294\3
rRNA	69-638 1093-2665
tRNA	1-68 1026-1092 2666-2740 3699-3836 3839-3907 4950-5016 5018-5086 5088-5160 5194-5329 6876-6941 6949-7016 7705-7771 9395-9463 9811-9879 11548-11678 11680-11749 14082-14150 15298-15432
D_loop	15433-16647
Remainder	639-1025 2741-2742 3837-3838 5017 5087 5161-5193 5330 6942-6948 7017 7702-7704 7772 9810 11679 14151-14154 15295-15297

Appendix Table 3.2 - Population definitions used in AMOVA.

Structure - Neolithic	Structure - Chalcolithic/Bronze Age	Structure - Iron Age/Medieval/Modern
Group 1	**Group 1**	**Group 1**
Neolithic SE Europe	Chalcolithic/Bronze Age SE Europe	Modern East/Central Asia
Neolithic West Anatolia	Chalcolithic/Bronze Age Anatolia	Iron Age-Modern Iran/Caucasus
Neolithic Serbia	**Group 2**	**Group 2**
Group 2	Bronze Age Uzbekistan	Iron-Modern Levant
Neolithic East Iran	Chalcolithic Turkmenistan	**Group 3**
Neolithic West Iran	Chalcolithic/Bronze Age Caucasus	Modern Central Europe
Neolithic Turkmenistan	Chalcolithic/Bronze Age Iran	Modern Mediterranean
Group 3	**Group 3**	**Group 4**
Neolithic Levant	Chalcolithic/Bronze Age Levant	Modern Turkey

Appendix Table 3.3 - Group definitions used in AMOVA.

Neolithic	Chalcolithic/Bronze Age	Iron Age/Medieval/Modern	
Neolithic East Iran	**Bronze Age Uzbekistan**	**Iron Age-Modern Iran/Caucasus**	**Modern Central Europe**
Semnan1-2	Bulak1	Azer4	A1a_02
Semnan10	Bulak2	Darre2	A1a_05
Semnan13	Bulak4	Geor2	A3_22
Semnan17	Bulak5	Kazbeg1	A4_26
Semnan3	**Chalcolithic Turkmenistan**	Tac2	A4_27
Semnan7	Monjukli1	A_45	A5_29
Semnan8	Monjukli2	A_47	A7_36
Semnan9	Monjukli4	A_48	A7_37
Neolithic Levant	Monjukli6	A_62	A_43
Ainghazal1	**Chalcolithic/Bronze Age Caucasus**	A_63	A_50
Ainghazal2	Azer3-5	A_64	A_54
Ainghazal4	Azer6	G_69	C1a_81
Neolithic South East Europe	Kohneh2	G_72	C1a_82
Kov57	Tac1	G_73	**Modern Mediterranean**
Kov60	Tac3	G_74	A1_01
Ovc11	**Chalcolithic/Bronze Age Iran**	**Iron-Modern Levant**	A1a_03
Neolithic Serbia	Chalow1	Miqne5	A1a_04
Blagotin1	Darre1	A2_08	A2a_11
Blagotin16	Fars1	A2a1_15	A2a1_17
Blagotin2	Fars4	A5_30	A2a1_18
Blagotin3	Qazvin1	A_39	A2a1_19
Neolithic Turkmenistan	**Chalcolithic/Bronze Age SE Europe**	**Modern East/Central Asia**	A3_20
Monjukli7	Dra34	A35	A3_21
Monjukli8	Pie17	D1_67	A3_23
Monjukli9	**Chalcolithic/Bronze Age Anatolia**	D1_68	A3_24
Neolithic West Anatolia	Acem1	B1_78	A4_25

AP38	Acem2	**Modern Turkey**	A4_28
AP45	Kan19	A1a_06	A5_31
AP46	Kan23	A2_07	A5_32
AP49	Kan25	A2a1_13	A7_38
AP50	**Chalcolithic/Bronze Age Levant**	A2a1_16	A_41
Neolithic West Iran	Gilat10	A6_33	A_46
Fars2-5	Gilat2	A6_34	A_49
Lur12	Gilat8	A_42	A_51
	Safi2	A_55	A_52
	Shiqmim1	A_56	A_53
	Shiqmim9	A_57	A_61
	Yarmut1	A_58	A_65
	Yarmut7	A_59	C1a_83
	Yoqneam2	A_60	
		G_70	
		G_71	

Appendix Table 3.4 - Modeling information for ancient genomes. 3=Neolithic, 2=Bronze Age / Chalcolithic, 1 = Iron Age / Medieval. W=West, E=East, L=Levant.

Sample name	Population	mtDNA missing data	mtDNA modelling	Additional mtDNA modelling	Autosomal modelling
Kov57	3W	2104	yes	yes	no
Kov60	3W	1079	yes	yes	no
Blagotin16	3W	0	yes	yes	yes
Blagotin1	3W	0	yes	yes	yes
Blagotin2	3W	0	yes	yes	yes
Blagotin3	3W	0	yes	yes	yes
Ovc11	3W	930	yes	yes	no
Semnan10	3E	0	yes	yes	no
Semnan1-2	3E	0	yes	yes	yes
Semnan13	3E	0	yes	yes	yes
Semnan17	3E	0	yes	yes	no
Semnan3	3E	0	yes	yes	yes
Semnan7	3E	0	yes	yes	yes
Semnan8	3E	0	yes	yes	no
Semnan9	3E	0	yes	yes	yes
Monjukli7	3E	45	yes	yes	no
Monjukli8	3E	0	yes	yes	no
Monjukli9	3E	116	yes	yes	no
Lur12	3E	0	yes	yes	no
Fars2-5	3E	0	yes	yes	no
Ainghazal1	3L	3598	yes	no	no
Ainghazal2	3L	743	yes	no	no
Ainghazal4	3L	63	yes	no	no
Acem1	2W	0	no	yes	no
Acem2	2W	0	no	yes	no
Kan19	2W	1510	no	yes	no
Kan23	2W	511	no	yes	no
Kan25	2W	1421	no	yes	no

Dra34	2W	2701	no	yes	no
Pie17	2W	1828	no	yes	no
Azer3-5	2E	0	no	yes	no
Azer6	2E	11	no	yes	no
Bulak1-3	2E	0	no	yes	no
Bulak2	2E	0	no	yes	no
Bulak4	2E	52	no	yes	no
Bulak5	2E	0	no	yes	no
Chalow1	2E	0	no	yes	no
Darre1	2E	311	no	yes	no
Fars1	2E	309	no	yes	no
Fars4	2E	0	no	yes	no
Kohneh2	2E	415	no	yes	no
Monjukli1	2E	9	no	yes	no
Monjukli2	2E	0	no	yes	no
Monjukli4	2E	0	no	yes	no
Monjukli6	2E	0	no	yes	no
Qazvin1	2E	0	no	yes	no
Tac1	2E	493	no	yes	no
Tac3	2E	79	no	yes	no
Tac2	3E	2	no	yes	no
Kazbeg1	3E	2	no	yes	no
Darre2	3E	0	no	yes	no
Azer4	3E	0	no	yes	no

Appendix Table 3.5 - mtDNA diversity estimates for Neolithic goat modelling groups.

Both the number of usable sites and π were calculated as average per population.

Population	Acronym	Sample size	No. of usable sites	π per site(10^{-4})
Neolithic West	3W	7	14315.5	1.63
Neolithic East	3E	13	15404.2	22.4
Neolithic Levant	3L	3	12717.7	15.2

Appendix Table 3.6 - mtDNA F_{ST} between Neolithic goat modelling groups.

Population	Neolithic West (3W)	Neolithic East (3E)	Neolithic Levant (3L)
Neolithic West (3W)	/	0.16	0.97
Neolithic East (3E)	0.16	/	0.88
Neolithic Levant (3L)	0.97	0.88	/

Appendix Table 3.7 - Prior distributions for all parameters of model SINGLE_MT and MULTIPLE_MT. Time points are expressed in generations considering already that our Neolithic samples are placed at 8,000 years ago.

Model SINGLE_MT		**Model MULTIPLE_MT**	
Nneol	Uniform (10-50,000)	Nneol	Uniform (10-50,000)
Nneow	Uniform (10-50,000)	Nneow	Uniform (10-50,000)
Nneoe	Uniform (10-50,000)	Nneoe	Uniform (10-50,000)
Nbotl	Uniform (10-5,000)	Nbotl	Uniform (10-5,000)
Nbotw	Uniform (10-5,000)	Nbotw	Uniform (10-5,000)
Nbote	Uniform (10-5,000)	Nbote	Uniform (10-5,000)
Nanc1	Uniform (10-50,000)	Nanc1l	Uniform (1000-50,000)
Nanc2	Uniform (1000-50,000)	Nanc1w	Uniform (1000-50,000)
		Nanc1e	Uniform (1000-50,000)
Rules applied:		Nanc2	Uniform (1000-50,000)
Nbotl, Nbotw and Nbote < Nanc1		Nanc3	Uniform (1000-50,000)
Nanc1< Nanc2		Tsplit	Uniform (4400-36,000)
Nbot < Nneo for each population		Tlevant	Uniform (Tsplit-80,000)
		Rules applied:	
		Nbotl, Nbotw and Nbote < Nanc1	
		Nanc1w and Nanc1e < Nanc2	
		Nbot < Nneo for each population	
		Nanc2 < Nanc3	
		Nanc1l < Nanc3	

Appendix Table 3.8 - Prior distributions for all parameters of additional modelling experiment. Time points are expressed in generations.

Parameter	Prior Distribution
Nmodw	10-1000000
Nmode	10-1000000
Nancw	10-1000000
Nance	10-1000000
Nanc1	10-1000000
resize2	0-1000000
g1w	-0.008-0
g1e	-0.008-0
Tc1	1333-3500
Tsplit	3667-30000
Tmig	10-3000
m1	0-1
m2	0-1

Appendix Table 3.9 - Addition model posterior probabilities. Calculated by a weighted multinomial logistic regression using whole mitochondrial genomes.

Number of simulations retained	Model BOT	Model SPLIT	Model SPLIT-MIG
25000	0.05	0	0.95
50000	0.06	0	0.94

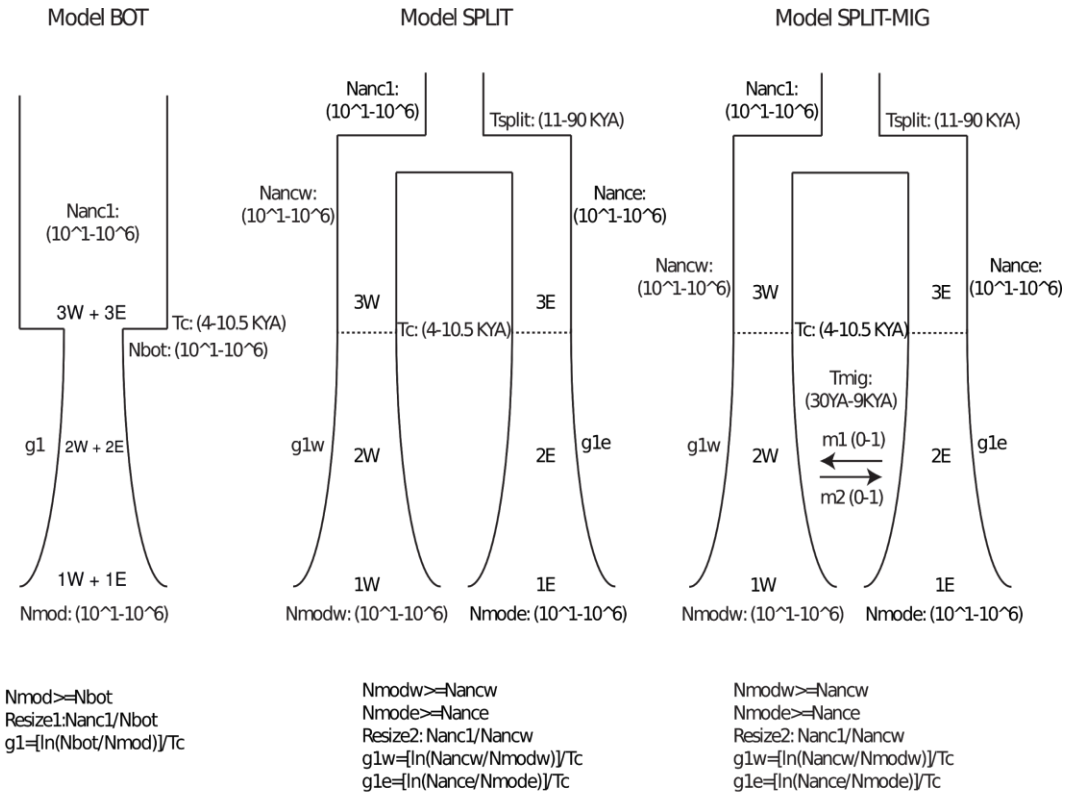
Appendix Table 3.10 - Nucleotide diversity (π) and Hudson's F_{ST} calculated for populations used in the additional mtDNA modelling experiment. Population abbreviations are the same used as in Appendix Table 3.6.

Fst_3W_3E	Fst_2W_2E	Fst_1W_1E	Fst_3E_2E	Fst_2E_1E	Fst_3W_2W	Fst_2W_1W
0.218	0.0365	0	0.1996	0.0755	0	0
π_3W	π_2W	π_1W	π_3E	π_2E	π_1E	
0.00178	0.00141	0.00113	0.00224	0.0012	0.00143	

Appendix Table 3.11 - Parameter estimation under the most supported model in second modelling experiment (SPLIT-MIG). Prior distributions and estimates of both Tsplit and Tmig have been converted in years using a generation time of 2.5 years.

Model SPLIT-MIG	Prior	Median	Mode	95% Credible Interval	0.025	0.975
Tsplit	9,167.5-75,000	30,695	30,405	13,757-56,318	13,757	56,318
Tmig	250-8,500	6,488	6,898	3,177-8,506	3,177	8,506

Appendix Figure 3.1 - Additional demographic models tested with whole mitochondrial genomes. 3E=Neolithic East, 3W=Neolithic West, 2E=Chalcolithic/Bronze Age East, 2W=Chalcolithic/Bronze Age West, 1E=Iron Age/Medieval East, 1W=Iron Age/Medieval West, KYA=thousand years ago.



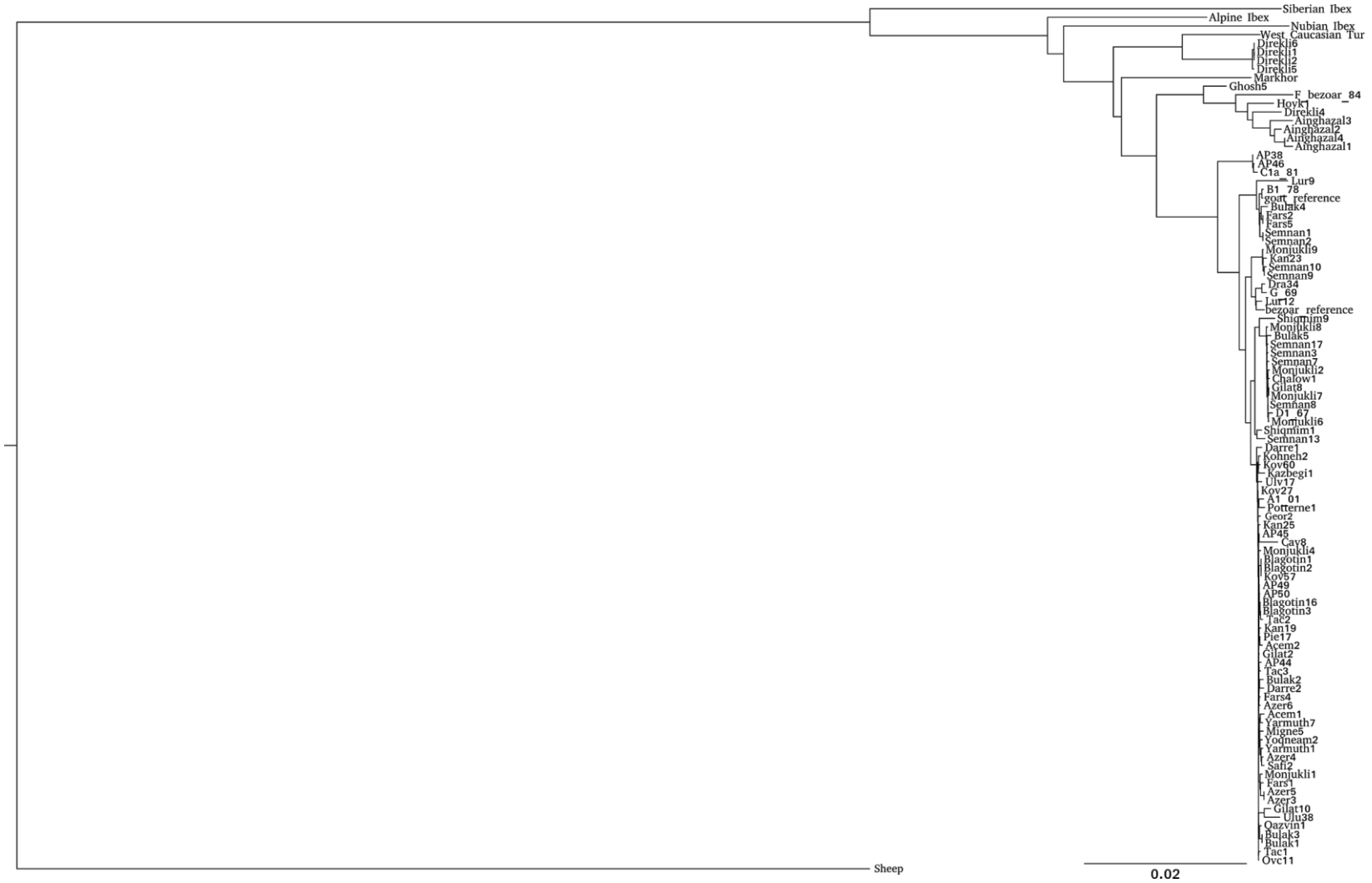
Nmod>=Nbot
Resize1:Nanc1/Nbot
g1=[ln(Nbot/Nmod)]/Tc

Nmodw>=Nancw
Nmode>=Nance
Resize2: Nanc1/Nancw
g1w=[ln(Nancw/Nmodw)]/Tc
g1e=[ln(Nance/Nmode)]/Tc

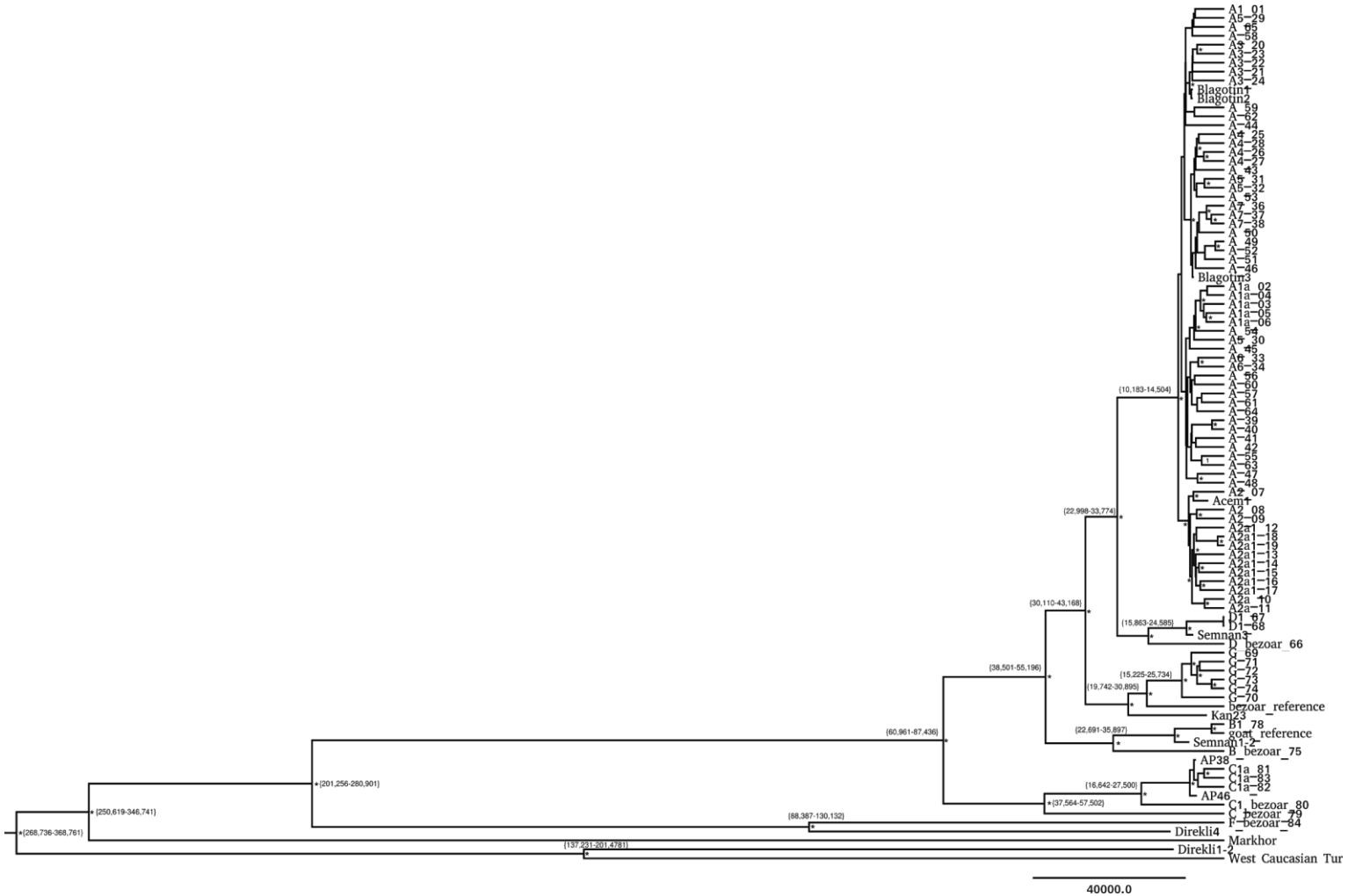
Nmodw>=Nancw
Nmode>=Nance
Resize2: Nanc1/Nancw
g1w=[ln(Nancw/Nmodw)]/Tc
g1e=[ln(Nance/Nmode)]/Tc

GROUP 1W: 15 samples (modern+Iron Age+Medieval) at 0 generations ago
GROUP 2W: 7 samples (Chalcolithic + Bronze Age) at 2000 generations ago, ~5 KYA
GROUP 3W: 11 samples (Neolithic) at 3200 generations ago, ~8 KYA

GROUP 1E: 12 samples (modern+Iron Age+Medieval) at 0 generations ago
GROUP 2E: 18 samples (Chalcolithic + Bronze Age) at 2000 generations ago, ~5 KYA
GROUP 3E: 11 samples (Neolithic) at 3200 generations ago, ~8 KYA



Appendix Figure 3.1 - Initial Maximum Likelihood phylogeny of ancient goat and modern *Capra* genus mitochondria.



Appendix Figure 3.2 - Phylogeny constructed using BEAST, using *Capra hircus, aegagrus, falconeri* and *caucasica*. Posterior support for high confidence (>0.6) nodes are marked by an asterix. 95% highest posterior density for key nodes are shown in braces, and rounded to whole numbers.

Appendix Material for Chapter 4

Appendix Table 4.1 - Modern Genomes used. Alignment was performed to CHIR_1.0. Mean genomic coverage calculated following mapping quality 30 filter. Moroccan goat originally published in (Benjelloun et al. 2015); French and Iranian domestic/bezoar genomes originally published in (Alberto et al. 2018a). Yak genomic reads originally published in (Qiu et al. 2012).

Sample ID	ENA Accession	Breed/Location	Genome Coverage	NGS admix	Grouping
Tog	ERR2534444	Togolese Village Goat, Togo	36.48	Yes	Modern Togo
IOG	ERR2534443	Old Irish Goat (feral), Ireland	41.93	Yes	Modern Ireland
French 1	ERR470103	French Alpine	14.52	Yes	Modern France
French 2	ERR470101	French Saanen	13.75	Yes	Modern France
French 3	ERR470102	French Saanen	13.47	Yes	Modern France
French 4	ERR470105	French Alpine	14.19	Yes	Modern France
Iranian 1	ERR313200	Iran	13.89	No	Modern Iran
Iranian 2	ERR313209	Iran	12.07	No	Modern Iran
Iranian 3	ERR313211	Iran	12.92	No	Modern Iran
Iranian 4	ERR313213	Iran	13.12	Yes	Modern Iran
Iranian 5	ERR340339	Iran	11.62	No	Modern Iran
Iranian 6	ERR299456	Iran	13.06	Yes	Modern Iran
Iranian 7	ERR313210	Iran	13.41	No	Modern Iran

Iranian 8	ERR340332	Iran	11.53	No	Modern Iran
Iranian 9	ERR340337	Iran	11.24	No	Modern Iran
Iranian 10	ERR313215	Iran	13	Yes	Modern Iran
Iranian 11	ERR297229	Iran	13.24	No	Modern Iran
Iranian 12	ERR313206	Iran	12.26	Yes	Modern Iran
Iranian 13	ERR299449	Iran	13.24	No	Modern Iran
Iranian 14	ERR313197	Iran	12.27	Yes	Modern Iran
Iranian 15	ERR313198	Iran	12.32	No	Modern Iran
Iranian 16	ERR313199	Iran	13.75	No	Modern Iran
Iranian 17	ERR313202	Iran	12.87	No	Modern Iran
Iranian 18	ERR313204	Iran	12.99	No	Modern Iran
Iranian 19	ERR313207	Iran	13.67	No	Modern Iran
Iranian 20	ERR313212	Iran	13.3	No	Modern Iran
Iranian bezoar 1	ERR470100	Qazvin Province	14.96	No	Modern Qazvin Bezoar
Iranian bezoar 2	ERR340330	Qazvin Province	12.79	No	Modern Qazvin Bezoar
Iranian bezoar 3	ERR340340	Iranian Azerbaijan	12.42	No	Modern Azerbaijan Bezoar
Iranian bezoar 4	ERR340328	Qazvin Province	6.87	Yes	Modern Qazvin Bezoar
Iranian bezoar 5	ERR340329	Khan Gomaz, West Hamedan	5.63	No	Modern Hamedan Bezoar
Iranian bezoar 6	ERR340331	Zanjan, Iranian Azerbaijan	11.26	No	Modern Azerbaijan Bezoar
Iranian bezoar 7	ERR340333	Khan Gomaz, West Hamedan	6.13	No	Modern Hamedan Bezoar
Iranian bezoar 8	ERR340334	Khoy, Iranian Azerbaijan	11.67	No	Modern Azerbaijan Bezoar

Iranian bezoar 9	ERR340335	Lashgardar, East Hamedan	6.85	Yes	Modern Hamedan Bezoar
Iranian bezoar 10	ERR340338	Qazvin Province	11.72	No	Modern Qazvin Bezoar
Iranian bezoar 11	ERR340336	Iranian Azerbaijan	6.81	No	Modern Azerbaijan Bezoar
Iranian bezoar 12	ERR340341	Khan Gomaz, West Hamedan	5.34	No	Modern Hamedan Bezoar
Iranian bezoar 13	ERR340342	Iranian Azerbaijan	6.69	No	Modern Azerbaijan Bezoar
Iranian bezoar 14	ERR340343	Azerbaijan	7.54	Yes	Modern Azerbaijan Bezoar
Iranian bezoar 15	ERR340344	Azerbaijan	11.52	No	Modern Azerbaijan Bezoar
Iranian bezoar 16	ERR340345	Iranian Azerbaijan	12.12	Yes	Modern Azerbaijan Bezoar
Iranian bezoar 17	ERR340347	Iranian Azerbaijan	12.93	No	Modern Azerbaijan Bezoar
Iranian bezoar 18	ERR340348	Haftad Qolleh, Markazi Province	10.07	No	Modern Hamedan Bezoar
Iranian bezoar 19	ERR340426	Iranian Azerbaijan	12.73	No	Modern Azerbaijan Bezoar
Iranian bezoar 20	ERR470104	Qazvin Province	13.55	Yes	Modern Qazvin Bezoar
Iranian bezoar 21	ERR470106	Qazvin Province	14.61	No	Modern Qazvin Bezoar
CHIR_1.0	PRJNA158393	Chinese Yunnan Black	20.86	No	Modern China
Moroccan 1	ERR219543	Moroccan Northern	12.71	No	Modern Morocco
Moroccan 2	ERR234315	Moroccan Black	13.84	Yes	Modern Morocco
Moroccan 3	ERR315500	Moroccan	13.59	Yes	Modern Morocco
Moroccan 4	ERR229484	Moroccan Black	15.43	No	Modern Morocco
Moroccan 5	ERR232497	Moroccan	12.61	Yes	Modern Morocco
Moroccan 6	ERR246139	Moroccan Black	15.08	No	Modern Morocco
Moroccan 7	ERR248932	Moroccan	15.82	No	Modern Morocco

Moroccan 8	ERR315498	Moroccan Draa	13.51	No	Modern Morocco
Moroccan 9	ERR340428	Moroccan	12.93	Yes	Modern Morocco
Yak	PRJNA74739	Reference Individual	20.83	N/A	N/A

Appendix Table 4.2 - Groupings used for ancient genomes.

Sample	Grouping	Sample	Grouping
Blagotin1	Neolithic West	Monjukli4	Chalcolithic Turkmenistan
Blagotin2	Neolithic West	Monjukli6	Chalcolithic Turkmenistan
Blagotin3	Neolithic West	Monjukli8	Neolithic East
Blagotin16	Neolithic West	Acem1	Bronze Age Turkey
Kov57	Neolithic West	Acem2	Bronze Age Turkey
AP45	Neolithic West	Tac3	Bronze Age Caucasus
AP49	Neolithic West	Kazbeg1	Iron Age/Medieval Caucasus
Direkli1-2	Turkish Ancient Wild	Geor2	Iron Age/Medieval Caucasus
Direkli5	Turkish Ancient Wild	Kohneh2	Bronze Age Caucasus
Direkli6	Turkish Ancient Wild	Azer3-5	Bronze Age Caucasus
Ainghazal1	Neolithic Levant	Azer4	Iron Age/Medieval Caucasus
Ainghazal2	Neolithic Levant	Azer6	Chalcolithic Caucasus
Ainghazal3	Neolithic Levant	Qazvin1	Bronze Age Iran
Ainghazal4	Neolithic Levant	Darre1	Chalcolithic Iran
Hovk1	Armenian Ancient Wild	Darre2	Medieval Iran
Lur12	Neolithic East	Fars4	Chalcolithic Iran
Semnan1-2	Neolithic East	Chalow1	Bronze Age Iran
Semnan3	Neolithic East	Bulak1	Bronze Age Uzbekistan
Semnan7	Neolithic East	Bulak2	Bronze Age Uzbekistan
Semnan8	Neolithic East	Bulak5	Bronze Age Uzbekistan
Semnan9	Neolithic East	Shiqmim1	Chalcolithic Levant
Semnan10	Neolithic East	Gilat2	Chalcolithic Levant

Semnan13	Neolithic East	Gilat8	Chalcolithic Levant
Semnan17	Neolithic East	Yarmut1	Bronze Age Levant
Fars1	Chalcolithic Iran	Yarmut7	Bronze Age Levant
Fars2-5	Neolithic East	Yoqneam2	Bronze Age Levant
Monjukli1	Chalcolithic Turkmenistan	Safi2	Bronze Age Levant
Monjukli2	Chalcolithic Turkmenistan	Potterne1	Bronze Age Britain

Appendix Table 4.3 - Assorted *D* statistics. Yak was used to define the reference allele. Significance is taken as $|Z| \geq 3$ and calculated using ANGSD.

