UAB
Universitat Autònoma de Barcelona

UNIVERSITY OF SASSARI

Department of Veterinary Medicine

PhD School in Veterinary Science

Cycle: XXXV

Director: Prof. Fiammetta Berlinguer

UNIVERSITAT AUTÒNOMA DE BARCELONA

Departament de Ciència Animal i dels Aliments, Facultat de Veterinària

PhD School in Animal Production

Director: Joaquín Casellas Vidal

CENTRE DE RECERCA EN AGRIGENÒMICA

Plant and Animals Genomics Program

Analyzing the impact of migration on the genetic diversity of cattle and goat breeds

Supervisor:
candidate:

Prof. Maria Luisa Dettori
Prof. Marcel Amills Eras

PhD

Elena Petretto

Academic year 2021/2022

Contents

Abstract

The aim of this thesis is to study the genetic diversity of goat and cattle populations, especially from the Mediterranean basin, to understand the impact of migration on the gene pools of these species. In **study 1**, we have investigated whether the post-domestication dispersal of goats in Europe and Africa left a genetic signature that can be recognized nowadays. To do so, we have retrieved high throughput genotyping data from 1,148 European and 1,187 African goats. In Europe, we have found a partial regional differentiation of breeds, with the Northern European breeds, and the British-Irish populations being clearly separated from the remaining European populations. In Africa, in contrast, we have identified five main clusters which coincide with the northern, southern, eastern, and western parts of the continent plus the island of Madagascar. Next, we have investigated whether the observed and expected heterozygosities of these goat populations correlate with distance between their spatial location and the Neolithic domestication center in Southeaster Anatolia. In doing so, we have considered datasets with and without insular populations because the diversity of insular populations can be greatly affected by geographic isolation, a process unrelated to post domestication dispersal. In fact, in European breeds we detected a negative correlation between distance and diversity only when insular populations were taken into consideration. In African breeds, on the other hand, significant negative correlations were found in both datasets (with and without insular populations). These results might indicate that the maritime diffusion of the Neolithic caprine stock was more relevant in Europe than in Africa, where post-domestication dispersal took place fundamentally overland. Moreover, geographic and biological barriers might have significantly restricted gene flow between African populations.

In **studies 2 and 3**, we have investigated the genetic consequences of the African introgression of Mediterranean cattle and goat breeds. More specifically, in **study 2**, we aimed to determine the origins and magnitude of the African introgression of the Murciano-Granadina dairy breed, which is mainly distributed in Southern Spain. To do so, we have genotyped 500 Murciano Granadina goats with the Goat SNP50BeadChip and we have performed supervised ADMIXTURE analysis using as reference populations (i) A Spanish breed (Bermeya), (ii) A Swiss cosmopolitan breed (Saanen) and (iii) North African breeds: Moroccan (Barcha, Draa, Ghazalia, Noire de Atlas, Nord, Moroccan), Egyptian (Barki, Oasis, Saidi), Algerian (Arabia, Makatia, M'Zabite, Kabyle) and Sudanese (Desert, Nilotic, Taggar). We have found that the African introgression of Murciano-Granadina goats has its roots in Morocco, with an average percentage of about 4% and a range between 0-12%. In contrast, goats from other Northern

African countries did not contribute alleles to Murciano-Granadina goats in a significant way. The presence of a Moroccan genetic component in Murciano-Granadina goats could be due to the conquest of Spain by Berber troops, originally from the Maghreb, who stayed in the Iberian Peninsula for eight centuries. However, other alternative historical scenarios making possible the gene flow between Moroccan and Spanish goat populations could be also envisioned.

In **study 3**, we were interested in characterizing the mitochondrial variation of local Sardinian cattle breeds (Sarda, Sardo Bruna and Sardo Modicana) to determine its magnitude as well as the presence of an African genetic signature. By doing so, we have observed a significant genetic differentiation between the three Sardinian breeds. Moreover, by constructing a Median Joining Network based on sequences from Sardinian cattle, a typical 'star-like' conformation likely produced by a founder effect was observed. In order to identify the main haplogroups of Sardinian bovine breeds and their relationship with other cattle breeds with a worldwide distribution, we made a second Median Joining Network including mitochondrial sequences from cattle of African, Italian, Spanish and Asian origin. This analysis revealed the presence of the T3 haplogroup (typically European) in the Sarda breed, and of the T1 and T1'2'3' haplogroups in the Sarda and Sardo Bruna breeds. The identification in Sardinian cattle of haplogroup T1, characteristic of African taurine breeds, may reflect the occurrence of an African introgression event. Such event could have occurred through direct exchanges of bovines between North Africa and Sardinia or indirectly through trade with Mediterranean populations already introgressed with African cattle. Such scenario is plausible given that Sardinia is located at the crossroads of many maritime trading routes traversing the Mediterranean Sea.

Resum

L'objectiu d'aquesta tesi és estudiar la diversitat genètica de les poblacions caprines i bovines, especialment de la conca mediterrània, per entendre l'impacte de la migració en el patrimoni genètic d'aquestes espècies. En l'**estudi 1**, hem investigat si la dispersió post-domesticació de cabres a Europa i Àfrica va deixar una empremta genètica que avui dia es pugui reconèixer. Per fer-ho, hem recuperat dades de genotipat d'alt rendiment de 1.148 cabres europees i 1.187 africanes. A Europa, hem trobat una diferenciació regional parcial de les races, sent evident que les races del nord d'Europa, i les poblacions britànico-irlandeses estan clarament separades de la resta de poblacions europees. A l'Àfrica, en canvi, hem identificat cinc grups principals que coincideixen amb les zones nord, sud, est i oest del continent i a més l'illa de Madagascar. A continuació, hem investigat si les heterozigositats observades i esperades d'aquestes poblacions de cabres es correlacionen amb la distància entre la seva localització geogràfica i el centre de domesticació neolític al sud-est d'Anatòlia. Per fer-ho, hem considerat conjunts de dades amb i sense poblacions insulars perquè la diversitat de poblacions insulars es pot veure molt afectada per l'aïllament geogràfic, un procés no relacionat amb la dispersió posterior a la domesticació. De fet, en les races europees vam detectar una correlació negativa entre distància i diversitat només quan es van tenir en compte les poblacions insulars. En les races africanes, en canvi, es van trobar correlacions negatives significatives en ambdós conjunts de dades (amb i sense poblacions insulars). Aquests resultats podrien indicar que la difusió marítima de les poblacions de caprins neolítics va ser més rellevant a Europa que a l'Àfrica, on la dispersió posterior a la domesticació va tenir lloc fonamentalment per terra. A més, les barreres geogràfiques i biològiques podrien haver restringit significativament el flux gènic entre poblacions africanes. En els **estudis 2** i **3**, hem investigat les conseqüències genètiques de la introgressió africana de les races bovines i caprines mediterrànies. Més concretament, en l'**estudi 2** ens vam proposar determinar l'origen i la magnitud de la introgressió africana de la raça lletera Murciano-Granadina, que es distribueix principalment al sud d'Espanya. Per fer-ho, hem genotipat 500 cabres d'aquesta raça amb el xip SNP50Beadchip i hem realitzat anàlisis supervisades de barreja poblacional utilitzant com a poblacions de referència (i) una raça espanyola (Bermeya), (ii) una raça cosmopolita suïssa (Saanen) i (iii) races nord-africanes: marroquines (Barcha, Draa, Ghazalia, Noire de Atlas, Nord, marroquines), egípcies (Barki, Oasis, Saidi), algerianes (Aràbia, Makatia, M'Zabite, Kabyle) i sudanesos (desert, nilòtic, taggar). Hem comprovat que la introgressió africana de cabres Murciano-Granadines té les seves arrels al Marroc, amb un percentatge mitjà de prop del 4% i un rang entre el 0-12%. En canvi, les cabres d'altres països

del nord d'Àfrica no van aportar al·lels a les cabres Murciano-Granadines de manera significativa. La presència d'un component genètic marroquí en les cabres Murciano-Granadines podria ser deguda a la conquesta d'Espanya per les tropes berbers, originàries del Magrib, que van romandre a la península Ibèrica durant vuit segles. No obstant això, també podrien estar implicats altres escenaris històrics alternatius que hagin possibilitat el flux gènic entre les poblacions de cabres marroquines i espanyoles. En l'**estudi 3**, ens interessava caracteritzar la variació mitocondrial de les races bovines locals de Sardenya (Sarda, Sardo Bruna i Sardo Modicana) per determinar-ne la magnitud així com la presència d'una signatura genètica africana. D'aquesta manera, hem observat una diferenciació genètica significativa entre les tres races sardes. A més, mitjançant la construcció d'una *median-joining network* basada en seqüències de bestiar sard, es va observar una conformació típica "semblant a una estrella" probablement produïda per un efecte fundador. Per tal d'identificar els principals haplogrups de races bovines de Sardenya i la seva relació amb altres races bovines de distribució mundial, vam realitzar una segona *median-joining network* incloent seqüències mitocondrials de bestiar boví d'origen africà, italià, espanyol i asiàtic. Aquesta anàlisi va revelar la presència de l'haplogrup T3 (típicament europeu) en la raça Sarda, i dels haplogrups T1 i T1'2'3' en les races Sardà i Sardo Bruna. La identificació en el bestiar sard de l'haplogrup T1, característic de les races taurines africanes, pot reflectir l'ocurrència d'un esdeveniment d'introgressió africana. Aquest succés podria haver-se produït a través d'intercanvis directes de bovins entre el nord d'Àfrica i Sardenya o indirectament a través del comerç amb poblacions mediterrànies ja introgressades amb bestiar africà. Aquest escenari és plausible, atès que Sardenya es troba a la cruïlla de moltes rutes comercials marítimes que travessen el mar Mediterrani.

Resumen

El objetivo de esta tesis es estudiar la diversidad genética de poblaciones caprinas y bovinas, especialmente de la zona mediterránea, para comprender el impacto de la migración en los acervos genéticos de estas especies. En el **estudio 1**, hemos investigado si la dispersión post domesticación de las cabras en Europa y África dejó una firma genética reconocible actualmente. Para ello, hemos recuperado datos de genotipado de alto rendimiento de 1.148 cabras europeas y 1.187 africanas. En Europa, hemos detectado una diferenciación regional parcial de las razas, estando las razas del norte de Europa y las poblaciones británico-irlandesas claramente separadas del resto de poblaciones europeas. En África, en cambio, hemos identificado cinco grupos principales que coinciden con las zonas septentrional, meridional, oriental y occidental del continente, más la isla de Madagascar. Seguidamente, hemos investigado si las heterocigosidades observadas y esperadas de estas poblaciones caprinas se correlacionan con la distancia entre su ubicación espacial y el centro de domesticación neolítico en el sudeste de Anatolia. De este modo, hemos considerado conjuntos de datos con y sin poblaciones insulares, porque la diversidad de las poblaciones insulares puede verse muy afectada por el aislamiento geográfico, un proceso no relacionado con la dispersión posterior a la domesticación. De hecho, en las razas europeas detectamos una correlación negativa entre distancia y diversidad sólo cuando se tuvieron en cuenta las poblaciones insulares. En las razas africanas, en cambio, se encontraron correlaciones negativas significativas en ambos conjuntos de datos (con y sin poblaciones insulares). Estos resultados podrían indicar que la difusión marítima de la estirpe caprina neolítica fue más relevante en Europa que en África, donde la dispersión post-domesticación tuvo lugar fundamentalmente por vía terrestre. Además, las barreras geográficas y biológicas podrían haber restringido significativamente el flujo genético entre las poblaciones africanas. En los **estudios 2** y **3** hemos investigado las consecuencias genéticas de la introgresión africana de razas bovinas y caprinas mediterráneas. Más concretamente, en el **estudio 2**, hemos tenido como objetivo determinar los orígenes y la magnitud de la introgresión africana de la raza lechera Murciano-Granadina, distribuida principalmente en el sur de España. Para ello, hemos genotipado 500 cabras Murciano-Granadinas con el Goat SNP50BeadChip y hemos realizado un análisis ADMIXTURE supervisado utilizando como poblaciones de referencia (i) una raza española (Bermeya), (ii) una raza cosmopolita suiza (Saanen) y (iii) razas norteafricanas: Marroquí (Barcha, Draa, Ghazalia, Noire de Atlas, Nord, Moroccan), Egipcia (Barki, Oasis, Saidi), Argelina (Arabia, Makatia, M'Zabite, Kabyle) y Sudanesa (Desert, Nilotic, Taggar). Hemos constatado que la

introgresión africana de la cabra Murciano-Granadina tiene sus raíces en Marruecos, con un porcentaje medio de alrededor del 4% y un rango entre el 0 y el 12%. En cambio, las cabras originarias de otros países del norte de África no aportaron alelos a las cabras Murciano-Granadinas de forma significativa. La presencia de un componente genético marroquí en las cabras Murciano-Granadinas podría deberse a la conquista de España por tropas bereberes, originarias del Magreb, que permanecieron en la Península Ibérica durante ocho siglos. Sin embargo, también podrían estar implicados otros escenarios históricos alternativos que hicieran posible el flujo genético entre poblaciones caprinas marroquíes y españolas.

En el **estudio 3**, nos interesaba caracterizar la variación mitocondrial de las razas bovinas sardas locales (Sarda, Sardo Bruna y Sardo Modicana) para determinar su magnitud, así como la presencia de una firma genética africana. Al hacerlo, hemos observado una diferenciación genética significativa entre las tres razas sardas. Además, al construir una Median Joining Network basada en las secuencias del ganado sardo, se observó una conformación típica " estrellada", probablemente producida por un efecto fundador. Con el fin de identificar los principales haplogrupos de las razas bovinas sardas y su relación con otras razas bovinas de distribución mundial, realizamos una segunda Median Joining Network incluyendo secuencias mitocondriales de bovinos de origen africano, italiano, español y asiático. Este análisis reveló la presencia del haplogrupo T3 (típicamente europeo) en la raza Sarda, y de los haplogrupos T1 y T1'2'3' en las razas Sarda y Sardo Bruna. La identificación en el ganado sardo del haplogrupo T1, característico de las razas taurinas africanas, puede reflejar la ocurrencia de un evento de introgresión africana. Dicho evento podría haberse producido a través de intercambios directos de bovinos entre el norte de África y Cerdeña o indirectamente a través del comercio con poblaciones mediterráneas ya introgresadas con ganado africano. Esta hipótesis es plausible, dado que Cerdeña está situada en la confluencia de muchas rutas comerciales marítimas que atraviesan el Mar Mediterráneo.

List of figures

List of publications

The present thesis is based on the work contained in the following articles:

Paper 1: Petretto E., Dettori M.L., Luigi-Sierra M.G., Pazzola M., Vacca G.M., Molina A., Martínez A., Goyache F., Carolan S., Amills M. & AdaptMap Consortium (2023). Footprints of post-domestication dispersal on the diversity of modern European and African goats. (in preparation).

Paper 2: Petretto E., Luigi-Sierra M.G., Vacca G.M., Martínez A., Delgado J.V., Fernández Álvarez J., Castelló A., Pazzola M., Jordana J., Dettori M.L. & Amills M. The African introgression of Murciano Granadina goats has a Moroccan origin and displays remarkable levels of inter-individual variability (2023). Animal Genetics. (submitted).

Paper 3: Petretto E., Dettori M. L., Pazzola M., Manca F., Amills M., & Vacca G. M. (2022). Mitochondrial DNA diversity of the Sardinian local cattle stock. Scientific Reports. 12(1), 2486. https://doi.org/10.1038/s41598-022-06420-3.

1 General introduction

1.1 Domestication and post-domestication dispersal of cattle and goats.

1.1.1 Cattle Domestication: From Aurochs to Cosmopolitan Breeds.

The domestication of cattle is dated to around 10,000 years ago from the wild auroch *Bos primigenius* that developed in the middle Pleistocene interglacial period 500,000 YBP (Guintard, 1999) and became extinct in 1627 (Zhang et al., 2020). From such ancestor, domestic cattle originated through two independent domestication events involving two different auroch subspecies i.e., *Bos taurus* (taurine) and *Bos indicus* (zebu) (Beja-Pereira et al., 2006). Taurine cattle, as evidenced by various archaeological findings, was domesticated in the Fertile Crescent (Beja-Pereira et al., 2006), while the zebu originated in the Indus Valley, in the present-day Pakistan (Meadow et al., 1993). Afterwards, Neolithic populations migrated into Europe carrying cattle as well as other livestock species and crops. Based on previous studies (Beja-Pereira et al., 2006, Zeder, 2008; Colli et al., 2018), the spread of cattle and goats into Europe occurred via two routes: the maritime or Mediterranean route, and the continental or Danubian route. Archaeological evidence (Zeder, 2008) has allowed to reconstruct the geographic coordinates of the maritime route traversed by the earliest Neolithic populations, who transported by boat livestock from the northern Levant during their first migrations. They first reached Cyprus (8,600 YBP), the Greek coast and the Aegean islands, then settled on coastal south-eastern Italy (Puglia) (8,000 YBP) and reached northern Italy around 7,800-7,600 YBP. Some Neolithic farming populations from mainland Italy reached southern France (7,700-7,600 YBP) and evidence of Neolithic settlements has been found on the east and south coasts of Spain dating to the same period (Zeder, 2008). Between 7,400 and 7,300 YBP, a Neolithic culture reached the Atlantic coasts of Portugal. This migration was not rapid and constant, but instead took place at different temporal scales, forming effective agricultural settlements on the Mediterranean coast without completely replacing the indigenous hunter-gatherer populations that directly or indirectly adopted Neolithic lifeways (Zeder, 2008). Regarding the continental route, the first Neolithic migrations began around 7,650 YBP in Poland and Germany with a subsequent expansion into northern Europe. Neolithic peoples arrived in north-western France (7,150 YBP) and in southern Scandinavia, England, and Ireland in 6,000 YBP (Magee et al., 2014).