H1	H2	H3	nABBA	nBABA	D jackEst	SE	Z
Semnan3	Lur12	Blagotin3	90531	90299	0.001282973	0.003590864	0.357288
Semnan3	Fars2-5	Blagotin3	4561	4485	0.008401503	0.0109927	0.7642805
Semnan3	Iranian 11	Blagotin3	173152	164723	0.0249471	0.003560577	7.006475
Semnan3	Iranian 1	Blagotin3	172781	163255	0.02834815	0.003269891	8.669448
Semnan3	Iranian 12	Blagotin3	161082	162791	-0.00527676	0.003339096	-1.580296
Semnan3	Iranian 7	Blagotin3	170882	163370	0.02247406	0.003402284	6.605582
Semnan3	Iranian 10	Blagotin3	173392	164379	0.02668376	0.003358238	7.945761
French2	French3	Direkli1-2	132155	128976	0.01217397	0.0033131	3.674494
French2	French4	Direkli1-2	153793	152065	0.005649681	0.003385153	1.668959
French3	French4	Direkli1-2	151741	153216	-0.004836747	0.003332339	-1.451457
French2	French1	Direkli1-2	154438	152020	0.007890151	0.003410438	2.31353
French3	French1	Direkli1-2	151251	151957	-0.002328435	0.003340198	-0.6970947
French4	French1	Direkli1-2	125725	124935	0.00315168	0.003120129	1.010112
French2	Blagotin3	Direkli1-2	166986	150229	0.05282537	0.003446586	15.32687
French3	Blagotin3	Direkli1-2	164640	151000	0.04321379	0.003570149	12.1042
French4	Blagotin3	Direkli1-2	165628	150523	0.0477778	0.003581599	13.3398
French1	Blagotin3	Direkli1-2	164305	149938	0.0457194	0.003690416	12.38868
French2	Potterne1	Direkli1-2	150069	139535	0.03637381	0.003751404	9.696051
French3	Potterne1	Direkli1-2	147580	140170	0.02575152	0.003465676	7.430447
French4	Potterne1	Direkli1-2	147536	138646	0.03106415	0.003508202	8.854722

French1	Potterne1	Direkli1-2	146995	138688	0.02907768	0.003645965	7.975305
Blagotin3	Potterne1	Direkli1-2	143151	147986	-0.0166073	0.003625681	-4.580465
French2	Semnan3	Direkli1-2	169915	186146	-0.04558489	0.003556095	-12.81881
French3	Semnan3	Direkli1-2	167451	186825	-0.05468618	0.003382434	-16.1677
French4	Semnan3	Direkli1-2	169153	187171	-0.05056634	0.003500675	-14.44474
French1	Semnan3	Direkli1-2	168356	187052	-0.05260433	0.003528999	-14.9063
Blagotin3	Semnan3	Direkli1-2	156911	190579	-0.09688912	0.003332698	-29.07228
Potterne1	Semnan3	Direkli1-2	149440	176616	-0.08334765	0.003285069	-25.37166
Blagotin3	Potterne1	Direkli1-2	142474	148026	-0.01911188	0.003538385	-5.401298
Blagotin3	Semnan3	Direkli1-2	156586	190721	-0.0982848	0.003297027	-29.81013
Potterne1	Semnan3	Direkli1-2	149802	176855	-0.08281776	0.003318845	-24.95379
Blagotin3	IOG	Direkli1-2	149565	171755	-0.06905888	0.003477345	-19.85966
Potterne1	IOG	Direkli1-2	134440	149645	-0.05352271	0.003725182	-14.36781
Semnan3	IOG	Direkli1-2	185947	175002	0.03032284	0.003439599	8.815808
Acem1	AP45	Semnan3	2446	2566	-0.02394254	0.01469121	-1.629719
Acem2	AP45	Semnan3	2338	2619	-0.05668751	0.01487752	-3.810279
Acem1	AP49	Semnan3	2454	2496	-0.008484848	0.01473113	-0.5759808
Acem2	AP49	Semnan3	2440	2573	-0.02653102	0.01480548	-1.791973
Acem1	Blagotin16	Semnan3	137452	148743	-0.03945212	0.0034203	-11.5347
Acem2	Blagotin16	Semnan3	135403	152100	-0.05807592	0.003622817	-16.03059
Acem1	Blagotin1	Semnan3	151830	165776	-0.04390975	0.003304099	-13.28948
Acem2	Blagotin1	Semnan3	148846	170252	-0.06708284	0.003492531	-19.20751
Acem1	Blagotin2	Semnan3	142216	155236	-0.04377177	0.003325738	-13.16152
Acem2	Blagotin2	Semnan3	139646	158901	-0.06449571	0.00349084	-18.4757
Acem1	Blagotin3	Semnan3	154575	169667	-0.04654548	0.003196415	-14.56178
Acem2	Blagotin3	Semnan3	151474	174279	-0.07000703	0.003293545	-21.25583
Acem1	Kov57	Semnan3	9176	10000	-0.04297038	0.007779949	-5.523222
Acem2	Kov57	Semnan3	9134	10139	-0.05214549	0.007953276	-6.556479
Acem1	Semnan3	Blagotin3	154575	174833	-0.0614982	0.003210358	-19.15618
Acem2	Semnan3	Blagotin3	151474	182385	-0.09258699	0.003362565	-27.53463

Appendix Table 4.4 - qpGraph SNP counts.

Figure	Number of SNPs	Description
4.1a	134566	Rejected baseline
4.1b	134566	Wild admixture from Direkli
4.1c	12023	Previous graph with nL
4.1d	12023	nL is admixed with wild
4.1e	9009	Previous graph with aAR
4.1f	16040	All moderns
4.1g	11740	All moderns with aAR
4.2a	9908	Bronze Age Levant
4.2b	146711	BA Turkish with no nL; admixed with East receiving additional wild
4.2c	94585	Potterne, no nL, addition wild into nW
4.2d	28432	All Caucasus - East admixed, with nE wild admixed
4.2e	86099	Chalcolithic East
4.2f	65597	BA Uzbek as BA East Iran and western admixture
4.2g	113979	Darre2 skeleton

Appendix Table 4.5 - Prior distributions for all parameters of model SINGLE_AU and BINARY_AU. Time points are expressed in generations considering already that our Neolithic samples are placed at 8,000 years ago.

Model SINGLE_AU		**Model BINARY_AU**	
Nneow	Uniform (10-50,000)	Nneow	Uniform (10-50,000)
Nneoe	Uniform (10-50,000)	Nneoe	Uniform (10-50,000)
Nbotw	Uniform (10-5,000)	Nbotw	Uniform (10-5,000)
Nbote	Uniform (10-5,000)	Nbote	Uniform (10-5,000)
Nanc1	Uniform (10-50,000)	Nanc1w	Uniform (10-50,000)
Nanc2	Uniform (1,000-50,000)	Nanc2e	Uniform (1,000-50,000)
		Nanc2e	Uniform (1,000-50,000)
		Tsplit	Uniform (4,400-36,000)
Rules applied:		**Rules applied:**	
Nbotw and Nbote < Nanc1		Nbot < Nanc1 for each population	
Nanc1< Nanc2		Nanc1w and Nanc1e < Nanc2	
Nbot < Nneo for each population		Nbot < Nneo for each population	

Appendix Table 4.6 - lcMLkin results for top 30 pi_HAT comparisons.

Ind1	Ind2	k0_hat	k1_hat	k2_hat	pi_HAT	nbSNP
Direkli1	Direkli2	0.002	0.055	0.943	0.971	7700
Azer3	Azer5	0.001	0.098	0.902	0.95	5569
Semnan1	Semnan2	0.004	0.104	0.892	0.944	7601
Bulak1	Bulak3	0.036	0.052	0.912	0.938	2300
Ap49	Blagotin16	0.48	0.236	0.284	0.402	167
Fars1	Fars4	0.407	0.398	0.195	0.394	134
Azer3	Darre1	0.22	0.778	0.002	0.391	212
Ap45	Blagotin1	0.605	0.055	0.34	0.368	135
Darre2	Kohneh2	0.387	0.548	0.065	0.339	300
Fars5	Geor2	0.344	0.653	0.003	0.329	166
Direkli5	Kov58	0.571	0.234	0.195	0.312	107
Fars1	Yoqneam2	0.548	0.298	0.154	0.303	158
Azer4	Fars5	0.443	0.548	0.01	0.283	221
Monjukli1	Tac3	0.56	0.322	0.119	0.28	203
Ainghazal1	Blagotin3	0.682	0.093	0.225	0.271	260
Direkli1	Direkli6	0.518	0.435	0.047	0.265	5482
Semnan17	Semnan9	0.618	0.259	0.122	0.252	861
Semnan7	Semnan9	0.579	0.35	0.072	0.247	8248
Monjukli8	Semnan8	0.629	0.256	0.115	0.243	1354
Fars1	Qazvin1	0.74	0.045	0.215	0.238	173
Fars5	Semnan13	0.531	0.463	0.005	0.237	201
Chalow1	Lur12	0.741	0.055	0.204	0.231	228
Semnan13	Semnan7	0.587	0.365	0.048	0.231	7450
Semnan17	Semnan8	0.563	0.42	0.016	0.227	177
Ap45	Blagotin2	0.56	0.428	0.013	0.226	128
Direkli2	Direkli6	0.607	0.336	0.057	0.225	6785
Blagotin16	Blagotin1	0.636	0.282	0.082	0.223	8705
Bulak2	Chalow1	0.748	0.065	0.187	0.22	365
Blagotin2	Blagotin3	0.61	0.342	0.048	0.219	8990
Blagotin1	Blagotin3	0.626	0.311	0.063	0.218	9461

Appendix Table 4.7 - f_3 outgroup results. Qazvin bezoar were used as an outgroup.

Source 1	Source 2	Target	f_3	std err	Z	SNPs
Neolithic Levant	Bronze Age Britain	Qazvin Bezoar	0.120223	0.002708	44.395	80884
Neolithic Levant	Bronze Age Caucasus	Qazvin Bezoar	0.089245	0.00222	40.203	103808
Neolithic Levant	Bronze Age Iran	Qazvin Bezoar	0.082391	0.002328	35.398	85115
Neolithic Levant	Bronze Age Israel	Qazvin Bezoar	0.099812	0.002379	41.951	93174
Neolithic Levant	Bronze Age Turkey	Qazvin Bezoar	0.108056	0.001835	58.887	131079
Neolithic Levant	Bronze Age Uzbekistan	Qazvin Bezoar	0.07509	0.001842	40.756	102464
Neolithic Levant	Chalcolithic Caucasus	Qazvin Bezoar	0.089444	0.003676	24.329	25874
Neolithic Levant	Chalcolithic Iran	Qazvin Bezoar	0.079048	0.00232	34.072	75034
Neolithic Levant	Chalcolithic Israel	Qazvin Bezoar	0.125655	0.015103	8.32	1649
Neolithic Levant	Chalcolithic Turkmenistan	Qazvin Bezoar	0.073351	0.002264	32.405	72073
Neolithic Levant	Modern France	Qazvin Bezoar	0.123137	0.001917	64.225	145740
Neolithic Levant	Modern Iran	Qazvin Bezoar	0.08833	0.001428	61.875	203100
Neolithic Levant	Iron Age/Medieval Caucasus	Qazvin Bezoar	0.095171	0.001864	51.068	121725
Neolithic Levant	Medieval Iran	Qazvin Bezoar	0.08672	0.002423	35.797	96871
Neolithic Levant	Modern China	Qazvin Bezoar	0.084568	0.001812	46.661	124081
Neolithic Levant	Modern Ireland	Qazvin Bezoar	0.12493	0.002399	52.072	125923
Neolithic Levant	Modern Togo	Qazvin Bezoar	0.135533	0.002453	55.256	125458
Neolithic Levant	Modern Morocco	Qazvin Bezoar	0.127492	0.001911	66.708	169265
Neolithic Levant	Neolithic West	Qazvin Bezoar	0.131618	0.002029	64.859	133583
Neolithic Levant	Neolithic East	Qazvin Bezoar	0.06898	0.001388	49.709	138950
Neolithic West	Bronze Age Britain	Qazvin Bezoar	0.137809	0.001571	87.703	1168300
Neolithic West	Bronze Age Caucasus	Qazvin Bezoar	0.081779	0.001116	73.275	1457112
Neolithic West	Bronze Age Iran	Qazvin Bezoar	0.081207	0.001103	73.614	1189062
Neolithic West	Bronze Age Israel	Qazvin Bezoar	0.086265	0.001133	76.164	1311383
Neolithic West	Bronze Age Turkey	Qazvin Bezoar	0.094064	0.001149	81.841	1808429
Neolithic West	Bronze Age Uzbekistan	Qazvin Bezoar	0.077015	0.001028	74.947	1415273
Neolithic West	Chalcolithic Caucasus	Qazvin Bezoar	0.080707	0.001256	64.248	356323
Neolithic West	Chalcolithic Iran	Qazvin Bezoar	0.076714	0.001097	69.943	1012144
Neolithic West	Chalcolithic Israel	Qazvin Bezoar	0.094858	0.003482	27.244	21272
Neolithic West	Chalcolithic Turkmenistan	Qazvin Bezoar	0.075812	0.001047	72.397	979387
Neolithic West	Modern France	Qazvin Bezoar	0.142171	0.001343	105.849	1909881
Neolithic West	Modern Iran	Qazvin Bezoar	0.082665	0.000954	86.642	2583509

Neolithic West	Iron Age/Medieval Caucasus	Qazvin Bezoar	0.084664	0.001022	82.829	1699436
Neolithic West	Medieval Iran	Qazvin Bezoar	0.08261	0.001164	70.948	1364760
Neolithic West	Modern China	Qazvin Bezoar	0.083247	0.001099	75.769	1749026
Neolithic West	Modern Ireland	Qazvin Bezoar	0.142536	0.001563	91.172	1731442
Neolithic West	Modern Togo	Qazvin Bezoar	0.101061	0.001313	76.998	1762925
Neolithic West	Modern Morocco	Qazvin Bezoar	0.105899	0.00112	94.575	2199203
Neolithic West	Neolithic East	Qazvin Bezoar	0.071531	0.000928	77.116	1889360
Neolithic West	Neolithic Levant	Qazvin Bezoar	0.131618	0.002029	64.859	133583
Neolithic East	Bronze Age Britain	Qazvin Bezoar	0.072615	0.001071	67.819	1210942
Neolithic East	Bronze Age Caucasus	Qazvin Bezoar	0.099809	0.001173	85.076	1477378
Neolithic East	Bronze Age Iran	Qazvin Bezoar	0.100801	0.001194	84.392	1205238
Neolithic East	Bronze Age Israel	Qazvin Bezoar	0.0925	0.001126	82.18	1335558
Neolithic East	Bronze Age Turkey	Qazvin Bezoar	0.087199	0.001027	84.92	1838747
Neolithic East	Bronze Age Uzbekistan	Qazvin Bezoar	0.111537	0.001141	97.729	1422851
Neolithic East	Chalcolithic Caucasus	Qazvin Bezoar	0.100374	0.00137	73.243	361550
Neolithic East	Chalcolithic Iran	Qazvin Bezoar	0.106459	0.001292	82.423	1022234
Neolithic East	Chalcolithic Israel	Qazvin Bezoar	0.095309	0.003227	29.539	21831
Neolithic East	Chalcolithic Turkmenistan	Qazvin Bezoar	0.111141	0.001165	95.367	987110
Neolithic East	Modern France	Qazvin Bezoar	0.072781	0.000963	75.601	2021590
Neolithic East	Modern Iran	Qazvin Bezoar	0.092483	0.000976	94.799	2545830
Neolithic East	Iron Age/Medieval Caucasus	Qazvin Bezoar	0.095396	0.001038	91.915	1718328
Neolithic East	Medieval Iran	Qazvin Bezoar	0.102441	0.001221	83.918	1384760
Neolithic East	Modern China	Qazvin Bezoar	0.120792	0.001316	91.756	1749547
Neolithic East	Modern Ireland	Qazvin Bezoar	0.073212	0.001094	66.921	1809422
Neolithic East	Modern Togo	Qazvin Bezoar	0.077839	0.001144	68.07	1808559
Neolithic East	Modern Morocco	Qazvin Bezoar	0.078724	0.000943	83.519	2254383
Neolithic East	Neolithic West	Qazvin Bezoar	0.071531	0.000928	77.116	1889360
Neolithic East	Neolithic Levant	Qazvin Bezoar	0.06898	0.001388	49.709	138950
Modern Togo	Modern Morocco	Qazvin Bezoar	0.17266	0.001602	107.761	2122842
Modern Togo	Neolithic Levant	Qazvin Bezoar	0.135533	0.002453	55.256	125458
Modern Togo	Chalcolithic Israel	Qazvin Bezoar	0.130525	0.004597	28.396	19901
Modern Togo	Modern France	Qazvin Bezoar	0.116866	0.001455	80.294	1896901
Modern Togo	Bronze Age Israel	Qazvin Bezoar	0.114547	0.001668	68.66	1229411
Modern Togo	Modern Ireland	Qazvin Bezoar	0.112915	0.001655	68.227	1673346

Modern Togo	Bronze Age Turkey	Qazvin Bezoar	0.111328	0.001354	82.225	1723560
Modern Togo	Modern Iran	Qazvin Bezoar	0.108545	0.001258	86.304	2535448
Modern Togo	Bronze Age Britain	Qazvin Bezoar	0.10641	0.001687	63.073	1114884
Modern Togo	Neolithic West	Qazvin Bezoar	0.101061	0.001313	76.998	1762925
Modern Togo	Iron Age/Medieval Caucasus	Qazvin Bezoar	0.100912	0.001285	78.516	1607111
Modern Togo	Medieval Iran	Qazvin Bezoar	0.099316	0.00156	63.662	1282235
Modern Togo	Bronze Age Caucasus	Qazvin Bezoar	0.098769	0.001495	66.058	1369666
Modern Togo	Bronze Age Iran	Qazvin Bezoar	0.097137	0.00154	63.067	1116262
Modern Togo	Chalcolithic Caucasus	Qazvin Bezoar	0.09095	0.001713	53.098	333759
Modern Togo	Modern China	Qazvin Bezoar	0.087899	0.001402	62.715	1654433
Modern Togo	Bronze Age Uzbekistan	Qazvin Bezoar	0.087584	0.001249	70.12	1332134
Modern Togo	Chalcolithic Turkmenistan	Qazvin Bezoar	0.08508	0.001293	65.791	919285
Modern Togo	Chalcolithic Iran	Qazvin Bezoar	0.084111	0.001407	59.791	950036
Modern Togo	Neolithic East	Qazvin Bezoar	0.077839	0.001144	68.07	1808559
Modern Ireland	Bronze Age Britain	Qazvin Bezoar	0.171995	0.002186	78.666	1101246
Modern Ireland	Modern France	Qazvin Bezoar	0.166055	0.001699	97.725	1856259
Modern Ireland	Neolithic West	Qazvin Bezoar	0.142536	0.001563	91.172	1731442
Modern Ireland	Neolithic Levant	Qazvin Bezoar	0.12493	0.002399	52.072	125923
Modern Ireland	Modern Morocco	Qazvin Bezoar	0.118265	0.001345	87.948	2160536
Modern Ireland	Modern Togo	Qazvin Bezoar	0.112915	0.001655	68.227	1673346
Modern Ireland	Bronze Age Turkey	Qazvin Bezoar	0.098719	0.001315	75.049	1727577
Modern Ireland	Chalcolithic Israel	Qazvin Bezoar	0.096436	0.00423	22.796	19935
Modern Ireland	Modern Iran	Qazvin Bezoar	0.092115	0.001133	81.291	2552909
Modern Ireland	Bronze Age Israel	Qazvin Bezoar	0.090791	0.001391	65.282	1233715
Modern Ireland	Iron Age/Medieval Caucasus	Qazvin Bezoar	0.090223	0.001226	73.591	1609458
Modern Ireland	Bronze Age Caucasus	Qazvin Bezoar	0.086591	0.001326	65.306	1370412
Modern Ireland	Bronze Age Iran	Qazvin Bezoar	0.0858	0.001332	64.396	1117573
Modern Ireland	Medieval Iran	Qazvin Bezoar	0.085683	0.001412	60.678	1283031
Modern Ireland	Chalcolithic Caucasus	Qazvin Bezoar	0.083885	0.001552	54.048	334064
Modern Ireland	Modern China	Qazvin Bezoar	0.080351	0.001382	58.16	1654529
Modern Ireland	Bronze Age Uzbekistan	Qazvin Bezoar	0.07987	0.001268	62.972	1332297
Modern Ireland	Chalcolithic Iran	Qazvin Bezoar	0.079403	0.001368	58.04	949342
Modern Ireland	Chalcolithic Turkmenistan	Qazvin Bezoar	0.079189	0.001247	63.517	918969
Modern Ireland	Neolithic East	Qazvin Bezoar	0.073212	0.001094	66.921	1809422

Appendix Table 4.8 - Significant results of regression between shared drift f_3 with Neolithic populations. Linear regression formula was shared drift with population Y regressed on shared drift with population X. Studentized residuals were computed. Outlier threshold was taken to be 2.12, based on the t-distribution with 16 degrees of freedom, alpha of 0.05, two-tailed test.

Y_X	Significant Population	Studentized Residual
Levant_West	Chalcolithic Levant	3.18
Levant_East	Chalcolithic Levant	4.82
East_West	Modern Chinese	3.40

Appendix Table 4.9 - f_4 ratio estimates for individual genomes. A |Z| score > 3 is taken as significant.

f_4 Numerator	f_4 Denominator	Ratio (α)	Error	Z
Proportion of Ancient Turkish Wild ancestry in western Neolithics				
Yak, Ancient Turkish Wild; Blagotin1, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.498576	0.022625	22.036
Yak, Ancient Turkish Wild; Blagotin2, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.486352	0.023568	20.636
Yak, Ancient Turkish Wild; Blagotin3, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.505037	0.018259	27.659
Yak, Ancient Turkish Wild; Blagotin16, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.502074	0.026098	19.238
Yak, Ancient Turkish Wild; Kov58, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.604866	0.082448	7.336
Yak, Ancient Turkish Wild; AP45, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.524484	0.140041	3.745
Yak, Ancient Turkish Wild; AP49, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.484271	0.191292	2.532
Yak, Ancient Turkish Wild; Neolithic West, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.500783	0.016381	30.571
Yak, Ancient Turkish Wild; Potterne1, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.461403	0.026913	17.145
Proportion of Ancient Turkish Wild ancestry in Levantine Neolithics				
Yak, Ancient Turkish Wild; Ainghazal1, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.65208	0.109275	5.967
Yak, Ancient Turkish Wild; Ainghazal2, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.503388	0.079223	6.354
Yak, Ancient Turkish Wild; Ainghazal4, Neolithic East	Yak, Ancient Turkish Wild; Direkli1-2, Neolithic East	0.558153	0.313416	1.781
Proportion of Ancient Turkish Wild ancestry in Modern European				
Yak, Neolithic East; french_modern2, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.410438	0.018732	21.911
Yak, Neolithic East; french_modern3, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.424716	0.018093	23.475
Yak, Neolithic East; french_modern1, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.439957	0.019011	23.142
Yak, Neolithic East; french_modern4, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.438974	0.018352	23.92

Yak, Neolithic East; IOG, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.416987	0.020232	20.61
Yak, Neolithic East; Modern Europe, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	0.426214	0.014847	28.706
Proportion of eastern Neolithic ancestry in Bronze Age Turkey				
Yak, Neolithic East; Acem1, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	-0.122054	0.04436	-2.751
Yak, Neolithic East; Acem2, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	-0.19306	0.035404	-5.453
Yak, Neolithic East; Bronze Age Turkey, Neolithic West	Yak, Neolithic East; Semnan3, Neolithic West	-0.166694	0.033185	-5.023
Yak, Neolithic West; Acem1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.296097	0.019933	14.855
Yak, Neolithic West; Acem2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.261313	0.014636	17.854
Yak, Neolithic West; Bronze Age Turkey, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.2729	0.0134	20.365
Proportion of eastern Neolithic ancestry in post-Neolithic Levant				
Yak, Neolithic East; Yoqneam2, Neolithic Levant	Yak, Neolithic East; Semnan3, Neolithic Levant	-0.040998	0.137178	-0.299
Yak, Neolithic East; Safi2, Neolithic Levant	Yak, Neolithic East; Semnan3, Neolithic Levant	0.255274	0.294152	0.868
Yak, Neolithic East; Bronze Age Israel, Neolithic Levant	Yak, Neolithic East; Semnan3, Neolithic Levant	-0.076243	0.139717	-0.546
Yak, Neolithic East; Gilat8, Neolithic Levant	Yak, Neolithic East; Semnan3, Neolithic Levant	0.404999	0.639156	0.634
Proportion of western Neolithic ancestry in Levantine post-Neolithic				
Yak, Neolithic West; Yoqneam2, Neolithic Levant	Yak, Neolithic West; Blagotin3, Neolithic Levant	-3.637875	2.984348	-1.219
Yak, Neolithic West; Safi2, Neolithic Levant	Yak, Neolithic West; Blagotin3, Neolithic Levant	5.793007	7.166927	0.808
Yak, Neolithic West; Bronze Age Israel, Neolithic Levant	Yak, Neolithic West; Blagotin3, Neolithic Levant	-3.72947	2.96882	-1.256
Yak, Neolithic West; Gilat8, Neolithic Levant	Yak, Neolithic West; Blagotin3, Neolithic Levant	0.79763	0.81561	0.978

Proportion of western Neolithic ancestry in eastern post-Neolithics				
Yak, Neolithic West; Fars1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.143167	0.124915	-1.146
Yak, Neolithic West; Fars4, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.114862	0.022775	-5.043
Yak, Neolithic West; Chalcolithic Iran, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.112181	0.022049	-5.088
Yak, Neolithic West; Monjukli1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.135165	0.042151	-3.207
Yak, Neolithic West; Monjukli2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.069055	0.040124	-1.721
Yak, Neolithic West; Monjukli4, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.110281	0.028707	-3.842
Yak, Neolithic West; Monjukli6, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.219998	0.106655	-2.063
Yak, Neolithic West; Chal. Turkmenistan, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.11759	0.022542	-5.217
Yak, Neolithic West; Bulak1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.197253	0.016481	11.969
Yak, Neolithic West; Bulak2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.096305	0.022343	-4.31
Yak, Neolithic West; Bulak5, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.163174	0.037363	-4.367
Yak, Neo. West; Bronze Age Uzbekistan, Neo. East	Yak, Neolithic West; Blagotin3, Neolithic East	0.081125	0.015157	5.352
Yak, Neolithic West; Chalow1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	-0.10174	0.083599	-1.217
Yak, Neolithic West; Qazvin1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.023124	0.021358	1.083
Yak, Neolithic West; Bronze Age Iran, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.018827	0.021046	0.895
Yak, Neolithic West; Kohneh2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.024231	0.102278	0.237
Yak, Neolithic West; Azer3-5, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.046622	0.020094	2.32
Yak, Neolithic West; Bronze Age Georgia, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.055295	0.019323	2.862
Yak, Neolithic West; Kazbeg1, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.149822	0.020984	7.14
Yak, Neolithic West; Azer4, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.039664	0.022583	1.756

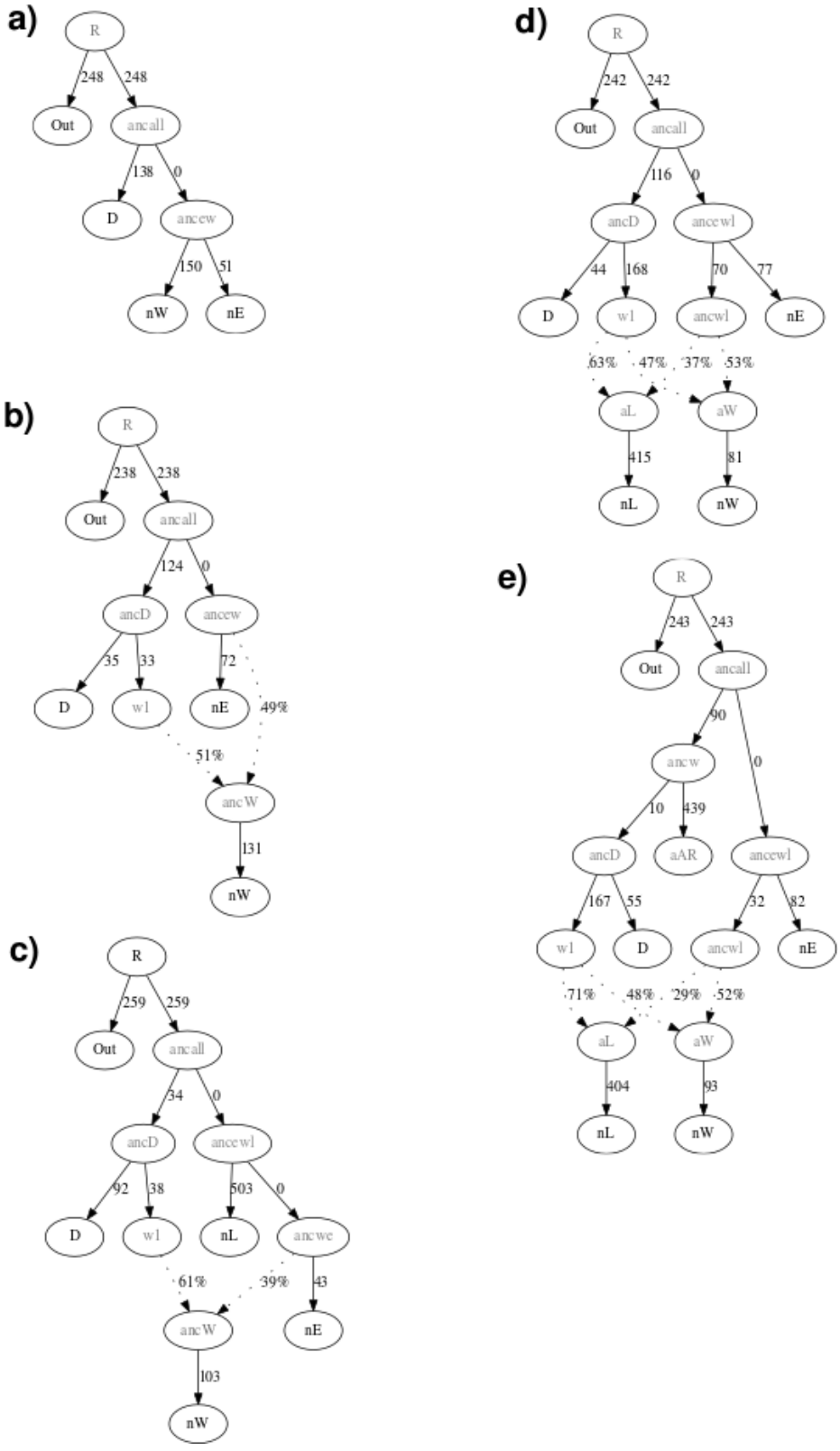
Yak, Neolithic West; Tac3, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.155776	0.051089	3.049
Yak, Neolithic West; Geor2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.069057	0.02283	3.025
Yak, Neo. West; Iron Age/Medieval Caucasus, Neo. East	Yak, Neolithic West; Blagotin3, Neolithic East	0.095781	0.015674	6.111
Yak, Neolithic West; Modern Iranian, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.11212	0.009598	11.682
Yak, Neolithic West; Azer6, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.060637	0.039426	1.538
Yak, Neolithic West; Darre2, Neolithic East	Yak, Neolithic West; Blagotin3, Neolithic East	0.00925	0.021208	0.436

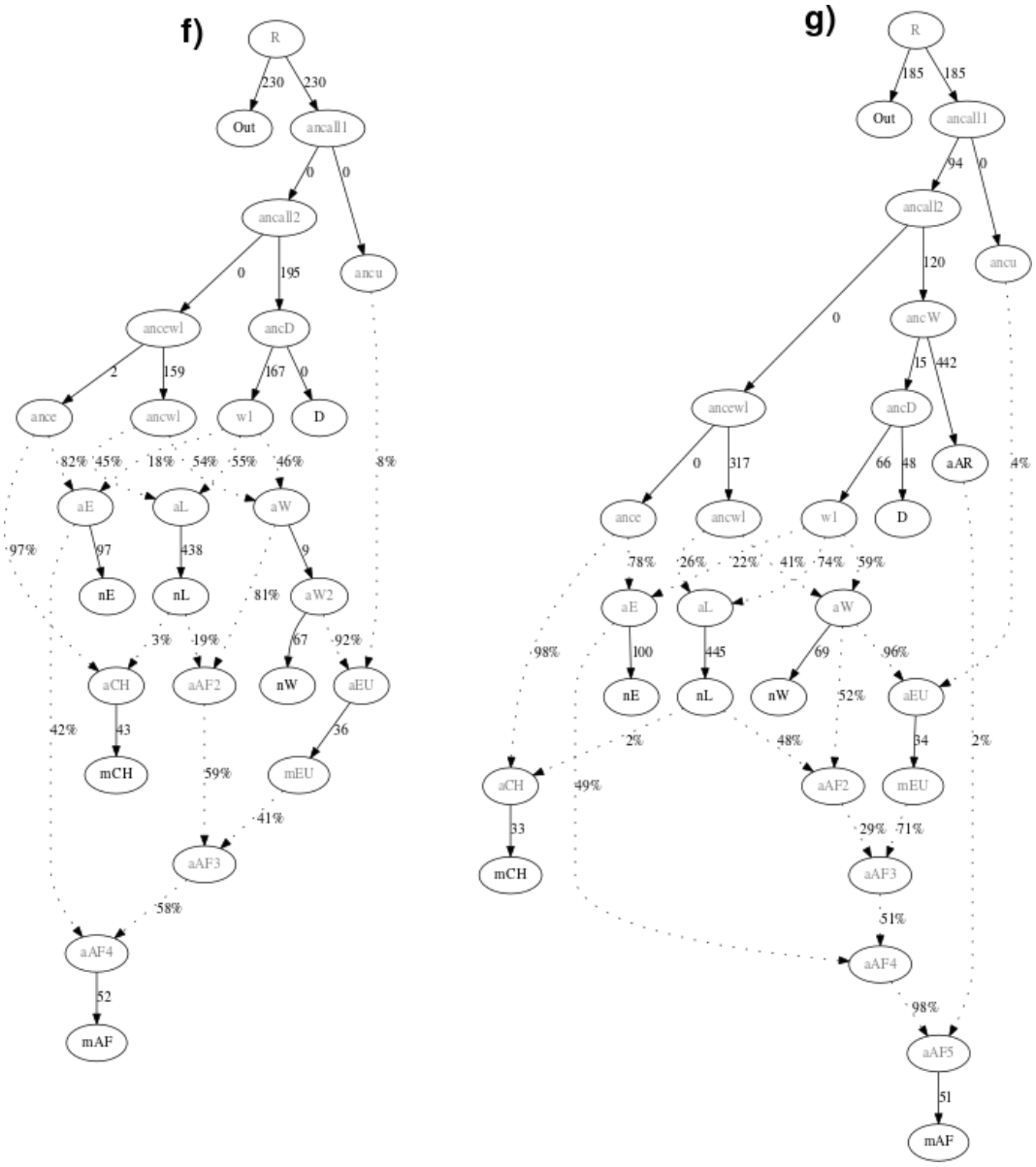
Appendix Table 4.10 - Hudson's pairwise Fst based on whole genome sequences.
E=East, W=West, L=Levant, 3=Neolithic, 2=Chalcolithic and Bronze Age, 1=Iron Age/Medieval.

	1E	2E	2W	3E	3W
1E	-	0	0.01	0.07	0.12
2E	0	-	0.02	0.08	0.13
2W	0.01	0.02	-	0.09	0.1
3E	0.07	0.08	0.09	-	0.17
3W	0.12	0.13	0.1	0.17	-

Appendix Table 4.11 - Select Group *D* statistics reproduced from (Daly et al. 2018).

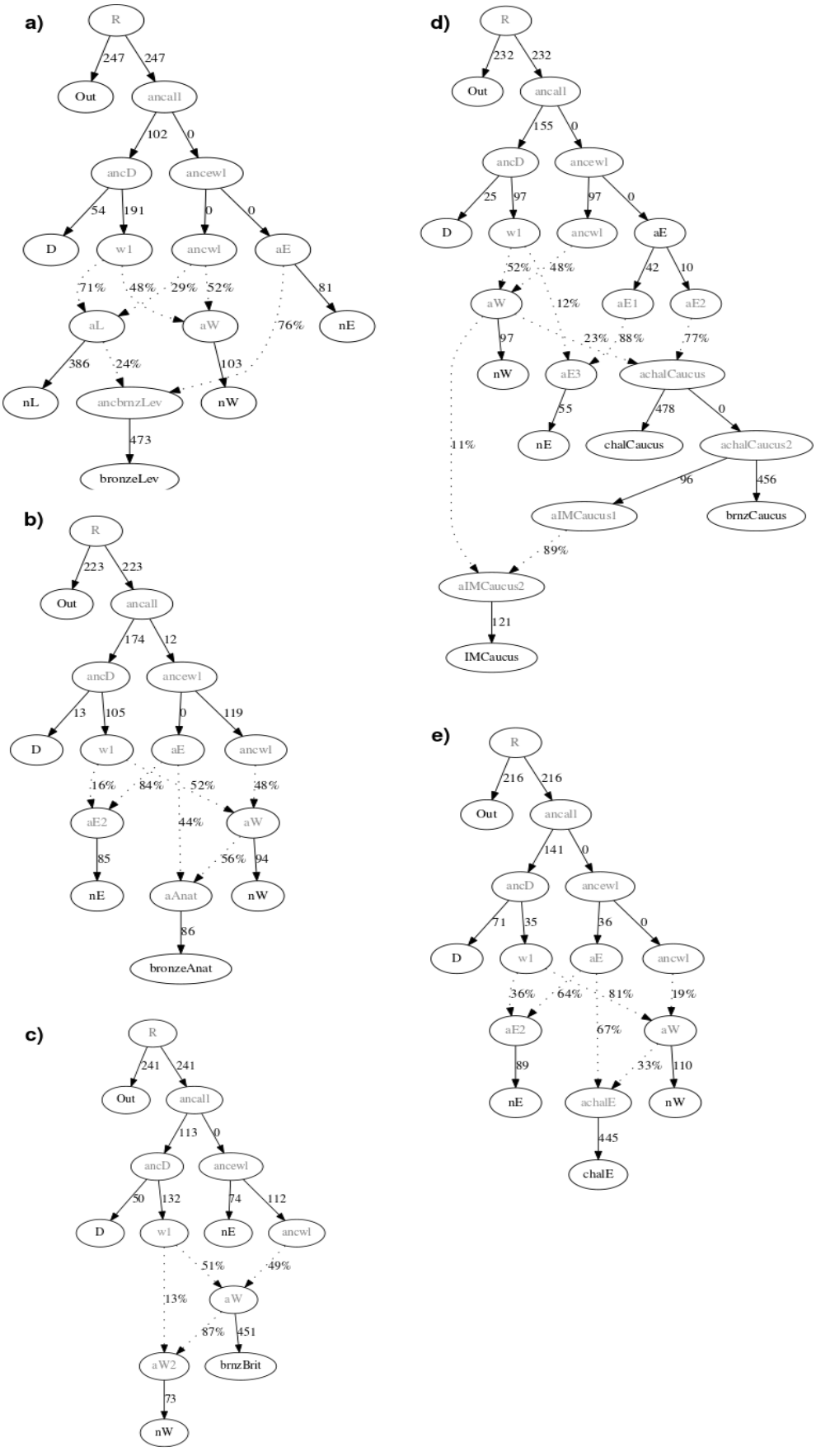
H1	H2	H3	*D*	Z	nABBA	nBABA
Neolithic West	Modern Europe	Neolithic East	-0.035	-16.4	151956.0235	163113.9868
Neolithic East	Modern China	Neolithic Levant	0.01	2.2	15343.5787	15031.8242
Neolithic East	Neolithic West	Qazvin Wild	-0.001	-0.8	173994.4549	174484.4962

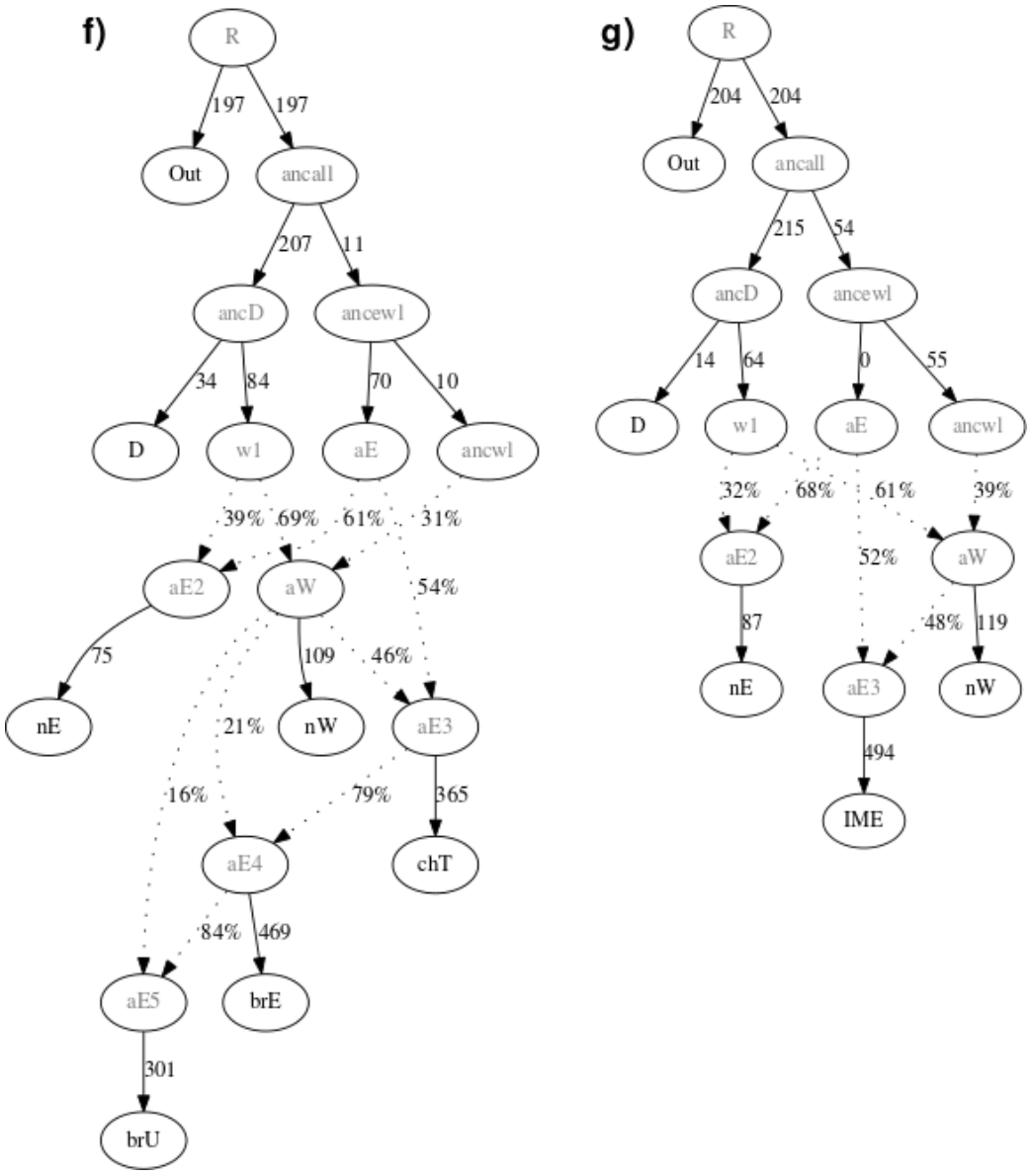

a)
R
248
248
Out
ancall
138
0
D
ancew
150
51
nW
nE
b)
R
238
238
Out
ancall
124
0
ancD
ancew
35
33
72
D
w1
nE
49%
51%
ancW
131
nW
c)
R
259
259
Out
ancall
34
0
ancD
ancewl
92
38
503
0
D
w1
nL
ancwe
61%
39%
43
ancW
nE
103
nW
d)
R
242
242
Out
ancall
116
0
ancD
ancewl
44
168
70
77
D
w1
ancwl
nE
63%
47%
37%
53%
aL
aW
415
81
nL
nW
e)
R
243
243
Out
ancall
90
0
ancw
10
439
ancD
aAR
ancewl
167
55
32
82
w1
D
ancwl
nE
71%
48%
29%
52%
aL
aW
404
93
nL
nW



Appendix Figure 4.1 - Admixture graph models for ancient and modern domestic goats. a) Base graph used, which was rejected. b) A modified version of the previous graph, allowing admixture from a wild population related to Turkish Ancient Wilds into the ancestors of western Neolithics, which was not rejected. c) Neolithic Levant modelled as the outgroup of eastern and western Neolithic goat, but the graph is rejected. d) A model in which Neolithic Levant and Neolithic West share ancestry, and both subsequently admix with a wild population, is not rejected. e) Addition of the Armenian Ancient Wild genome to the root of the wild clade, which is not rejected. f) Model d, with the addition of modern African, Chinese, and European genomes, fits the data. Modern populations are modelled as a mixture of ancestries. g) The previous model with the addition of the ancient Armenian wild goat. This model results in three outlier f_4 statistics (ranging 3-3.3). Intermediate, theoretical

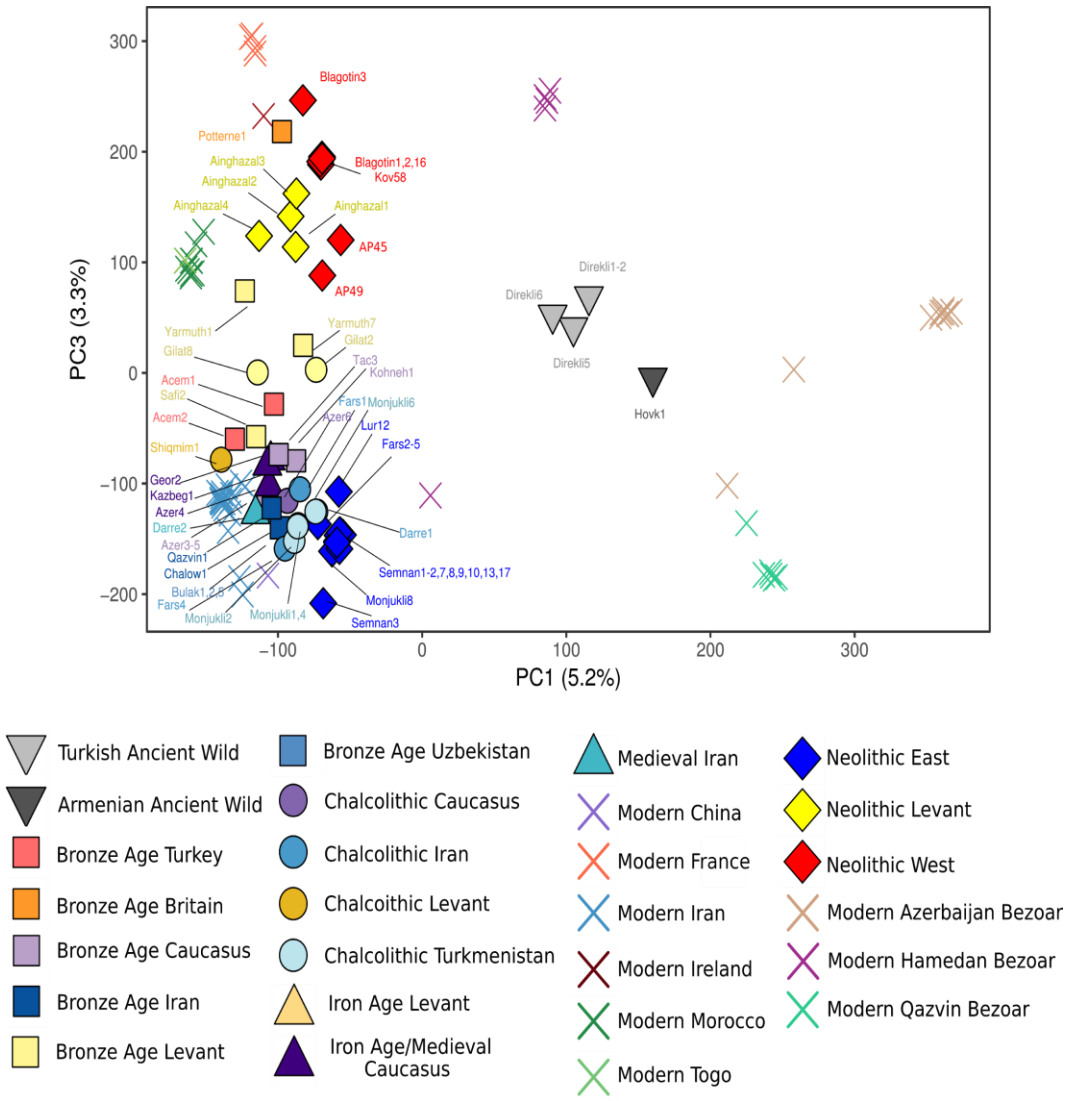
populations are denoted in grey. Edge drift values = Fst x 1000. nW=Neolithic West, nE=Neolithic East, nL=Neolithic Levant, D=Turkish Ancient Wild (Direkli), aAR=Armenian Ancient Wild (Hovk1), mEU=Modern Europe, mCH=Modern China, mAF=Modern Africa.


a)
R
247
247
Out
ancall
102
0
ancD
ancewl
54
191
0
0
D
w1
ancwl
aE
71%
48%
29%
52%
81
aL
aW
76%
nE
386
24%
103
nL
ancbrnzLev
nW
473
bronzeLev
b)
R
223
223
Out
ancall
174
12
ancD
ancewl
13
105
0
119
D
w1
aE
ancwl
16%
84%
52%
48%
aE2
44%
aW
85
56%
94
nE
aAnat
nW
86
bronzeAnat
c)
R
241
241
Out
ancall
113
0
ancD
ancewl
50
132
74
112
D
w1
nE
ancwl
51%
49%
13%
aW
87%
451
aW2
brnzBrit
73
nW
d)
R
232
232
Out
ancall
155
0
ancD
ancewl
25
97
97
0
D
w1
ancwl
aE
52%
48%
42
10
aW
12%
aE1
aE2
97
23%
88%
77%
nW
aE3
achalCaucus
55
478
0
11%
nE
chalCaucus
achalCaucus2
96
456
aIMCaucus1
brnzCaucus
89%
aIMCaucus2
121
IMCaucus
e)
R
216
216
Out
ancall
141
0
ancD
ancewl
71
35
36
0
D
w1
aE
ancwl
36%
64%
81%
19%
aE2
67%
aW
89
33%
110
nE
achalE
nW
445
chalE

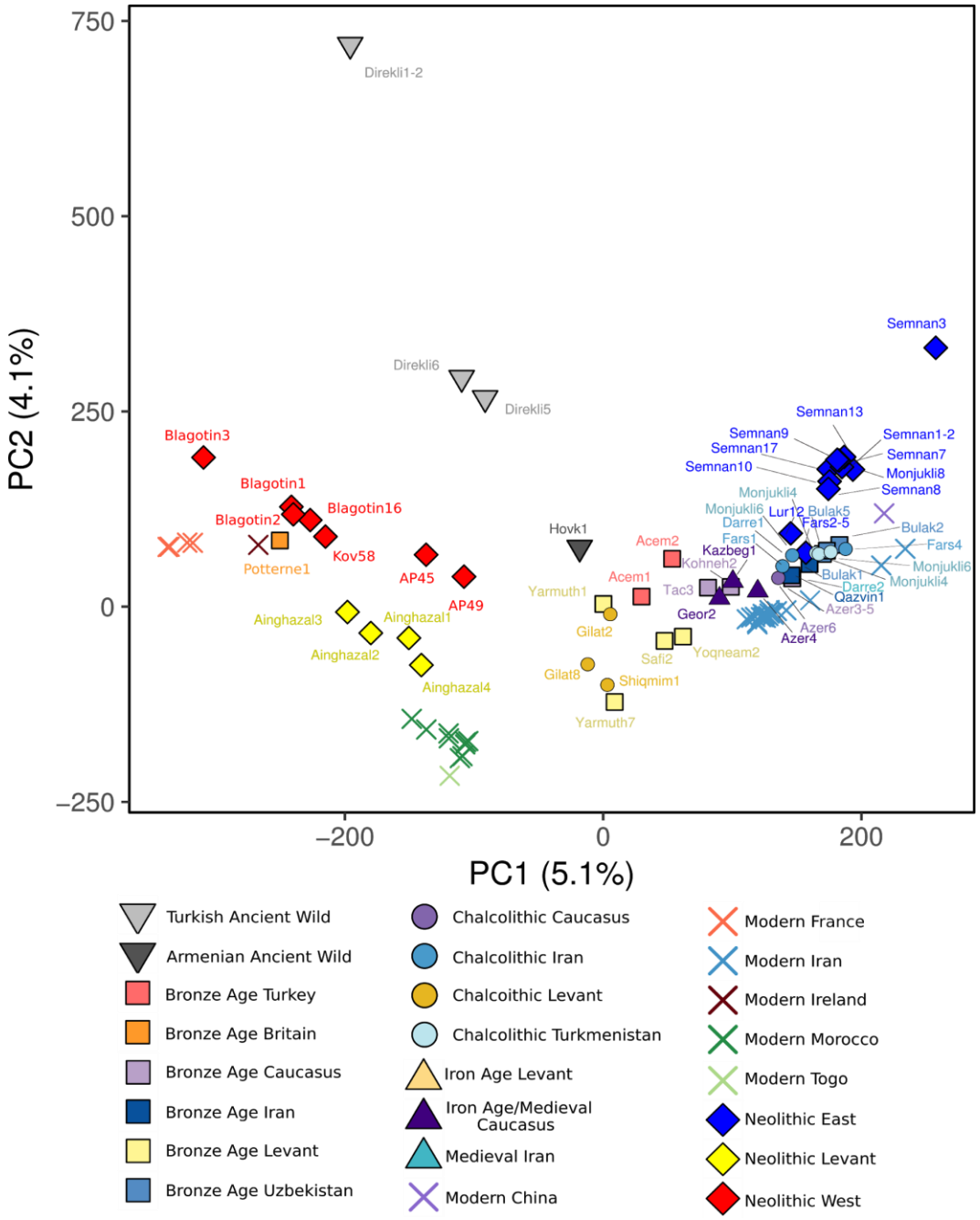


Appendix Figure 4.2 - Admixture graph models for Post-Neolithic ancient domestic goats. a) Bronze Age Levant ancestry modelled as predominantly deriving from an Eastern Neolithic-like population, but with some contribution from a Neolithic Levant-related population. b) Bronze Age Turkey modelled as roughly equal mixtures of populations related to both western and eastern Neolithics. c) Bronze Age Britain is modelled as a sister group to Neolithic West, which requires an additional wild input to fit. d), e), f), g) Post-Neolithic eastern populations relate to Neolithic East but require an input from a Neolithic West-like population, as well as additional wild ancestry in Neolithic East. Intermediate, theoretical populations are denoted in grey. Edge drift values = Fst x 1000. nW=Neolithic West, nE=Neolithic East, nL=Neolithic Levant, D=Turkish Ancient Wild (Direkli),

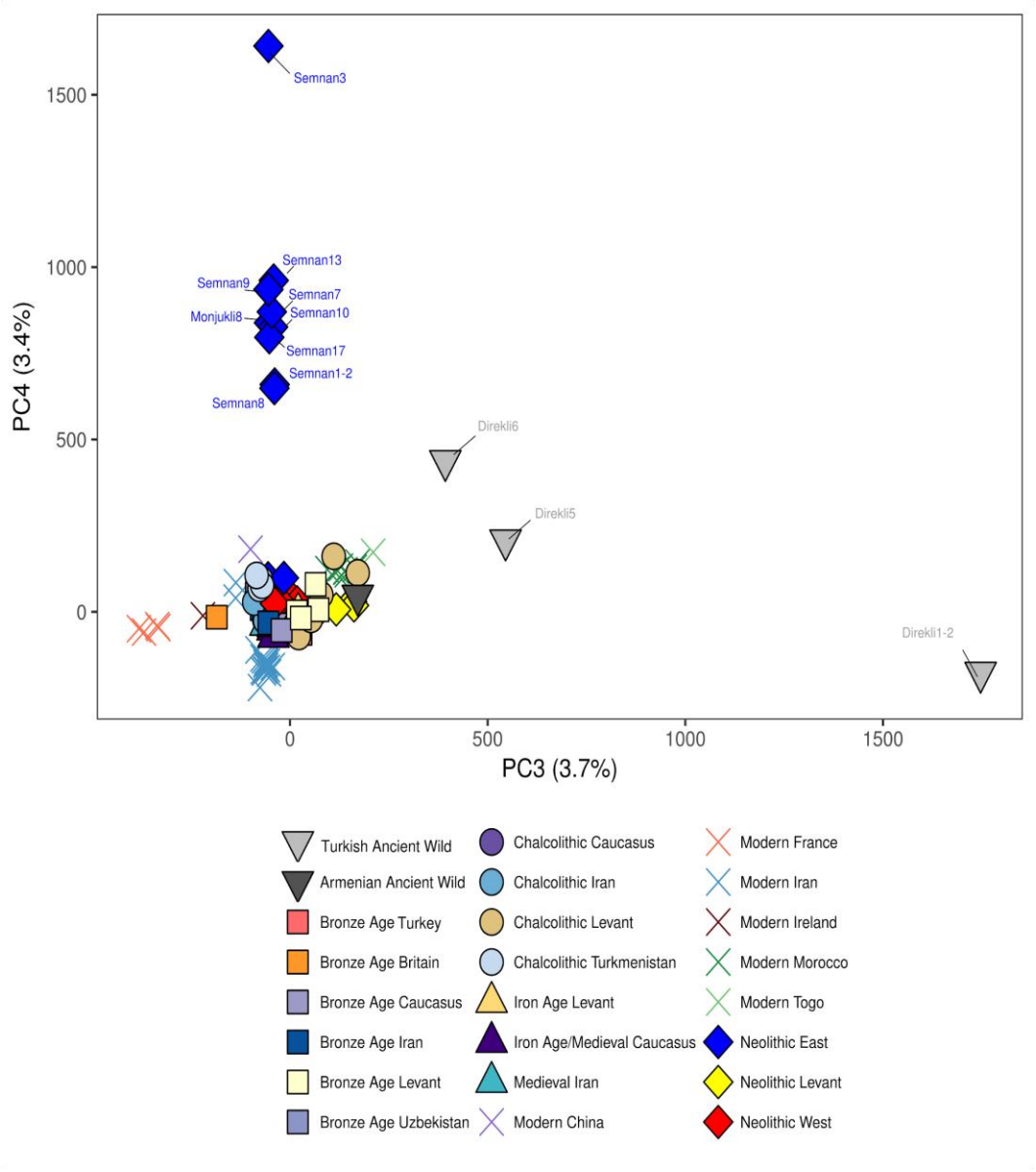
bronzeLev=Bronze Age Levant, bronzeAnat=Bronze Age Turkey, brnzBrit=Bronze Age Britain, chalCaucus=Chalcolithic Caucacus, brnzCaucus=Bronze Age Caucasus, IMCaucus=Iron Age/Medieval Caucasus, chalE=Chalcolithic East, brE=Bronze Age Iran, chT=Chalcolithic Turkmenistan, brU=Bronze Age Uzbekistan, IME=Iron Age/Medieval Iran.



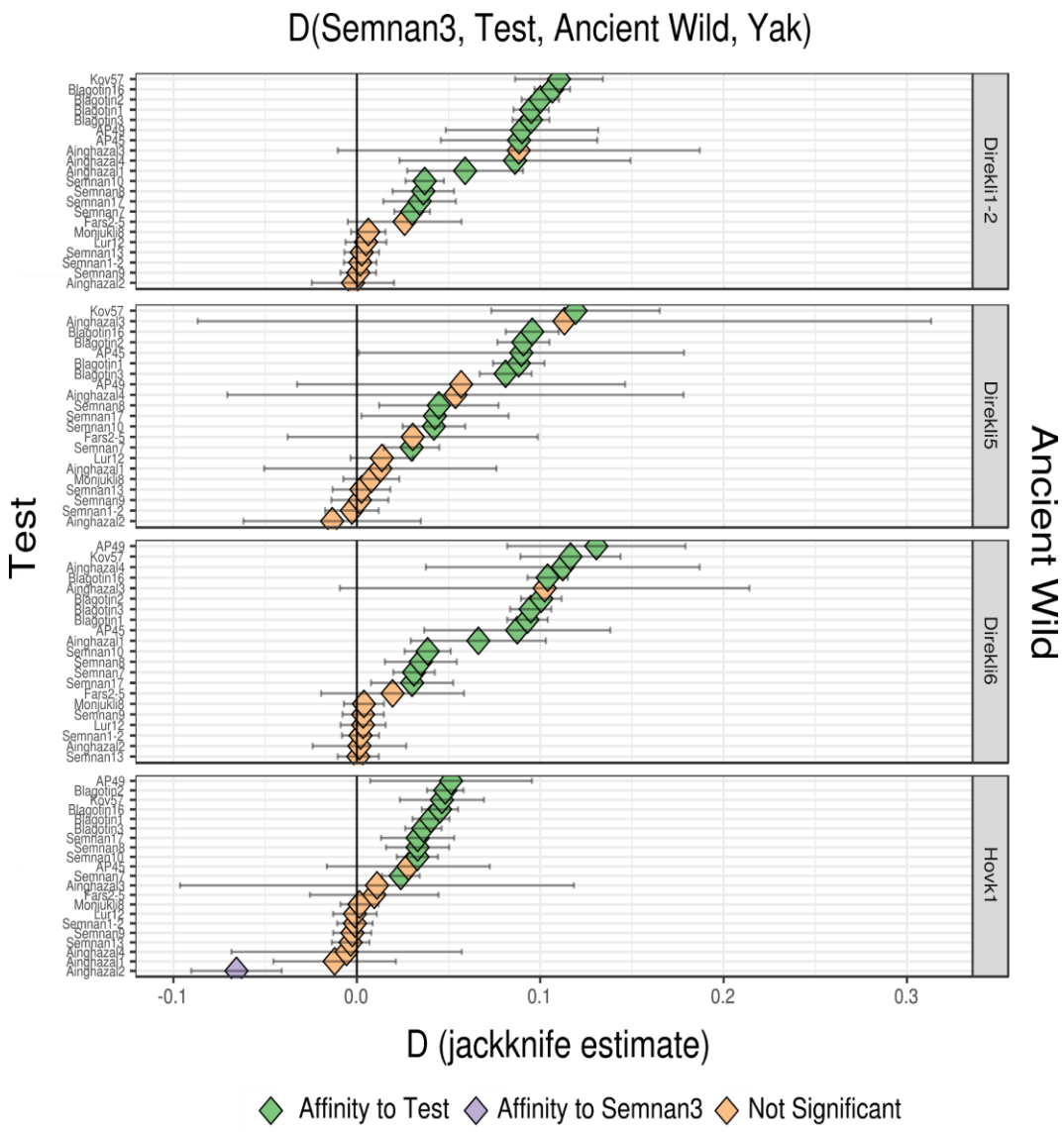
Appendix Figure 4.3 - LASER projection PCA of all ancient samples and modern goat/bezoar, plotting PC1 vs PC3, using LD-pruned sites. Values in parenthesis represent the percentage of variance explained by a given PC, as estimated by LASER.