The first signs of taurine cattle on the African continent can be traced back to around 8,000 YBP, being later influenced by the local aurochs and *Bos indicus*, thus giving rise to the '

taurindicine' hybrids, which appeared to be better adapted to tropical conditions than taurine cattle (Zhang et al., 2020). Zebu cattle arrived in Africa around 1,300 years ago, hybridizing with the founder taurine populations and originating the mixed *Bos taurus × Bos indicus* populations of East, Central and Southern Africa (Decker et al., 2014). The arrival of cattle in America coincided with the discovery of the New World in 1492, with cattle breeds from Europe (e.g., Spanish and Portuguese) imported in Central and South America and, after England's triumph over the Spanish power in 1588, the North-Western European breeds in North America (McTavish et al., 2013). From the 18th century onwards, the advent of breeding programmes promoted an increase in the differences between local cattle breeds (Felius et al., 2015). From the 19th and 20th centuries, 'cosmopolitan' breeds, selected for milk and meat production and mainly of European origin, began to spread all over the world, among which the most important today are, for example, Angus and Hereford for high quality meat and Holstein-Friesian and Jersey for milk production (Felius et al., 2015).

1.1.2   Goat domestication and dispersal: from the Fertile Crescent to global expansion.

Bezoar (*Capra aegagrus*), the wild ancestor of domestic goats, was domesticated around 10,000 YBP in western Asia in least two different locations of the Fertile Crescent: (1) The South Zagros Plateau/central Iran, and (2) Eastern Anatolia, which was probably the main domestication center (Naderi et al., 2008). Similar to other livestock species domesticated in the Neolithic period, the expansion of goats occurred gradually, accompanying humans in their migrations around the world. Goats and sheep arrived in Europe (5,000 YBP) via the Danubian and Mediterranean routes, moving eastwards to Asia and southwards to Africa (7,000 YBP) (Smith, 1992; Colli et al. 2018). As previously described, goats also reached Bulgaria (6,500 YBP), Scandinavia and the British Isles (4,000 YBP) via the Danubian route (Porter et al., 2016) and Greece (8,500YBP), Italy (7,600- 8,100 YBP), Spain (7,700 YBP), Libya and Algeria (7,000 YBP) following the Mediterranean route (Zeder, 2008). In addition, two main routes of dispersal through Asia have been reported: one via the Indian subcontinent and spreading to Southeast Asia through land and sea; and the other one across the Eurasian steppe connecting the Middle East, Mongolia, and northern China (Porter et al., 2016).

The introduction of goats into North Africa is dated to 7,000 YBP, with the arrival of the first pastoral populations by boat across the Mediterranean Sea and overland through the Sinai Peninsula (Smith, 1992). In the African continent, expansion occurred gradually, with evidence of the presence of goats in sub-Saharan Africa dating back to 4,000 YBP and in southern Africa around 2,000 YBP (Smith et al., 1992; Colli et al., 2018). This is supposed to be mainly due to

two factors: firstly, the presence of the natural barrier of the Sahara Desert, where upon the arrival of the first herder-gatherer groups there was a grazing niche, a favourable environment for agriculture and pastoralism in which these populations thrived for a long time. Around 4,500 YBP, the desertification of this area forced human and livestock populations to move southwards in the search of more suitable environments (Smith, 1992). The second factor is the presence of a so-called 'biological barrier', which coincides with the distribution area of the Tsetse fly located at latitudes between 12°N and 25°S, representing about one third of the African continent and limiting the movement of livestock to infested areas. Indeed, the low moisture of North Africa and the low temperatures of South Africa in winter limit the proliferation of Tsetse flies, which are largely confined to sub-Saharan Africa (Dorn et al., 2017). These flies are the biological vectors of a protozoan disease called trypanosomiasis that causes anaemia, fever, weight loss and, sometimes, even death (Ebhodaghe et al., 2018). Based on findings of goat and sheep remains at archaeological sites in Zambia and Zimbabwe, the arrival of small ruminants into southern Africa might be quite recent (2,400 YBP) (Smith, 1992).

Goats and sheep became part of the Neolithic package typical of the Mediterranean expansion, given their small size, being a source of milk and meat, and their ability of surviving in arid and hostile environments (Zeder, 2017). Gradually, small ruminants spread throughout the world and with the discovery of the New World, small ruminants reached America through Spanish and Portuguese colonisers' expeditions between the 15th and 18th century (Amills et al., 2017). During the colonisation of America, the arrival of the Spanish and Portuguese led to the spread of domesticated animals and the practice of pastoralism, unknown to the native peoples (Rodero et al., 1992). The conquerors came mainly from ports located in the coastal areas between the Bay of Cadiz in Castilla and San Vincente cape in Portugal and whose trade was centred in Seville in Andalusia (southern Spain). The trips to America were made from ports in southern Spain with a stop in the Canary Islands or/and a stop in the Antilles, often livestock (including goats) were not transported directly from Spain. The animals were recovered from the Canary Islands for transport to America, but the cattle were found to have Andalusian origins; for example, pre-Hispanic cattle did not exist, and present-day cattle breeds have origins in northern Spanish breeds (e.g., Gallega breed) introduced after the discovery of America (Rodero, 1992).

Goats were first introduced to Oceania in the 18th century, probably by British colonists (Porter et al., 2016). In fact, there is evidence of an earlier introduction of goats to the oceanic island of St Helena, much earlier in 1513 (Dunbar, 1984). Well known is the expedition of the famous

British Captain Cook, who released goats in New Zealand, some South Pacific islands and Hawaii at the end of the 18th century (Tomich, 1986).

1.2 The formation of cattle and goats breeds and specific insights into Sardinian cattle and Murciano-Granadina goat breeds.

Knowledge the demographic history of cattle and goats, through archaeological sources and genetic analysis, can shed light on the way these species have adapted to natural and artificial selection. The development of agriculture and the domestication of animals in the Neolithic period was one of the most innovative stages of human cultural development (Ahmad et al., 2020); in particular, the process of domestication was a long process with a transition of animals from wild to domesticated forms (Uerpmann, 1996). With the first human migrations, the process of domestication was followed by the spread of livestock to all inhabited continents (Felius et al., 2014). This caused livestock to adapt to extremely hot/cold or new pathogenic environments and generated a diversity of breeds that we can still appreciate today.

Among the animal species that dominate global livestock production, the most important are cattle, with a global population of about 1.5 billion heads, and sheep and goats with 1.3 and 1.1 billion heads respectively (FAO, 2022; https://www.fao.org/dad-is/en/). Given the rapid growth of the human population and the increasing demand for meat and milk, the number of livestock continues to increase especially in Africa, Asia, and South America.

According to the *Status and Trends of Animal Genetic Resources* (2022) based on data reported
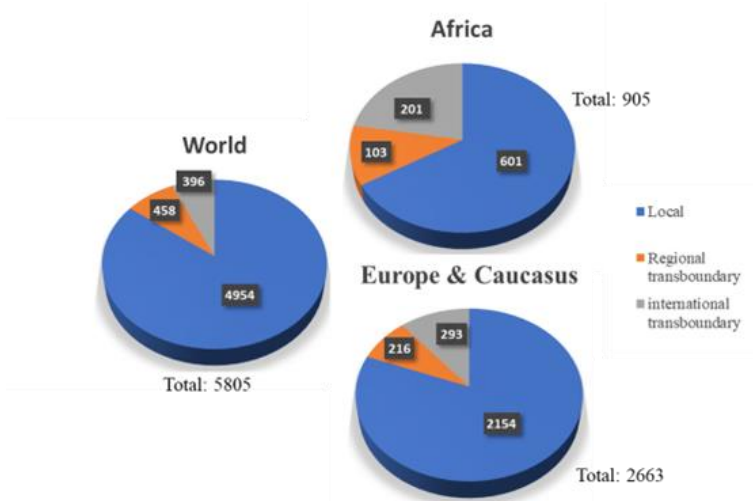
Figure 1. Number of local and transboundary (regional and international) mammalian breeds. The numbers indicated correspond to the number of local, regional, and international transboundary breeds. The graphs are based on data in the FAO's Status and Trends of Animal Genetic Resources report (2022).

by National Coordinators for the Management of Animal Genetic Resources to DAD-IS by September 2022 (FAO, 2022; https://www.fao.org/dad-is/en/), mammalian breeds (they include local breeds and regional and international transboundary breeds) registered in the world are 5805, of which 2663 are from Europe and the Caucasus breeds and 905 are African (**Figure 1**). According to Food and Agriculture Organization (FAO), local breeds are defined as breeds found in only a single country, regional transboundary breeds as those occurring in several countries in a single region, and international transboundary breeds as those present in several regions of the world. Specifically, Europe and the Caucasus have the highest number of local breeds, along with Asia, and transboundary regional breeds in most livestock species, with 379 cattle breeds and 207 goat breeds (**Figure 2**). Africa has the highest number of transboundary regional breeds of cattle and goats, which clearly indicates the necessity for urgent regional and sub-regional action to manage and conserve animal genetic resources for agriculture and food (FAO, 2022). With the exponential growth of human populations, there is an increasing demand for animal products that can have a negative impact on livestock diversity (Yaro et al., 2017). However, these factors differ depending on the degree of development of the geographic area under consideration: in highly developed regions, specialised breeding in response to socio-economic pressures leads to a loss of genetic diversity associated with excessive exploitation (Groeneveld et al., 2010); whereas in developing countries, the lack of financial resources primarily results in a general abandonment of livestock and unstructured breeding programmes causing a neglect of animal genetic resources (Philipsson et al., 2011; Biscarini et al., 2015). According to *FAO's Status and Trends of Animal Genetic Resources report*, 27% of breeds are at risk of extinction and about 7% of reported livestock breeds have become extinct (FAO,2022; https://www.fao.org/3/nl432en/nl432en.pdf). The largest number of breeds threatened by extinction among mammal species are those involving sheep, cattle and horses. The species with the largest number of breeds reported as extinct are cattle (156 breeds), sheep (107), horses (101) and pigs (70) (FAO, 2022). Loss of disease resistance and fertility, as well as the emergence of recessive genetic diseases, represent some of the alarming consequences of the decline in genetic diversity (Taberlet et al., 2018).
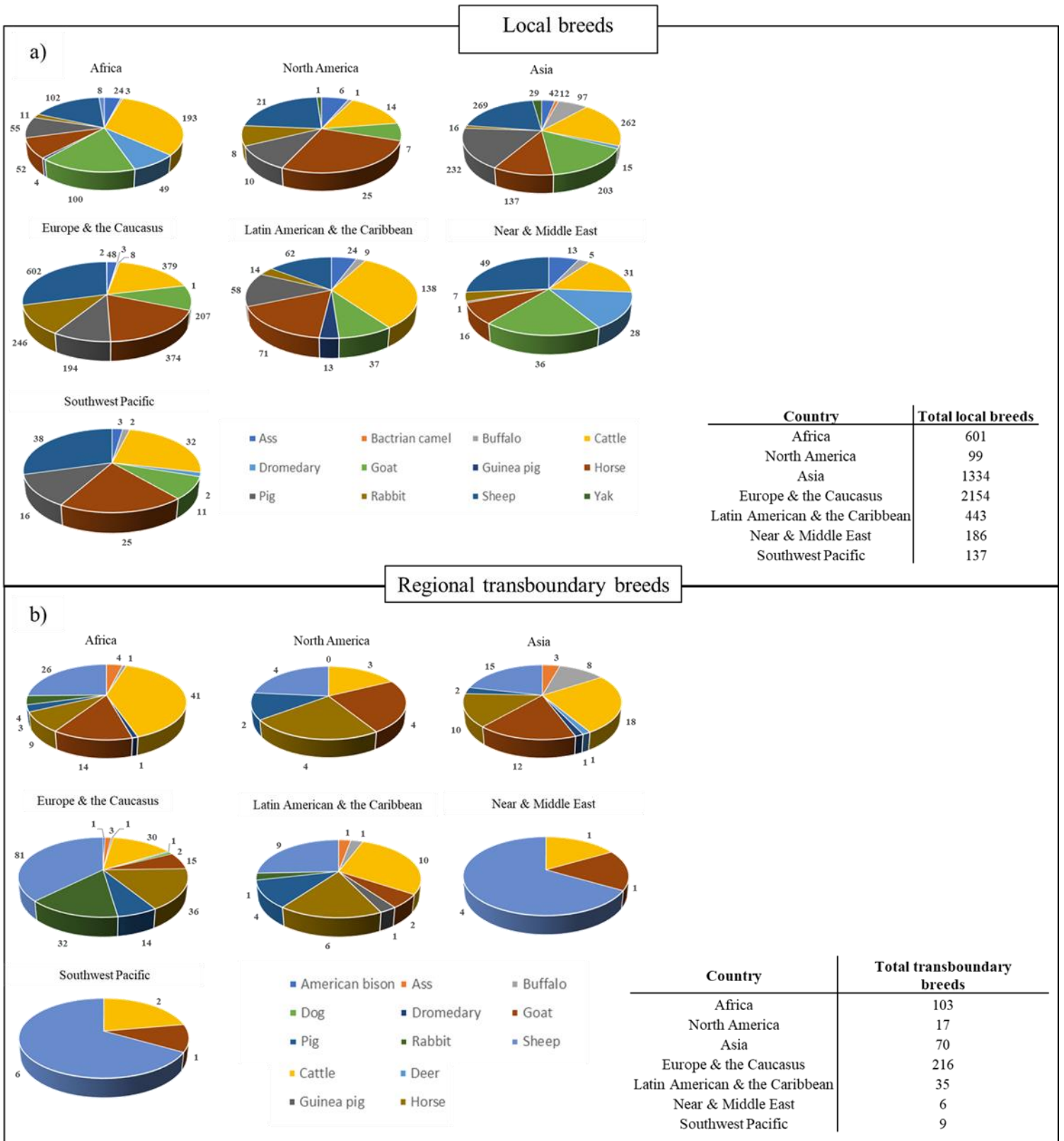
Figure 2. Number of reported local (**a**) and regional transboundary (**b**) mammalian breeds. Numbers reported correspond to the number of local breeds (a) and transboundary (b) per species on each continent. The graphs are based on data in the FAO's Status and Trends of Animal Genetic Resources report (2022).

The study of genetic diversity in livestock is useful for the sustained management of available genetic resources, conserving biodiversity (Boettcher et al., 2010), for implementing sustainable agriculture and for enhancing disease resilience, as a diverse gene pool reduces the risk of widespread epidemics by providing some individuals with natural resistance to specific diseases.

In the following paragraphs we detail the formation, history, and characteristics of the two breeds studied in this thesis: the Murciano-Granadina goat breed (**study 2**) and Sardinian cattle breeds (**study 3**).

1.2.1 Murciano-Granadina goat breed

The Murciano-Granadina (MG) breed originated in Southern Spain and has become the most important and widespread dairy goat breed on the Spanish country. As its name suggests, it originated from the crossing of two different southern Spanish goat populations, Murciana and Granadina, in 1975 (Delgado et al., 2018). A few years later, the herd book of the Murciano-Granadina breed was established, in which the ancestry of each registered animal is recorded for the correct development of breeding programmes. The MURCIGRAN federation, born from the fusion of two MG breeder associations (CAPRIGRAN and ACRIMUR) has been involved in monitoring and optimising the available genetic resources since 2011 (Delgado et al., 2018). Nowadays, the Murciano-Granadina breed is widespread in Spain with most of the herds located in Andalusia (39.30%) and Murcia (21.99%) and it is also present in other European and African countries, such as Greece, Morocco, Algeria, Portugal, France and also in South America (MAPA, https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo_razas/goat/murciano-granadina/iframe-ejemplo-arca.aspx). In 2022, the herd book of the MG breed included 113,735 animals from 201 farms, making MG the most important goat breed for milk production in Spain. This breed has two main features that are relevant in genetic improvement programmes. First, it has an excellent milk production performance of around 530 kg per lactation, with fat and protein contents of 5.6% and 3.6% respectively. Moreover, MG has a high capacity to adapt to arid environments, characterised by dry and hot climates. The improvement program of the MG breed mostly aims to increase milk production and quality, but morphological and conformation traits are also used as selection criteria.

The Murciano-Granadina goat is medium-sized, hornless, generally short-haired and with a coat colour ranging from mahogany to black. A distinctive feature of this breed is the long, dropping

ears, a sign of adaptation to the hot and arid climate which helps to regulate body temperature (**Figure 3**).
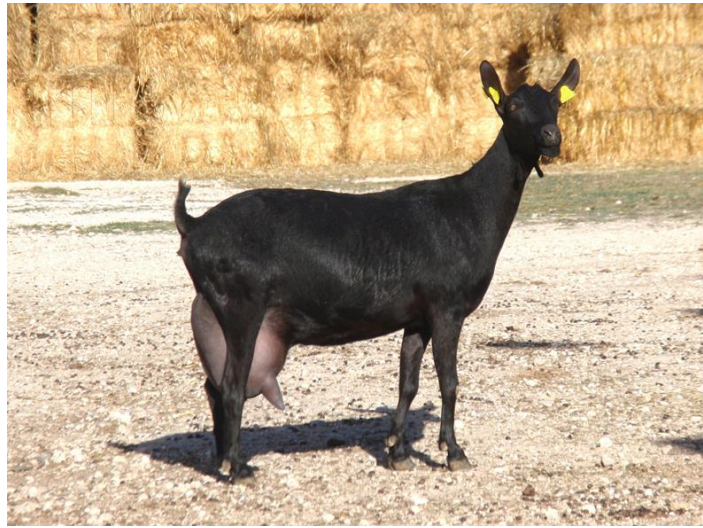
Figure 3. A typical Murciano-Granadina goat. Figure retrieved from MAPA (https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo razas/goat/murciano-granadina/iframe-ejemplo-arca.aspx).