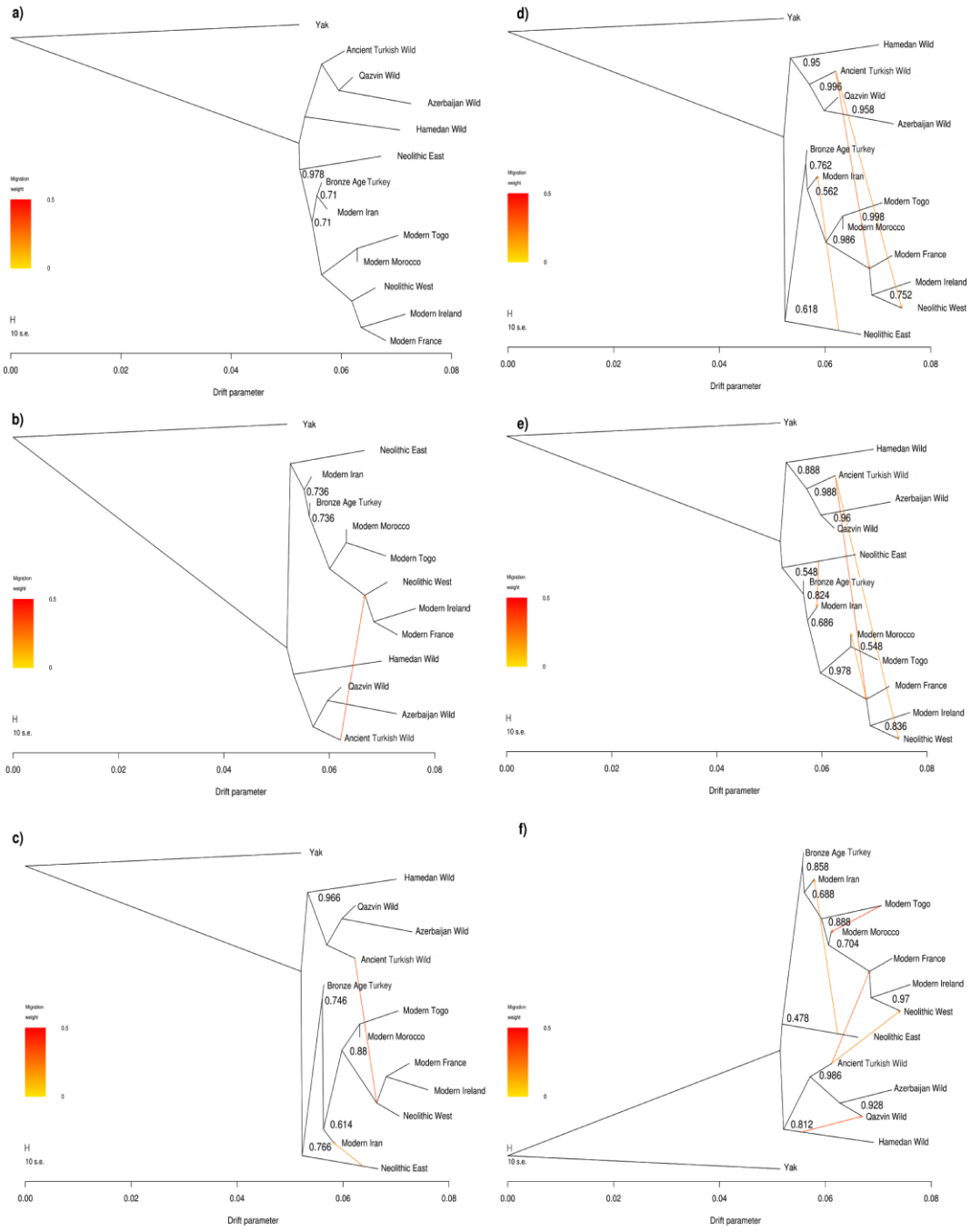
Appendix Figure 4.4 - LASER projection PCA of all ancient samples and modern domestic goat, plotting PC1 vs PC2, using LD-pruned sites. Values in parenthesis represent the percentage of variance explained by a given PC, as estimated by LASER.



Appendix Figure 4.5 - LASER projection PCA of all ancient samples and modern domestic goat, plotting PC3 vs PC4, using LD-pruned sites. Values in parenthesis represent the percentage of variance explained by a given PC, as estimated by LASER.



Appendix Figure 4.6 - *D* statistic test for relative affinity (derived allele sharing) of a Test genome to an Ancient Wild individual, relative to an Ancient Eastern (Semnan3). Colour indicates relative affinity of the Ancient Wild to either Semnan3 or the Test, with non-significance ($|Z| < 3$) indicated by the colour orange. Tests genomes are ordered by their *D* value for each Ancient Wild individually.



Appendix Figure 4.7 - Treemix analysis of high coverage (>8×) samples. Bootstrap support (500 iterations) for branches is displayed with the bootstrap score was less than 1. Migration edges were varied from 0 to 5, shown in a) no migration edge, b) one migration edge, c) two migration edges, d) three migration edges, e) four migration edges, f) five migration edges.

Appendix Text 4.1 - qpGraph procedure.

The following is a reproduction of the "Admixture Graph Construction" Supplementary Information section of (Daly et al. 2018), updated to refer to the figures and tables as presented in this thesis.

Based on Treemix (Appendix Figure 4.7) and IBS (Figure 4.7) results, Turkish Ancient Wild was placed as the outgroup to Neolithic East and Neolithic West (Appendix Figure 4.1a), but this model was rejected with 17 f_4 outliers with $|Z| \geq 3$. As Treemix and *D* statistics (Figure 4.9) suggest that ancestors of Neolithic West to have admixed with Turkish bezoar, we modelled Neolithic West as being a mixture of a population related to Turkish Ancient Wild and a population leading to Neolithic Eastern goat (Appendix Figure 4.1b). This model fits the data with no f_4 outliers.

We then added the Neolithic Levant population, comprising of three individuals of low coverage (average ~0.03X). Based on IBS results (Figure 4.7), we excluded Neolithic Levant as being an outgroup to all other populations modelled. We found that modelling Neolithic Levant as an outgroup to domestics (Neolithic East and West) was rejected with 42 f_4 outliers (Appendix Figure 4.1c). Based on f_3 outgroup (Figure 4.19) and ancestry estimation (Figure 4.18), Neolithic Levant and Neolithic West show relatively high affinity. We investigated topologies consistent with this and found that a graph in which Neolithic Levant and Neolithic West were composed of separate mixtures between a Turkish-like population and a population sister to Neolithic East fits the data (Appendix Figure 4.1d). This topology was supported by IBS tree building (Figure 4.17), and the affinity of Neolithic Levant and Neolithic West in principal component analysis (Figure 4.6).

We then introduced a single genome, Hovk1, at least 47,000 years old (Table 4.1) and representing Armenian Ancient Wilds, to the graph. When placed as the root of sampled wild and domestic goat, the graph is rejected with 57 f_4 outliers. Modelling ancient Armenians as the sister clade of Turkish Ancient Wild and related populations results in a graph that with no f_4 outliers, which is displayed in Appendix Figure 4.1e. This graph topology is in line with PCA analyses (Figure 4.6) and IBS (Figure 4.7) which suggest an affinity of Armenian Ancient Wild and Turkish Ancient Wild.

To investigate how these Neolithic populations contributed to modern goat populations, we sequentially added three modern populations (Chinese/China, Europe, and Africa) to the graph. We first removed Armenian Ancient Wild due to it being represented by a single pseudo-diploid sample, starting instead with the model depicted in Appendix Figure 4.1d. Modern Chinese required admixture between a population ancestral to Neolithic East, and Neolithic Levant, based on f_4 outliers such as:

D(Neolithic West, Neolithic Levant; Neolithic East, Modern Chinese), Z=4.1
D(Neolithic East, Modern Chinese; Ancient Turkish, Neolithic Levant), Z=3

This affinity between Modern Chinese and Neolithic Levant was estimated as a contribution of ~2% from Neolithic Levant to the ancestor of Chinese goat. Additionally, fitting Modern Chinese required Neolithic East to be modelled as containing an additional source of wild ancestry. We note that the *D* statistic Neolithic Levant(Modern Chinese, Neolithic East) is not significant (Z=2.2) (Appendix Table 4.11), and that Neolithic Levant is represented by a small number of low coverage samples. Therefore, we cannot exclude that there might be additional unsampled populations which better represent ancestral populations which contributed to the genomes of modern Chinese goat, or that samples with greater sequencing depth would fit a different model of Chinese goat ancestry.

Modelling Modern Europe as descending from the same ancestral population to Neolithic West resulted in three f_4 outliers which did not clearly indicate a single unmodelled event. We hypothesized that the ancestors of modern European domestic goat may have undergone admixture with a European wild caprid population, and introduced an outgroup population to the model which mixed with the ancestors of modern Europe. The resulting model fit the data with no f_4 outliers. This admixture event is supported by the *D* statistic Neolithic East (Neolithic West, Modern Europe), Z=16.4 (Appendix Table 4.3 and 4.11), which can be interpreted as an increase of ancestral alleles in Modern Europe. Alternatively, there may be unsampled structure in ancient European goat, despite the high affinity of Neolithic West with modern Europe (Table 4.6).

We then added Modern African to the model, which did not fit in a clade with either Neolithic Levant or Neolithic West despite IBS (Table 4.7), Treemix (Appendix Figure 4.7), and outgroup f_3 values (Figure 4.19) suggesting an affinity of Modern Africa with these populations. Modelling modern Africans as a three-way mixture between modern Europeans,

Neolithic Levant, and a population basal to Neolithic West and Modern Europeans resulted in a model with two f_4 outliers. The larger of these outliers, (Neolithic West, Neolithic East; Neolithic Levant, Modern Africa), Z=3.4 suggested unmodelled shared drift between Neolithic West and Neolithic Levant or Neolithic East and Modern Africa. We then modelled an additional mixture event from a population ancestral to Neolithic East, to the ancestors of modern African goat, resulting in no f_4 outliers (Appendix Figure 4.1f). We note that f_3 outgroup values (Figure 4.19) suggests a greater affinity of Neolithic East with Modern Africa than with Modern Europe, as does the *D* statistic Neolithic Iran(Modern Africa, Modern Europe), Z=11.3 (Appendix Table 4.11).

Finally, we attempted to fit the ancient Armenian sample Hovk1 into the graph with these modern populations fitted. Modelling Hovk1 as a sister branch to Turkish Ancient Wild was rejected with eight f_4 outliers, despite fitting in the case of Fig S14e. Several of these outlier statistics suggested unmodelled affinity between Hovk1 and Modern Africa, for example (Neolithic East, Armenian Ancient Wild; Modern Europe, Modern Africa), Z=4.1. Adding an additional admixture event from Hovk1 to the ancestors of Modern Africa resulted in three outlier Z values, all within the range of 3-3.2, and suggested a minor (2%) contribution to Modern Africa (Appendix Figure 4.1g). Additional admixture events or alterations to graph increased the number of *f* statistic outliers. Given that Armenian Ancient Wild was represented by a single pseudo-diploid individual, and the uncertainty of modelling modern populations with ancient samples unevenly distributed across time and space, we did not further search the graph space to fit Armenian Ancient Wild.

We then investigated if other ancient goat populations could be modelled using Neolithic and Pre-Neolithic samples. Due to the quality and number of genomes for many time periods and regions, a skeleton graph of Neolithic East, Neolithic West and Turkish Ancient Wild was used to fit single populations.

In fitting Bronze Age Levant, Neolithic Levant was included in order to investigate how local Neolithic ancestry contributed to later populations. Due to low coverage of Chalcolithic Levant samples, this population was not modelled. Bronze Age Levant could not be modelled as a sister clade to Neolithic Levant (55 f_4 outliers), with highest f_4 outlier (Neolithic West, Neolithic East; Neolithic Levant, Bronze Age Levant) (Z=11.95) implying unmodelled ancestry between Neolithic East and Bronze Age Levant. Similarly, Bronze Age Levant could not be modelled as a sister population to Neolithic East (24 f_4 outliers). Modelling

Bronze Age Levant as a mixture of Neolithic Levant and Neolithic East-like ancestry results in a single outlier, with 24% and 76% ancestry contributions respectively (Appendix Figure 4.2a). The remaining f_4 outliers (Neolithic West, Neolithic Levant; Turkish Ancient Wild, Bronze Age Levant), suggests additional affinity between the Levantine populations that is not explained by this model; modelling an additional contribution from a Turkish-like population to the ancestors of Neolithic West did not resolve this outlier.

Bronze Age Turkish could not be fit as a sister group to either Neolithic East or Neolithic West (38 and 54 outliers respectively), or as an outgroup to both (59 outliers). Fitting Bronze Age Turkish as a mixture of Neolithic East and West resulted in 15 outliers, which strongly suggested an additional wild contribution to Neolithic East by the f_2 (Neolithic East, Turkish Ancient Wild) producing a Z score of 7. Allowing this additional Turkish Wild-like ancestry resulted in the model fitting the data with no outliers (Appendix Figure 4.2b), which describes Bronze Age Turkey as approximately even mixes of Neolithic East and West-like ancestry (44% and 56%), with a 16% Turkish Wild-like contribution to Neolithic East.

Fitting Bronze Age Britain as a sister group to Neolithic West resulted in two f_4 outliers, both suggestive of additional unmodelled ancestry present in Neolithic West but not Bronze Age Britain. Including this additional ancestry in Neolithic West results in no outlier statistics (Appendix Figure 4.2c). Interestingly, this result held when Neolithic West was represented only by high coverage individuals from Blagotin-Poljna, Serbia, suggesting that these early European goats have a population history that is distinct from the ancestors of Bronze Age British goat. This model was consistent with Treemix (Appendix 4.7), which suggested additional Turkish Ancient Wild ancestry in Neolithic West that was absent in modern European (French and Irish) goat.

To fit populations from the Caucasus region (Georgia and Iranian Azerbaijan), Chalcolithic, Bronze Age, and Iron Age/Medieval populations were sequentially added to the skeleton graph. Fitting all populations with no outliers (Appendix Figure 4.2d) suggested the Caucasus populations share the majority of ancestry with Neolithic East, with some admixture from Neolithic Western-like source that increases over time (23% for Chalcolithic and Bronze Age populations, and an additional influx of 11% to the ancestors of Iron Age/Medieval populations). Similar to previous models, this required a small (12%) wild input to the ancestors of Neolithic East.

To model the ancestry of Iranian, Turkmen, and Uzbek goat, Chalcolithic Iran was first fit to the skeleton graph as a sister group to Neolithic East, which was rejected with 21 f_4 outliers. We added additional admixture from an Ancient Turkish-like population to Neolithic East, and from a Neolithic West-like population to Chalcolithic Iran, which results in no outliers (Appendix Figure 4.2e). Notably, this model suggested a substantial contribution from the West to Chalcolithic Iran (33%) that is not detected in other analyses (NGSadmix, *D* statistics). To fit additional post-Neolithic eastern populations, Chalcolithic Iran was removed due to low SNP count (Appendix Table 4.4). The resulting graph, which fit with no outliers (Appendix Figure 4.2f), models these post-Neolithic populations (Chalcolithic Turkmenistan, Bronze Age Iran, Bronze Age Uzbekistan) as containing substantial Western-derived ancestry which increases through time. This is only partially consistent with other analyses; though a change in ancestry is observed in the PCA (Figure 4.6), *D* statistics (Figure 4.13) and NGSadmix (Figure 4.18) detect a similar signal only in some populations. When adding Iron Age/Medieval Iran to this graph, a small number of f_4 outliers persisted which could not easily be resolved. As such we reduced the samples down to the skeleton graph and fit Iron Age/Medieval Iran as a mixture of Eastern (52%) and Western-like (48%) ancestries (Figure S18g).

Appendix Material for Chapter 5

Appendix Table 5.1 - Outlier Fst windows. Highest Fst window within each outlier regions of selection analysis. Chrom = Chromosome. Populations are defined in Table 5.1. θ calculated was Waterson's θ.

Population	Chrom	midPos	F_{ST} East - Bezoar	F_{ST} West - Bezoar	θ_{East}	θ_{West}	θ_{Bezoar}	$\log(\theta_{Bezoar}/\theta_{East})$	$\log(\theta_{Bezoar}/\theta_{West})$
Neolithic West	1	133175000	0.41703	0.599238	101.1809	15.74633	108.2031	-0.02914108424	-0.8370605813
	1	144055000	0.178221	0.632396	46.47697	22.56825	48.215	-0.01594444642	-0.3296842743
	5	18115000	0.617585	0.561898	30.69526	32.8044	66.15599	-0.3334978793	-0.3046370411
	6	68245000	0.405734	0.561014	35.17289	15.11657	40.83073	-0.06477907159	-0.4315338415
	17	20785000	0.607584	0.733631	41.48599	23.08997	50.14706	-0.08234402357	-0.3368220254
	28	22235000	0.416527	0.637407	54.94243	18.69951	41.5688	0.1211403446	-0.3469372692
	29	28695000	0.229033	0.620538	86.97662	23.24215	53.64968	0.2098354068	-0.36329082
Neolithic East	2	78885000	0.543475	0.393643	31.13867	57.81501	61.96681	-0.2988590597	-0.03011851147
	2	127985000	0.527998	0.107472	28.66108	56.9025	83.26636	-0.4631770327	-0.1653382445
	3	51335000	0.553355	0.362574	16.81623	58.26427	68.85851	-0.6122288721	-0.07255533784
	3	102885000	0.545601	0.244399	33.65774	57.36828	54.2362	-0.2072043428	0.02438259207
	4	73545000	0.559825	0.201437	18.05807	50.20387	55.89372	-0.4906915781	-0.04662580826

	4	92495000	0.630302	0.206057	26.61261	47.96346	46.04378	-0.2380833976	0.01773952709
	5	18115000	0.617585	0.561898	30.69526	32.8044	66.15599	-0.3334978793	-0.3046370411
	6	68345000	0.594616	0.218253	37.62755	46.84444	52.55979	-0.1451476485	-0.04999552445
	6	86515000	0.635134	0.393283	37.78558	97.13771	91.97311	-0.3863348278	0.02372695674
	8	38535000	0.603036	0.59514	27.85608	33.73282	45.05761	-0.2088481157	-0.1257154706
	8	38675000	0.54578	0.203788	21.25614	43.6645	39.491	-0.2690136445	0.04363035452
	9	11625000	0.594317	0.283803	28.0798	49.88961	44.09345	-0.1959801408	0.05363600635
	10	38485000	0.542376	0.333615	36.29548	51.69413	65.12629	-0.2539038491	-0.1003151484
	26	15015000	0.870403	0.65401	13.60916	51.67329	31.87432	-0.3696096403	0.2098251936

Appendix Material for Chapter 6

Appendix Table 6.1 - Photo attributions and Creative Commons License for Figure 6.2. Photographs were obtained from https://commons.wikimedia.org/, and modified by cropping and dimension rescaling.

Species	URL	Creative Commons License	Author
Bezoar ibex	https://commons.wikimedia.org/wiki/File:Athina_Ogrod_Narodowy_kozy.jpg	Attribution-Share Alike 3.0 Unported	Andrzej Otrębski
Walia Ibex	https://commons.wikimedia.org/wiki/File:Male_Walia_Ibex_in_the_fog,_Simien_Mts,_Ethiopia_(7155872550).jpg	Attribution-Share Alike 2.0 Generic	Rod Waddington
East Caucasian Tur	https://commons.wikimedia.org/wiki/File:Capra_cylindricornis_2.JPG	Public Domain	Altaipanther
West Caucasian Tur	https://commons.wikimedia.org/wiki/File:West_Caucasian_Tur.jpg	Attribution-Share Alike 3.0 Unported	Pavanravela
Markhor	https://commons.wikimedia.org/wiki/File:National_Animal_of_Pakistan.jpg	Attribution-Share Alike 4.0 International	MonomoyDas
Iberian Ibex	https://commons.wikimedia.org/wiki/File:Cabra_mont%C3%A9s_2.jpg	Attribution-Share Alike 3.0 Unported	Juan Lacruz
Siberian Ibex	https://commons.wikimedia.org/wiki/File:Steinbock1000943.JPG	Attribution-Share Alike 2.5 Generic	Gunnar Ries
Alpine Ibex	https://commons.wikimedia.org/wiki/File:Steinbock_at_Dachstein.jpg	Attribution-Share Alike 4.0 International	Stefan Kasberger
Nubian Ibex	https://commons.wikimedia.org/wiki/File:Male_Nubian_Ibex.jpg	Attribution 2.0 Generic	Rennett Stowe

Appendix Table 6.2 - New modern goat genomes included in analysis. These samples are in addition to those presented in Appendix Table 4.1. Genome coverage presented after mapping quality 30 filter. Italian modern genomes were initially published in (Alberto et al. 2018b), while the Cashmere goat genomes were first published in (Li et al. 2017).

Sample ID	ENA Accession	Breed/Location	Genome Coverage	Grouping
Italian 1	ERR405774	Italian Saanen	13.86	Modern Italy
Italian 2	ERR405775	Italian Saanen	13.72	Modern Italy
Italian 4	ERR405776	Italian Saanen	12.96	Modern Italy
Italian 3	ERR405777	Italian Saanen	13.45	Modern Italy
Italian 5	ERR405778	Italian Saanen	12.82	Modern Italy
Cashmere 1	SAMN05592064	Alashan, Inner Mongolia Cashmere	11.49	Modern China
Cashmere 2	SAMN05592070	Alashan, Inner Mongolia Cashmere	3.18	Modern China
Cashmere 3	SAMN05592075	Alashan, Inner Mongolia Cashmere	6.01	Modern China
Cashmere 4	SAMN05592080	Alashan, Inner Mongolia Cashmere	3.23	Modern China
Cashmere 5	SAMN05592082	Alashan, Inner Mongolia Cashmere	11.23	Modern China
Cashmere 6	SAMN05601115	Liaoning, Inner Mongolia Cashmere	7.83	Modern China
Cashmere 7	SAMN05601117	Liaoning, Inner Mongolia Cashmere	3.65	Modern China
Cashmere 8	SAMN05601123	Liaoning, Inner Mongolia Cashmere	3.33	Modern China
Cashmere 9	SAMN05601126	Liaoning, Inner Mongolia Cashmere	6.98	Modern China
Cashmere 10	SAMN05601128	Aerbasi, Inner Mongolia Cashmere	2.06	Modern China
Cashmere 11	SAMN05601591	Aerbasi, Inner Mongolia Cashmere	2.95	Modern China
Cashmere 12	SAMN05601592	Aerbasi, Inner Mongolia Cashmere	8.18	Modern China
Cashmere 13	SAMN05601602	Aerbasi, Inner Mongolia Cashmere	2.56	Modern China
Cashmere 14	SAMN05601605	Aerbasi, Inner Mongolia Cashmere	11.82	Modern China
Cashmere 15	SAMN05601608	Aerbasi, Inner Mongolia Cashmere	0.5	Modern China
Cashmere 16	SAMN05601625	Erlangshan, Inner Mongolia Cashmere	4.14	Modern China
Cashmere 17	SAMN05601632	Erlangshan, Inner Mongolia Cashmere	2.97	Modern China
Cashmere 18	SAMN05601633	Erlangshan, Inner Mongolia Cashmere	8.66	Modern China
Cashmere 19	SAMN05601636	Erlangshan, Inner Mongolia Cashmere	6.37	Modern China
Cashmere 20	SAMN05601638	Erlangshan, Inner Mongolia Cashmere	2.53	Modern China

Appendix Table 6.3 - mtDNA sequences used in historical *Capra* mtDNA alignment.

Sequence Name	Accession No.	Species	Sequence Name	Accession No.	Species
A1a_04	KR059149.1	*Capra hircus*	Nubian Ibex Ref.	NC_020624.1	*Capra nubiana*
A2a1_14	KR059159.1	*Capra hircus*	Markhor/C. falconeri Ref.	NC_020622.1	*Capra falconeri*
A7_38	KR059182.1	*Capra hircus*	Goat_Ref.	NC_005044.2	*Capra hircus*
D_bezoar_66	KR059210.1	*Capra aegagrus*	Bezoar_Ref.	NC_028161.1	*Capra aegagrus*
D1_67	KR059211.1	*Capra hircus*	Alpine Ibex Ref.	NC_020623.1	*Capra ibex*
D1_68	KR059212.1	*Capra hircus*	Arabian Tahr	NC_020621.1	*Hemitragus jayakari*
G_69	KR059213.1	*Capra hircus*	Himalayan Tahr	NC_020628.1	*Hemitragus jemlahicus*
G_74	KR059218.1	*Capra hircus*	Bharal	NC_020632.1	*Pseudois nayaur*
B_bezoar_75	KR059219.1	*Capra aegagrus*	Pyrenean Chamois	NC_020789.1	*Rupicapra pyrenaica*
B1_78	KR059220.1	*Capra hircus*	Chamois	NC_020633.1	*Rupicapra rupicapra*
C_bezoar_79	KR059221.1	*Capra aegagrus*	Takin	NC_013069.1	*Budorcas taxicolor*
C1_bezoar_80	KR059222.1	*Capra aegagrus*	Rocky Mountain Goat	NC_020630.1	*Oreamnos americanus*
C1a_83	KR059225.1	*Capra hircus*	Barbary Sheep	NC_009510.1	*Ammotragus lervia*
F_bezoar_84	KR059226.1	*Capra aegagrus*	Sheep Ref.	NC_001941.1	*Ovis aries*
West Caucasian Tur Ref.	NC_020683.1	*Capra caucasica*	Yak	NC_006380.3	*Bos grunniens*
Siberian Ibex Ref.	NC_020626.1	*Capra sibirica*			

Appendix Table 6.4 - Mitochondrial references used for *Capra* sample realignment.

Accession number and binomial species names are provided in Appendix Table 6.2.

***Capra* sample**	**mtDNA reference used**	***Capra* sample**	**mtDNA reference used**
Bouc1	G_69	Pyrenica1	Alpine Ibex Ref.
Falconeri1	Barbary Sheep	Pyrenica2	Alpine Ibex Ref.
Falconeri2	Markhor/*C. falconeri* Ref.	Sibirica1	Nubian Ibex Ref.
Ibex1	Alpine Ibex Ref.	Sibirica2	Markhor/*C. falconeri* Ref.
Ibex2	Alpine Ibex Ref.	Tur1	West Caucasian Tur Ref.
Nubiana1	Nubian Ibex Ref.	Tur2	West Caucasian Tur Ref.
Nubiana2	Nubian Ibex Ref.	Walie1	Nubian Ibex Ref.

Appendix Table 6.5 - Modern and ancient *Capra aegagrus*/*hircus* included in IBS calculations.

Samples		
Ainghazal1	Monjukli8	CHIR_1.0
Ainghazal2	Potterne1	Cashmere 1
Azer3-5	Qazvin1	Cashmere 5
Blagotin1	Semnan1-2	Cashmere 14
Blagotin3	Semnan3	Iranian modern 5
Direkli1-2	Yoqneam2	Iranian modern 18
Direkli5	Moroccan 3	Iranian bezoar 4
Direkli6	Moroccan 7	Iranian bezoar 7
Geor2	Tog	Iranian bezoar 19
Hovk1	Italian modern 2	Iranian bezoar 20
Kazbeg1	French modern 1	Yak
Lur12	IOG	

Appendix Table 6.6 - Historic *Capra* Illumina HiSeq sequencing and whole genome alignment results. Filtered reads are following cutadapt trimming and minimum read length 30bp filter. Endogenous was calculated as reads aligned following mapQ 20 filter divided by the filtered read count. Samples were aligned to ARS1.

Sample Name	Raw Reads	Filtered reads	Aligned reads	Rmdup total reads	Rmdup aligned reads	>q20 Aligned rmdup reads	Endog. %age	Coverage
Bouc1	142818189	137331014	113372552	117549599	93591137	66879083	48.70	1.19
Falconeri1	182537854	181326994	107781316	142292569	52648172	17335980	9.56	0.51
Falconeri2	44646734	43877030	14771365	33965473	4859808	1898172	4.33	0.05
Ibex1	224934434	224033309	173852483	210082698	159901872	133564342	59.62	3.53
Ibex2	12051498	11894653	3963190	11000937	3069474	2193867	18.44	0.05
Nubiana1	397040312	388001792	223487192	239516967	75002367	37455095	9.65	1.10
Nubiana2	173761	173501	137398	168961	132858	50423	29.06	0.001
Pyrenica1	8929995	8596593	47845	8592312	43564	11287	0.13	0.0002
Pyrenica2	71109841	70758448	11189389	68503640	8934581	5338055	7.54	0.14
Sibirica1	8509565	8442164	3867413	7660096	3085345	995597	11.79	0.03
Sibirica2	115731840	115399258	88110057	94698707	67409506	43571001	37.76	1.34
Tur1	139994905	139908781	117455113	129384434	106930766	75499049	53.96	2.48
Tur2	7288450	7246195	978124	7168006	899935	657728	9.08	0.02
Walie1	127190988	126735760	99472128	65811302	38547670	21753295	17.16	0.66

Appendix Table 6.7 - ARS1 realignment statistics for ancient genomes. Final number of aligned reads to ARS1 was computed after removal of duplicates and mapping quality filter (minimum 30).

Sample	q30 rmdup aligned reads	Coverage	Sample	q30 rmdup aligned reads	Coverage	Sample	q30 rmdup aligned reads	Coverage
Acem1	216383993	3.88	Darre2	166484043	3.41	Monjukli4	26749390	0.52
Acem2	410791061	7.53	Direkli1-2	499370201	10	Monjukli6	1358535	0.02
Ainghazal1	1699170	0.03	Direkli5	13449345	0.23	Monjukli8	110878235	2.22
Ainghazal2	3145996	0.05	Direkli6	130428598	2.18	Potterne1	154641839	3.19
Ainghazal3	182022	0.002	Fars1	1061649	0.02	Qazin1	146829894	2.79
Ainghazal4	390552	0.01	Fars2-5	1622229	0.03	Safi2	1976753	0.04
AP45	714305	0.01	Fars4	47231510	0.91	Semnan1-2	314193038	5.91
AP49	983953	0.02	Geor2	57276726	1.31	Semnan3	640499787	12.9
Azer3-5	221975203	4.05	Ghosh5	8512	0.0001	Semnan7	185622636	2.84
Azer4	113630701	2.23	Gilat2	77805	0.0012	Semnan8	12378245	0.18
Azer6	13149155	0.25	Gilat8	822826	0.01	Semnan9	138000788	2.65
Blagotin1	377109260	6.04	Gilat10	39563	0.0006	Semnan10	78385159	1.24
Blagotin2	219442397	3.48	Hovk1	147473380	2.87	Semnan13	118904073	2.53
Blagotin3	634230293	9.93	Kazbeg1	158546114	3.34	Semnan17	5705565	0.11
Blagotin16	207095649	3.03	Kohneh2	2335065	0.04	Shiqmim1	151753	0.0027
Bulak1	38508419	0.76	Kov57	3350768	0.06	Shiqmim9	36531	0.0006

Bulak2	136455211	2.23	Lur9	7992	0.0001	Tac3	5579565	0.11
Bulak3	29881709	0.53	Lur12	50672962	0.9	Yarmut1	91530	0.0013
Bulak5	14005152	0.23	Miqne5	32461	0.00045	Yarmut7	53705	0.008
Chalow1	2095205	0.04	Monjukli1	10555806	0.21	Yoqneam2	103069174	1.9
Darre1	2033209	0.03	Monjukli2	8507662	0.18			

Appendix Table 6.8 - ARS1 realignment statistics for modern genomes. Final number of aligned reads to ARS1 was computed after removal of duplicates and mapping quality filter (minimum 30). An exception is the Yak, for which a mapping quality filter of 20 was used.

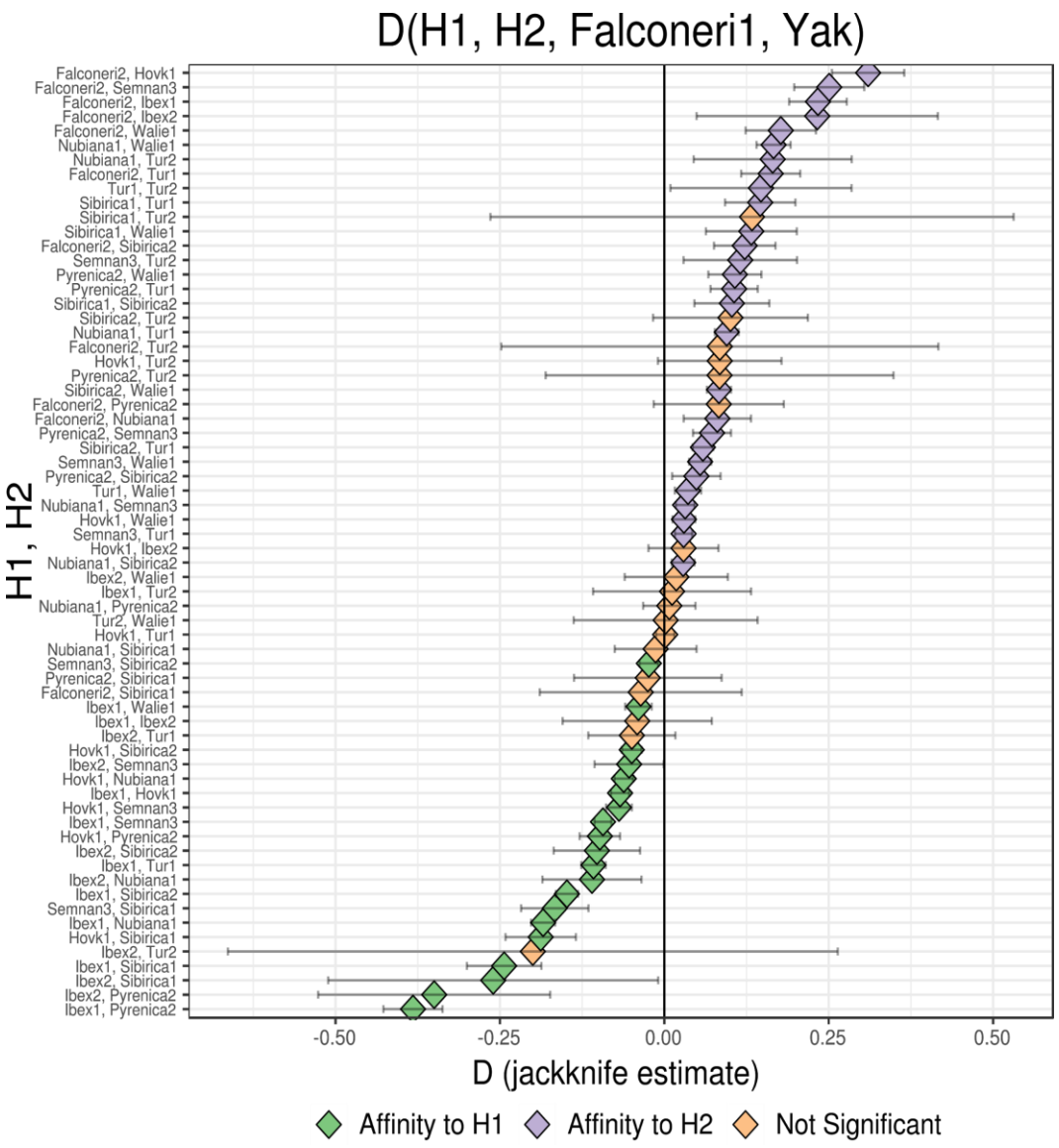
Sample	Coverage	Sample	Coverage	Sample	Coverage	Sample	Coverage
IOG	37.8	Iranian 9	10.08	Moroccan 4	13.75	Iranian bezoar 9	6.07
Tog	34.32	Iranian 10	11.57	Moroccan 5	11.29	Iranian bezoar 10	10.41
Yak	43.01	Iranian 11	11.81	Moroccan 6	13.45	Iranian bezoar 11	6.05
French 1	12.97	Iranian 12	10.93	Moroccan 7	14.15	Iranian bezoar 12	4.73
French 2	12.28	Iranian 13	11.81	Moroccan 8	12.08	Iranian bezoar 13	6.72
French 3	12.03	Iranian 14	10.93	Moroccan 9	11.6	Iranian bezoar 14	10.23
French 4	12.67	Iranian 15	10.99	CHIR_1.0	32.99	Iranian bezoar 15	10.78
Iranian 1	12.43	Iranian 16	12.28	Iranian bezoar 1	13.27	Iranian bezoar 16	11.54
Iranian 2	10.77	Iranian 17	11.5	Iranian bezoar 2	11.36	Iranian bezoar 17	11.29
Iranian 3	11.56	Iranian 18	11.58	Iranian bezoar 3	5.93	Iranian bezoar 18	9.64
Iranian 4	11.75	Iranian 19	12.21	Iranian bezoar 4	6.12	Iranian bezoar 19	11.03
Iranian 5	10.41	Iranian 20	11.87	Iranian bezoar 5	5.03	Iranian bezoar 20	13.07
Iranian 6	11.67	Moroccan 1	11.38	Iranian bezoar 6	10.84	Iranian bezoar 21	12.95
Iranian 7	11.99	Moroccan 2	12.4	Iranian bezoar 7	5.81		
Iranian 8	10.32	Moroccan 3	12.13	Iranian bezoar 8	11.07		

Appendix Table 6.9 - Mitochondrial capture sequencing and mtDNA alignment results for historic *Capra* samples. Alignment statistics presented following realigned to a closer reference sequence as defined in Appendix Table 6.4.

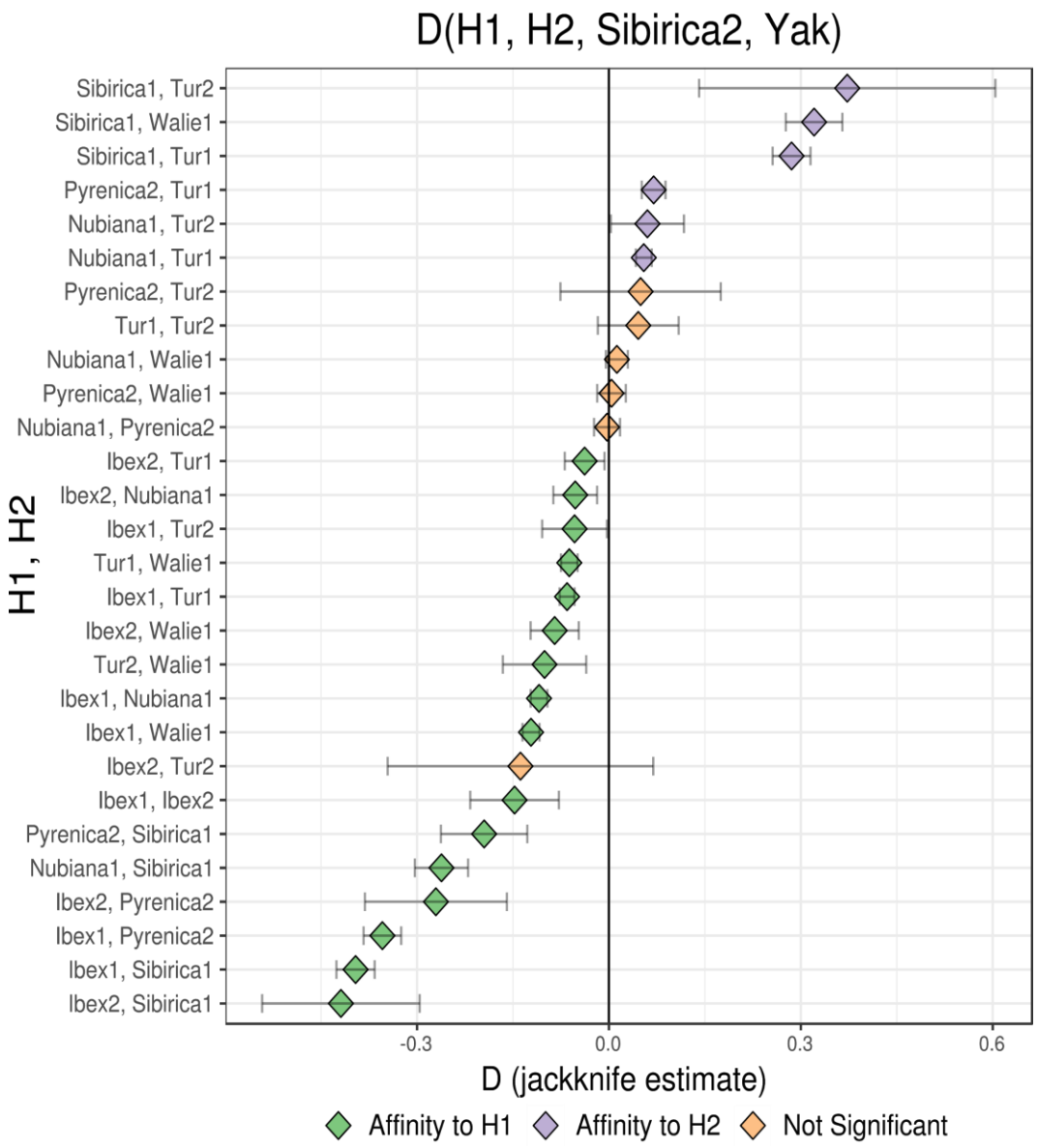
Sample	Raw reads	Filtered reads	>q30 Aligned reads	mtDNA coverage	Called Sites	%age called
Bouc1	142818189	137331014	29791	150.76	16641	100.00
Falconeri1	182537854	181326994	14523	72.11	14088	85.23
Falconeri2	44646734	43877030	8722	45.78	16640	100.00
Ibex1	224934434	224033309	34508	179.03	16716	100.00
Ibex2	12051498	11894653	4658	21.4	16665	99.69
Nubiana1	397040312	388001792	41357	211.32	16705	100.00
Nubiana2	173761	173501	34	0.18	2779	16.64
Pyrenica1	8929995	8596593	54	0.16	2474	14.81
Pyrenica2	71109841	70758448	10802	51.2	16157	96.66
Sibirica1	8509565	8442164	28924	163.16	16705	100.00
Sibirica2	115731840	115399258	37725	205.78	16458	98.91
Tur1	139994905	139908781	10374	59.22	16624	100.00
Tur2	7288450	7246195	1029	4.96	15687	94.36
Walie1	127190988	126735760	18160	103.12	16705	100.00

Appendix Table 6.10 - Additional information of Muséum National d'Histoire Naturelle (MNHN) historic *Capra* samples.

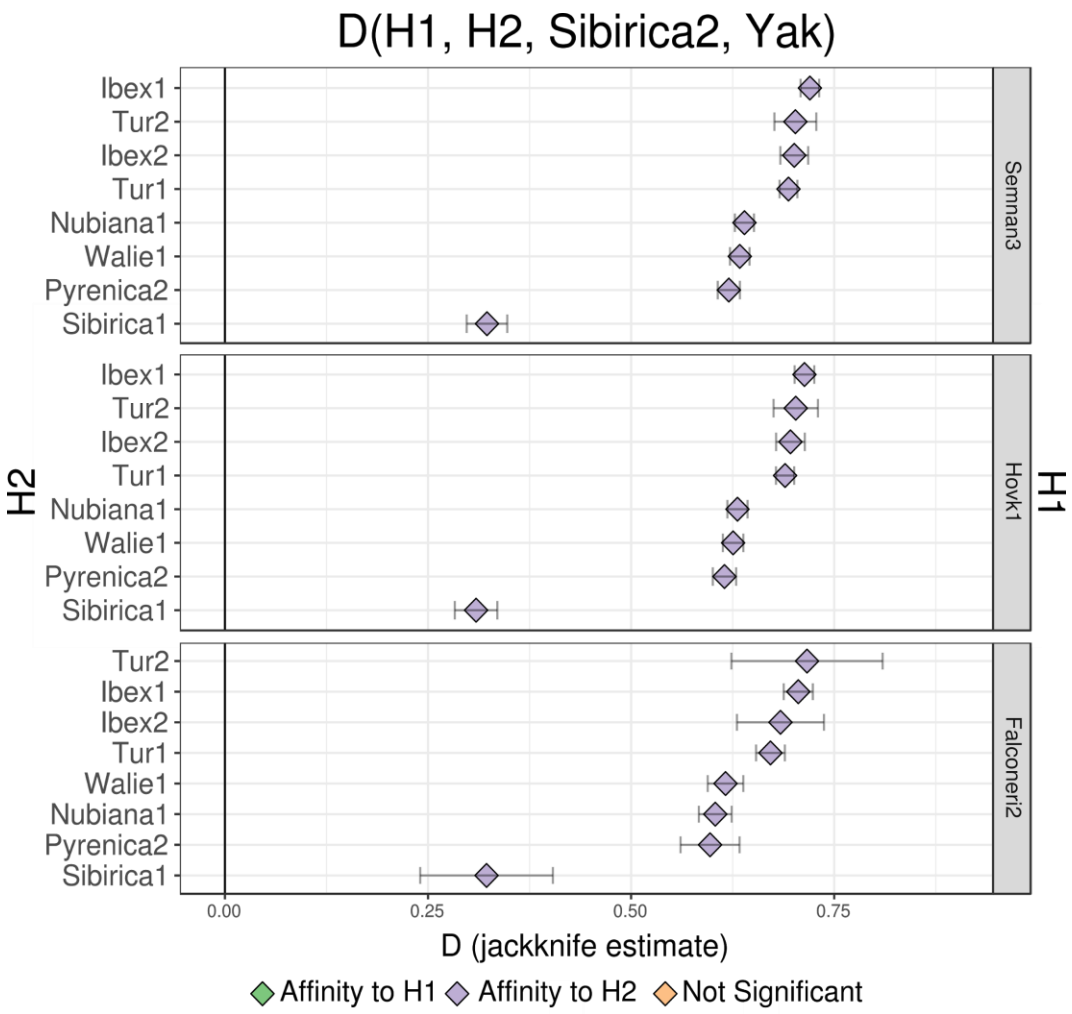
Sample	Origin	Molecular Sex
Bouc1	Male from Palestine. The head was given to MNHN in 1906.	M
Falconeri1	Male coming from the Parc de la Haute-Touche (France).	M
Falconeri2	Born at the museum zoo.	M
Ibex1	Bought from M. Pons in 1899 but no more info on its origin.	F
Ibex2	Male from Savoie. Killed at the Pointe de Calabre.	M
Nubiana1	Male born and died at the museum zoo.	M
Nubiana2	Female entered at 2 years old at the museum.	F
Pyrenica1	Bones given by the museum of Toulouse but no clear origin.	M
Pyrenica2	Bones given by the museum of Toulouse but no clear origin.	M
Sibirica1	Born at the museum zoo. Believed to be a Siberian ibex / unknown ibex hybrid.	M
Sibirica2	Born at the museum zoo.	F
Tur1	Born at the museum zoo.	M
Tur2	Female provided by the zoo of Vincennes.	F
Walie1	Born at the museum zoo but originated from a specimen (unknown gender) kept in Germany but born in Nubia.	M



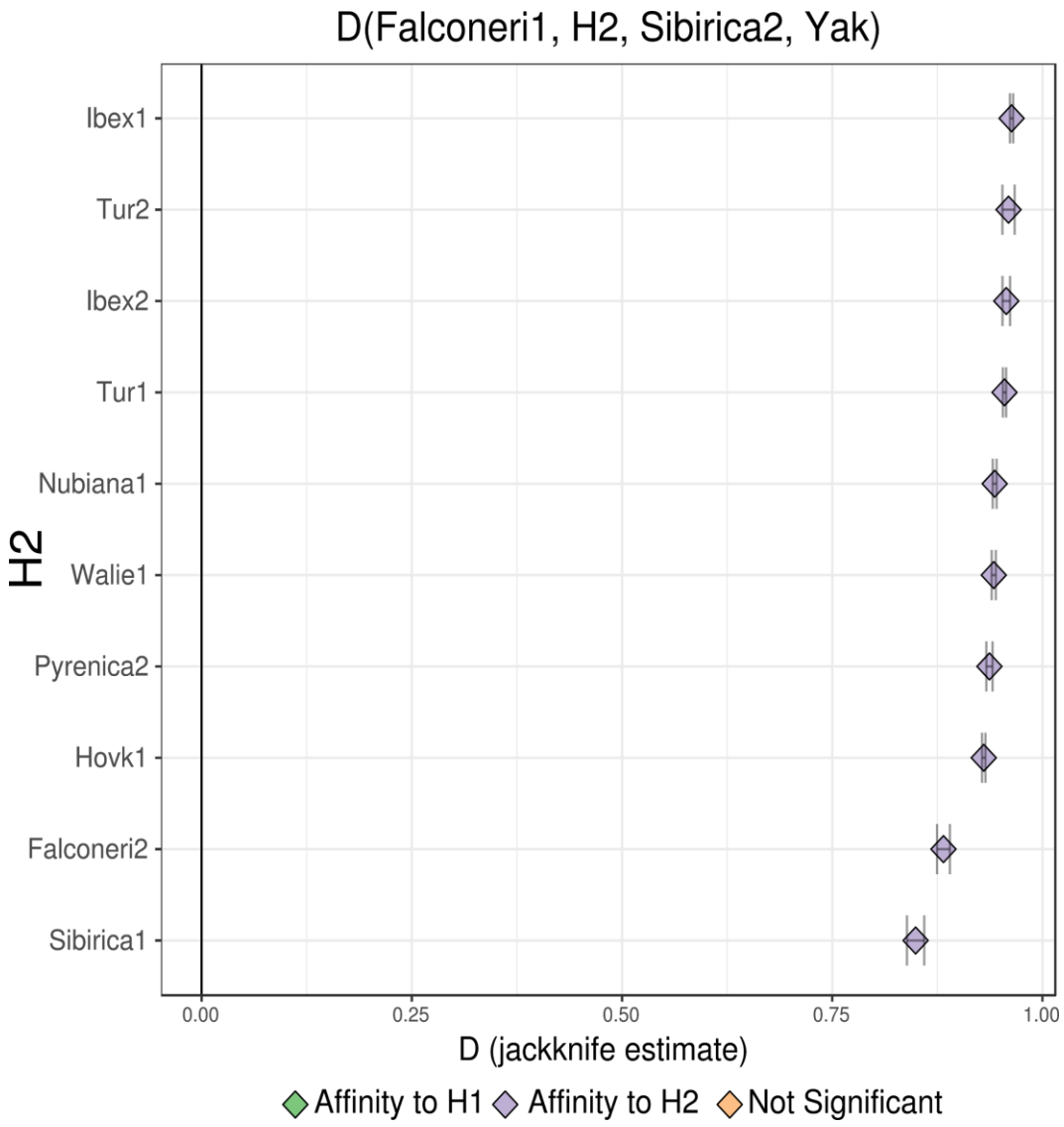
Appendix Figure 6.1 - *D* statistic test for relative affinity (derived allele sharing) of the markhor Falconeri1 to members of pairs of *Capra* genomes, H1 and H2. Colour indicates Falconeri1 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange. Semnan3, a Neolithic Iranian genome, is used to represent domestic goat.



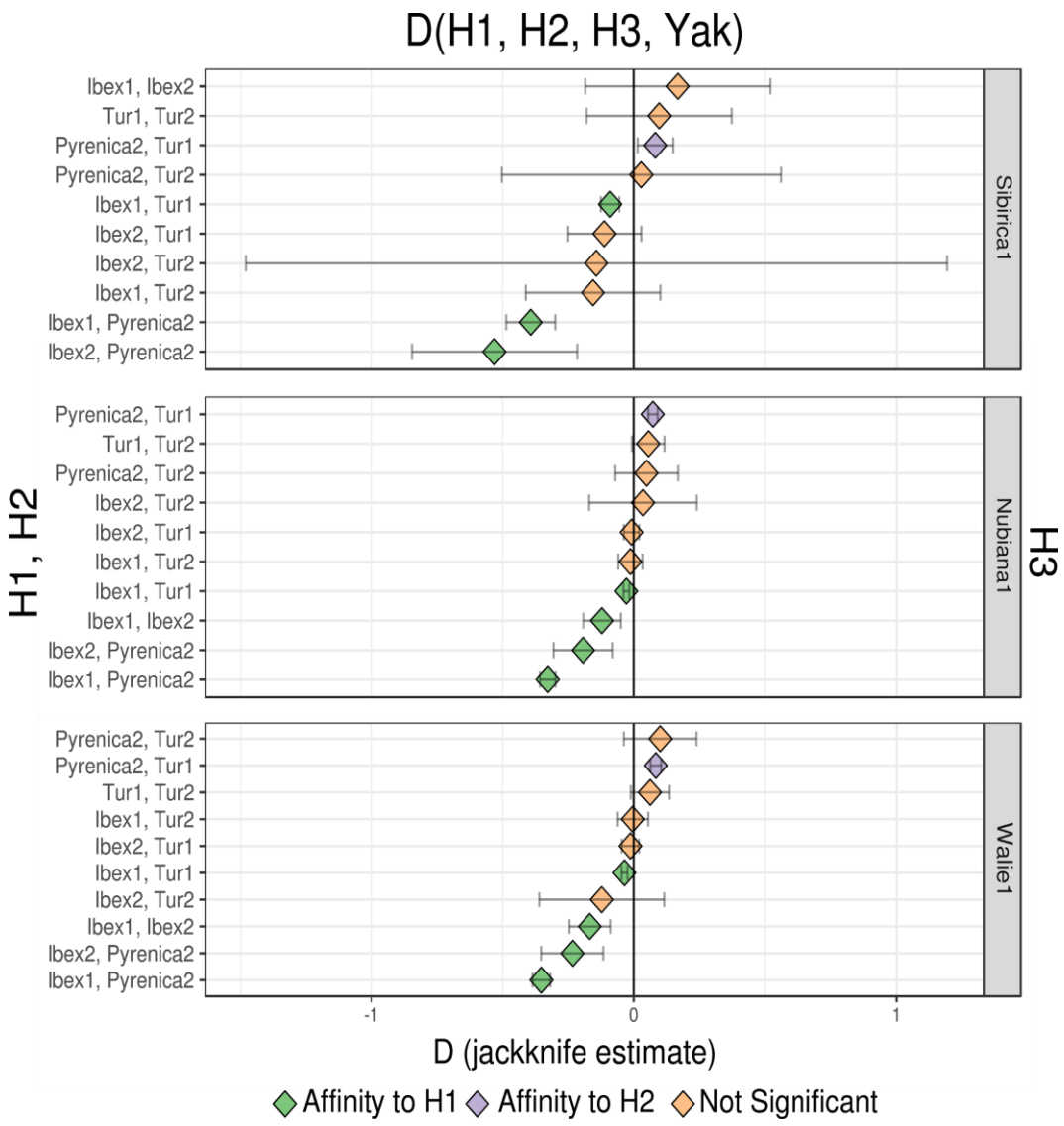
Appendix Figure 6.2 - *D* statistic test for relative affinity (derived allele sharing) of the Siberian ibex Sibirica2 to members of pairs of ibex *Capra* genomes, H1 and H2. Colour indicates Sibirica1 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



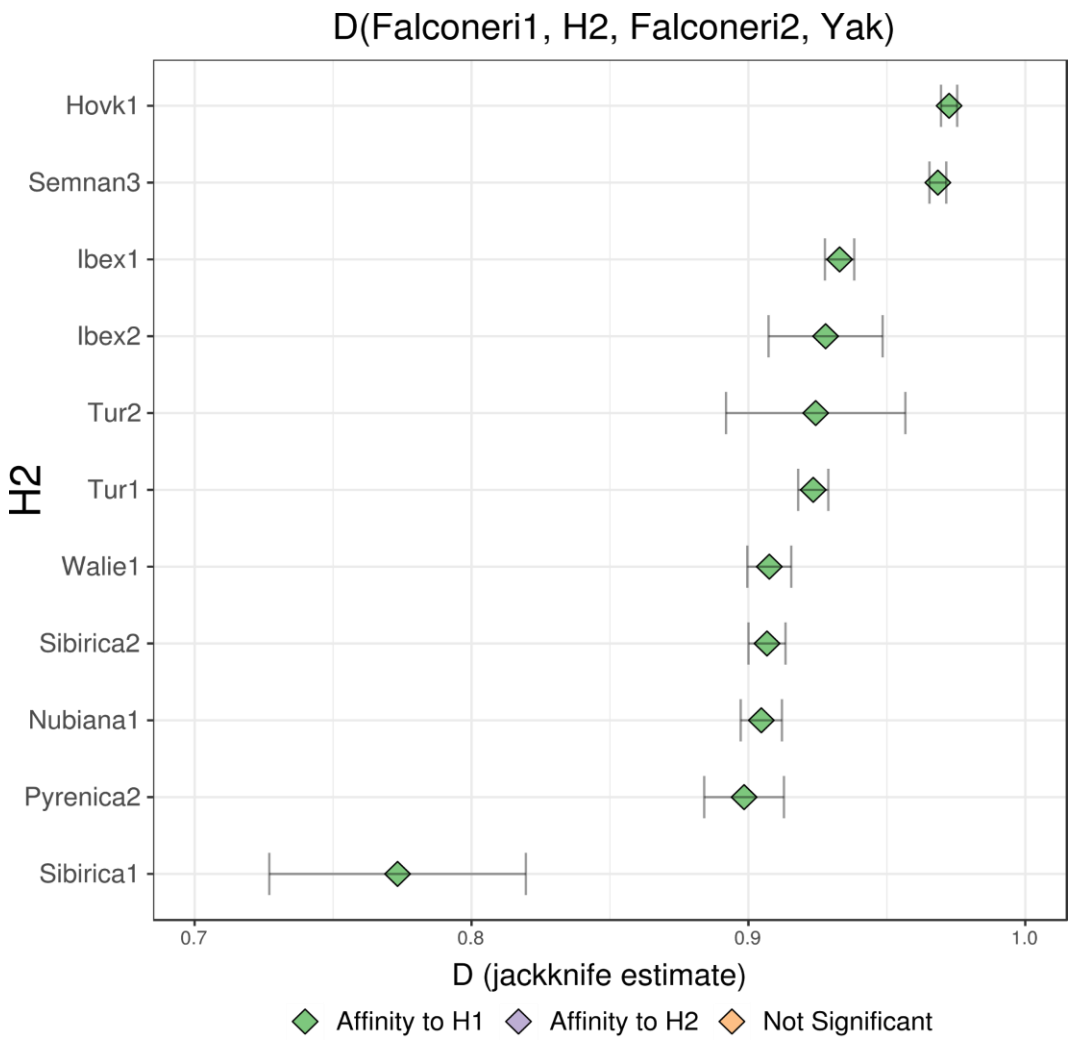
Appendix Figure 6.3 - *D* statistic test for relative affinity (derived allele sharing) of the Siberian ibex Sibirica2 to domestic goat or wild bezoar genome (H1) compared to an ibex genome (H2). Colour indicates Sibirica1 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



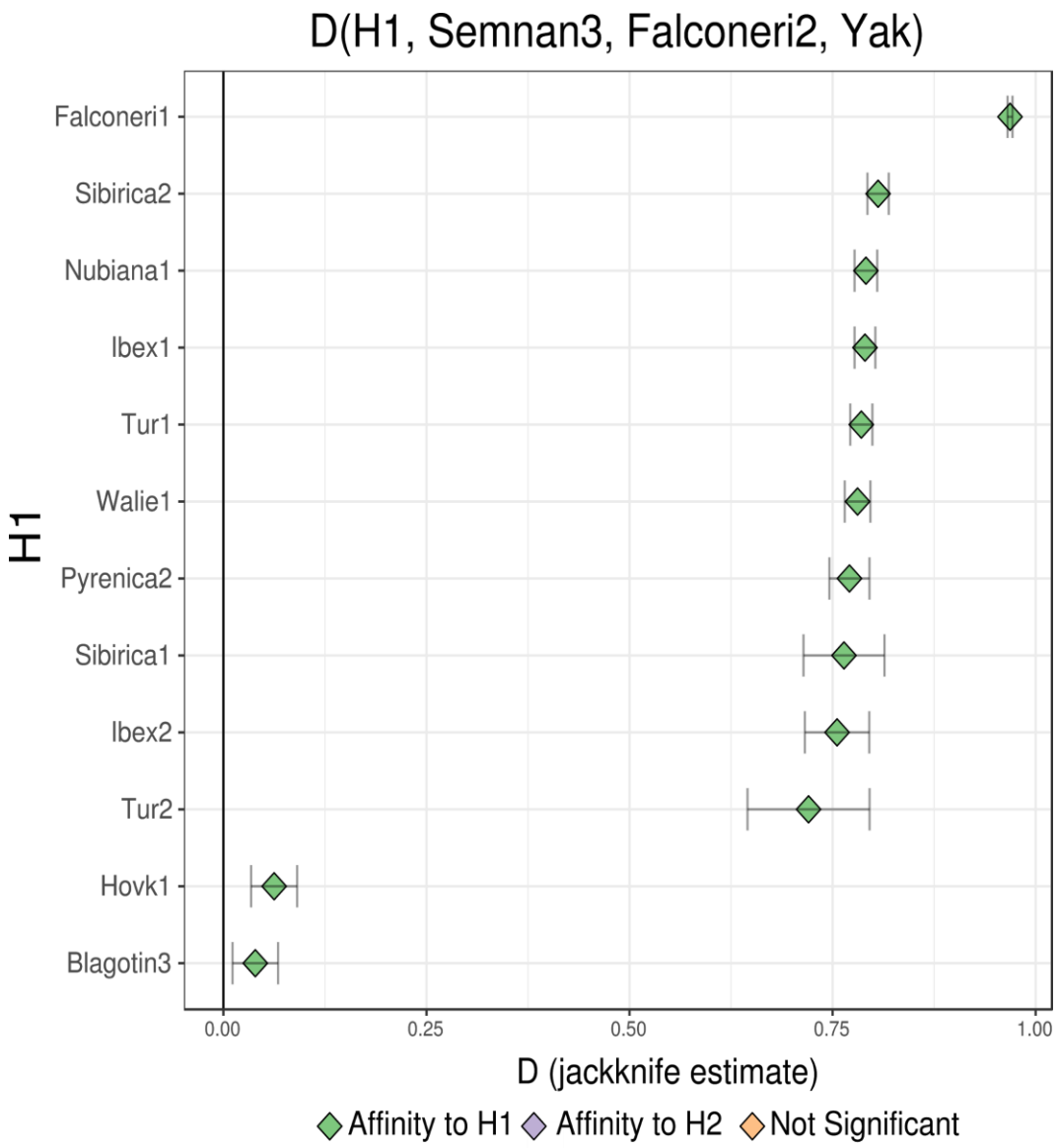
Appendix Figure 6.4 - *D* statistic test for relative affinity (derived allele sharing) of the Siberian ibex Sibirica2 to the markhor Falconeri1 (H1) compared to other *Capra* genomes (H2). Colour indicates Sibirica1 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