Luigi-Sierra et al. (2022) demonstrated that the MG breed has low levels of homozygosity and inbreeding, probably due to the large census of the breed and the absence of historical bottlenecks. Regarding genetic diversity and population structure, Manunza et al. (2016) explored the demographic history of seven Spanish breeds, including the Murciano-Granadina, based on genotyping data of more than 50,000 SNPs. This study highlighted the presence of moderate levels of African introgression in breeds from southern Spain (Andalusia), including the MG breed. Several studies have confirmed the presence of such African introgression, especially among goat breeds from North Africa and Southern Europe (Martinez et al., 2016; Colli et al., 2018), and in other livestock species such as cattle (Decker et al., 2014) and sheep (Ben Jemaa et al., 2019). Colli and colleagues (2018) analysed whole-genome SNP data of goat populations from around the world in which the presence of introgression from African breeds into Spanish goat breeds was evidenced (Colli et al., 2018); such signs of African introgression were also evidenced in Mediterranean (Italian and Spanish) cattle breeds (Decker et al., 2014). In sheep, there is evidence of gene flow between Spanish and North African sheep breeds, especially for the African counterpart breeds from Morocco (Ben Jemaa et al., 2019). The genetic evidence described above might be explained by several factors. The easiness of reaching North Africa through the Strait of Gibraltar by Spanish populations and vice versa

may have allowed the exchange between animals. This hypothesis is supported by evidence of contact between northern Africa and southern Spain since around 3000 B.C. with the discovery of bovine remains in archaeological sites in northern Spain (Atapuerca), of which one individual possessed the haplogroup T1, presumably typically found in Africa (Anderung et., 2015). In addition, Spain has undergone various colonisations throughout history, including by people from Africa. e.g., the introduction of animals from Africa by the North African Berbers during the Muslim conquest in 710 A.D., where Iberian cattle were introduced by North African breeds (Cymbron et al., 1999).

1.2.2 Sardinian cattle breeds

One of the major Italian islands located in the western part of the Mediterranean Sea is Sardinia, an ancient land full of history, whose economy has been based throughout the centuries, mainly on dairy sheep farming (Vacca et al., 2010; Noce et al., 2016). Anyway, goats and cattle are reared in the island under intensive and extensive methods. Extensive cattle farming relies on foreign specialized breeds (such as Limousine and Charolaise), and three autochthonous breeds, officially recognized, which are Sarda, Sardo-Bruna and Sardo-Modicana (FAO, 2022). Each breed is registered in the Herd Book of Breeds and the Registry of Limited Native Cattle Breeds managed by the Italian Breeders Association (AIA, 2020) (**Figure 4**). From the latest regional report based on data from the National Data Bank of the Livestock Registry (2019), the cattle compartment on the island ranks sixth in terms of number of heads after Lombardy, Piedmont, Veneto, Emilia-Romagna and Sicily in terms of meat and milk production (https://www.sardegnaagricoltura.it/). The Sarda is the local breed reared since early times as sustenance for the first farmers. After 1880, this breed was crossbred with the Alpine Brown Swiss with the aim of improving dairy traits. These cattle are small sized and well adapted to both plain and mountain environments typically arid. The Sardo-Modicana breed, is the result of crossbreeding, around 1800, of cows from central southern Sardinia and bulls of the Modicana breed imported from Sicily to improve their aptitude for work in the fields. The advent of agricultural technology has resulted in a conversion of this breed towards meat production, with a consequent increase in transverse diameters and a reduction in size, and a discrete production of milk for food consumption as well as for the production of typical Sardinian cheeses such as 'Cansizolu' or 'Fresa'. Currently, the number of Sarda, Sardo-Bruna and Sardo-Modicana cattle breeds reared in Sardinia are 22,991, 29,220 and 2221 individuals, respectively (FAO, 2021 Domestic Animal Diversity Information System (DAD-IS), https://www.fao.org/dad-is/browse-by-country-and-species/en/).

Figure 4. Specimens of the Sarda (a), Sardo Bruna (b) and (c) Sardo Modicana cattle breeds. The figures come from the information sheets of the AIA (Italian Breeders' Association).

From a historical perspective, there is evidence that between the 8th and 7th millennium B.C., when the first Neolithic hunter-gatherer settlers arrived in Sardinia, wild mammals were almost non-existent. The Neolithic populations, however, carried some of their domesticated species with them, including cattle, which probably arrived on the island in this way (Wilkens, 2012). This is consistent with the greater spread of trade in the north-western Mediterranean basin (Manen and Sabatier, 2003). With regard to the evolutionary history of the three local Sardinian breeds, a study based on the analysis of SNP data of the 29 autosomal chromosomes in Italian insular breeds, including the Sarda, Sardo-Bruna and Sardo-Modicana cattle, highlighted that the Sardo-Bruna breed is genetically closer to the Sarda than to the Bruna Italiana. In contrast, the Sardo-Modicana is more similar to the Modicana breed of Sicilian origin (Cesarani et al, 2018). However, no in-depth studies exist on the potential African introgression detection of these autochthonous breeds (**study 3** of this thesis).

1.3 The role of migration in shaping the diversity of cattle and goats

1.3.1   Genetic markers

In the context of genetics and genomics, genetic markers are specific identifiable and heritable sequences or variations in the DNA sequence of an organism, used to assess genetic diversity within and between populations, species, or groups of organisms.

1.3.1.1 Mitochondrial markers in cattle and goats

Y chromosome and mitochondrial DNA (mtDNA) have certain characteristics in common i.e., recombination does not occur, they have a high mutation rate, and they are heritable through

maternal (mtDNA) and paternal (Y-chromosome) lineages. These features have made the evolution of mtDNA and Y-chromosome markers less complex than that of their autosomal counterparts. They permit the reconstruction of evolutionary relationships between and within species and allow the identification of relationships between, for example, domesticated species and their wild ancestors, and the identification of different haplogroups. In general, a haplotype is a set of alleles in an organism which are inherited together from a single progenitor (Cox et al., 2016) whereas a haplogroup is a group of haplotypes that share a common ancestor. In the case of mtDNA, the first molecular phylogeny studies were based on the sequencing of the hypervariable region called D-loop or hypervariable region (HVR), through which different haplogroups could be differentiated owing to the high rate of variability (e.g., in goats Luikart et al., 2001; Naderi et al., 2007). However, it has been shown that studies focusing on mitochondrial D-loop variability are not sufficiently informative in revealing the phylogenetic structure of livestock because of the existence of nuclear mitochondrial transpositions (NUMTs), which precludes to infer the true degree of divergence and similarity amongst maternal lineages (Hassanin et al., 2010). Therefore, more precise insights have been obtained by analysing the entire mitochondrial genome, as reported for cattle (Achilli et al., 2009; Bonfiglio et al., 2012; Olivieri et al., 2015) and goats (Colli et al., 2015).

1.3.1.2 Y-chromosome markers in cattle and goats

The study of Y-chromosome variation is confined to the identification of SNP markers in or near male-specific regions, including the most common ones: SRY (Sex Determining Region Y), AMELY (Amelogenin Y-Linked), ZFY (Zinc Finger Protein Y-Linked) and DDX3Y (DEAD-Box Helicase 3 Y-Linked). The genotyping of Y-chromosome SNPs has allowed the identification of haplotypes in goats (Pereira et al., 2009; Vidal et al., 2017) and cattle (Edwards et al., 2011). However, it is not clear whether such haplotypes are major haplotypes or if they merely represent local variants. In a recent study based on complete goat genome data generated in the VarGoats (https://www.goatgenome.org/vargoats.html) goat genome project, Y-chromosomal variants were extracted from more than 380 domestic and wild goats. The distribution of the haplogroups of 186 domestic goat breeds studied by Nijman et al. (2022) highlights the bottlenecks and expansions of ancestral paternal populations that occurred during migrations in northern Europe, sub-Saharan Africa and eastern and southern Asia. According to Nijman and colleagues, (2022) the sharing of haplogroups indicates introgression events of anthropogenic nature: (i) presence of gene flow in Madagascar of Asian goats, (ii) origin of the Boer and Anglo-Nubian breeds due to crossbreeding in the 19th century, and (iii) recent

introgressions by European goats into Korean breeds and by the Boer breed into goats originating in East and South East Africa (Nijman et al., 2022-VarGoats Consortium).

It is well known that in most domesticated species a reduction in the size of male populations has historically been observed, making Y-chromosome variants highly informative to differentiate breeds and populations (Nijman et al., 2022-VarGoats Consortium).

1.3.1.3 Autosomal markers in cattle and goats

With the continuous progress of high-throughput sequencing technologies, microsatellites were replaced by single nucleotide polymorphisms (SNPs) as autosomal markers of choice. Indeed, massive amounts of SNPs can be discovered by whole genome sequencing (WGS), making it possible to identify not only SNPs but also other types of genetic variants such as insertions and deletions (indels) or copy number variations (CNVs).

SNPs are DNA sequence variations due to the substitution of nucleotides carrying different nitrogenous bases at a specific position of the genome. They are the most frequent genomic variants in all species, occurring with a frequency of one SNP every 300-1000 bp in most mammalian genomes (Aitken et al., 2004). Single nucleotide polymorphisms have become the markers of choice for the study of genetic diversity and the assessment of population structure and relationships, because of their broad distribution throughout the genome and their excellent amenability to high-throughput genotyping (Ajmone-Marsan et al. (FAO), 2023). Several techniques exist for genotyping SNPs:

    a.   *Single SNP genotyping assays*: these are based on conventional procedures such as polymerase chain reaction (PCR). Several examples of such approach are restriction fragment length polymorphism combined with PCR (PCR-RFLP; Jarcho, 2001), mass spectrometry-based primer extension detection (Sequenom) (Bradić et al., 2011) and the use of hydrolysis fluorescent probes (Taqman assay; Holland et al., 1991). The Taqman genotyping assay is based on the PCR reaction that includes two probes each specific for one of the two SNP alleles. Each probe has a different fluorescent called a reporter (e.g., 6-FAM or VIC dye) that binds with one of the PCR product strands during amplification depending on the alleles. The probe is hydrolysed by Taq polymerase and generates fluorescence of a different colour depending on the allele. This allows discrimination of homozygous alleles and heterozygous alleles in a single tube without an additional post-PCA process (https://www.thermofisher.com/it/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays/custom-snp-assays.html).

b.   *Multiple SNP genotyping*: it consists of the simultaneous genotyping of 10,000 o 1,000,000 SNPs (Nicolazzi et al., 2015) through SNP array platforms. Several types of SNP microarrays are commercially available, but the most well-known and broadly used platform is the "Illumina's Infinium iSelect Microarray". This microarray is also called a 'bead chip' because the oligonucleotide probes cover the surface of silica microbeads placed inside minuscule microwells. The Infinium assay starts with the amplification of genomic DNA overnight and then subjected to enzymatic fragmentation. The bead chip is prepared for hybridisation in the capillary flow chamber after alcohol precipitation and DNA resuspension. DNA fragments are incubated overnight with millions of types of beads, each corresponding to each SNP locus allele, and DNA- locus-specific hybridisation takes place. Afterwards, the bases are subjected to enzymatic extension with release of fluorescent colouration. The iScan system detects the fluorescence intensity of the beads and the associated software (Illumina) performs the analysis and genotype calling automatically (https://www.illumina.com/Documents/products/workflows/workflow_infinium_ii.pdf). In this PhD thesis, genotyping data of several goat breeds obtained through GoatSNP50 Illumina BeadChip (http://snp.toulouse.inra.fr/~sigenae/50K_goat_snp_chip/index.html) were used. This genotyping array contains approximately 54,000 SNPs evenly distributed along the goat genome and covering all chromosomes (Tosser-Klopp et al., 2014). These SNPs were identified from whole genome sequences of 97 individuals from six caprine breeds, i.e., Alpine, Boer, Creole, Katjang, Saanen and Savannah, and subsequently validated in 10 breeds represented by 285 individuals. A new version of this microarray containing 59,000 SNPs (https://www.goatgenome.org/projects.html) has been available since January 2021.

c.   *Genotype-by-sequencing (GBS)*: there are several methods belonging to this category such as low-depth whole genome sequencing (WGS), but in this case GBS refers to reduced-representation sequencing, in other words sequencing a specific fraction of the genome. This can be achieved, for example, by obtaining DNA fragments generated by the digestion of total DNA with restriction enzymes with a subsequent selection step by fragment size (ddRADseq; Peterson et al., 2012). After obtaining DNA fragments, they are sequenced and reads are then mapped against a reference genome, with polymorphic sites being called from the reads. The SNPs and indels obtained in this way pass through several qualitative and quantitative controls, using software that

takes into account depth (average number of times each fragment is sequenced) and sequencing error (Lou et al., 2021). These two features give reliability to the calling of allele genotypes (Nielsen et al., 2012). This approach is mainly used when no species-specific standard SNP chip is available. Generally, for large numbers of samples it is less expensive to use SNP chips than GBS and data analysis is also easier. The fact that SNP arrays evaluate pre-selected variants (SNPs) and the combination of different types of analysis to achieve the desired goal is often useful.

In recent years, the increased accessibility of whole-genome sequencing (WGS) (Eusebi, Martinez and Cortes, 2020) has resulted in the production of a multitude of genetic markers leading to the development of new software for massive data management and new statistical analyses (Biscarini et al., 2018).

1.3.1.4 Diversity parameters based on genetic data in goats and cattle

On the basis of the selection of markers, it is possible to quantify genetic diversity through certain parameters such as observed (H_o) and expected (H_e) heterozygosities, nucleotide (π) and haplotype (H_d) diversities, effective population size (N_e), and Wright's F-statistics.

1.3.1.5 Observed (H_o) and expected (H_e) heterozygosities

Observed heterozygosity (H_o) is the effective proportion of heterozygous genotypes in a population, meaning individuals possessing two different alleles at a specific locus, while expected heterozygosity (H_e) is calculated based on the population allele frequencies (Nei, 1973; 1978). These parameters are often compared: generally, H_o is lower than H_e, due to sampling bias or non-random mattings occurring sporadically. Large differences between both indices might suggest either a recent population interbreeding (H_o values higher than H_e) or a population partition (H_o values lower than H_e).

1.3.1.6 Wright's F-statistics

Three fixation indices were invented by Wright et al (1951) to describe population structure i.e., F_{st}, F_{is} and F_{it} (Wright et al., 1951). Wright defined the basic inbreeding coefficient, F_{it} or F, as the correlation between genes on uniting gametes relative to the total array of those in random derivatives of the foundation stock. In contrast, the F_{st} index represents the correlation between uniting gametes relative to those across all subdivisions. Its value can range from 0 to 1, which corresponds to the total absence or presence of differences in the allele frequencies of the two populations under consideration. The F_{is} index or inbreeding coefficient, on the other

hand, is used to measure the excess of homozygotes or heterozygote deficiency. Values range from -1 to +1 indicating an excess (positive values) or a deficit (negative values) of heterozygotes.

1.3.1.7 Nucleotide diversity (π) and haplotype diversity (H_d)

Nucleotide diversity (π) can be defined as the average number of nucleotide differences per site between two randomly selected DNA sequences from the studied population (Nei and Li, 1979), while haplotype diversity (H_d), is the probability that two randomly selected haplotypes are different (Nei, 1987).

1.3.1.8 Effective population size (N_e)

A very important parameter in population genetics is the estimation of the effective population size (N_e), which corresponds to the number of individuals of an ideal population (reproduction by random mating, non-overlapping generations, absence of migration, selection and mutation, etc.) that would have the same rate of loss of genetic diversity as the real population under study (Frankham, 2010). Obviously, the census dimension of a population is larger than Ne, as there is no population that strictly matches the conditions of an idealised Wright-Fisher population. In domestic breeds, non-random mating is common, e.g., in populations subjected to strong artificial selection or by admixture events, whereby the Ne value of a breed may be reduced, even if the population is large. This may lead to allele fixation and loss of genetic variability (Kristensen et al., 2015).

1.3.1.9 Analysis of molecular variance (AMOVA)

This is one of the methods to detect population differentiation based on estimates of F-statistics across genetic datasets (Excoffier et al., 1992; Michalakis and Excoffier, 1996; Meirmans, 2012). What makes AMOVA flexible is the fact that a specific distance metric can be used depending on the type of data being analysed, whether it is SNPs, sequence data, haplotype data, microsatellite loci (Michalakis & Excoffier, 1996; Whitlock, 2011).
There are a number of software packages that enable the estimation of summary statistics of genetic diversity within a breed and/or between breeds, these include: Arlequin (Excoffier et al., 2010); DnaSP (http://www.ub.edu/dnasp/), PLINK (Purcell et al., 2007; Chang et al., 2015).