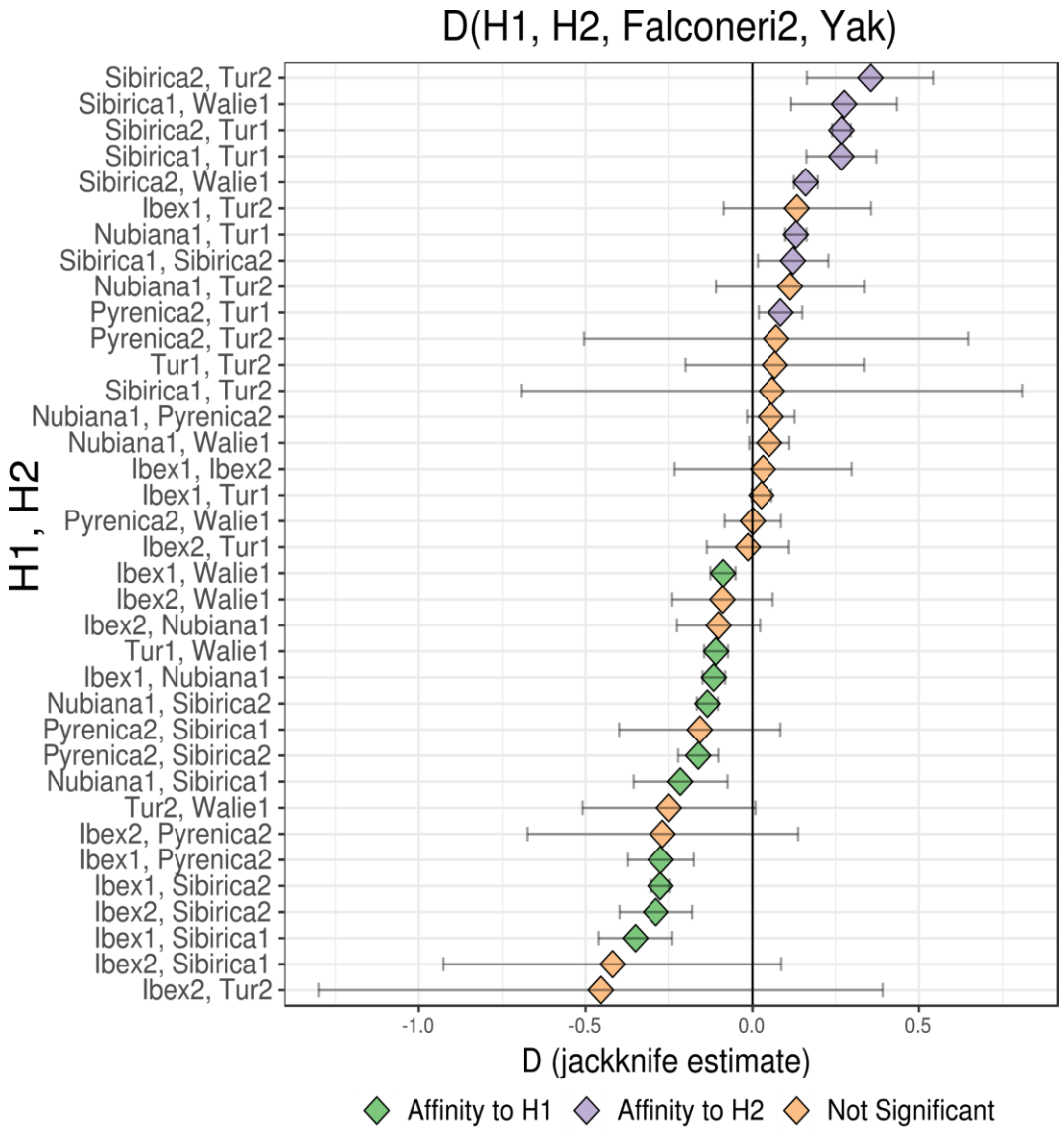
Appendix Figure 6.5 - *D* statistic test for relative affinity (derived allele sharing) of the test genomes Sibirica1, Nubiana1, and Walie1 to members of pairs of other ibex genomes, H1 and H2. Colour indicates test affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



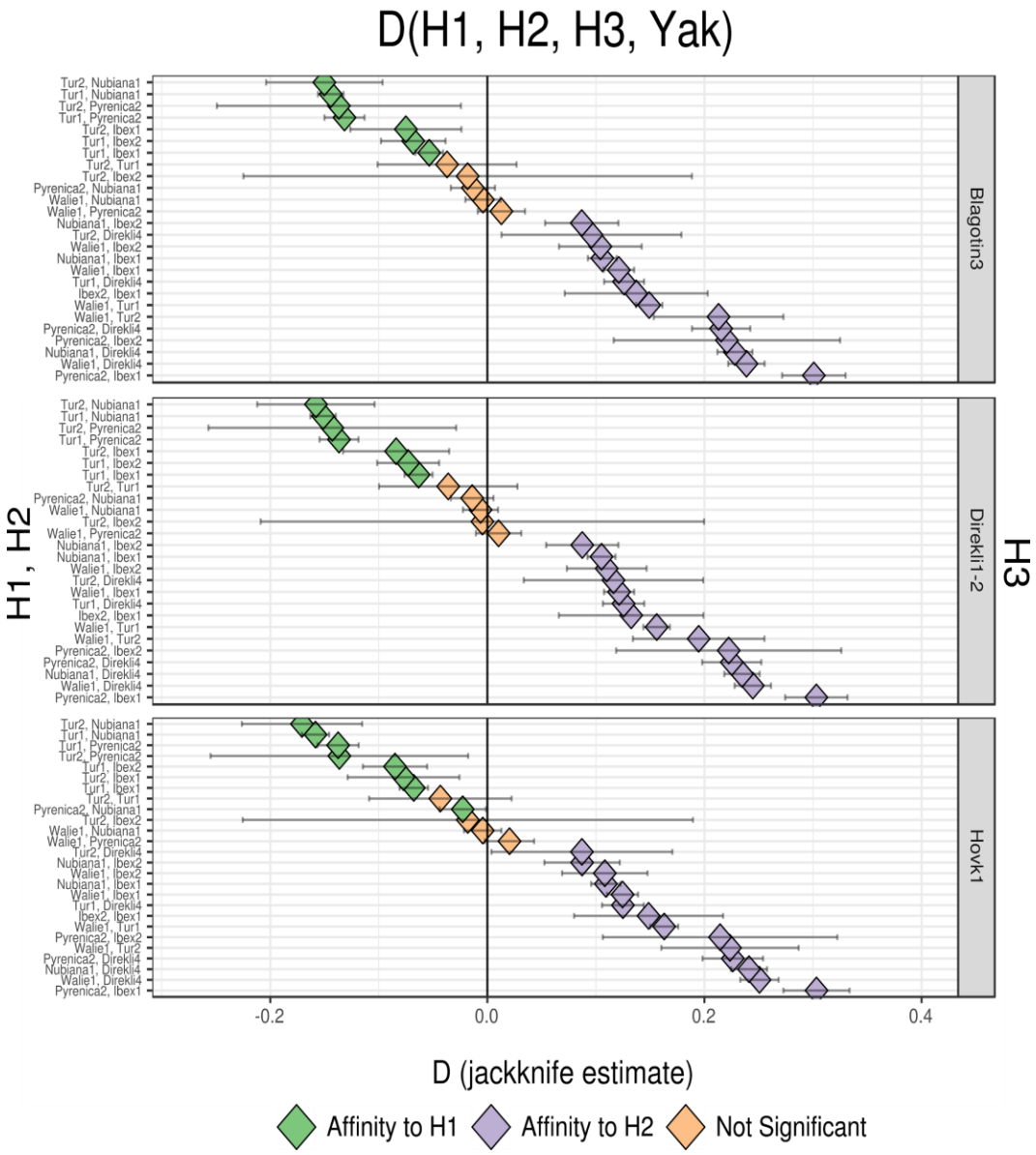
Appendix Figure 6.6 - *D* statistic test for relative affinity (derived allele sharing) of the markhor Falconeri2 to *Capra* genomes (H2) relative to Falconeri1 (H1). Colour indicates Falconeri2 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



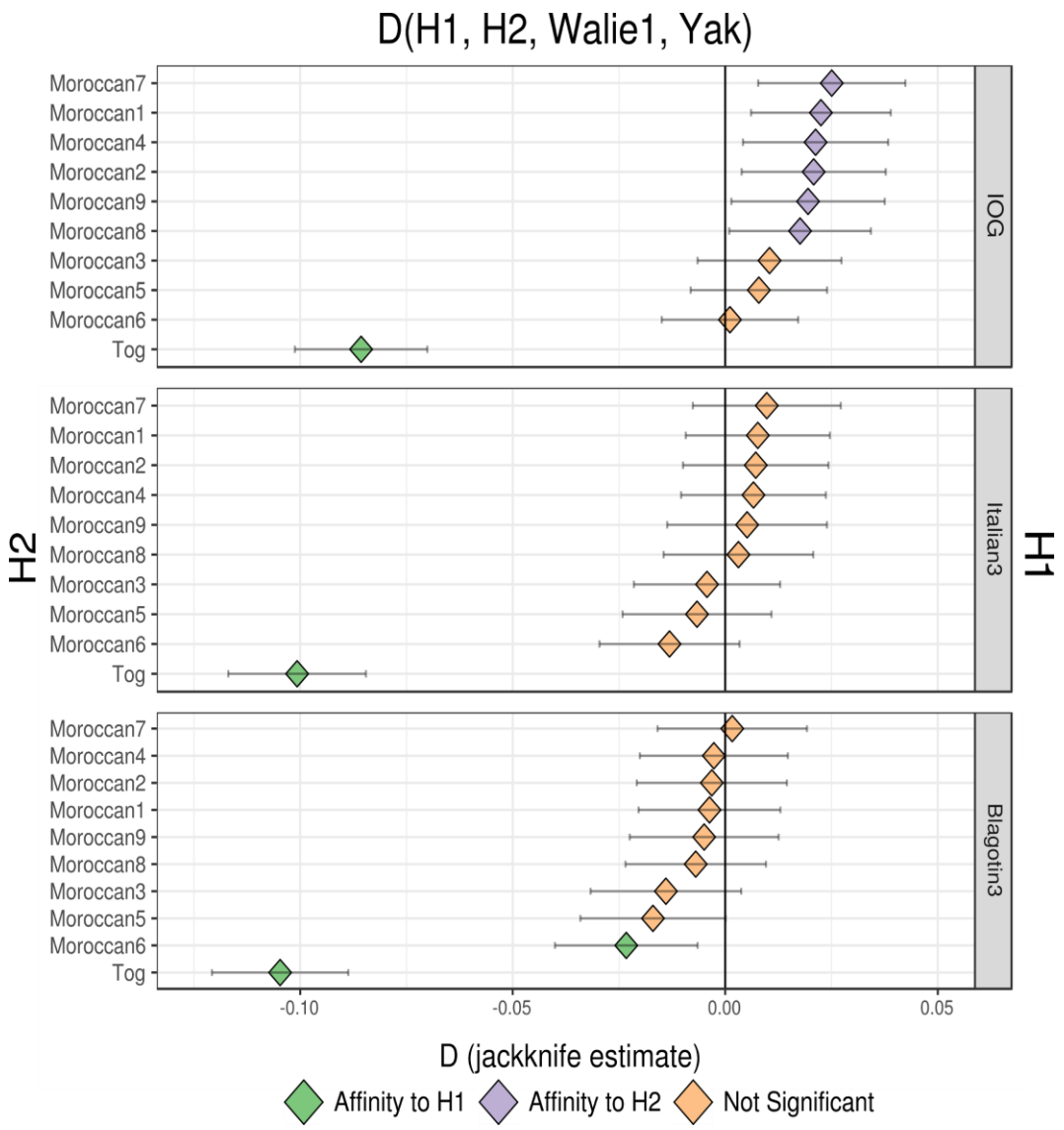
Appendix Figure 6.7 - *D* statistic test for relative affinity (derived allele sharing) of the markhor Falconeri2 to *Capra* genomes (H1) relative to the Neolithic Eastern domestic goat Semnan3 (H2). Colour indicates Falconeri2 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange. Note that for visual purposes the form of this test is reversed to who it was presented in the test (Semnan3, H2, Falconeri2, Yak).



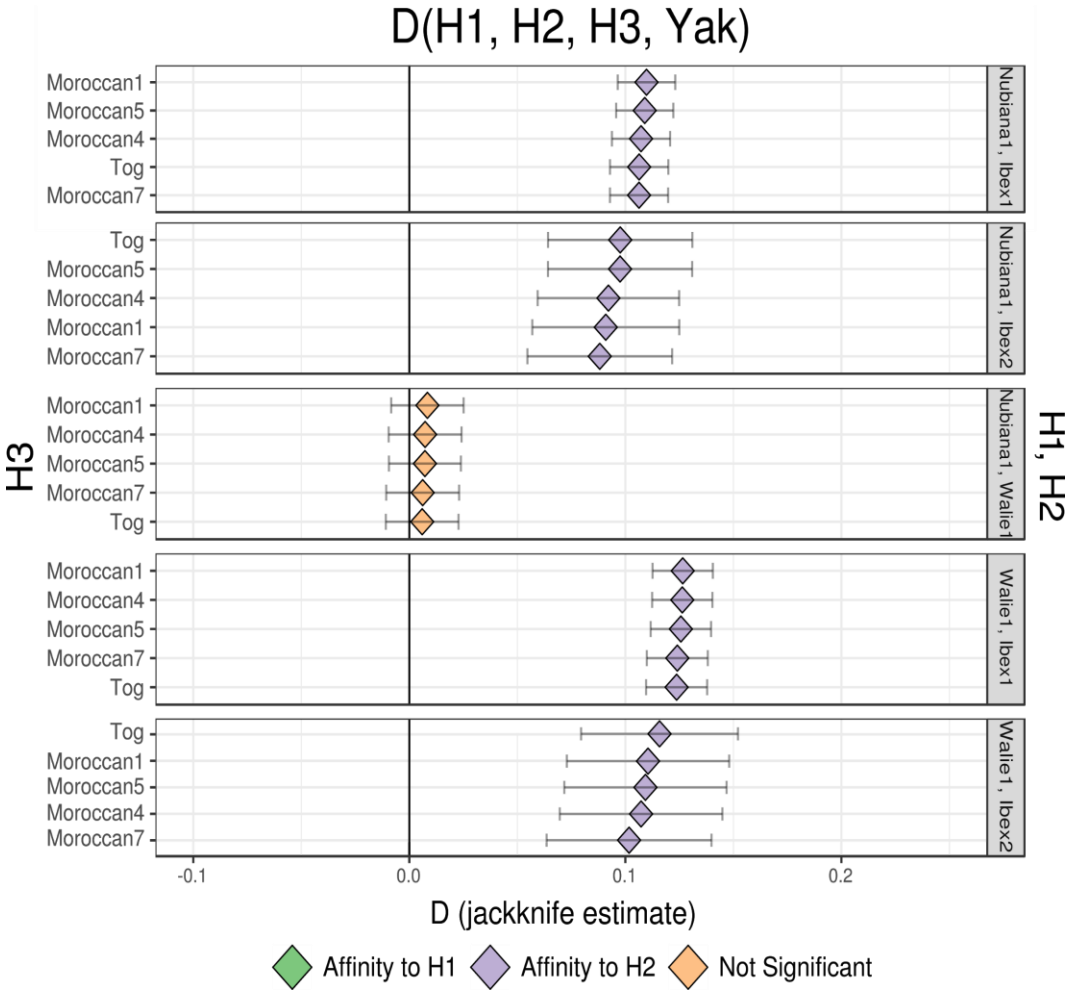
Appendix Figure 6.8 - *D* statistic test for relative affinity (derived allele sharing) of the markhor Falconeri2 to members of pairs of *Capra* genomes, H1 and H2. Colour indicates Falconeri2 affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



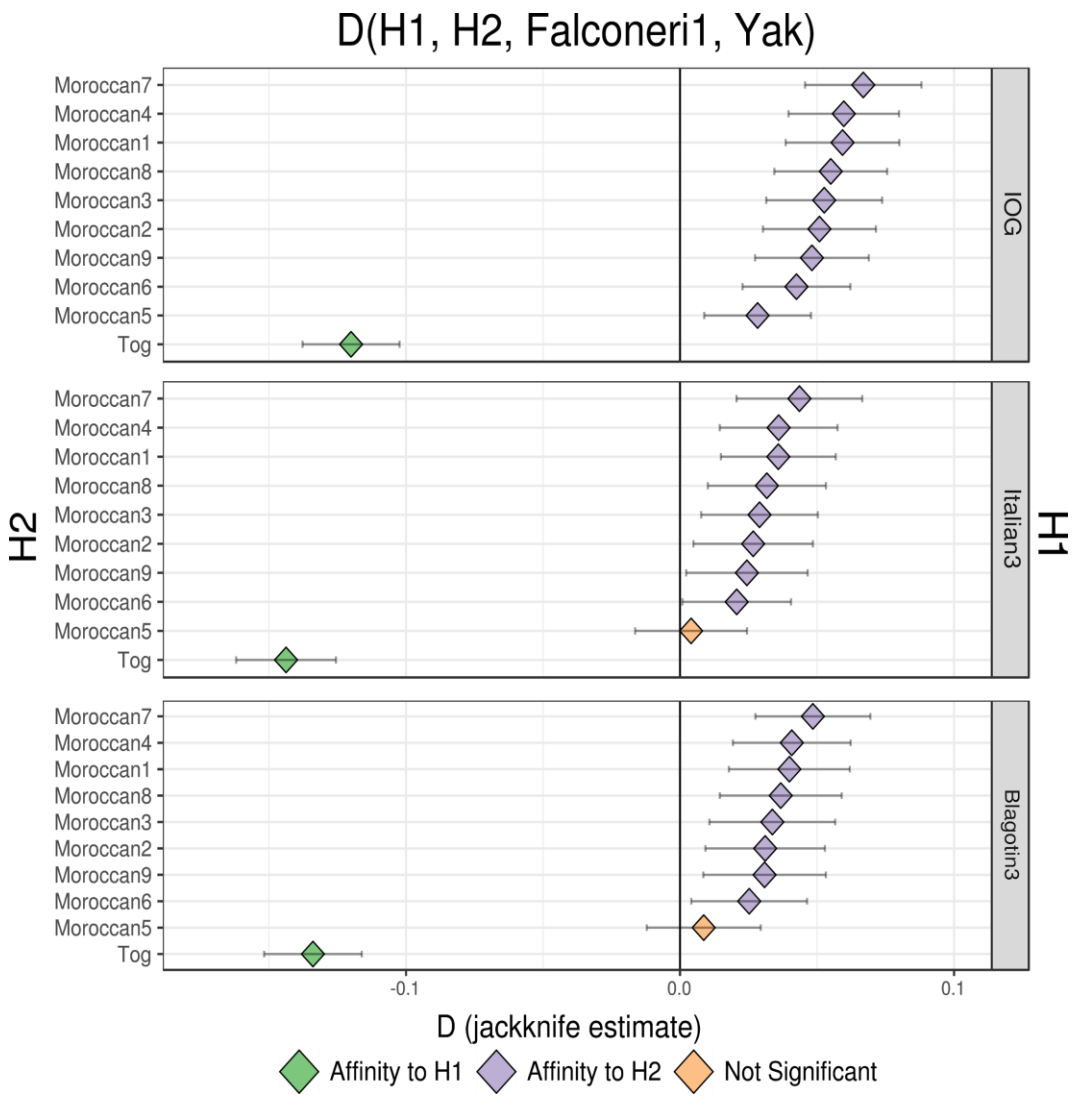
Appendix Figure 6.9 - *D* statistic test for relative affinity (derived allele sharing) of an ancient domestic or wild goat genome (H3) to members of pairs of *Capra* genomes, H1 and H2. Colour indicates H3 affinity (excess of derived alleles shared) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



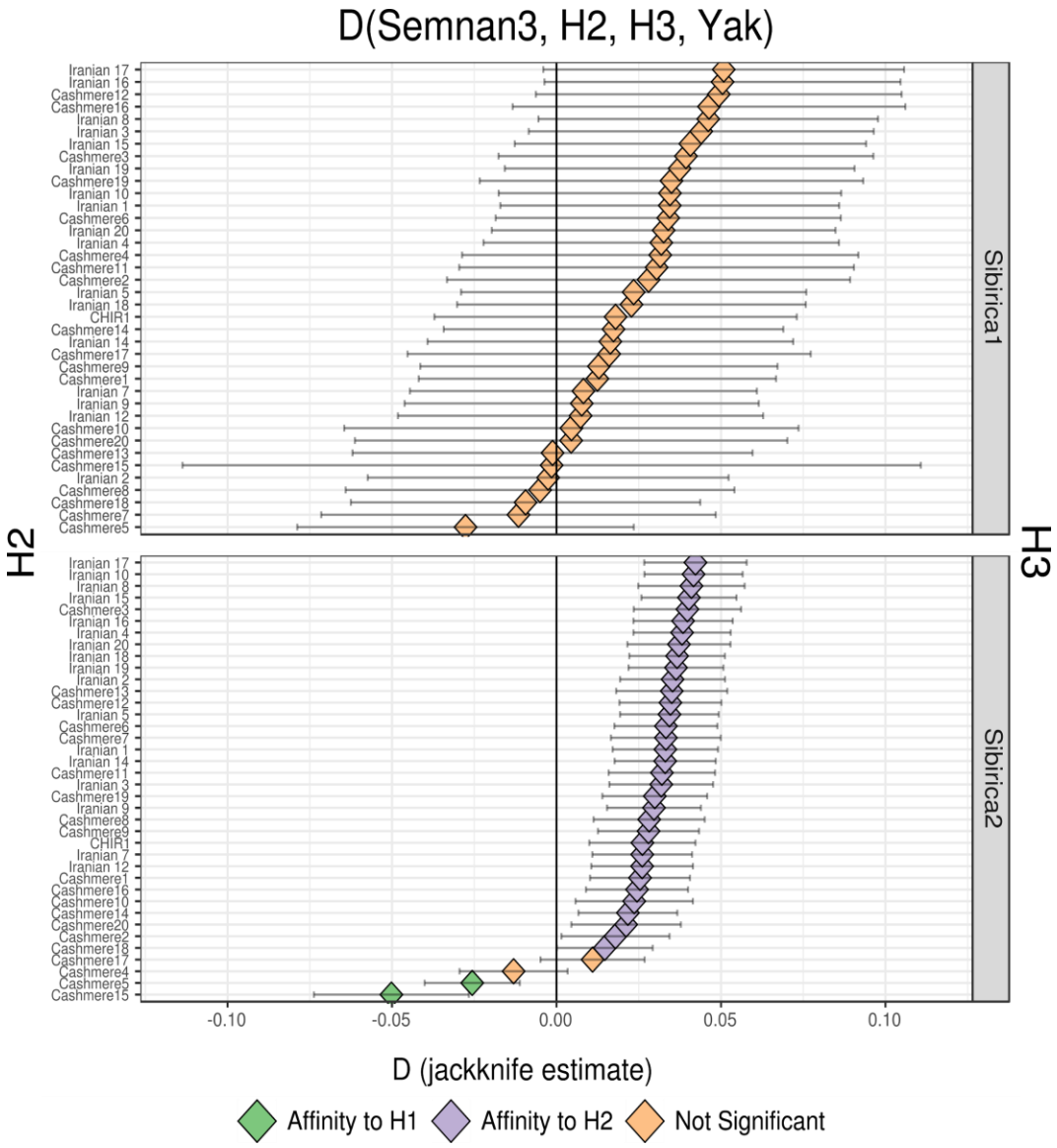
Appendix Figure 6.10 - *D* statistic test for admixture from the Walia ibex Walie1 (H3) to modern African goat (H2) relative to several European genomes (H1). Colour indicates Walie1 affinity (degree of derived allele sharing) to either the European genome H1 or the modern African goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



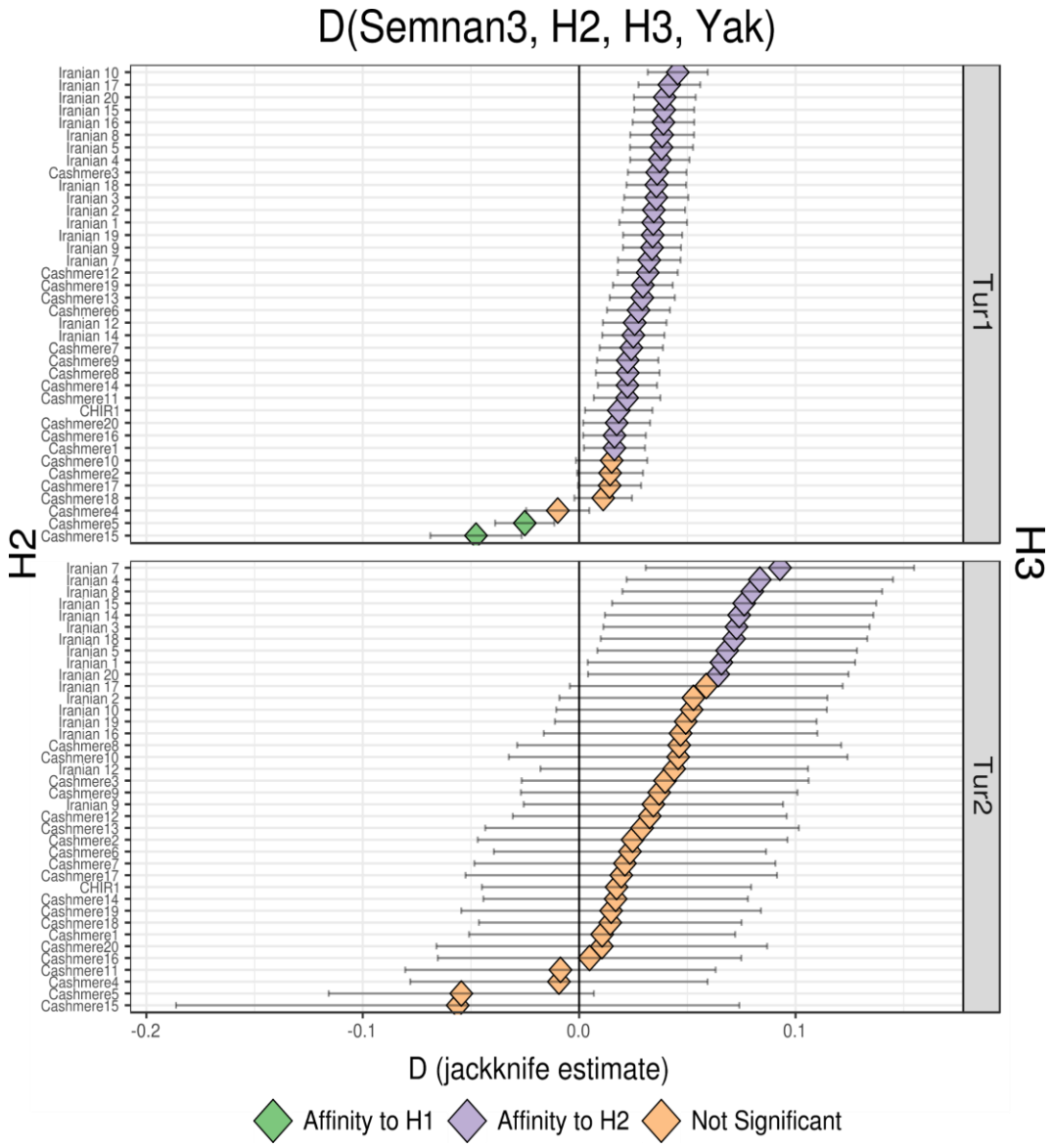
Appendix Figure 6.11 - *D* statistic for derived allele sharing of modern African goat genomes (H3) with either pairs of Nubiana1, Ibex1, Ibex2, and Walie1 (H1, H2). Colour indicates modern African goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



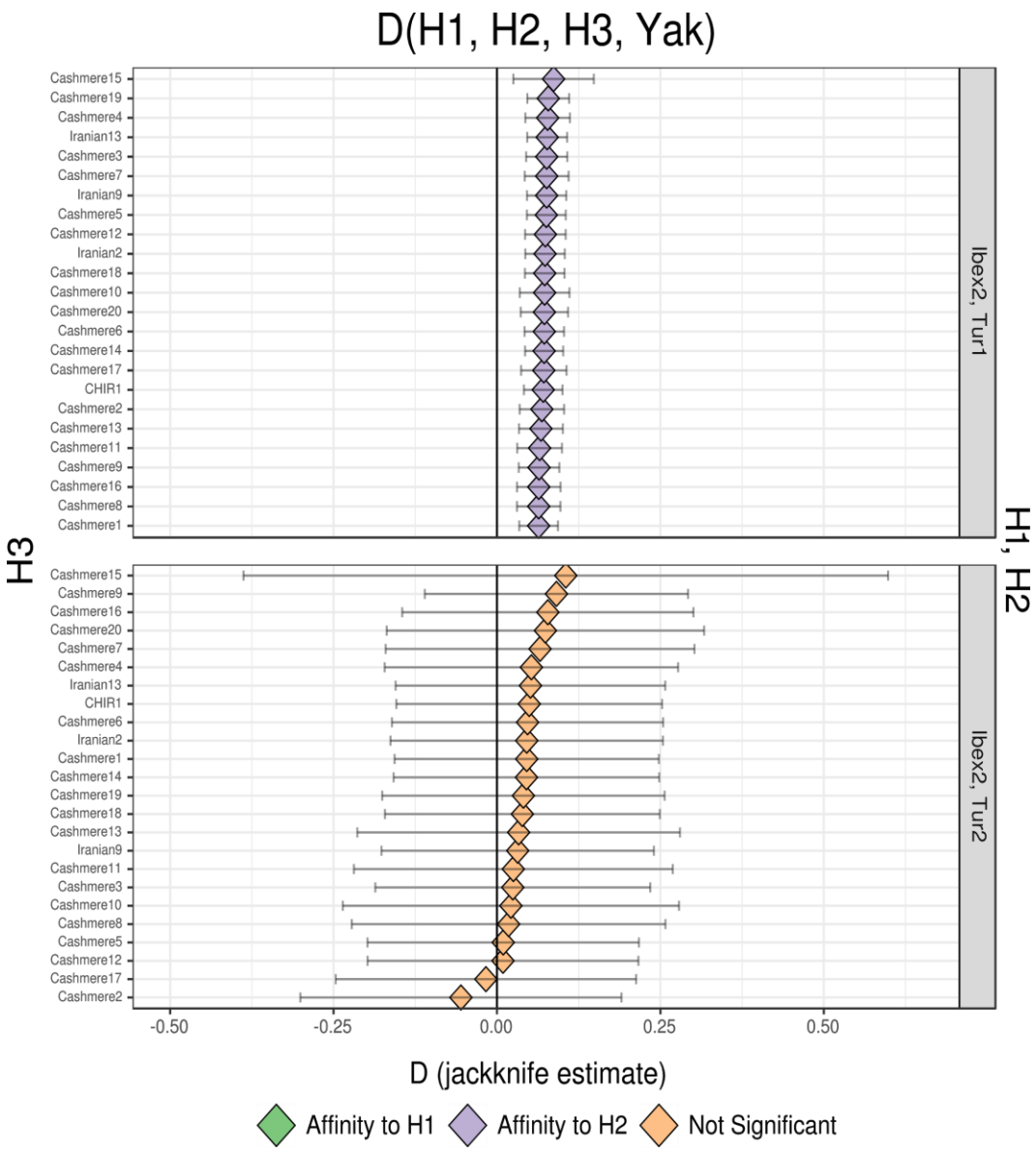
Appendix Figure 6.12 - *D* statistic test for admixture from the markhor Falconeri1 (H3) to modern African goat (H2) relative to several European genomes (H1). Colour indicates markhor affinity (degree of derived allele sharing) to either the European genome H1 or the modern African goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