1.3.3.6. Principal Component Analysis (PCA)

One of the most widely used methods to dissect the population structure of breeds by using multilocus genotypes is principal component analysis (PCA). This analysis owes its origins to Cavalli Sforza et al. (1993) who, around forty years ago, applied this method to the study of the variation of human gene frequencies in continental regions aiming to summarise them and render them interpretable to explore the migratory history of populations (Menozzi et al., 1978; Cavalli-Sforza et al., 1993; Cavalli-Sforza et al., 1994). PCA is a statistical method that transforms high-dimensionality data or variables (marker genotypes) into a low-dimensionality form but preserving the variance of such data as much as possible. What results are variables called 'principal components (PCs)' which are reported according to the amount of variance they explain, generally in descending order whereby principal component 1 generally explains most of the variance, followed by component 2 etc... After obtaining the principal components, they are projected onto a two-dimensional plane (e.g., PC1 vs PC2 or PC1 vs PC3), providing an overview of the entire dataset. An example is shown in **Figure 5**, based on genotyping data with GoatSNP50K of European goat breeds, specifically from North Africa, Spain, and the cosmopolitan Swiss Saanen breed. Here, breeds cluster according to their geographical origin and PC1 differentiates the Saanen (cosmopolitan) breed from the other breeds, while PC2 separates the Spanish, African and Saanen breeds. Principal components are variables that have no biological significance but through them it is possible to obtain a logical interpretation of highly complex data; in general, the proportion of variance explained by principal component 1 corresponds to the F_{st} (McVean, 2009). It has been shown that the principal components often show genetic clines correlated to geographical dispersion, resulting in a graph that likely corresponds to the geographical map of the breeds under study (Novembre et al., 2008).
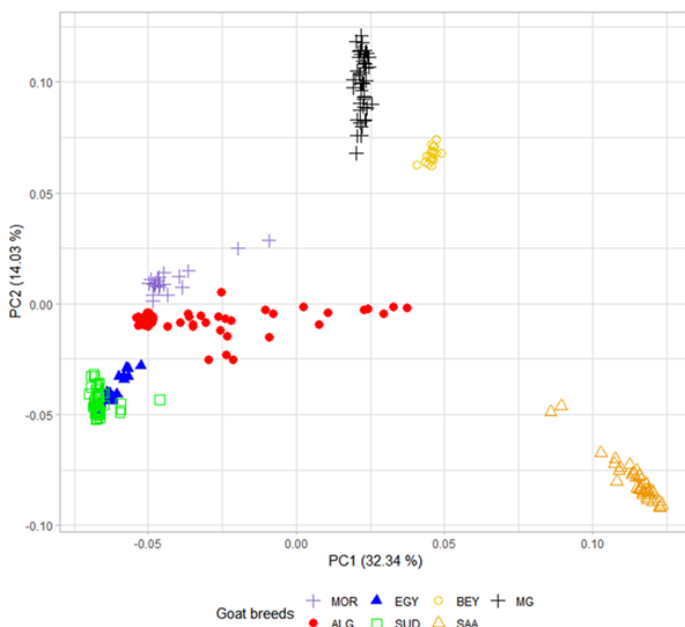
Figure 5. An example of a PCA plot of different goat breeds: Spanish breeds (Murciano Granadina= MG and Bermeya=BEY), Swiss Cosmopolitan (Saanen=SAA) and North African breeds from Morocco=MOR, Egypt=EGY, Algeria=ALG and Sudan=SUD. It is evident how SAA, Spanish and North African goats differ from each other by clustering according to geographical origin.

The PCA method has limitations: the data in PCA graphs are often quite dense, not making the distribution and differentiation of breeds clearly detectable (in this case, averages of the principal components are usually calculated and projected in two dimensions). Moreover, the number of samples within populations should be balanced, as populations with larger sample sizes are usually pressed into the centre of the graph (McVean, 2009). Finally, the projections of the components occur on a two-dimensional plane despite coming from a multi-dimensional environment whereby only part of the total variation is appreciated. These problems could lead to a distorted interpretation of the structure and relationships between populations. Software used for the calculation of principal components are, for instance, PLINK (Chang et al., 2015) or EIGENSOFT (Price et al., 2006), while several RStudio packages can be employed for plot creation (RStudio Team, 2020).

1.3.3.7. Clustering methods

Another method for studying population structure is based on the inference of admixture proportions (Pritchard, Stephens & Donnelly, 2000). The aim of this method is to infer the proportions of ancestry in each individual from each population and report them on graphs that visually reveal the existence of population structure and admixture or introgression events. Underlying this approach is the consideration of the existence of genetic groups or clusters (K) whose polymorphic sites are in HWE and linkage equilibrium. The general formula on which the model is based can be found in Pritchard et al. (2000) and Alexander and Lange (2011). More specifically, each individual (i) turns out to be a mixture of groups (K) with specific proportions (q_{ij}) equal to 1 ($\Sigma K\ j = 1\ q_{ij} = 1$, here j indicates cluster). These proportions are calculated using specific algorithms, e.g., the Structure software uses the Monte Carlo Markov Chain (MCMC) algorithm, on the basis of the observed genotypes. Hence, when an individual (i), has a value of q_{ij} close to 1, it implies that this individual most probably pertains to cluster j. There are two main applications for this model: (i) standard or unsupervised clustering (ii) supervised clustering. In standard clustering, the number of clusters and the corresponding allele frequencies are deduced, and the number of clusters (K) is identified from the set of all individuals afterwards. When choosing the number of clusters which best represents the structure of the investigated populations, the cross-validation error (cv) **(Figure 6a)** defined by Alexander and Lange (2011) as "based on the prediction of systematically omitted data points" is calculated. The lowest value of cv indicates a minimum estimated prediction error and a best-fit corresponding K-value. In contrast, in supervised clustering it is possible to select ancestral clusters a priori and infer the mixing proportions of the other target individuals of the analysis

directly. More efficient approaches to model population structure have been developed, such as in the Admixture software (Alexander, November and Lange, 2009), which allows larger datasets to be analysed with larger values of K, under the assumption that the loci are independent of each other and are in linkage equilibrium.
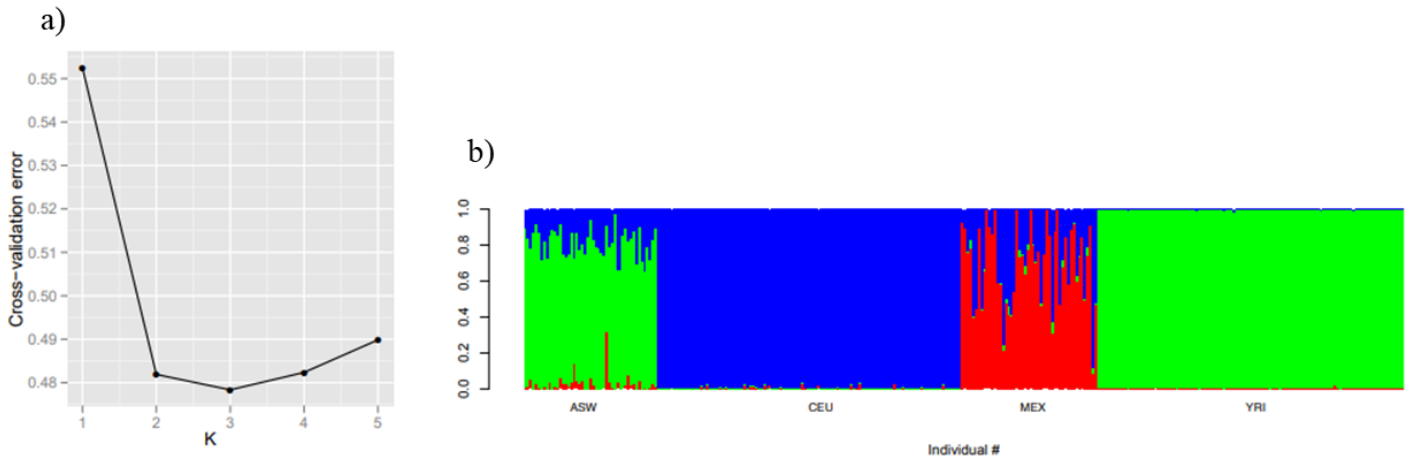
Figure 6. a) Cross-validation plot for selecting the number of clusters (K) which best represents the population structure of the dataset under consideration. In this case K=3 is the optimal one. b) Example of a bar plot based on results obtained using Admixture software (P and Q output files). Bars represent individuals and colours are the proportions of the ancestral components within each individual. Source: Admixture software manual (https://vcru.wisc.edu/simonlab/bioinformatics/programs/admixture/admixture-manual.pdf).

The graphical representation of the results obtained with the Admixture software is done via the RStudio software with the dedicated "pophelper" package (Francis et al., 2017), using the output files Q and P that contain the ancestry fractions and allele frequencies deduced from the ancestral populations, respectively. An example can be seen in **Figure 6b**, where the individuals are placed on the horizontal axis and each individual is represented by a bar, while the different colours correspond to the ancestry percentage, represented on the vertical axis.

1.3.3.8.Phylogenetic trees and haplotype networks

A further approach to visualising genetic structure is to use matrices of genetic distances. An example for visualising such distances are phylogenetic trees (**Figure 7a**). These are graphs, with or without a root, whose 'branches' correspond to the distances between populations/individuals under consideration while the nodes joining branches correspond to their ancestries. The construction of phylogenetic trees is traditionally based on the neighbour-joining (NJ) algorithm and on tree searches that employ optimality criteria such as parsimony and maximum likelihood (ML). In the NJ, protein or DNA sequences are converted into a

distance matrix, which is the representation of the estimated evolutionary distance between sequences. This algorithm is fast and works well with recently differentiated sequences but has difficulty deducing older relationships (Holder and Lewis, 2003). Parsimony is not an approach based on distance matrices but scores each tree based on the minimum number of mutations that could produce the data, so the use of dense sampling is essential (Holder and Lewis, 2003). Maximum likelihood like parsimony, is a method based on sequences and not on distance matrices. With this method, hypothesis is tested according to its ability to predict the observed data, whereby the tree with the highest probability of justifying the observed sequences is chosen (Holder and Lewis, 2003). The statistical significance of the various nodes that make up the tree and thus give reliability to the phylogenetic analysis is from the so-called bootstrap. Through this test, the original data matrix is randomly resampled with replacement, producing pseudo-replicated data on which the tree construction algorithm is run (Holder & Lewis, 2003). The bootstrap values indicate the number of times the same branch is observed whenever a new phylogenetic tree is generated on resampled data (Ojha et al., 2022). If such an observation is obtained 100 times out of 100, this supports our result. Values less than 50 out of 100 are not considered for the construction of the tree (Ojha et al., 2022).
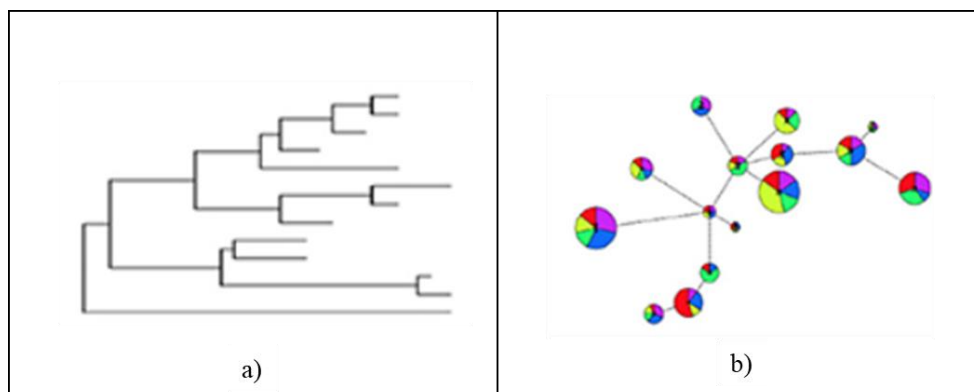
Figure 7. Graphical visualisation of a phylogenetic tree based on distance matrices (a) and a haplotype network using the NJ algorithm (b). In the network, the size of the circles corresponds to the number of individuals and the colours show the percentage of corresponding haplotypes.

Phylogenetic networks, unlike trees, are graphs connected by cycles (**Figure 7b**). Among the different types of networks there are haplotype or allelic networks, generated by median-joining analysis (Bandelt et al., 1999). Median-Joining (MJ) networks use intraspecific data from a large number of samples to produce a network that is the result of encapsulating the most plausible trees in a single network figure (Bandelt et al., 1999). This algorithm is the basis of

the Network software (https://www.fluxus-engineering.com/sharenet.htm), with which phylogenetic networks can be easily constructed.

1.3.2. The impact of migration on the genetic diversity of cattle.

The reconstruction of the phylogenetic history of livestock is based on both archaeological evidence and the analysis of mitochondrial and nuclear genomic data. Previous studies on variable portions and complete sequences of mitochondrial DNA in taurine cattle (Beja-Pereira et al., 2006; Achilli et al., 2009; Bonfiglio et al., 2012; Zhang et al., 2020), highlight the presence of a T macro haplogroup, which includes haplogroups T1'2'3', T2, T3 and T4 plus a T5 haplogroup (**Figure 8**). Haplogroups T1, T2, T3 originate from the earliest domesticated cattle in south-west Asia. Haplogroup T3 is predominant in European cattle, occurs with variable frequency in Near-Eastern cattle (Achilli et al., 2009; Magee et al., 2014). Haplogroup T4 derives from T3, and it is present in East Asia (Korea, China and Japan). Haplogroup T2 can be found in Asia, the Balkans, Italy, and sporadically in Central, Eastern and Western Europe. Haplogroup T1 is predominant in African cattle breeds and in southern Europe (Portugal, Spain, Italy and Greece). As pointed out by Bonfiglio et al. (2012), haplogroup T1 is subdivided into six sub-haplogroups: T1a, T1b, T1c, T1d, T1e, T1f. Haplogroups T1a, T1b and T1c spread mainly in Africa, Europe, and Asia, and T1d is present exclusively in African breeds. Haplogroup T1c is more common in Egyptian cattle, while T1b is predominant in Ethiopian cattle (Bonfiglio et al., 2012). The reason for this might be that taurine cattle arrived in Africa from Middle East via two alternative routes that were traversed at different times: one route from the north through the Suez Canal to Egypt and another route from the south via Arabia to Somalia and Ethiopia (Caramelli, 2006). The presence of the T1 African haplogroup in Southern European cattle breeds suggests a probable gene flow between African and Mediterranean populations (via the Mediterranean Sea). An additional T1c sub-haplogroup called T1c1 was discovered in cattle breeds from Brazil and Paraguay, confirming migrations from Africa to South America (Ginja et al., 2019). Haplogroup T5, which splits from haplogroup T1'2'3', has been identified in Italian Valdostana cattle breeds and is considered a rare haplogroup along with T6, found in China. Rare haplogroups of *Bos taurus* include haplogroups identified on the basis of ancient mtDNA: haplogroup E (in German samples taken from early Neolithic sites) and haplogroup P (in most aurochs) (Achilli et al., 2009). This latter haplogroup was actually also detected through the mtDNA analysis of samples collected from taurine cattle fossils in Switzerland (Schlumbaum et al., 2006) as well as in modern Korean

Hanwoo animals (Noda et al., 2018), which might suggest that there was introgression between auroch and domestic cattle.
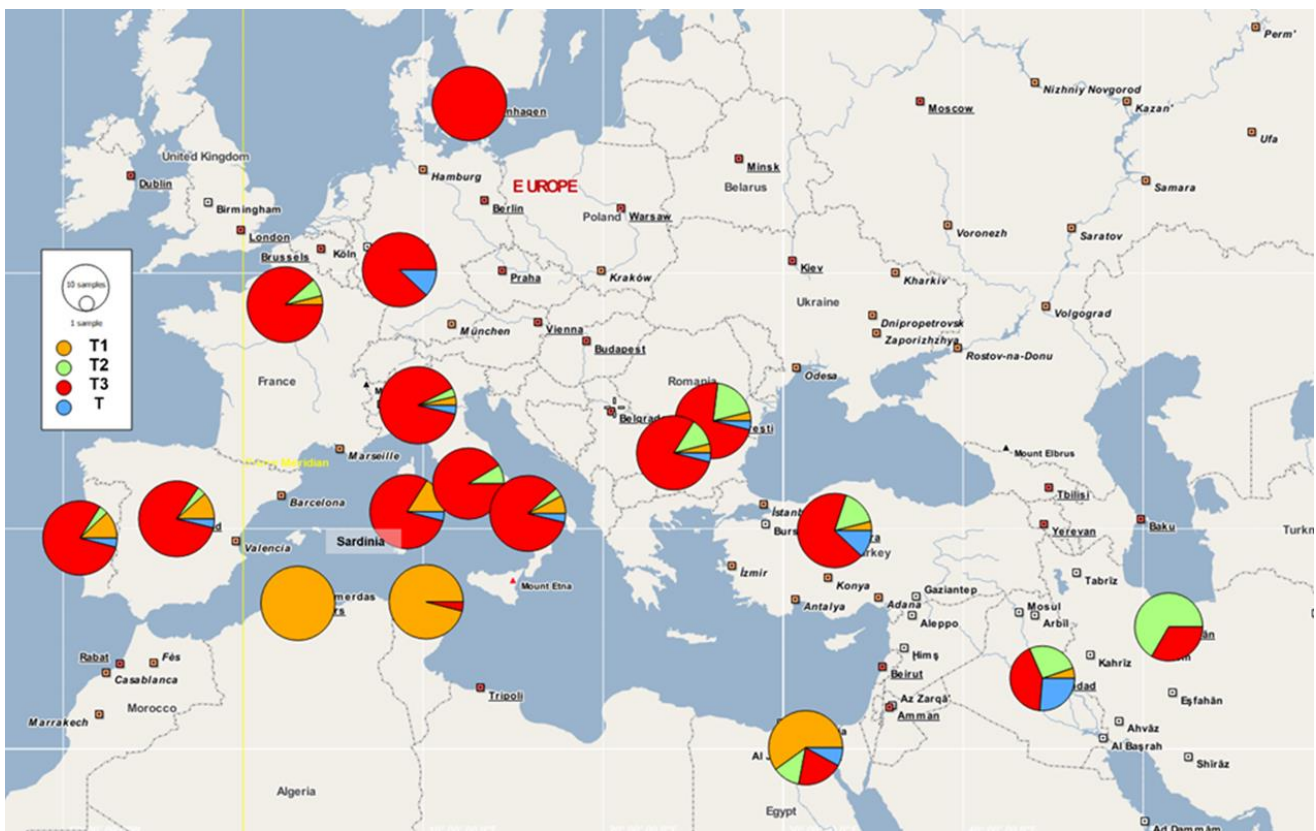
Figure 8. Geographical distribution of taurine haplogroups in Europe, North Africa and the Middle East. The T macro-haplogroup comprises T, T1'2'3 and T5. Haplogroup frequencies were plotted using POPART v.1.7 software (http://popart.otago.ac.nz/index.shtml). This figure is part of the supplementary material of study 3 of the thesis.

The rare haplogroup Q has been identified in samples of ancient and modern cattle from different parts of Europe and North Africa (Bonfiglio et al., 2012; Olivieri et al., 2015), while haplogroup R has only been detected in some Italian cattle of the Agerolese, Cinisara, Romagnola and Marchigiana breeds (Bonfiglio et al., 2010).

With regard to the phylogenetic history of *Bos indicus*, a macro haplogroup I subdivided into haplogroups I1 and I2 was identified through mtDNA analysis. These haplogroups are typical of South Asia, I1 prevailing in cattle from the primary domestication site (Indus Valley) to Southeast Asia, while haplogroup I2 is dominant in the Indian subcontinent. Y-chromosome markers, as well as autosomal markers, confirm the remarkable divergence between taurus and indicus taxa suggested by mtDNA analysis. The study of Y chromosome SNP has led to the

identification of two haplogroups for taurine cattle called Y1, typical of northern Europe and northern Spain, and Y2 found in central and southern Europe, Africa and western Asia (Edwards et al., 2011). A haplogroup derived from Y1, called Y1A has been found in cattle from China, while, in zebu, haplogroup Y3 has been identified (Chen et al., 2018; Xia et al., 2019).

Focusing on European cattle breeds, the predominant gene pool is *Bos taurus* with some exceptions in breeds showing introgression with *Bos indicus* such as Turkish and a few Italian breeds. Besides such introgression, in Iberian and southern European taurus, cattle breeds show admixture with African breeds, suggesting the existence of gene flow between both continents. In a recent study, more than 3,000 cattle from 205 populations distributed worldwide were genetically characterized through SNP data genotyping, with a focus on European and Italian breeds in order to identify relationships and admixture among the investigated breeds (Mastrangelo et al., 2020). In this study, through an MDS analysis, one component (PC1) distinguished *Bos t. taurus* and *Bos t. indicus*, while for the second component (PC2), the differentiation of African taurine cattle breeds from European, American and Asian cattle breeds was evident. Mediterranean breeds clustered together and appeared to be closer to African breeds than northern European breeds (Mastrangelo et al., 2020). This further supports the presence of gene flow between cattle populations from Africa and Southern Europe. Italian cattle breeds generally cluster with Taurine populations, and they include two separate groups: (1) Northern Italian breeds, which share ancestry with French and Swiss breeds, evidencing that northern European breeds contributed to their gene pool (Mastrangelo et al., 2018); and (2) Podolian breeds, characterized by a gray coat and long horns (Di Lorenzo et al., 2018), Sicilian breeds (Cinisara, Rossa Siciliana, and Modicana) and Sardo-Modicana breed. This second group showed minor African contributions compared to other Italian breeds. However, Podolian and Sicilian breeds share an indicine origin with Balkan cattle, suggesting that they were influenced by breeds of non-European origin (Upadhyay et al., 2019; Mastrangelo et al., 2020).

1.3.3. The impact of migration on the genetic diversity of goats.

Initial studies on the molecular phylogeny of livestock species were mainly based on the characterization of hypervariable portions of mitochondrial DNA, through which six haplogroups were identified: A, B, C, D, F, G (Luikart et al., 2001; Naderi et al., 2007; Colli et al., 2018) (**Figure 9**). Haplogroup A is the most predominant in domestic goats and it is distributed worldwide from Europe to Asia, Africa and America, being rare in wild goats which

mostly harbour haplogroup C. Based on studies of archaeological remains of ancient goats, it was observed that these haplogroups (A and C) were equally represented in Europe (specifically in southern France), whereas today haplogroup A predominates by far (Naderi et al., 2008).
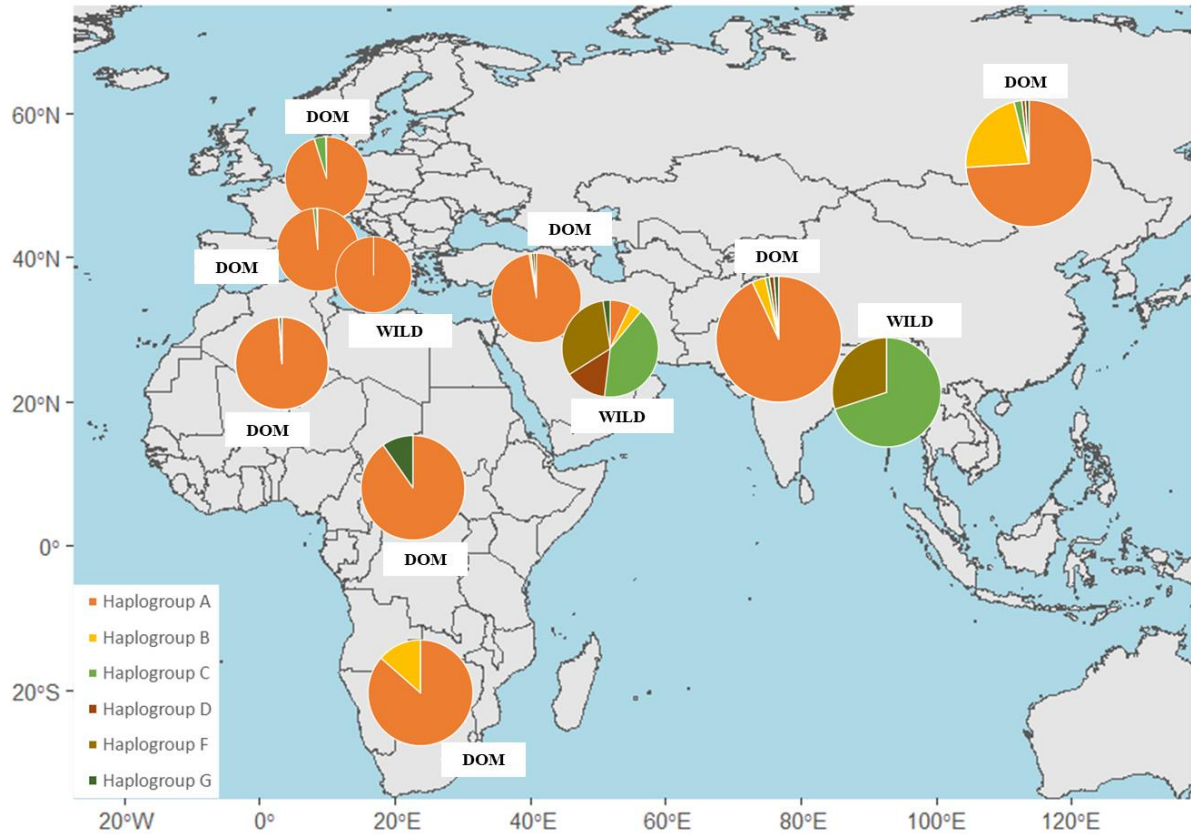
Figure 9. Distribution of goat haplogroups obtained using mitochondrial DNA in European, African, Asian and Middle Eastern breeds. Data from modern domestic (DOM, C. hircus) and wild (WILD, C.aegagrus) goats are reported. This figure was adapted from Colli et al. (2015). Haplogroup frequencies were plotted using RStudio (RStudio Team, 2020).

This suggests the presence of the two haplogroups A and C in goats transported in the first Neolithic migratory waves. Furthermore, haplogroup F also appears to be present in wild goats but not in domestic goats, where it is very rare and has only been identified in some animals originating in Sicily (Italy) (Colli et al., 2015). Haplogroup B is present in southern Africa and Asia (Chen et al., 2005), and haplogroups C and D segregate in Europe and northern Europe, respectively. Haplogroup G was detected in the Middle East, North Africa and Asia (Colli et al., 2015). Haplogroup E was initially identified, but upon comparison with a larger and more complete dataset, the latter was found to be a subgroup of Haplogroup A. A study on the hypervariable region of mitochondrial DNA in more than 1600 goats of the Sardinian breed (Italy) showed the presence of 11 subgroups of haplogroup A (A1-A11) (Piras et al., 2012).

These results were further confirmed by Dettori et al. (2020) who investigated mitochondrial sequences from Sardinian goats as well as mtDNA sequences available on GenBank from African, Asian and European goat breeds (Dettori et al., 2020). In general, and in contrast to cattle, a weak geographical structure of matrilineal diversity has been found in goats, partly due to the predominant presence of haplogroup A, which masks the structure of the other, albeit rare, haplogroups, and partly due to the characteristic easiness of transportation of goats compared to cattle, which are much larger in size.

Based on the study of markers in the SRY, ZFY and DDX3Y genes present on the goat Y chromosome, six haplotypes were characterised: Y1A, Y1B, Y1C, Y2A, Y2B and Y2C. The Y1A haplotype has been found in goat breeds from central and southern Europe, Iran, Asia, and North Africa (Morocco), while the Y1B haplotype was detected in Swiss goats, Portuguese and North African (Pereira et al., 2009; Cinar Kul et al., 2015; Vidal et al., 2017). The Y1C haplotype was mainly detected in goat breeds from Turkey and Switzerland (Cinar Kul et a2015; Vidal et al., 2017). The Y2A haplotype is predominant in Spanish, Portuguese and Italian breeds and is the only one that segregates in sub-Saharan and West African breeds. The Y2B haplotype is typical of breeds from East and South-East Asia (Pereira et al., 2009; Waki et al., 2015), while the Y2C haplotype is a minority cluster detected in breeds from Turkey (Cinar Kul et al., 2015). Based on these results, evidence of a strong paternal population structure exists in goats, contrary to the findings from mitochondrial markers. For instance, differentiation between Mediterranean and Central European breeds has also been shown in microsatellite studies on European and Near Eastern goat breeds (Cañon et al., 2006), but also through the study of autosomal SNPs. The study by Colli and colleagues (2018) was carried out in a large number of goat breeds distributed worldwide and a strong phylogeographic structure was appreciated with 3 distinct gene pools corresponding to breeds from Europe, Africa and Western Asia, respectively. Recent research based on goat whole-genome sequencing data first confirmed the differentiation between Y1 and Y2 haplogroups (Zheng et al., 2020; Xiao et al., 2021) and the presence of the Y1B, Y2A, Y2B, Y1C and Y2C haplotypes already identified in previous studies based on SNP data (Lenstra & Econogene Consortium, 2005; Pereira et al., 2009; Cinar Kul et al., 2015; Waki et al., 2015; Vidal et al., 2017; Tabata et al., 2019). In 2022, Nijman et al. (2022) used whole-genome sequences of several goat breeds obtained through the VarGoats project (https://www.goatgenome.org/vargoats.html) by extracting SNPs of the Y chromosome. In this case, as well, they reconfirmed the presence of the haplotypes already identified, however, they pointed out that the Y1A haplotype splits into Y1AA haplotype, occurring in goats from South Asia, South-East Africa, Southern China and

in several Italian breeds (e.g., Ciociara and Montecristo), and Y1AB, prevalent in Northern China.

2 Goals

The general goal of this thesis is to investigate the role of migration in shaping the genetic diversity of cattle and goat populations, with a special emphasis on those from the Mediterranean Basin.

The specific goals of this thesis are:

• To assess the existence of genetic clines associated with the post-domestic dispersal of goats from the Fertile Crescent domestication site into Europe and Africa based on a large collection of high throughput genotyping data obtained from 1,148 modern European and 1,187 African goat populations belonging to 38 and 43 different breeds, respectively.

• To investigate the genome-wide diversity and population structure of Spanish Murciano Granadina goats and compare it with that of Moroccan, Egyptian, Algerian, Sudanese goats to determine the extent of the African introgression of Murciano-Granadina goats as well as its most likely geographic origin.

• To characterize the mitochondrial variation of three Sardinian cattle breeds (Sarda, Sardo Bruna and Sardo Modicana) and to elucidate whether African populations contributed to it.

3 Papers and Studies

3.1 PAPER 1

Footprints of post-domestication dispersal on the diversity of modern European and African goats

Elena Petretto[1,2], Maria Luisa Dettori[2], Maria Gracia Luigi-Sierra[1], Michele Pazzola[2], Giuseppe Massimo Vacca[2], Antonio Molina[4], Amparo Martínez[4], Félix Goyache[5], Sean Carolan [6], The AdaptMap Consortium, Marcel Amills[1,3] *

[1] Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [2] Department of Veterinary Medicine, University of Sassari,07100, Sassari, Italy. [3] Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [4] Department of Genetics, University of Cordoba 14071, Spain. [5]Área de Genética y Reproducción Animal, SERIDA-Deva, Camino de Rioseco 1225, Gijón, 33394, Spain. [6] Old Irish Goat Society, Mulranny, Ireland

*Corresponding author: marcel.amills@uab.cat
E-mail addresses:
EP, elenapetretto@outlook.it
MLD, mldettori@uniss.it
MGLS, marigls.ml@gmail.com
MP, pazzola@uniss.it
GMV, gmvacca@uniss.it
AMo: ge1moala@uco.es
AMa: ib2mamaa@uco.es
FG: fgoyache@serida.org
SC: carolan.sean1@gmail.com
MA, marcel.amills@uab.cat

Abstract

Background: Goats were domesticated in the Fertile Crescent about 10,000 years before present (YBP) and subsequently spread across Eurasia and Africa. This dispersal is expected to generate a gradient of declining genetic diversity with increasing distance from the domestication center. Previous studies have reported the existence of such genetic cline in European goat populations, but these were based on a limited number of microsatellite markers. Here, we have analyzed data generated by the AdaptMap project and other studies. More specifically, we have used the geographic coordinates and estimates of the observed (H_o) and expected (H_e) heterozygosities of 1,148 European and 1,187 African goats belonging to 38 and 43 different breeds, respectively, to find out whether genetic diversity and distance to the domestication site in the Fertile Crescent are significantly correlated.

Results: Principal component and Admixture analyses revealed an incomplete regional differentiation of European breeds, but two genetic clusters representing Northern Europe and the British-Irish Isles were remarkably differentiated from the remaining European populations. In African breeds, we observed five main clusters: (1) North Africa, (2) West Africa, (3) East Africa, (4) South Africa, and (5) Madagascar. For European goats, no strong evidence of significant correlations between H_o and H_e and distance to Nevali Çori, the archeological site representing the Southern Anatolian domestication center, was found. In contrast, in African breeds we detected a significant gradient of diversity, which decreased with distance to Nevali Çori.

Conclusion: The detection of a genetic cline associated with distance to the domestication center in African but not in European goat breeds might reflect differences in the post-domestication dispersal process and subsequent migratory movements associated with the management of caprine populations from these two continents.

Background

Goats were domesticated 10,000 years before present (YBP) in the Fertile Crescent from distinct bezoar populations, a process that was dispersed in time and space but featured by connected human communities [1,2]. Neolithic goats showed considerable genetic structure associated with geography, so different gene pools were established when human populations with their livestock migrated to Europe, Asia, and Africa [2]. The potential routes of the post-domestication spread of livestock across Europe [3], Asia [4] and Africa [5] have been reported by several authors. Such a dispersal process may cause genetic clines characterized by a decrease in genetic diversity of livestock populations over geographical distance to the domestication center. In European goats, a gradual reduction of genetic diversity with increasing distance to the Fertile Crescent was observed, with two different clines being the south-east to north-west cline steeper than the Mediterranean east-to-west cline [7,8]. However, these studies used a limited number of microsatellite markers to investigate the patterns of genetic variation, which does not allow estimations of genetic diversity of individuals. In addition, they were exclusively focused on European goats, whereas combining of datasets from different laboratories is problematic [6]. Here, we have used Illumina Goat SNP50 BeadChip [9] data generated in the AdaptMap project [10] and other studies [11,12,13] to assess the existence of genetic clines associated with the post-domestication dispersal of goats in Europe and Africa.

Materials and methods

Genotype data

We used published data from Illumina Goat SNP50 BeadChip [9] of European and African goats generated in the Adaptmap project [14]. We excluded from our study crossbred populations, and we maintained the number of animals per breed in a range between 15 to 50 individuals (with the only exception of the Carpathian goat, N = 14) by using the "bite.representative.sampling" function of the BITE R package v.2 [15]. This tool preserves the variance structure of the original dataset, despite reducing the sample size to a user-defined number. We obtained a final dataset of 681 European and 953 African goats. The number of Asian goat populations investigated in the AdaptMap project is very limited, so they were not considered in our study.