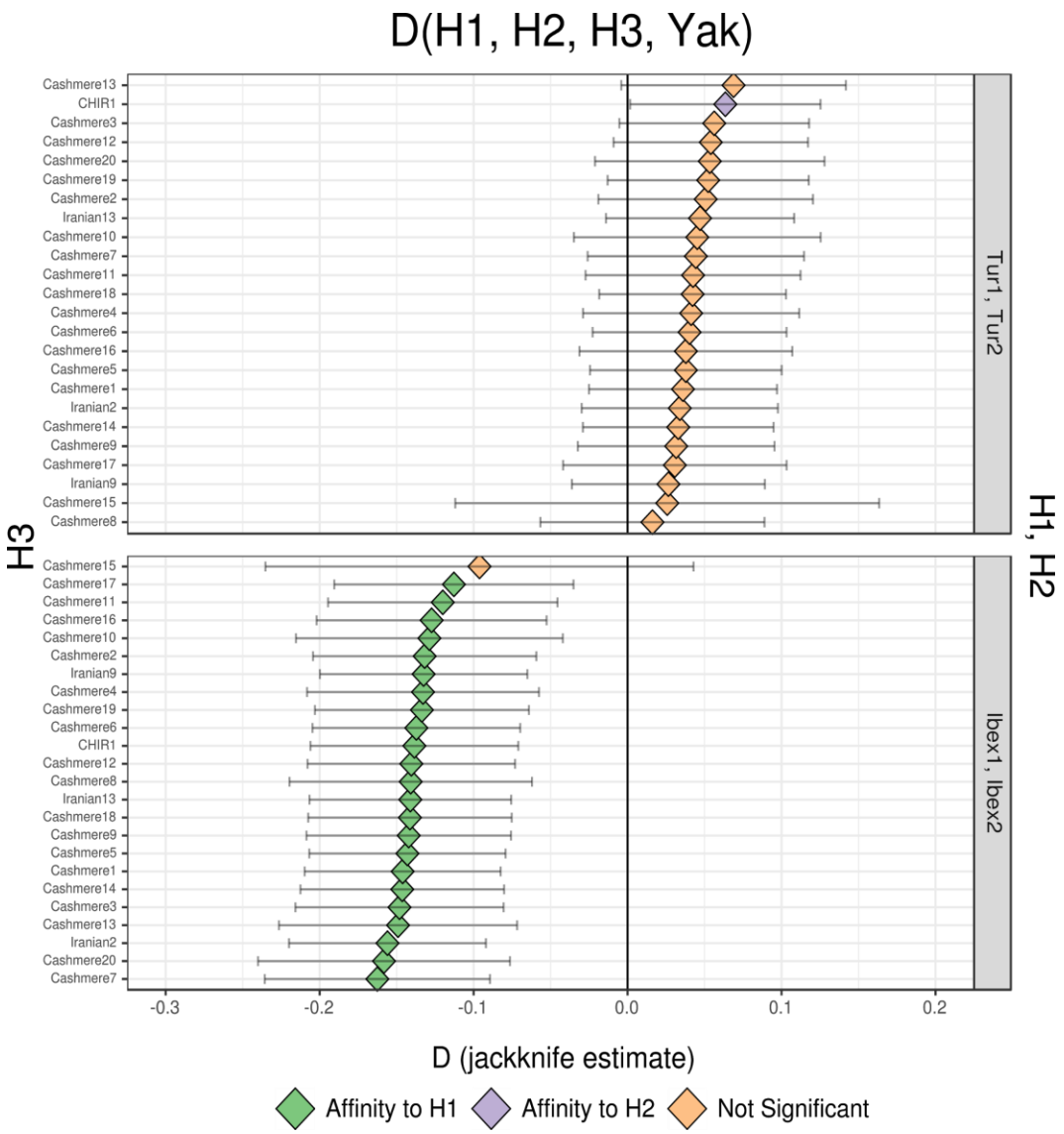
Appendix Figure 6.13 - *D* statistic test for admixture from *Capra sibirica* genomes (H3) into modern Asian goat (H2). Colour indicates Siberian ibex affinity (degree of derived allele sharing) to either Semnan3 or the modern Asian goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



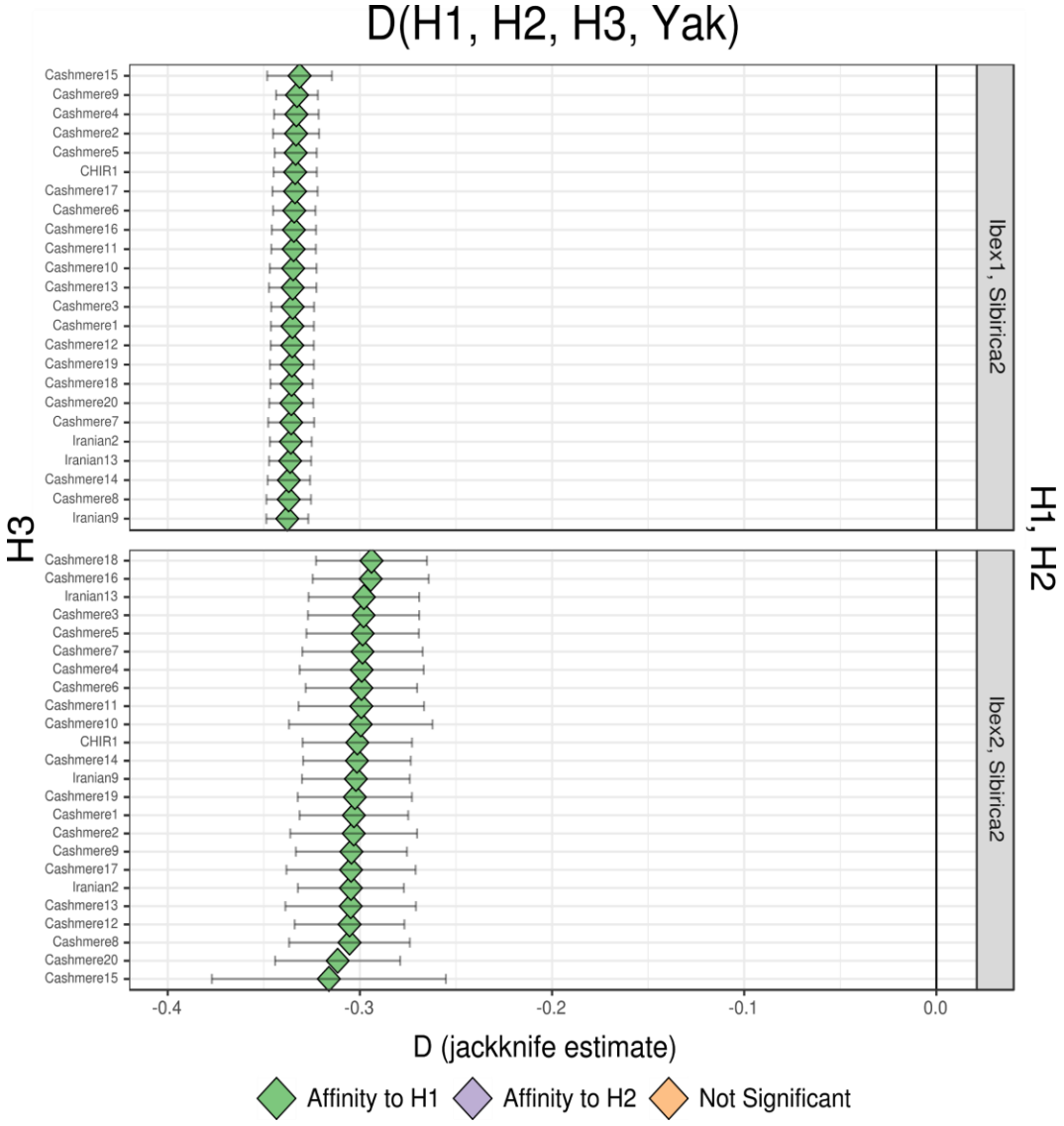
Appendix Figure 6.14 - *D* statistic test for admixture from Caucasian ibex/tur genomes (H3) into modern Asian goat (H2). Colour indicates tur affinity (degree of derived allele sharing) to either Semnan3 or the modern Asian goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



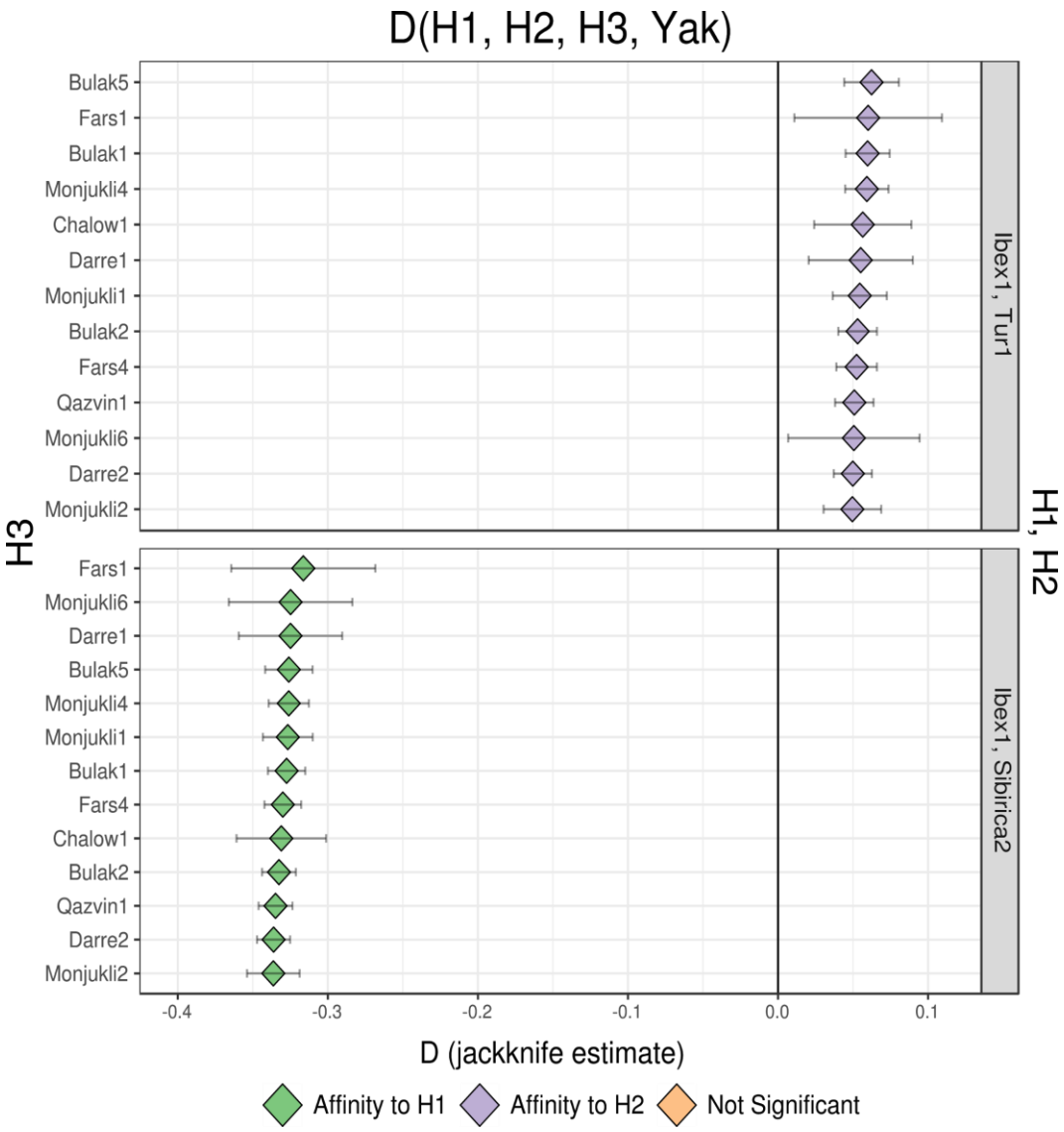
Appendix Figure 6.15 - *D* statistic for derived allele sharing of modern Asian goat genomes (H3) with the Alpine ibex Ibex2 or Caucasian ibex (H1, H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



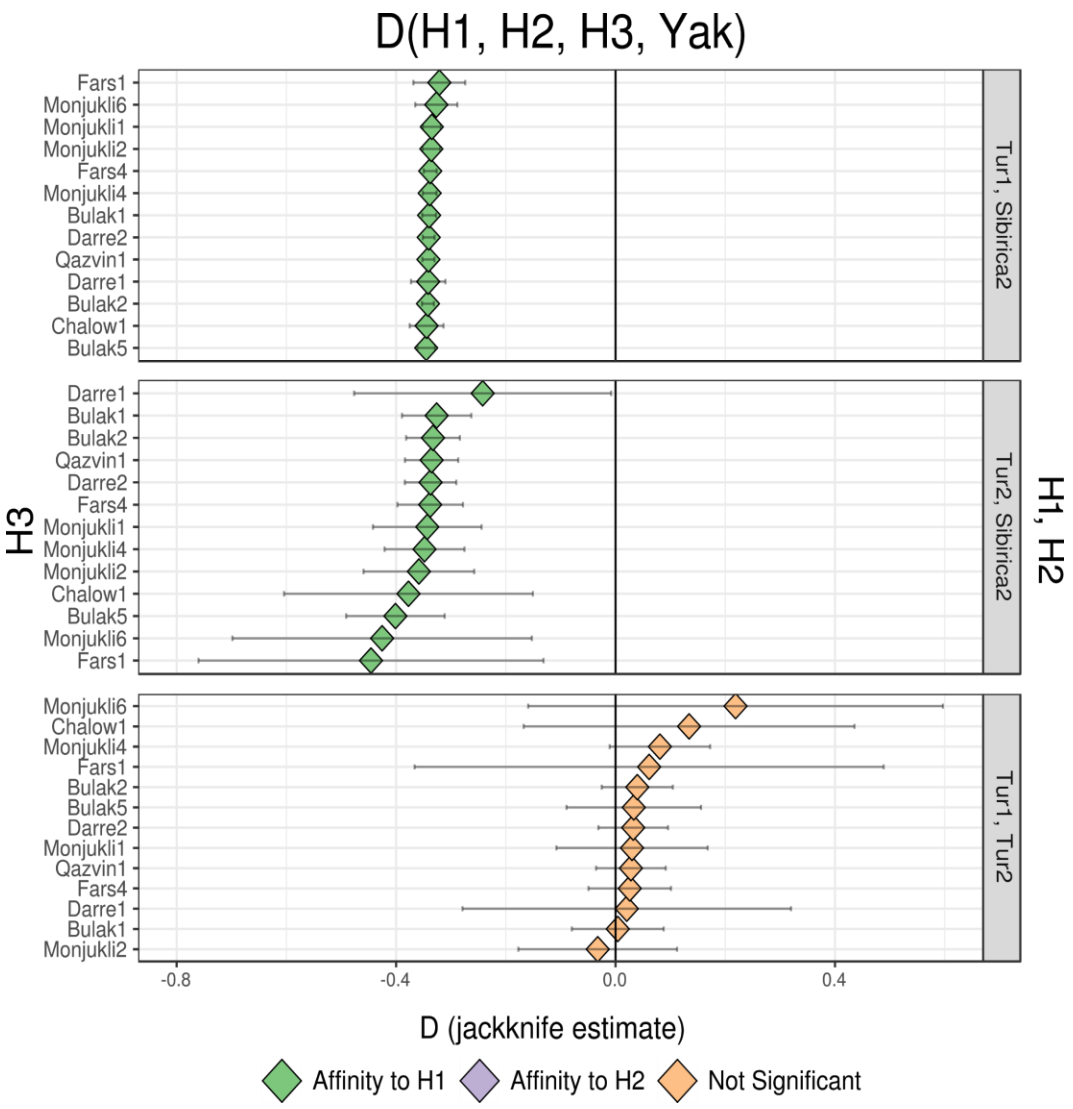
Appendix Figure 6.16 - *D* statistic for derived allele sharing of modern Asian goat genomes (H3) with either Caucasian ibex genomes (H1, H2), and with either Alpine ibex genomes (H1, H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



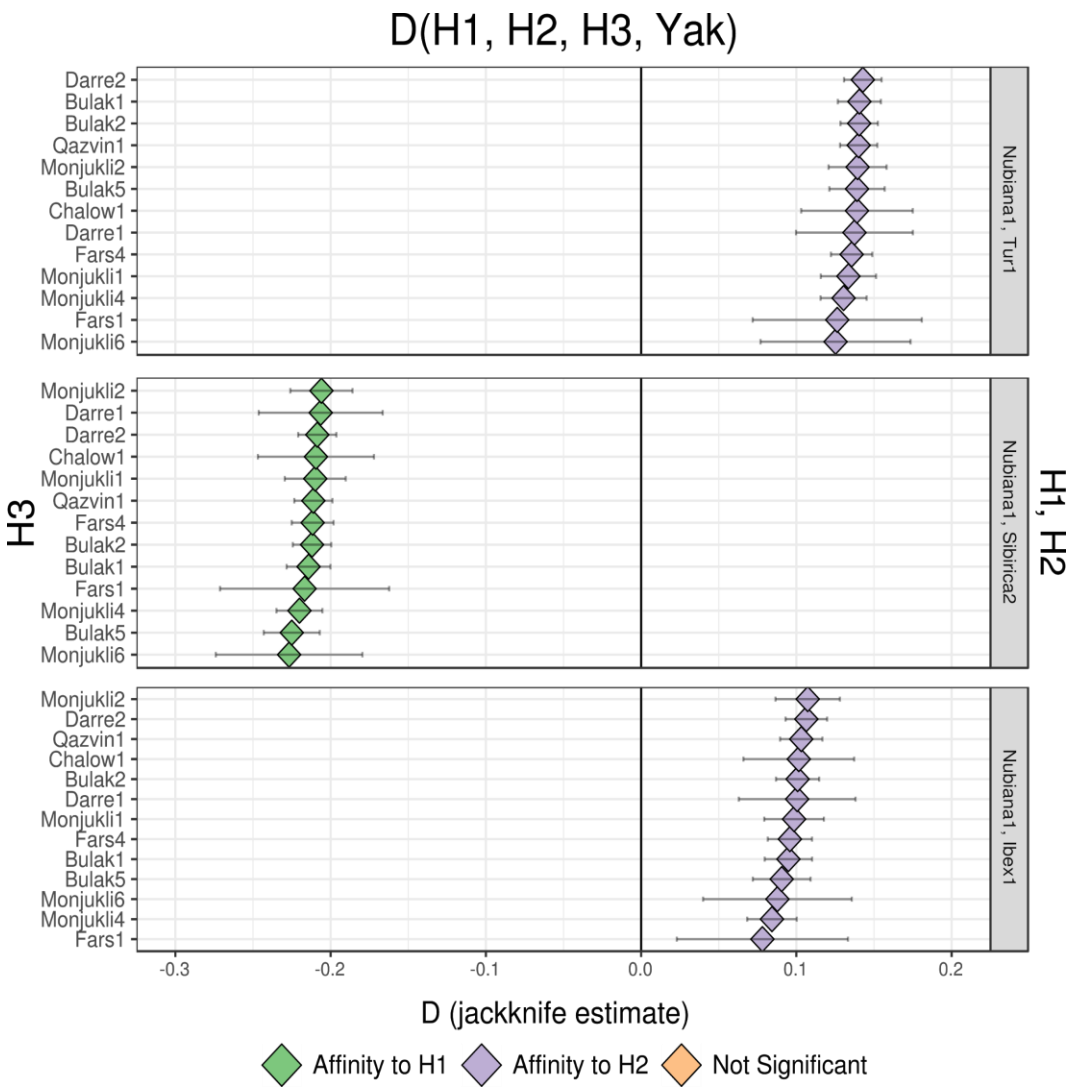
Appendix Figure 6.17 - *D* statistic for derived allele sharing of modern Asian goat genomes (H3) with the Siberian ibex Sibirica2 or Alpine ibex (H1, H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



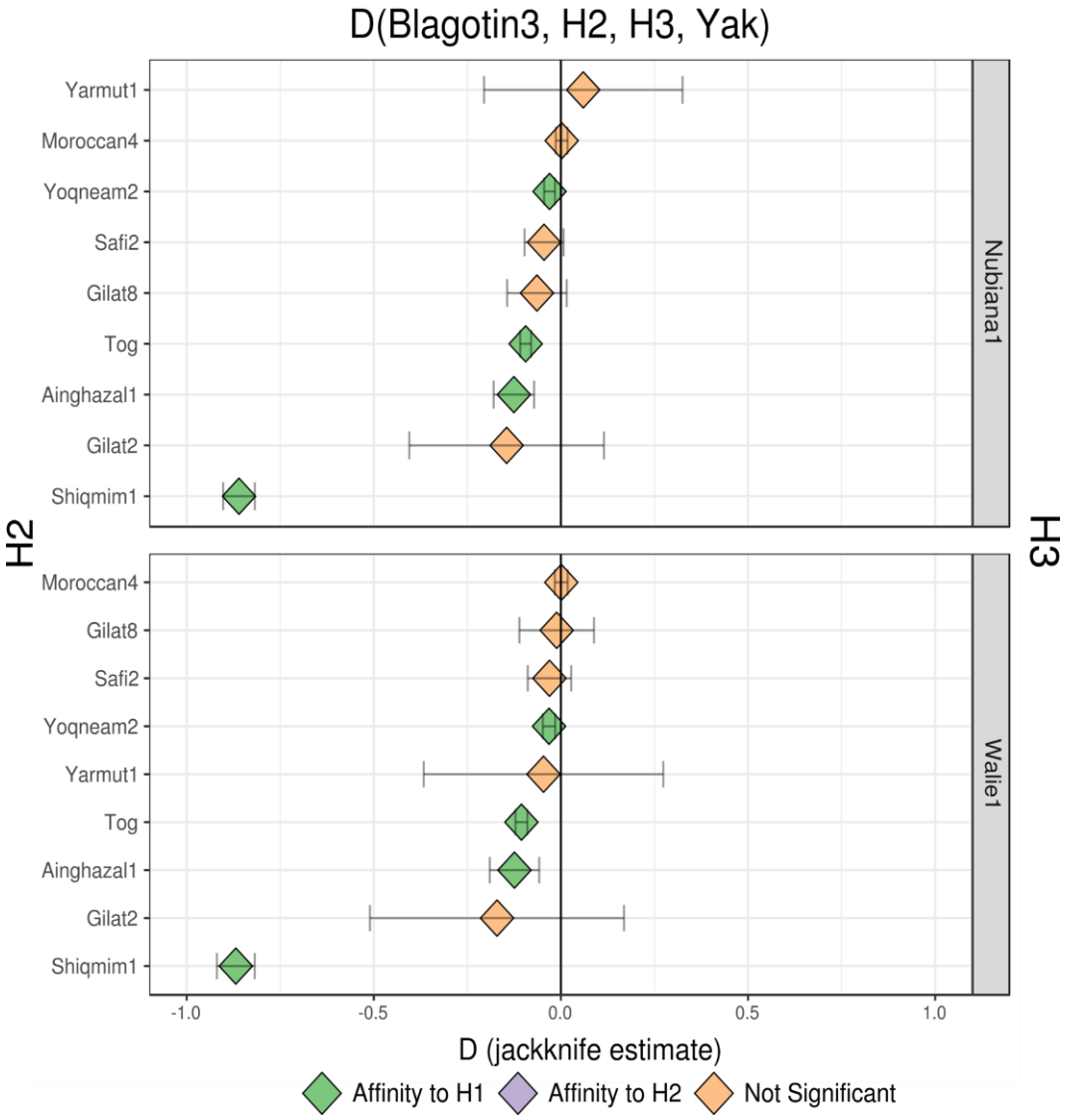
Appendix Figure 6.18 - *D* statistic for relative derived allele sharing of Post-Neolithic eastern goat genomes (H3) with either Ibex1 (H1) and Sibirica2 or Tur1 (H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



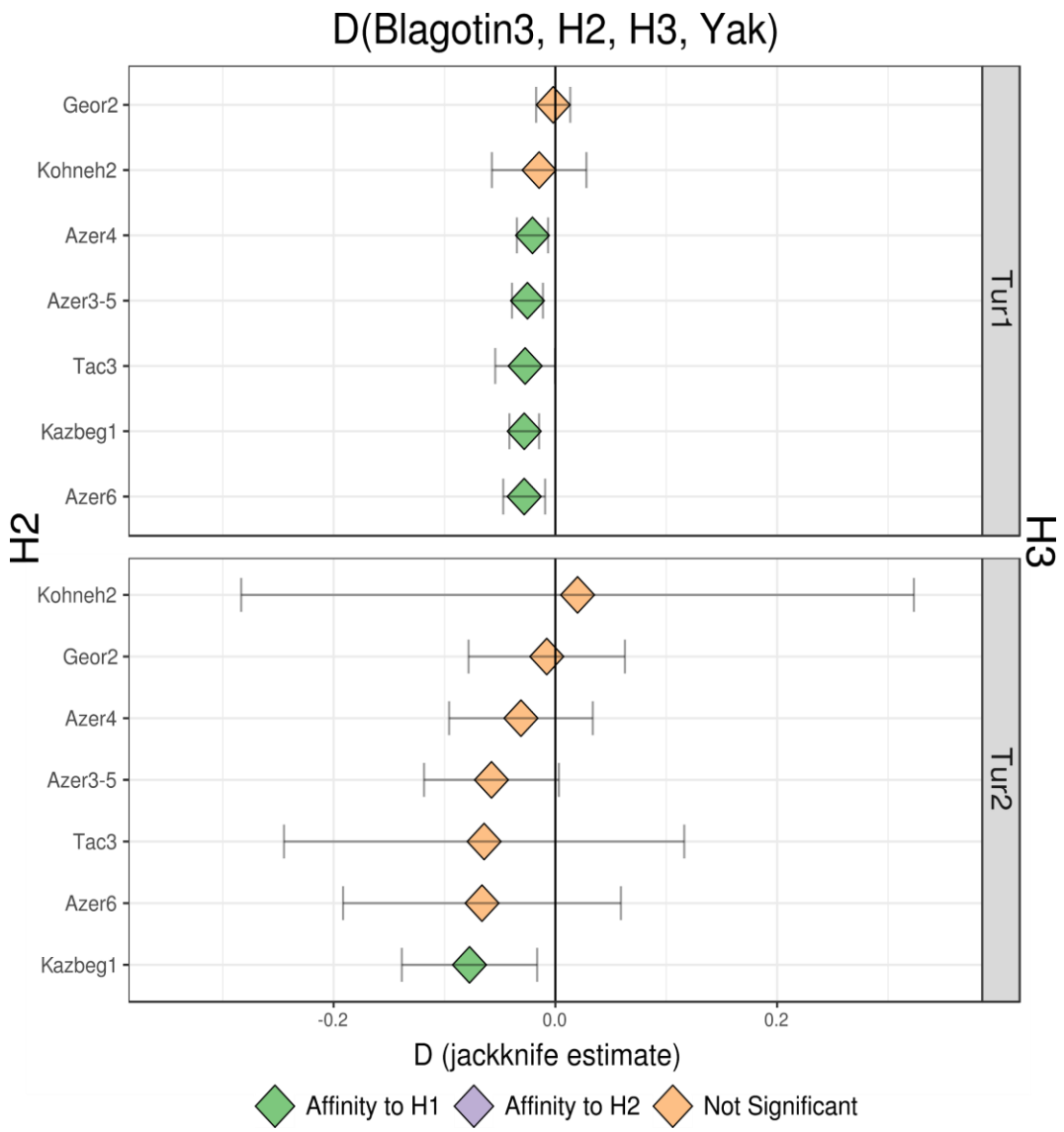
Appendix Figure 6.19 - *D* statistic for derived allele sharing of Post-Neolithic eastern goat genomes (H3) between pairs (H1, H2) of the three samples Sibirica2, Ibex1, or Ibex2. Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



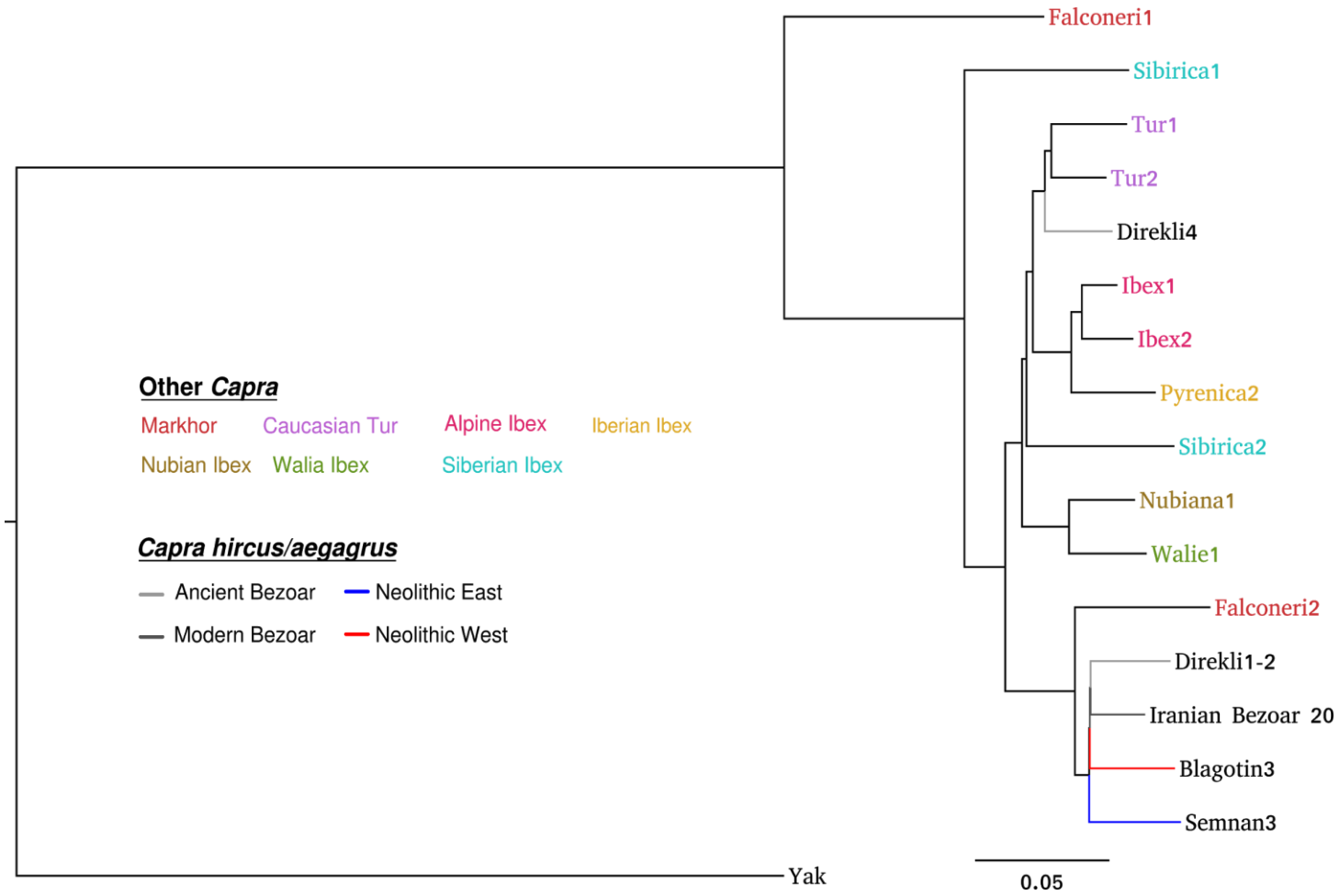
Appendix Figure 6.20 - *D* statistic for relative derived allele sharing of Post-Neolithic eastern goat genomes (H3) with either Nubiana1 (H1) and Sibirica2, Ibex1, or Tur1 (H2). Colour indicates modern Asian goat affinity (degree of derived allele sharing) to either H1 or H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



Appendix Figure 6.21 - *D* statistic test for admixture from the African ibex Nubiana 1 or Walie1 (H3) to ancient post-Neolithic Levantine goat (H2) relative to the Neolithic Serbian goat Blagotin3 (H1). Colour indicates African ibex affinity (degree of derived allele sharing) to either Blagotin3 H1 or the post-Neolithic Levantine goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



Appendix Figure 6.22 - *D* statistic test for admixture from Caucasian ibex/tur genomes (H3) into ancient Caucasian goat (H2) relative to the Serbian Neolithic Blagotin3 (H1). Colour indicates tur affinity (degree of derived allele sharing) to either Blagotin3 or the ancient Caucasian goat H2, with non-significance ($|Z| < 3$) indicated by the colour orange.