In addition to the Adaptmap data, we also retrieved SNP50 [9] data from 473 Swiss goats from 10 different breeds [16]. Moreover, the Old Irish Goat Society based on Mulranny (https://oldirishgoat.ie) provided data from 383 Old Irish and Old English goats. With regard to African breeds, we retrieved previously published SNP50 data from Algerian (N=48; [11]), Sudanese (N=72; [12]), and South African (commercial and local breeds N=114; [13]) goats. In total, our data set contained genotype data from 1148 European and 1187 African goats belonging to 38 and 43 populations, respectively. Observed and expected heterozygosity measurements and geographic coordinates of all goat populations included in the current work are described in **Tables 1 and 2 and Fig. S1.** By using the PLINK v 1.9 software [17] and taking as a reference the goat ARS1 genome [18], the chromosome number, genomic position and name of each SNP were updated, resulting in the obtention of 49,376 single nucleotide polymorphisms (SNPs) for European goats and 49,056 SNPs for African goats. The PLINK v 1.9 software [17] was also used to merge different datasets and filter out uninformative markers *i.e.* (1) SNPs with minor allele frequencies (MAF) lower than 0.05, (2) SNPs with missing call rates higher than 0.05, (3) SNPs that did not adjust to the Hardy–Weinberg expectation (*P*-value ≤0.001) and (4) unmapped SNPs. Moreover, individuals with missing call rates higher than 0.1 were also excluded. After these filtering steps, the African and European data sets comprised 25,990 and 21,152 SNPs respectively. The final total data set (combined data sets of African and European breeds) contained, after filtering, 27,005 SNPs genotyped in 2,335 goats from 81 breeds.

Population structure analysis

We assessed population structure using PLINK v. 1.9 [17] to carry out a principal component analysis (PCA) and the R software v.4.1.3. was employed for visualizing the resulting plot. Moreover, population structure was investigated with the ADMIXTURE v.1.3.0 package [19] with number of clusters (K) varying from 2 to 10. To assess the quality of the clustering process and thus infer the most likely K, we estimated the cross-validation error for each K. To visualize the results of the Admixture analysis, we used the *pophelper* R package [20].

Correlating genome-wide diversity with distance to the domestication center

We used Arlequin v. 3.5.2.2 [21] to calculate observed (H_o) and expected (H_e) heterozygosities. We have chosen Nevali Çori on the middle Euphrates as a location representative of the geographic coordinates of the goat domestication center in the Fertile Crescent because substantial archaeological and genetic evidence of ancient goat domestication (10,500–10,000

Y BP) has been collected in this early Neolithic settlement. Besides, Southeastern Anatolia has been proposed to be the likely origin of almost all modern domestic goats [22]. To calculate geographic distances (in kilometers) from the sampling site of each breed to Nevali Çori (Turkey latitude = 37.52º N and longitude= 38.61º E), we have used the latitude and longitude coordinates provided by Colli et al. [14], Stella et al.[10], Ouchene-Khelifi et al. [11]. The sampling site lists of the South African and Algerian populations are available in Chokoe et al., [13] and Rahmatalla et al., [12], respectively, and we have searched for the corresponding coordinates in the opensource databases available online (https://www.latlong.net/). For Swiss [16], Irish and British (Old Irish Goat Society, https://oldirishgoat.ie) breeds, we used coordinates of the capital cities of respective countries of origin. Geographical distances were obtained with the *geosphere* package [23] of the R software v.4.1.3. using the "*distVincentyEllipsoid*" method which considers the earth as an ellipsoid flattened at the poles, thus providing a very accurate calculation of distances [24]. Pearson correlation coefficients (*r*) were computed to assess if there is a linear relationship between H_o and H_e estimates and geographical distances to the Nevali Çori domestication site by using the *stats* package included in the R software v.4.1.3 [25]. Linear regressions were plotted with the *ggplot2* package of R software v.4.1.3. For each continent, we did two separate analyses including or not including insular populations. The reason for not including insular populations is that they usually have reduced levels of diversity due to geographic isolation rather than to ancient post-domestication events [26]. In the case of African populations, we excluded from our analysis goats from the Boer, Savanna and Kalahari Red breeds because there is evidence that they have been strongly introgressed with Asian blood, so they are not representative of South African indigenous local goats [27].

In addition, expected and observed heterozygosity values computed for each population were used to construct interpolation maps drawn using the inverse distance weighted (IDW) option implemented in the GIS software ArcGIS v. 3.0.3 (https://www.arcgis.com/index.html ESRI, Redlands, CA, United States). This deterministic method of multivariate interpolation considers a set of scattered points with known values for a variable and calculates the values of the variable for points with missing values by taking into account the weighted average of the values available at the known points. The measured values closest to the location to be predicted have more influence on the predicted value than those farther away. The sampling area of each population was used as geographic coordinates and interpolation surfaces were divided into ten equal classes.

Results and Discussion

Population structure analysis

We analyzed the population structure of the European and African goats through Admixture (**Fig. 1**) and principal component analysis (PCA, **Fig. 2**). Regarding European breeds, we observed an incomplete regional differentiation, but a clear regional clustering with separate positions for Northern Europe (Denmark, The Netherlands and Finland), Great Britain and Ireland. Strong differences in autosomal SNP as well as chromosome Y haplotype frequencies have been observed when comparing Northern and Southern European goats [27,14] and we detect the same trend in the PCA shown in **Fig. S2** with the 50° latitude dividing northern and southern European goats.. This pattern can be explained partially by the post domestic spreading across Europe through two main routes: the Mediterranean route, which involved the maritime transportation of livestock along the Mediterranean basin until reaching the Iberian Peninsula 7,300-7,700 YBP, and the Danubian route, which traversed the European mainland and reached Scandinavia and the British Isles 4,000 YBP [3]. This may very well have established the well-differentiated gene pools in Southern and Northern Europe.

For African goats, we observed five main clusters representing populations from South, West, North and East Africa plus a fifth Madagascar group **(Fig.S3),** which was supported by the Admixture analysis (**Fig.1**) and agrees with the findings of Colli et al. [14]. The important geographic (e.g. Sahara and Kalahari deserts) and biological (e.g. Tsetse fly belt) barriers may have contributed substantially to the genetic differentiation of goat populations from West, East, North and South Africa. In the case of Malagasy goats, their genetic differentiation from continental populations is probably explained not only by their insular origin but also because their ancestry might be, at least in part, Austronesian [27]. Finally, Palmera goats cluster with the West African breeds because they were transported to the Canary Islands by settlers of Amazigh origin 2,000-2,500 YBP [29].

Diversity of European goat populations is not correlated with distance to the domestication site.

We investigated whether genetic diversity (expressed as observed and expected heterozygosity) of African and European populations correlates with distance from their sampling location to Nevali Cori. When analyzing goat populations from Europe (**Figs. 3, Figs. S4**), we obtained negative and significant correlations (H_o: r = -0.44, *P*-value= 0.005, **Fig. 3b**) for observed and expected heterozygosities (H_e: r = -0.4, *P*-value= 0.011, **Fig. S4b**). However, both correlations

became non-significant (H_o: $r = -0.23$, P-value = 0.21, **Fig. 3a**; H_e: $r = -0.16$, P-value = 0.37, **Fig. S4a**) when British and Irish populations were removed from the European data set. Indeed, the majority of European breeds displayed moderate to high H_e (**Fig. 3b**), with the exception of the populations from United Kingdom ($H_e = 0.317$) and Ireland ($H_e = 0.359$). Even the Spanish Bermeya and Malagueña breeds, which are located very far apart from Nevali Çori, displayed high expected heterozygosity values ($H_e = 0.4$ in Bermeya and $H_e = 0.42$ in Malagueña). This result is do not agree with Cañón et al. [7] describing a decrease in caprine genetic diversity from the south-east to the north-west of Europe. These discrepancies could be due to experimental factors, but also by a different breed representation within the previous study a better representation of Greek and Albanian breeds. The significant gradient that we observe when British and Irish populations are included in the analysis is might be due to their strong demographic recession [30], which is reflected in a high levels of homozygosity [26]. However, we cannot rule out the possibility that the low diversity of British and Irish cattle is partly explained by one or more founder effects associated with the arrival of livestock to the United Kingdom and Ireland 5,800-6,000 YBP, as suggested for British cattle [37].

Although the post-domestication spread of goats through the European continent might have generated an initial gradient of diversity, our data from modern breeds do not allow us to detect it. This may very well be due to post-domestication migratory movements associated with trading and herding. Indeed, the Mediterranean Sea has throughout the millennia been a facilitated exchange of goods and livestock, via a dense network of commercial maritime routes connecting distant port cities within and outside Europe. Indeed, Cardoso et al. [26] reported that goats from Mediterranean islands have lower levels of homozygosity than those from remote islands as Iceland, La Palma or Madagascar. In addition, the Great European Plain, which is one of the largest continuous expanses of plain on the Earth's surface, also may have facilitated the exchange of goats and other livestock amongst distant locations within Europe. More recently, the widespread use of improved breeds (e.g., Saanen, Toggenburg and Alpine), and artificial insemination may also have contributed to increasing gene flow between distant European populations. Besides, there is evidence that these highly productive cosmopolitan breeds have introgressed many local breeds in Europe [27].

Detection of a significant gradient of diversity associated to distance to the domestication center in African goats.

In contrast with European goats, significant negative correlations between the diversity of African caprine populations and distances to Nevali Çori were observed in the datasets with

(Madagascar and La Palma) and without islands (**Figs. 3, Fig.S4**). Indeed, we obtained r coefficients of -0.35 (H_o, *P*-value = 0.033) and -0.39 (H_e, *P*-value = 0.018) in the dataset with no islands (**Fig.3c, Fig.S4c**) and r coefficients of -47 (H_o, *P*-value = 0.0022) and -0.49 (H_e, *P*-value = 0.0012) in the dataset with islands (**Fig.3d, Fig.S4d**). The Egyptian, Algerian, and Sudanese populations, which are closest to the center of domestication, show the highest heterozygosity values (see **Table 2**). When proceeding southwards and particularly south-eastwards, diversity decreases as evidenced in goat breeds from Mozambique (H_o=0.319; H_e=0.335) and Malawi (H_o=0.339; H_e=0.368), and particularly in the island of Madagascar (H_o=0.29; H_e=0.30). With regard to indigenous South African breeds, their diversity is high (H_o=0.409; H_e=0.43), probably because many of these breeds have been introgressed by Boer goats. The Boer breed as a mixed Asian and African ancestry [27], and there is evidence that Anglo-Nubian bucks contributed to its foundation [4].

The detection of a gradient of diversity associated with distance to the domestication site in African goats might be because the spread of goats throughout the African continent probably took place mostly by land rather than by sea, with the only exception of the North African shoreline where maritime diffusion throughout the Mediterranean Sea was important [34]. Indeed, archaeological sites in Lybia, Algeria and Morocco contain remains of impressed pottery, crop plants and sheep, goats and cattle [38]. Notably, the surfaces of Europe and Africa are about 30,370,000 km^2 and 10,180,000 km^2, respectively, while their coastal lines are 30,000 km (Africa) and 143,000 km (Europe) long [35]. This means that the inner parts of the European continent are more easily accessible by sea than African inland. While nowadays almost 50% of the population of the European Union lives less than 50 km from the sea (https://www.eea.europa.eu/themes/water/europes-seas-and-coasts/europes-seas-and-coasts#:~:text=The%20EU%20coastline%20is%2068,to%20185%20000%20km%20long).

One third of African countries are landlocked. While the overland dispersal of livestock is expected to take place through a series of founder effects, thus generating a clear gradient of diversity, this is not necessarily true when domestic animals are transported by sea. In such case, it is more likely to observe a leap-frog pattern of diffusion that does not necessarily result in a genetic cline.

There is evidence that goats and sheep entered North Africa through the Sinai Peninsula as well as through the Mediterranean Sea [5, 27], coinciding with the opening of a grassland niche in the Sahara that was gradually occupied by pastoral communities [5]. The increasing aridity of the Sahara around 4,500 YBP and the consequent southward retreat of the Tsetse fly belt favored the migration of herders towards the Sahel. However, the entry of livestock into West

and East Africa takes place not before than 3,500 BP and even later [**Errore. L'origine riferimento non è stata trovata.**], possibly because of lack of immunity to native diseases. Goat and sheep remains dating back to 2,400 YBP and 2,100 YBP have been found at the sites of Salumano (Zambia) and Bamba (Zimbabwe), proving that the arrival of small ruminants to southern Africa is quite recent [5]. This might have involved migrations through and along the coastal areas of the Congo Basin or facilitated by the opening of tsetse corridors along the highland of the Rift Valley [38].

As shown in **Figs. 4**, goats from Central and East Africa are less diverse than their Northern counterparts, possibly because the Sahara Desert, which covers 9.1 million km^2, constitutes a formidable geographical barrier to the southwards spread of pastoral communities [36] and their livestock. Moreover, Central Africa overlaps with the Tsetse fly belt, which covers a geographic area of 10 million km^2, between latitudes 14° N and 20° S, representing about one third of the African continent. Trypanosomiasis is a protozoan disease which causes anemia, fever, and weight loss and sometimes can be fatal [31]. This may have limited the diffusion and exchange of caprine stocks susceptible to this parasite in Tsetse fly infested areas. Interestingly, Traorè and coworkers showed that the presence of the Tsetse fly influences the genetic variability of goats from Burkina-Faso, and they demonstrated that trypanosomiasis might have acted as a landscape boundary both for the spread of trypanosensitive goats and for strong selection pressure on trypanotolerant goats in infested areas [32].

We have detected a high variability of several South African indigenous breeds even though this region remained considerably isolated from Asia and Europe [33]. We excluded from the gradient analysis South African commercial goats (Boer, Kalahari Red and Savanna) because it is well known that Boer goats have a mixed African and Asian ancestry [14, 27], and that Kalahari and Savanna goats have a strong Boer component. We kept in our analysis indigenous communal populations sampled in the main goat-producing provinces of South Africa (Limpopo, Freestate, Gauteng, Northwest), which happened to have high levels of heterozygosity. This could be done to the fact that these South African populations have been also introgressed to some extent by Boer goats as well as by goats from European origin. Indeed, the establishment, in South Africa, of British and Dutch farmers, during the 17th-19th centuries, promoted the development or importation of highly productive breeds to improve the local stocks.

Conclusions

Our data indicate that a genetic cline associated with distance to the domestication center in Southern Anatolia can be observed in African goats but not in their European counterparts. Given the physical and geographic features of Europe and Africa, it is reasonable to assume that the maritime post-domestication diffusion of goats and other livestock was much more important in Europe than in Africa. Besides, about two thirds of the African continent are occupied by two formidable geographic (Sahara Desert) and biological (Tsetse fly belt) barriers that restrict substantially the transportation of livestock from one area to another. In this content, the migratory movements of goats and other domestic animals that took place after domestication were probably much more intense, sustained, and recurrent in Europe than in Africa, leading to the erasure of the post-domestication gradient of diversity in the former and to its preservation in the latter.

List of abbreviations

YBP: Years Before Present
SNP: Single Nucleotide Polymorphism
MAF: Minor Allele Frequency
PCA: Principal Component Analysis
H_o: observed heterozygosity
H_e: expected heterozygosity
IDW: Inverse Distance Weighted

Declarations

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All relevant data are included in the manuscript and its additional file. Genotype datasets can be accessed at: https://bridgeurl.com/goat-genotype-data-2
Genotype data from Old Irish and Old English goats can be requested at https://oldirishgoat.ie.

Competing interests

Funding

Authors' contributions

MA, MLD, MP and GMV designed the experiment. AdaptMap Consortium members, Ama, FG and AMo performed sampling activities. All members of the AdaptMap Consortium, AMM, FG and AM were involved in the genotyping of goats. EP and MGL analyzed the data. EP and MA drafted the manuscript with the help of FG. All authors revised and approved the manuscript.

Acknowledgements

Goat Consortium, Italy; Donata Marletta, University of Catania, Italian Goat Consortium, Italy; Paola Crepaldi, University of Milan, Italian Goat Consortium, Italy; Michele Ottino, Parco Nazionale del Gran Paradiso, Italy; Ettore Randi, ISPRA Istituto Superiore per la Protezione e la Ricerca Ambientale, Italy; Badr Benjelloun, INRA Maroc, Morocco; Hans Lenstra, University of Utrecht, The Netherlands; Muhammad Moaeen-ud-Din and Jim Reecy, PMAS-Arid Agriculture University, Pakistan; Isabel Alvarez, Área de Genética y Reproducción Animal del Serida, Spain; Marcel Amills and Armand Sànchez, Centre for Research in Agricultural Genomics (CRAG), Spain; Juan Capote, Instituto Canario de Investigaciones Agrarias (ICIA), Spain; Jordi Jordana, Universitat Autònoma de Barcelona, Spain; Agueda Pons, Serveis de Millora Agrària i Pesquera (SEMILLA), Illes Balears, Spain; Amparo Martínez and Antonio Molina, University of Córdoba, Spain; Benjamin Rosen, USDA/ARS, USA; Carina Visser, Faculty of Natural and Agricultural Sciences, South Africa; Cord Drögemüller, University of Bern, Switzerland; Gordon Luikart, Montana Conservation Genomics lab (MCGL), Division of Biological Sciences, University of Montana, Missoula, USA; Clet Wandui Masiga, Tropical Institute of Development Innovations (TRIDI), Uganda; Denis Fidalis Mujibi, International Livestock Research Institute (ILRI), Nairobi Kenya; Hassan Ally Mruttu, Ministry of Livestock and Fisheries Development, Tanzania; Timothy Gondwe, Department of Animal Science, Lilongwe University of Agriculture and Natural Resources, Malawi; Joseph Sikosana, Department of Research and Specialist Services, Division of Livestock Research, Zimbabwe; Maria Da Gloria Taela, Animal Production Institute, Ministry of Agriculture, Maputo, Mozambique; Oyekan Nash, National Biotechnology Development Agency, Nigeria.

References

1. Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. Science. 2000;287(5461):2254-7.

2. Daly KG, Maisano Delser P, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science. 2018;361(6397):85-8.

3. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc Natl Acad Sci USA. 2008;105(33):11597-604.

4. Porter V, Alderson L, Hall SJG. & Sponenberg DP. Goats. In: Mason's World Encyclopedia of Livestock Breeds and Breeding (volume 1). CABI. Oxfordshire:UK; 2016. p. 350-352.

5. Smith AB. Origins and Spread of Pastoralism in Africa. Annu. Rev. Anthropol. 1992; 21, 125-141.

6. FAO. Genomic characterization of animal genetic resources. Rome: FAO Animal Production and Health Guidelines. No. 32; 2023

7. Cañón J, García D, García-Atance MA, Obexer-Ruff G, Lenstra JA, Ajmone-Marsan P, et al. Geographical partitioning of goat diversity in Europe and the Middle East. Anim Genet. 2006;37(4):327-34.

8. Lenstra JA, Tigchelaar J, Biebach I, Hallsson JH, Kantanen J, Nielsen VH, et al. Microsatellite diversity of the Nordic type of goats in relation to breed conservation: how relevant is pure ancestry? J Anim Breed Genet. 2017;134(1):78-84.

9. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52K SNP chip for goats. PLoS One. 2014;9(1): e86227.

10. Stella A, Nicolazzi EL, Van Tassell CP, Rothschild MF, Colli L, Rosen BD, et al. AdaptMap: exploring goat diversity and adaptation. Genet Sel Evol. 2018;50(1):61.

11. Ouchene-Khelifi N-A, Lafri M, Pompanon F, Ouhrouch A, Ouchene N, Blanquet V, et al. Genetic homogeneity of North-African goats. PLoS One. 2018;13(8): e0202196.

12. Rahmatalla SA, Arends D, Reissmann M, Said Ahmed A, Wimmers K, Reyer H, et al. Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits. BMC Genet. 2017;18(1):92.

13. Chokoe TC, Mdladla-Hadebe K, Muchadeyi F, Dzomba E, Matelele T, Mphahlele T, et al. Genetic diversity of South African indigenous goat population from four provinces using genome-wide SNP data. Sustainability. 2020;12(24):10361.

14. Colli L, Milanesi M, Talenti A, Bertolini F, Chen M, Crisà A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet Sel Evol. 2018;50(1):58.

15. Milanesi M, Capomaccio S, Vajana E, Bomba L, Garcia JF, Ajmone-Marsan P, et al. Bite: An R package for biodiversity analyses. bioRxiv 181610; 2017.

16. Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, et al. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. Anim Genet. 2016;47(6):727-39.

17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience. 2015; 4:7.

18. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the Domestic Goat Genome. Nat. Genet. 2017;49(4):643-650.

19. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-64.

20. Francis RM. POPHELPER: an R package and web app to analyse and visualize population structure. Mol Ecol Resour. 2016;17(1):27-32.

21. Excoffier L, Lischer HE. Arlequin Suite Ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564-7.

22. Naderi S, Rezaei HR, Pompanon F, Blum MG, Negrini R, Naghash HR, et al. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc Natl Acad Sci USA. 2008;105(46):17659-64.

23. Karney CF. Algorithms for geodesics. J Geod. 2012;87(1):43-55.

24. Vincenty T. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. Surv Rev. 1975; 23 (176): 88-93.

25. Freedman D, Pisani R & Purves R. Instructor's Manual for Statistics, third Edition. New York: W.W. Norton; 1998.

26. Cardoso TF, Amills M, Bertolini F, Rothschild M, Marras G, Boink G, et al. Patterns of homozygosity in insular and continental goat breeds. Genet Sel Evol. 2018; 50:56.

27. VarGoats Consortium, Nijman IJ, Rosen BD, Bardou P, Faraut T, Cumer T, Daly KG, et al. Geographical contrasts of Y-chromosomal haplogroups from wild and domestic goats reveal ancient migrations and recent introgressions. Mol Ecol. 2022;31(16):4364-80.

28. Pereira F, Amorim A. Origin and spread of Goat Pastoralism. eLS. 2010.

29. Fregel R, Ordóñez AC, Santana-Cabrera J, Cabrera VM, Velasco-Vázquez J, Alberto V, et al. Mitogenomes illuminate the origin and migration patterns of the indigenous people of the Canary Islands. PLoS ONE. 2019;14(3): e0209125.

30. Cassidy LM, Teasdale MD, Carolan S, Enright R, Werner R, Bradley DG, et al. Capturing goats: documenting two hundred years of mitochondrial DNA diversity among goat populations from Britain and Ireland. Biol Lett. 2017;13(3):20160876.

31. Ebhodaghe F, Ohiolei JA, Isaac C. A systematic review and meta-analysis of small ruminant and porcine trypanosomiasis prevalence in sub-Saharan Africa (1986 to 2018). Acta Trop. 2018; 188:118-31.

32. Traoré A, Álvarez I, Fernández I, Pérez-Pardal L, Kaboré A, Ouédraogo-Sanou GM, et al. Ascertaining gene flow patterns in livestock populations of developing countries: A case study in burkina faso goat. BMC Genet. 2012;13(1):35.

33. Martínez A, Manunza A, Delgado JV, Landi V, Adebambo A, Ismaila M, et al. Detecting the existence of gene flow between Spanish and North African goats through a coalescent approach. Sci Rep. 2016; 6:38935.

34. Pereira F, Queirós S, Gusmão L, Nijman IJ, Cuppen E, Lenstra JA et al. Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. Mol. Biol. Evol. 2009; 26(12): 2765-2773.

35. Stanners D & Bourdeau P. Europe's Environment: The Dobríš Assessments. European Environment Agency (EEA). Luxembourg: Office for Official Publications of the European Communities; 1995.

36. Nikita E, Mattingly D, Lahr MM. Sahara: Barrier or corridor? Nonmetric cranial traits and biological affinities of North African late Holocene populations. Am. J. Phys. Anthropol. 2012;147(2):280-292.

37. Cummings V & Morris J. Neolithic Explanations Revisited: Modelling the Arrival and Spread of Domesticated Cattle into Neolithic Britain. Environ. Archaeol. 2022; 27(1): 20-30.

38. Zeder MA. Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In: Boivin N, Crassard R, Petraglia M, editors. Human Dispersal and Species Movement: From Prehistory to the Present. Cambridge: Cambridge University Press; 2017. p. 261–303.

39. Muigai AWT & Hanotte O. The Origin of African Sheep: Archaeological and Genetic Perspectives. Afr. Archaeol. Rev. 2013; 30(1): 39–50.

Table 1. Observed and expected heterozygosities and geographic coordinates of the European goat breeds (ordered by country of origin) analyzed in the current work (distances between sampling locations and Nevali Çori are indicated in km).

ID_Breed	Breed	Country	Ho	He	Long.	Lat.	N	Distance (km)
LNR_DK	Landrace Goat	Denmark	0.366	0.389	11.44	55.56	50	3348.50
LNR_FI	Landrace Goat	Finland	0.373	0.381	22.56	62.77	20	2845.86
LNR_NL	Landrace Goat	Netherlands	0.366	0.359	5.12	52.09	15	3840.83
ENG	Old English Goat	United Kingdom	0.289	0.317	-0.13	51.50	32	4374.27
CRS	Corse	France	0.401	0.409	9.00	42.19	29	3179.32
FSS	Fosses	France	0.393	0.403	-1.12	47.97	24	4385.37
PTV	Poitevine	France	0.381	0.386	0.35	46.50	27	4194.53
PVC	Provençale	France	0.412	0.408	4.01	46.28	17	3792.22
PYR	Pyrenean	France	0.380	0.395	0.52	43.33	26	4125.19
IRL	Old Irish Goat	Ireland	0.342	0.359	-9.78	53.90	50	5473.27
ARG	Argentata	Italy	0.416	0.419	15.13	38.02	24	2481.34
ASP	Aspromontana	Italy	0.402	0.413	15.91	37.99	23	2395.12
BIO	Bionda dell'Adamello	Italy	0.405	0.412	10.36	46.03	24	3099.06
CCG	Ciociara Grigia	Italy	0.399	0.418	13.82	41.61	16	2642.46
DIT	Di Teramo	Italy	0.396	0.380	13.37	42.38	19	2701.53
GAR	Garganica	Italy	0.416	0.393	15.57	40.70	15	2440.72
GGT	Girgentana	Italy	0.371	0.378	14.17	37.61	24	2588.76
MLT	Maltese	Italy	0.371	0.389	14.36	37.60	16	2567.79
NIC	Nicastrese	Italy	0.400	0.416	16.45	38.93	20	2334.81
ORO	Orobica	Italy	0.370	0.371	9.50	46.04	22	3192.16
RME	Rossa Mediterranea	Italy	0.415	0.403	15.57	40.70	30	2440.72
SAR	Sarda	Italy	0.403	0.414	9.22	39.71	27	3136.42
VAL	Valdostana	Italy	0.376	0.391	7.38	45.71	24	3414.35
VSS	Valpassiria	Italy	0.402	0.415	11.21	46.80	24	3027.32
CRP	Carpathian goat	Romania	0.424	0.429	25.78	46.12	14	1482.60
BEY	Bermeya	Spain	0.405	0.404	-5.26	43.34	23	4761.09
MAL	Mallorquina	Spain	0.370	0.390	3.03	39.55	18	3820.21
MLG	Malagueña	Spain	0.415	0.417	-4.42	37.07	40	4645.51

|					Lon.	Lat.	N	Distance
RAS	Blanca de Rasquera	Spain	0.382	0.396	0.61	41.00	20	4093.98
ALP_CH	Alpine	Switzerland	0.395	0.395	7.67	46.95	50	3411.65
APP	Appenzell	Switzerland	0.368	0.363	7.88	46.88	29	3387.28
CHA	Swiss Chamois	Switzerland	0.396	0.394	7.88	46.88	50	3387.28
GST	Grisons Striped	Switzerland	0.395	0.390	7.88	46.88	49	3387.28
NVE	Nera Verzasca	Switzerland	0.386	0.388	7.88	46.88	42	3387.28
PEA	Peacock	Switzerland	0.400	0.391	7.88	46.88	31	3387.28
SAA	Saanen	Switzerland	0.382	0.378	7.88	46.88	50	3387.28
SGB	Booted goat	Switzerland	0.387	0.381	7.88	46.88	23	3387.28
TGR	Tessin grey goat	Switzerland	0.397	0.396	7.88	46.88	37	3387.28
TOG	Toggenburg	Switzerland	0.369	0.365	7.88	46.88	31	3387.28
VAG	Valais	Switzerland	0.366	0.366	7.88	46.88	43	3387.28

N. sample size of each breed; Ho. Observed heterozygosity; He. Expected heterozygosity; Lon. longitude in degrees; Lat. latitude in degrees; Distance. distance (in kilometres) between Nevali Çori) and European sampling locations calculated with the Vincenty ellipsoid-model method (Nevali Çori: lon 38.61°, lat 37.52°).

Table 2. Diversity and geographic coordinates of the African goat breeds (ordered by country of origin) analyzed in the current work (distances between sampling locations and Nevali Çori are indicated in km).

ID Breed	Breeds	Country	Ho	He	Long	Lat	N	Distance (km)
ALG	Arabia, Makatia, M'Zabite, Kabyle	Algeria	0.416	0.433	1.67	28.03	48	3591.08
SAH	Sahel	Burkina Faso	0.380	0.384	-0.39	14.73	15	4595.33
BUR	Burundi goat	Burundi	0.355	0.361	29.83	-2.91	50	4566.82
CAM	Cameroon Goat	Cameroon	0.369	0.377	14.39	10.11	37	3886.49
WAD_CM	West African Dwarf	Cameroon	0.351	0.361	10.27	5.9	31	4528.38
PAL	Palmera	Canary Islands	0.338	0.340	-17.69	28.66	15	5268.58
BRK	Barki	Egypt	0.398	0.407	26.9	29.89	50	1374.09
OSS	Oasis	Egypt	0.369	0.398	29.2	26.17	50	1539.47
SID	Saidi	Egypt	0.385	0.403	31.58	26.24	50	1415.15
ABR	Abergelle	Ethiopia	0.375	0.378	38.83	13.33	49	2679.91
GUM	Gumez	Ethiopia	0.380	0.383	36.2	12.97	39	2730.15
KEF	Keffa	Ethiopia	0.362	0.374	37	7.42	44	3337.48
WYG	Woyito Guji	Ethiopia	0.376	0.379	37.48	5.25	39	3575.34
GAL	Galla	Kenya	0.382	0.384	37.66	2.01	23	3932.97
SEA	Small East African	Kenya	0.351	0.378	36.97	0.61	30	4090.04
MEN	Malagasy goat (Menabe)	Madagascar	0.299	0.304	45.13	-20.16	19	6420.68
SOF	Malagasy goat (Sofia)	Madagascar	0.305	0.327	47.67	-16.74	22	6080.27
DZD	Dedza	Malawi	0.339	0.368	34.33	-14.37	15	5760.74
GUE	Guera	Mali	0.388	0.375	-9.19	14.54	16	5344.35
PEU	Peulh	Mali	0.381	0.378	-4.2	14.5	22	4926.02
SDN	Soudanaise	Mali	0.376	0.378	-6.27	13.45	22	5172.61
TAR	Targui	Mali	0.378	0.383	-0.05	16.27	19	4457.55
MOR	Barcha, Draa, Ghazalia, Morrocan, Noire de l'Atlas,Nord	Morocco	0.384	0.407	-7.17	31.09	30	4231.84

			Ho	He	Lon	Lat	N	Distance
LND	Landin	Mozambique	0.319	0.335	32.36	-25.5	29	7006.05
RSK	Red Sokoto	Nigeria	0.367	0.385	8.17	11.89	19	4152.58
SHL	Sahel	Nigeria	0.382	0.389	8.73	11.25	19	4165.76
WAD	West African Dwarf	Nigeria	0.357	0.367	3.74	7.59	19	4830.76
BOE	Boer	South Africa	0.397	0.405	28.19	-25.75	22	7087.49
KHAR	Kalahari Red	South Africa	0.401	0.406	20.15	-25.26	27	7212.41
SAV	Savanna	South Africa	0.412	0.411	23.63	-29.07	39	7534.84
SAFR	South Africa Local breeds (from Limpopo, Freestate, Gauteng, Northwest)	South Africa	0.409	0.430	26.22	-29.12	26	7488.93
DESE	Desert	Sudan	0.419	0.418	30.37	13.7	24	2762.43
NI	Nilotic	Sudan	0.408	0.416	32.67	13.17	24	2761.25
TAGG	Taggar	Sudan	0.411	0.409	29.65	12.05	24	2959.39
MAA	Maasai	Tanzania	0.378	0.381	36.65	-11.38	18	5416.39
PRW	Pare White	Tanzania	0.370	0.375	37.92	-4.25	19	4624.53
SNJ	Sonjo	Tanzania	0.382	0.375	36.32	-2.7	20	4458.85
TUN	Tunisian	Tunisia	0.400	0.409	9.14	35.74	21	2632.08
KAR	Karamonja	Uganda	0.382	0.385	34.67	2.53	19	3895.15
MUB	Mubende	Uganda	0.368	0.378	32.29	0.44	18	4156.35
SEB	Sebei	Uganda	0.380	0.378	34.45	1.4	21	4021.87
MSH	Mashona	Zimbabwe	0.345	0.359	31.1	-18.5	22	6250.11
MTB	Matebele	Zimbabwe	0.401	0.399	28.51	-20.55	22	6513.89

N, sample size of each breed; Ho, Observed heterozygosity; He, Expected heterozygosity; Lon, longitude in degrees; Lat, latitude in degrees; Distance, distance in kilometres between Nevali Çori: lon 38.61° lat 37.52° and African sampling locations
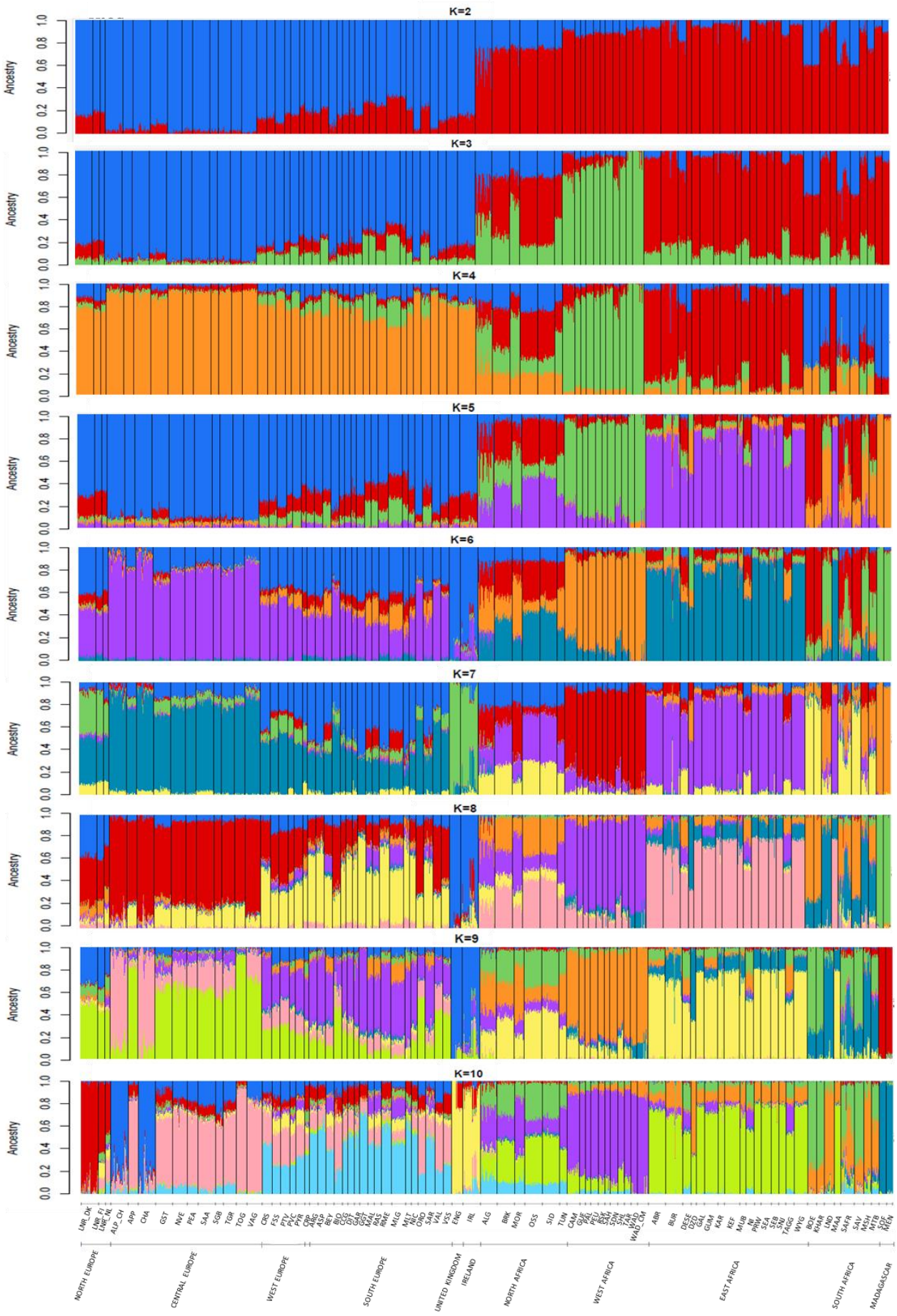
Paper 1-Figure 1. Admixture analysis of African and European goat breeds included in our study.