Appendix Figure 6.23 - Neighbour-Joining tree of IBS matrix of historic *Capra* samples and a reduced number of ancient and modern goat/bezoar.

Appendix Material for Chapter 7

Appendix Table 7.1 - All screened bone elements presented in Chapter 7. Samples are ordered by Internal Name. Genetic species was identified using Fastq Screen. Endogenous percentage was calculated as the number of reads aligning to the appropriate reference genome following MapQ 30 filtering, divided by the total number of reads following adaptor trimming. Samples with internal names beginning in AH or MV were screened by Andrew Hare or Marta Verdugo respectively.

Sample	Internal Name	Site	Region	Context/Date	Genetic Species	Skeletal Element	Collaborator	Endog. %age
David4	KD137	Jerusalem	Israel	Iron Age	Goat	Petrous	Liora Horwitz	18.15
Ulucak1	KD257	Ulucak Höyük	Turkey	Neolithic	-	Petrous	Marjan Mashkour	0.06
Ulucak2	KD258	Ulucak Höyük	Turkey	Neolithic	Pig	Petrous	Marjan Mashkour	56.26
Ulucak3	KD259	Ulucak Höyük	Turkey	Neolithic	Goat	Petrous	Marjan Mashkour	12.63
Ulucak4	KD260	Ulucak Höyük	Turkey	Neolithic	Sheep	Petrous	Marjan Mashkour	0.96
Aroer1	KD261	Tel'Aroer	Negev Desert, Israel	Iron Age - Hellenistic (Iron Age II)	Goat	Petrous	Liora Horwitz	0.45
Aroer2	KD262	Tel'Aroer	Negev Desert, Israel	Iron Age - Hellenistic (Iron Age II)	Goat	Petrous	Liora Horwitz	3.89
Nizzana1	KD263	Nizzana	Israel	Byzantine - Islamic	-	Petrous	Liora Horwitz	-
Nizzana2	KD264	Nizzana	Israel	Byzantine - Islamic	Goat	Petrous	Liora Horwitz	42.31
Nizzana3	KD265	Nizzana	Israel	Byzantine - Islamic	Goat	Petrous	Liora Horwitz	1.09
Nizzana4	KD266	Nizzana	Israel	Byzantine - Islamic	Sheep	Petrous	Liora Horwitz	44.4

Nizzana5	KD267	Nizzana	Israel	Byzantine - Islamic	Sheep	Petrous	Liora Horwitz	45.47
Sepphoris1	KD268	Sepphoris	Israel	Roman - Byzantine	Goat	Petrous	Liora Horwitz	65.51
Ashqel14	KD269	Tel Ashqelon	Israel	Late Hellenistic	Sheep	Petrous	Liora Horwitz	10.23
Yarmut8	KD270	Tel Yarmuth	Israel	Early Bronze Age	Sheep	Petrous	Liora Horwitz	0.44
Ulugh1	KD271	Ulugh Depe	Turkmenistan	Bronze Age	Goat	Humorous	Marjan Mashkour	21.25
Mushki1	KD272	Mushki	Iran	Neolithic (7450-7200 BP)	-	Humorous	Marjan Mashkour	-
Mianrud1	KD273	Mianrud	Iran	Likely Paleolithic	-	M3 in	Marjan Mashkour	-
Ghar1	KD274	Ghar e Cave	Iran	Paleolithic/Neolithic (9,000-8,000 BC)	-	Astragalus	Marjan Mashkour	-
Yafteh1	KD275	Yafteh Cave	Iran	Upper Paleolithic (35,000 - 20,000 BP)	-	Long Bone	Marjan Mashkour	-
Gorgan1	KD276	Gorgan Wall	Golestān, Iran	13/14th cent AD	Sheep	Petrous	Marjan Mashkour	0.02
Chenar1	KD277	Chenar Cave	Iran	Paleolithic/Early Holocene	-	Petrous	Marjan Mashkour	-
Chenar2	KD278	Chenar Cave	Iran	Paleolithic/Early Holocene	-	Petrous	Marjan Mashkour	-
Abdul1	KD279	Tepe Abdul Hosein	Luristan, Iran	Neolithic	Goat	Petrous	Marjan Mashkour	20.49
Abdul2	KD280	Tepe Abdul Hosein	Luristan, Iran	Neolithic	Goat	Petrous	Marjan Mashkour	24.03
Abdul3	KD281	Tepe Abdul Hosein	Luristan, Iran	Neolithic	-	Petrous	Marjan Mashkour	-
Abdul4	KD282	Tepe Abdul Hosein	Luristan, Iran	Neolithic	Goat	Petrous	Marjan Mashkour	18.30

Caucasus1	KD283	Tamara Fort	Kazbegi, Georgia	Medieval (15th-21st Cent AD)	East Caucasian Tur	Metacarpal	Marjan Mashkour	27.87
Caucasus2	KD284	Tamara Fort	Kazbegi, Georgia	Medieval (15th-21st Cent AD)	Goat/East Caucasian Tur	Tibia	Marjan Mashkour	0.87
Caucasus3	KD285	Tamara Fort	Kazbegi, Georgia	Medieval (15th-21st Cent AD)	Goat	Long bone	Marjan Mashkour	1.13
Montou1	KD286	Montou	France	Early Bronze Age (1900-1700 cal BC)	Sheep	Petrous	Jean-Dennis Vigne & Juliette Knockaert	27.58
Montou2	KD287	Montou	France	Early Bronze Age (1900-1700 cal BC)	-	Petrous	Jean-Dennis Vigne & Juliette Knockaert	-
Montou3	KD288	Montou	France	Early Bronze Age (1900-1700 cal BC)	Goat	Petrous	Jean-Dennis Vigne & Juliette Knockaert	30.83
Montou4	KD289	Montou	France	Late Bronze Age (1100-900 cal BC)	Goat	Petrous	Jean-Dennis Vigne & Juliette Knockaert	47.39
Atlityam3	KD290	Atlit Yam	Israel	PPNC	-	Petrous	Liora Horwitz	-
Zahara1	KD291	Tel Zahara	Israel	Roman-Islamic	Goat	Petrous	Liora Horwitz	0.97
Banias2	KD292	Banias	Israel	Roman-Islamic	Goat	Petrous	Liora Horwitz	10.15
Banias3	KD293	Banias	Israel	Roman-Islamic	-	Petrous	Liora Horwitz	-
Miqne7	KD294	Tel Miqne- Ekron	Shephelah, Israel	Iron Age	-	Petrous	Liora Horwitz	-
Miqne8	KD295	Tel Miqne- Ekron	Shephelah, Israel	Iron Age	-	Petrous	Liora Horwitz	-
Kortik7	KD296	Körtik Tepe	Turkey	~10,000 BC	-	Astragalus	Benjamin Arbuckle	-
Kortik8	KD297	Körtik Tepe	Turkey	~10,000 BC	-	Distal humerus	Benjamin Arbuckle	-
Kortik9	KD298	Körtik Tepe	Turkey	~10,000 BC	-	Astragalus	Benjamin Arbuckle	-
Kortik10	KD299	Körtik Tepe	Turkey	~10,000 BC	-	Astragalus	Benjamin Arbuckle	-
Kortik11	KD300	Körtik Tepe	Turkey	~10,000 BC	-	Distal tibia	Benjamin Arbuckle	-

Suberde1	KD301	Suberde	Turkey	7,500-7,000 BC	-	First phalanx	Benjamin Arbuckle	-
Suberde2	KD302	Suberde	Turkey	7,500-7,000 BC	-	Second phalanx	Benjamin Arbuckle	-
Suberde3	KD303	Suberde	Turkey	7,500-7,000 BC	-	Second phalanx	Benjamin Arbuckle	-
Erbaba1	KD304	Erbaba Höyük	Turkey	6,500-6,000 BC	-	Distal humerus	Benjamin Arbuckle	-
Erbaba2	KD305	Erbaba Höyük	Turkey	6,500-6,000 BC	Goat	Astragalus	Benjamin Arbuckle	0.66
Kosk1	KD306	Kosk Höyük	Turkey	Neolithic/Chalcolithic (~6,000 BC)	Sheep	Long bone shaft fragment	Benjamin Arbuckle	0.17
Kosk2	KD307	Kosk Höyük	Turkey	Neolithic/Chalcolithic(~6,000 BC)	-	Long bone shaft fragment	Benjamin Arbuckle	-
Gorgan2	AH077	Gorgan Wall	Iran	Sasanian, ~400 AD	Goat	Petrous	Marjan Mashkour	34.25
Abdul5	AH079	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	Goat	Petrous	Marjan Mashkour	3.09
Abdul6	AH080	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	Goat	Petrous	Marjan Mashkour	9.02
Abdul7	AH081	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	Goat	Petrous	Marjan Mashkour	0.78
Abdul8	AH082	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	Goat	Petrous	Marjan Mashkour	0.62
Abdul9	AH084	Tepe Abdul Hosein	Luristan, Iran	Early Neolithic	Goat	Petrous	Marjan Mashkour	2.62
Banias1	MV258	Banias	Israel	Roman-Islamic	Goat	Petrous	Liora Horwitz	14.18

Appendix Table 7.2 - Radiocarbon dating information for goat identified in Chapter 7. Two sigma calibration was performed using OxCal 4.3 (Bronk Ramsey 1994; Ramsey & Lee 2013) and IntCal 13 (Niu et al. 2013).

Sample	Internal Name	C14 Code	Conventional Age (BP)	Calibrated C14 date (95.4% Probability)
Ulucak3	KD259	UBA-38300	7362 +/- 45	6,366-6,094 cal BC
Erbaba2	KD305	UBA-38512	7592 +/- 48	6,567-6,379 cal BC
Abdul6	AH080	UBA-38301	8759 +/- 51	8,166-7,606 cal BC

Appendix Table 7.3 - Chapter 7 Illumina HiSeq sequencing, whole genome alignment results, and PCA groupings used. Filtered reads are following cutadapt trimming and minimum read length 30bp filter. Endogenous was calculated as reads aligned following mapQ 30 filter divided by the filtered read count. Samples were aligned to ARS1. Coverage statistics calculated after mapping quality 30 filter.

Sample	Raw reads	Trimmed reads	Aligned reads	Total rmdup reads	Aligned rmdup reads	q30 rmdup Aligned	Endog %age	Coverage	Population Grouping
Abdul1	71782023	67748178	13876836	67376655	13505313	9838427	14.52	0.14	Neolithic East
Abdul2	82929361	79490074	20147508	78420030	19077464	13110568	16.49	0.21	Neolithic East
Abdul4	66325680	63125471	13194166	62520983	12589678	9105122	14.42	0.13	Neolithic East
Abdul6	55217711	53981014	5101880	53639071	4759937	3877434	7.18	0.07	Neolithic East
Banias1	68768928	64763525	9370220	64245480	8852175	7229222	11.16	0.14	Iron Age /Medieval Israel
Caucasus1	67233165	66236781	39705627	35186807	27532563	18461811	27.87	0.49	Iron Age /Medieval Caucasus
Gorgan2	30093055	29049599	12444246	26390951	9785598	8529205	29.36	0.16	Iron Age /Medieval Iran
Montou3	77466006	73388728	66141190	70767198	63519660	46806015	63.78	0.88	Bronze Age France
Montou4	132051972	127505752	122357477	123719913	118571638	92158033	72.28	1.91	Bronze Age France
Nizzana2	91214489	81682264	49439005	76823058	44579799	29536644	36.16	0.46	Iron Age/ Medieval Israel
Sepphoris1	114632406	110475747	104522082	106701132	100747467	73278142	66.33	1.47	Iron Age /Medieval Israel
Ulucak3	224565704	216201379	35858838	213598161	33255620	23151881	10.71	0.34	Neolithic West
Ulugh1	77249579	74050070	20573369	72312560	18835859	12914417	17.44	0.25	Bronze Age Turkmenistan
David4	327102091	313344205	91536299	67140437	67140184	60349922	19.26	1.02	Iron Age /Medieval Israel

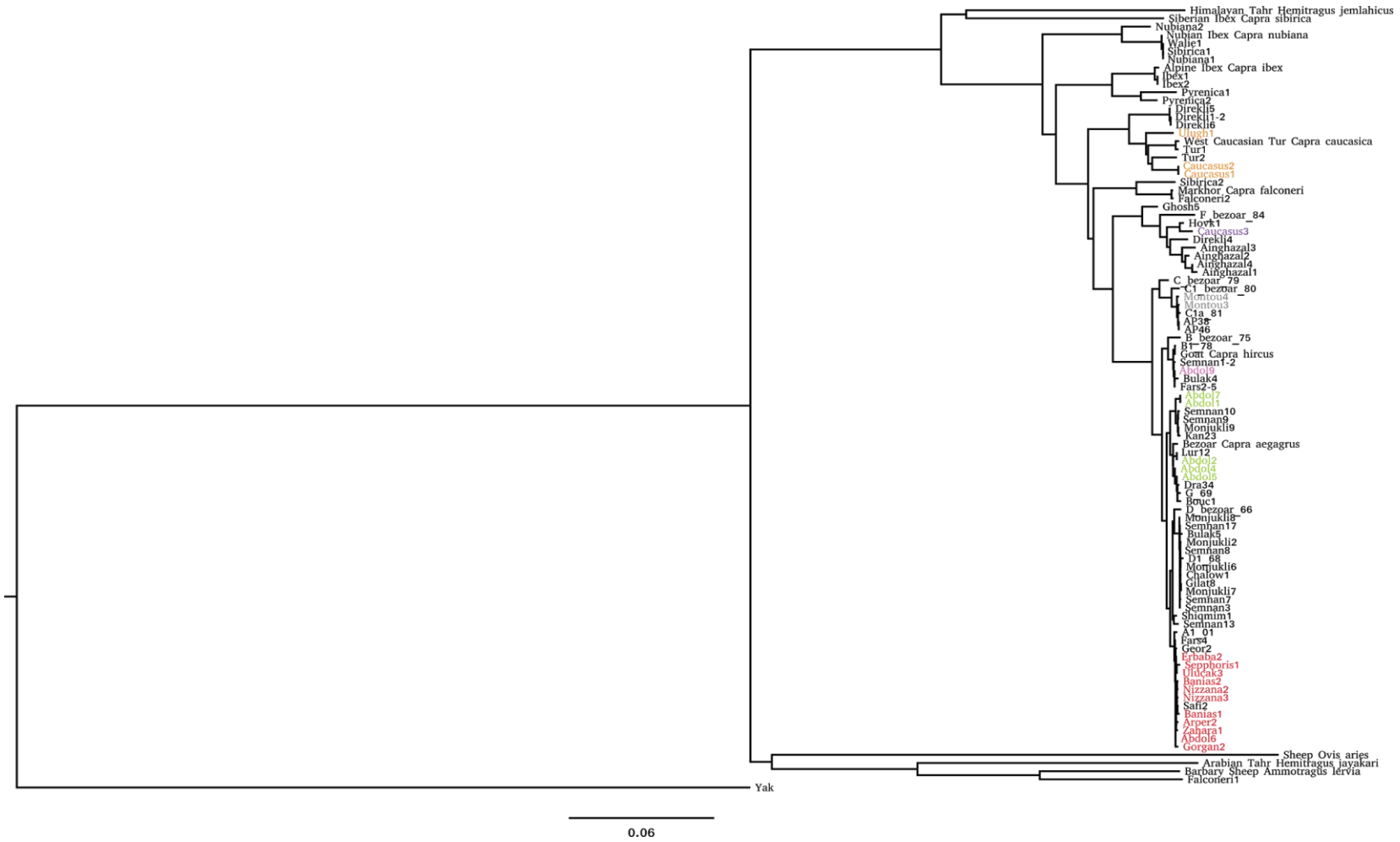
Appendix Table 7.4 - Chapter 7 mitochondrial capture sequencing and mtDNA alignment results. Alignment statistics presented following realigned to a closer reference sequence as defined in this table.

Sample	Raw Reads	Filtered Reads	Aligned Reads	Coverage	Called Sites	mtDNA hap.	Reference Used	Accession
Abdul1	71782023	67748178	14665	47.91	16625	G'	G_69	KR059213.1
Abdul2	82929361	79490074	25983	87.82	16641	G	G_69	KR059213.1
Abdul4	66325680	63125471	20898	63.52	16641	G	G_69	KR059213.1
Abdul5	4892347	4732487	918	3.36	10760	G	G_69	KR059213.1
Abdul6	55217711	53981014	12537	47.12	16642	A	A1_01	KR059146.1
Abdul7	5390383	4134491	1079	4.00	12365	G'	G_69	KR059213.1
Abdul9	6101653	5949151	1954	6.84	16158	B	Goat Reference	NC_005044.2
Aroer2	977628	940505	20384	64.31	16632	A	A1_01	KR059146.1
Banias1	68768928	64763525	8978	33.25	16642	A	A1_01	KR059146.1
Banias2	939084	924602	31292	118.04	16632	A	A1_01	KR059146.1
Caucasus1	30093055	29049599	12550	64.63	16398	T	West Caucasian Tur	NC_020683.1
Caucasus2	4273753	4172419	104027	379.71	16465	T	West Caucasian Tur	NC_020683.1
Caucasus3	3007832	2963763	127308	464.51	16534	F	F_bezoar_84	KR059226.1
David4	327102091	313344205	96730	245.69	16561	A	A1_01	KR059146.1
Erbaba2	4956156	4880170	111394	394.13	16642	A	A1_01	KR059146.1
Gorgan2	67233165	66236781	16984	71.07	16642	A	A1_01	KR059146.1
Montou3	77466006	73388728	35367	147.96	16638	C	C1a_81	KR059223.1

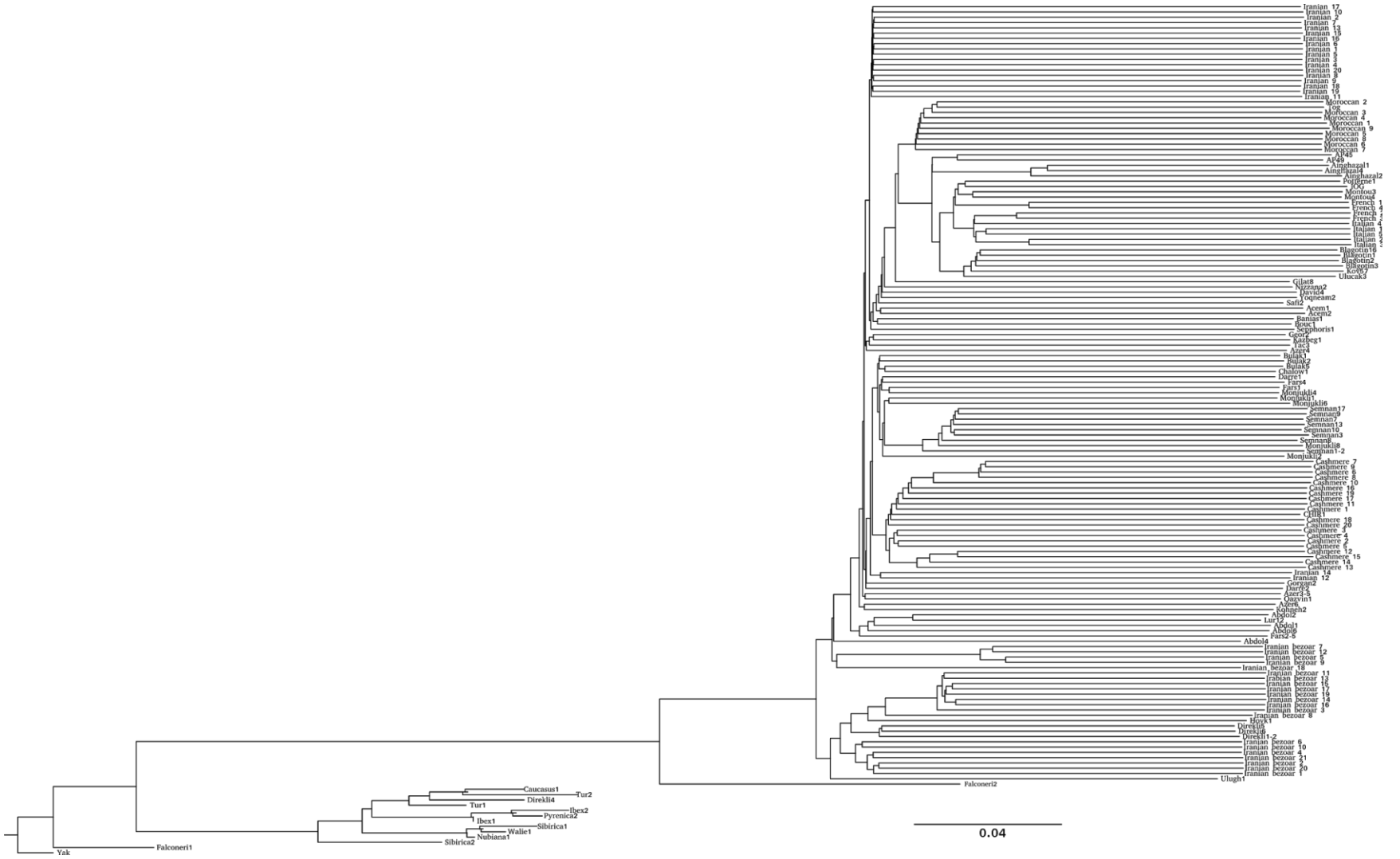
Montou4	132051972	127505752	38897	155.34	16638	C	C1a_81	KR059223.1
Nizzana2	91214489	81682264	48863	169.55	16642	A	A1_01	KR059146.1
Nizzana3	2027702	2008006	15872	46.10	16577	A	A1_01	KR059146.1
Sepphoris1	114632406	110475747	40862	154.39	16642	A	A1_01	KR059146.1
Ulucak3	224565704	216201379	45501	125.03	16642	A	A1_01	KR059146.1
Ulugh1	77249579	74050070	4690	14.08	16200	T	West Caucasian Tur	NC_020683.1
Zahara1	3953597	3856623	111227	377.66	16642	A	A1_01	KR059146.1

Appendix Table 7.5 - Chapter 7 archaeological sites which yielded *Capra* DNA.

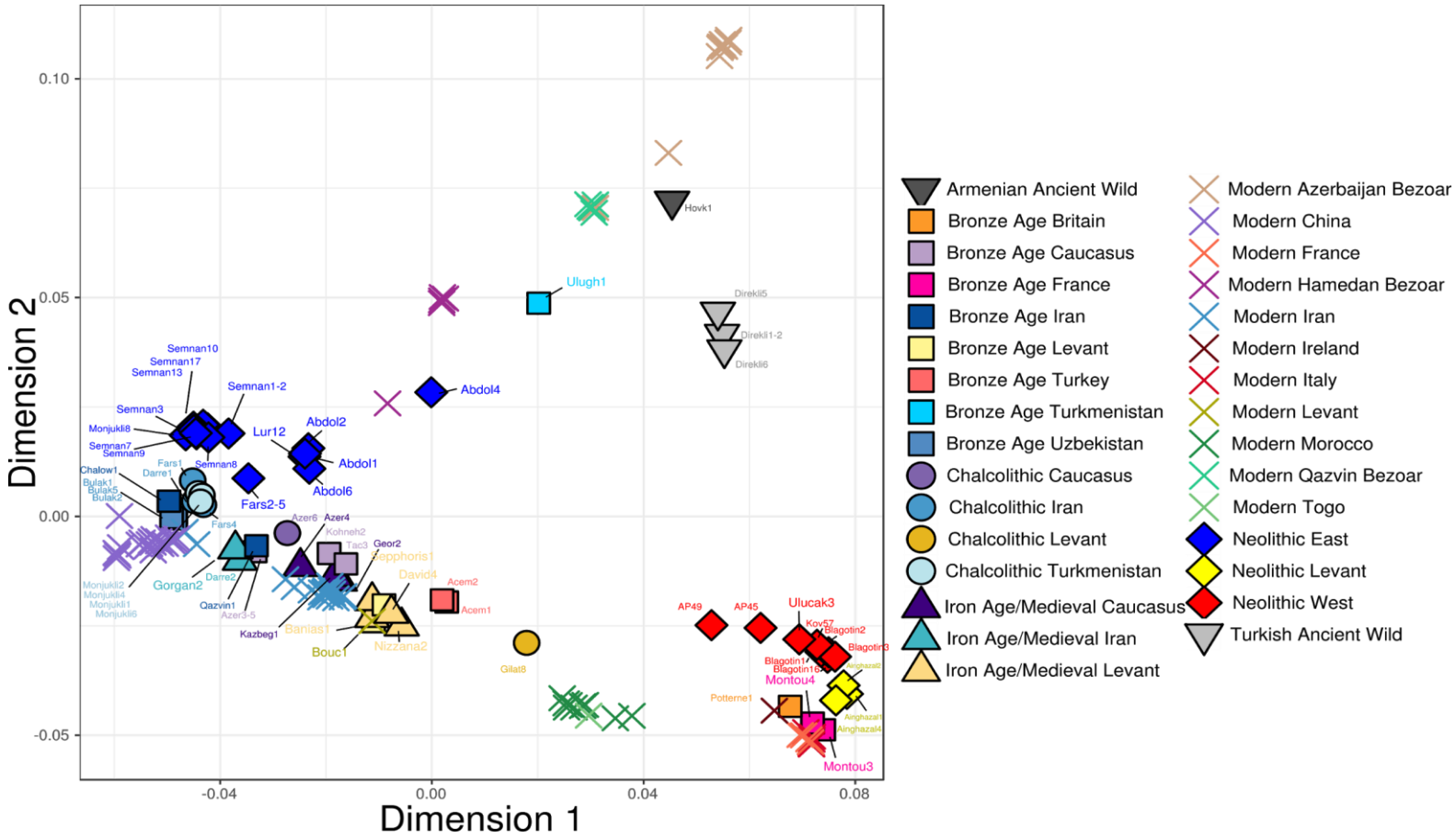
Site Name	Location	Latitude	Longitude	Period	Map Label
Erbaba Höyük	Turkey	37.938516	31.506575	Neolithic	30
Gorgan Wall	Golestān, Iran	37.127493	54.189516	Sasanian	31
Ulugh Depe	Turkmenistan	37.15479	60.029511	Bronze Age	15B
Tel'Aroer	Negev Desert, Israel	31.153678	34.984919	Iron Age - Hellenistic	28E
Banias	Israel	33.248611	35.694444	Roman - Islamic	28F
Jerusalem	Israel	31.774702	35.235637	Iron Age	28G
Nizzana	Israel	30.876119	34.432833	Byzantine - Islamic	28H
Sepphoris	Israel	32.745556	35.278611	Roman - Byzantine	28I
Tel Zahara	Israel	32.497421	35.474498	Roman - Islamic	28J
Montou	France	42.652778	2.68	Bronze Age	32



Appendix Figure 7.1 - Maximum Likelihood phylogeny of ancient and domestic *Capra* mtDNA, including Chapter 7 sequences. Newly sequenced samples are indicated by coloured names. Rooted on Yak.



Appendix Figure 7.2 - Neighbour-Joining tree of IBS matrix of all *Capra* genomes >0.01× mean coverage used in this thesis, rooted on Yak.



Appendix Figure 7.3 - IBS-MDS plot of all ancient and modern goat/bezoar, with all ancients labelled. Population labels used are those defined in Appendix Tables 4.1, 4.2, 6.2, and 7.3.

Appendix Text 7.1 - Multidimensional scaling of the chapter 7 IBS matrix.

Multidimensional scaling was performed on the IBS matrix generated in Chapter 7, following the removal of all genomes which showed greater affinity with non *Capra aegagrus/hircus* genomes: all Chapter 6 historic genomes except Bouc1; Direkli4; Caucasus1. The resulting IBS matrix was converted to a distance matrix using the R function as.dist(), and multidimensional scaling performed on the distance using the R function cmdscale(), default settings.

Appendix References

Alberto, F.J. et al., 2018a. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

Alberto, F.J. et al., 2018b. Convergent genomic signatures of domestication in sheep and goats. *Nature communications*, 9(1), p.813.

Benjelloun, B. et al., 2015. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (Capra hircus) using WGS data. *Frontiers in genetics*, 6, p.107.

Daly, K.G. et al., 2018. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science*, 361(6397), pp.85–88.

Li, X. et al., 2017. Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. *Scientific reports*, 7(1), p.15142.

Qiu, Q. et al., 2012. The yak genome and adaptation to life at high altitude. *Nature genetics*, 44(8), pp.946–949.