Legend: Each bar represents the percentages of global ancestries from one or more of K=2-10 genetically distinct sources for each individual. Continental subregions in Africa include the following countries: (1) Northern Africa: Morocco, Algeria, Tunisia, and Egypt, (2) Western Africa: Burkina Faso, Mali, Nigeria, Cameroon, and Canary Islands, (3) Eastern Africa: Sudan, Ethiopia, Uganda, Burundi, Kenya, Tanzania, and Malawi, (4) Southern Africa: Mozambique, Zimbabwe and South Africa, and (5) Madagascar. Continental subregions in Europe include the following countries: (1) Northern Europe: Denmark, Finland, and The Netherlands, (2) Central Europe: Switzerland, (3) Western Europe: France, (4) Eastern Europe: Romania, (5) Southern Europe: Italy and Spain, and (6) United Kingdom and Ireland. African and European breeds and subregions names are reported.

Subregions

North Europe
LNR_DK
LNR_FI
LNR_NL

Central Europe
ALP_CH   SAA
APP   SGB
CHA   TGR
GST   TOG
NVE   VAG
PEA

West Europe
CRS
FSS
PTV
PVC
PYR

East Europe
CRP

South Europe
ARG   RAS
ASP   RME
BEY   MLG
BIO   MLT
CCG   NIC
DIT   ORO
GAR   SAR
GGT   VAL
MAL   VSS

United Kingdom
ENG

Ireland
IRL

North Africa
ALG
BRK
MOR
OSS
SID
TUN

West Africa
CAM   SDN
GUE   SHL
PAL   TAR
PEU   WAD
RSK   SAH
WAD_CM

East Africa
ABR   MUB
BUR   NI
DESE PRW
DZD   SEA
GAL   SEB
GUM   SNJ
KAR   TAGG
KEF   WYG

South Africa
BOE   SAFR
KHAR SAV
LND   MSH
MAA   MTB

Madagascar
SOF
MEN

64

Paper 1-Figure 2. Principal Component Analysis plot of European and African breeds.

Legend: Principal components 1 and 2 and percentages of variance explained by them. Samples are coloured according to their continental subregion of sampling and represented by breed acronyms. The list of complete breeds names can be found in Tables 1 and 2.

a) European dataset without islands
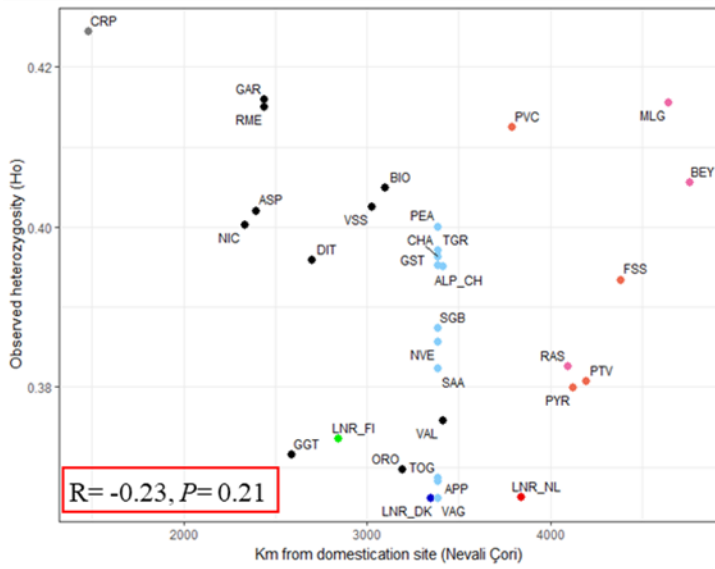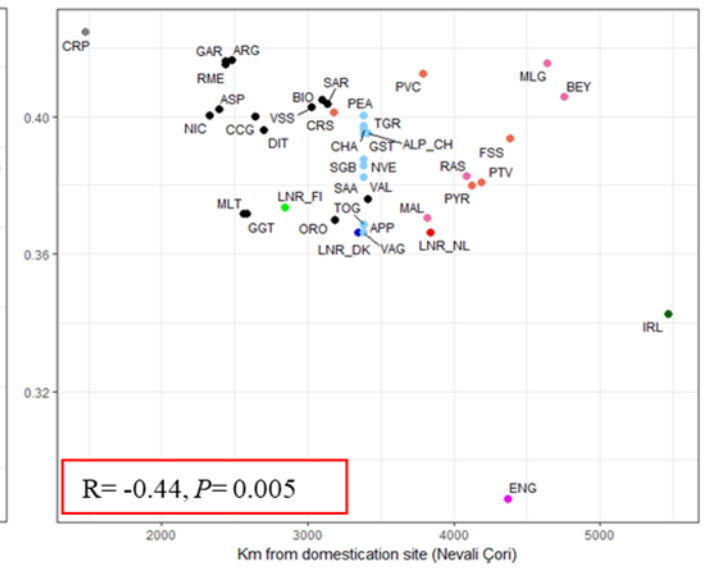
b) European dataset with islands

R= -0.23, *P*= 0.21

R= -0.44, *P*= 0.005

c) African dataset without islands

R= -0.35, *P*= 0.033

R= -0.47, *P*= 0.0022
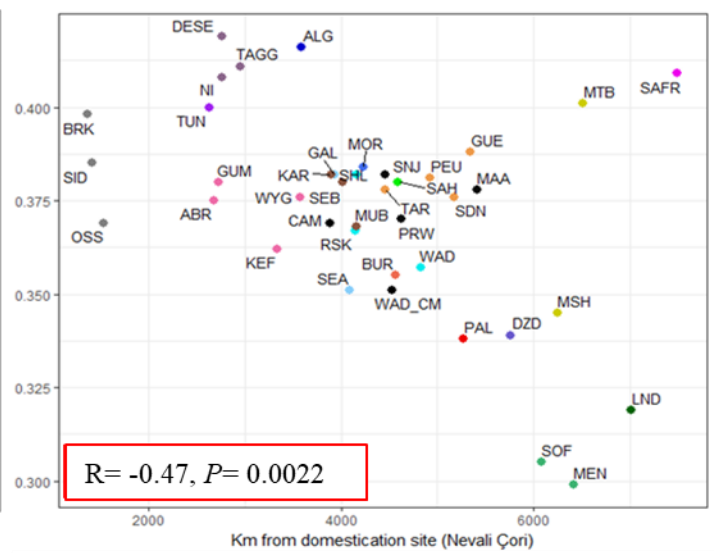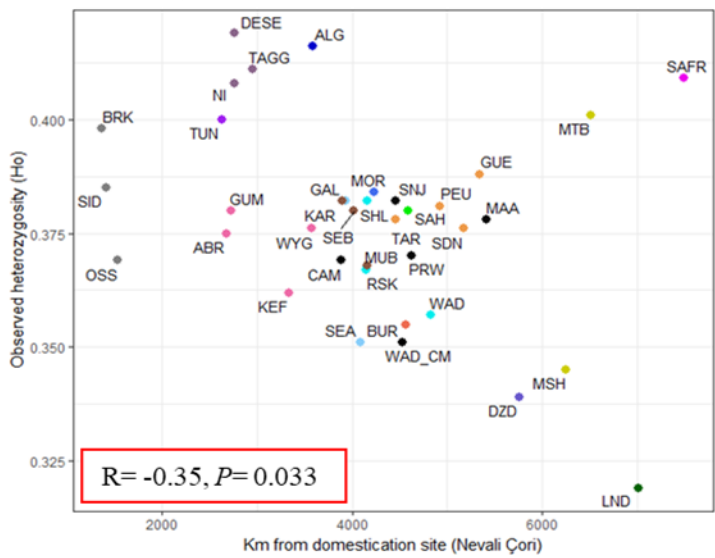
European sampling countries

- Denmark
- France
- Italy
- Romania
- Switzerland
- Finland
- Ireland
- Netherlands
- Spain
- United Kingdom

African sampling countries

- Algeria
- Canary Islands
- Madagascar
- Mozambique
- Tanzania
- Burkina Faso
- Egypt
- Malawi
- Nigeria
- Tunisia
- Burundi
- Ethiopia
- Mali
- South Africa
- Uganda
- Cameroon
- Kenya
- Morocco
- Sudan
- Zimbabwe

Paper 1-Figure 3. Graphs depicting the relationships between observed heterozygosities of European and African goat populations and distance between their sampling locations and Nevali Çori.

Legend Graphs depicting the relationships (expressed as Pearson correlations and their *P*-values) between observed heterozygosity and distance between Nevali Çori (early Neolithic settlement on the middle Euphrates selected as a geographic reference for the Fertile Crescent domestication site) and sampling locations of (a) European breeds not including insular populations, (b) European breeds including insular populations, (c) African populations not including insular populations, (d) African populations including insular populations. In all plots, country of origin is indicated with specific colours. Breed acronyms are listed in Tables 1 and 2.
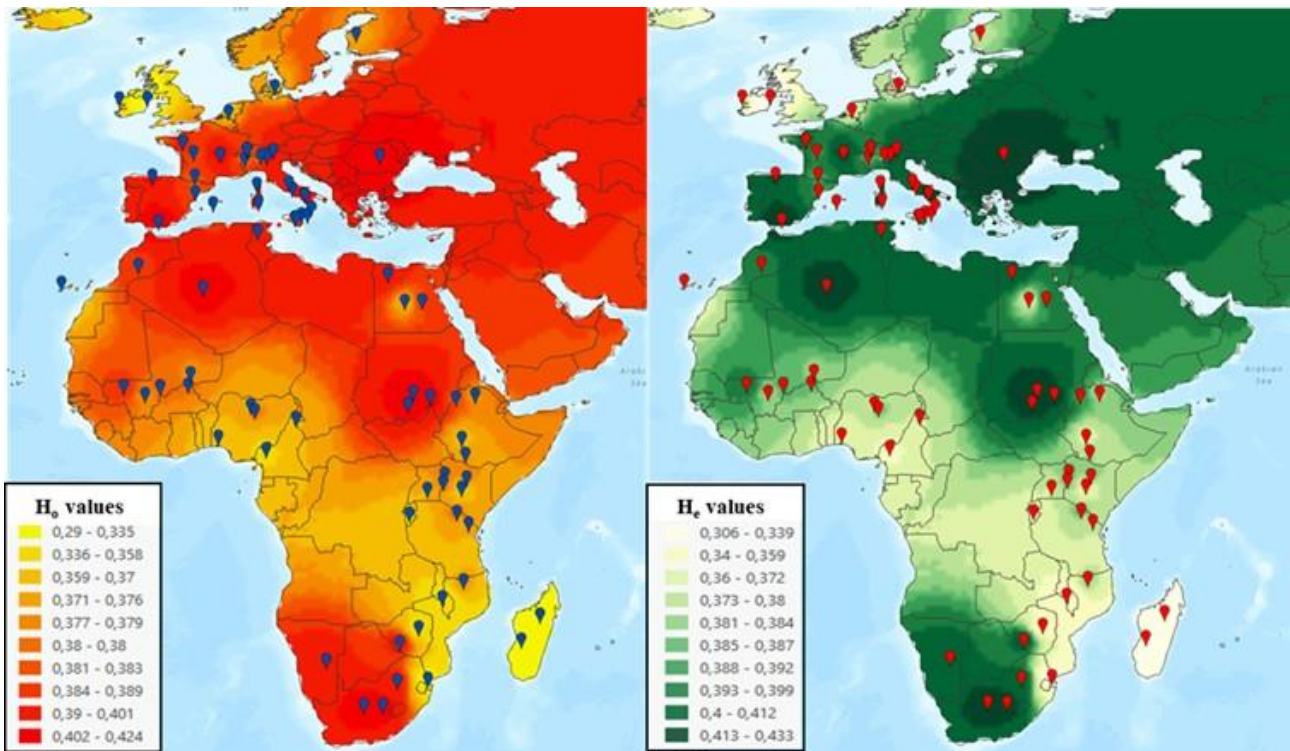
Paper 1-Figure 4. Interpolation maps showing the geographic distribution of observed and expected heterozygosities in African and European breeds.

Legend: Interpolation maps showing the distribution of genetic diversity in African and European breeds. a) observed heterozygosity, H_o. b) Expected heterozygosity, H_e. Blue and red points represent sampling localities in (a) and (b), respectively. In Europe, a reduction of diversity is evident in goats from the United Kingdom and Ireland, while in Africa low diversity coincides with the Tsetse fly belt (a geographic area comprised between latitudes 14° N and 20° S) and Madagascar.

Additional files

Additional file 1 Figures S1

Format: Additional file 1.ppt

Title: Geographical distribution of European (a) and African (b) breeds.

Description: It shows the points and the acronym name of the breed coloured according to the country of origin. The location of Nevali Çori is also indicated.

Additional file 1 Figure S2

Format: Additional file 1.ppt

Title: Principal Component Analysis (PCA) plot of the first two components for 38 European goat breeds.

Description: The percentage of variation explained by the two main components is shown in brackets. Individuals are coloured according to their subregion of sampling. Breed acronyms are as follows: ALP_CH= Alpine, APP=Appenzell, ARG= Argentata, ASP= Aspromontana, BEY= Bermeya, BIO= Bionda dell'Adamello, CCG= Ciociara Grigia, CHA= Swiss Chamois, CRP= Carpathian goat, CRS= Corse, MAL= Mallorquina, MLG= Malagueña, MLT=Maltese, NIC= Nicastrese, NVE= Nera Verzasca, ORO= Orobica, PEA= Peacock, PTV= Poitevine, PVC= Provençale, PYR= Pyrenean, DIT= Di Teramo, ENG= Old English Goat, FSS= Fosses, GAR= Garganica, GGT= Girgentana, GST= Grisons striped, IRL= Old Irish Goat, LNR_DK= Landrace Goat (Denmark), LNR_FI= Landrace Goat (Finland), LNR_NL= Landrace Goat (Netherlands), RAS= Blanca de Rasquera, RME= Rossa Mediterranea, SAA= Saanen, SAR= Sarda, SGB= Booted goat, TGR= Tessin grey goat, TOG= Toggenburg, VAG= Valais, VAL= Valdostana,VSS= Valpassiria.

Additional file 1 Figure S3

Format: Additional file 1.ppt

Title: Principal Component Analysis (PCA) plot of the first two components for the 43 African goat breeds.

Description: The respective percentage of variation explained by the two main components is shown in brackets. Individuals are coloured according to their subregions of sampling. Breed acronyms are as follows: ABR= Abergelle, ALG= Arabia,Makatia, and M'Zabite,Kabyle, BOE=Boer, BRK=Barki, BUR= Burundi goat, CAM= Cameroon Goat, DESE=Desert, DZD= Dedza, MAA= Maasai, MEN= Malagasy goat (Menabe), MOR= Barcha,Draa,Ghazalia, Moroccan goats, Noire de l'Atlas,Nord, MSH= Mashona, MTB= Matebele, MUB= Mubende, NI= Nilotic, OSS= Oasis, SDN= Soudanaise, SEA= Small East African, SEB= Sebei, SHL=

69

Sahel, SID= Saidi, SNJ=Sonjo, SOF= Malagasy goat (Sofia), TAGG= Taggar, GAL= Galla, GUE= Guera, GUM= Gumez, KAR= Karamonja, KEF= Keffa, KHAR= Kalahari Red, LND= Landin, PAL= Palmera, PEU= Peulh, PRW= Pare White, RSK= Red Sokoto, SAFR= South Africa Local breeds (from Limpopo, Freestate,Gauteng,Nortwest), SAH= Sahel, SAV= Savanna. TAR= Targui, TUN= Tunisian, WAD_CM= West African Dwarf (Cameroon), WAD= West African Dwarf (Nigeria), WYG= Woyito Guji.

Additional file 1 Figure S4

Format: Additional file 1.ppt

Title: Correlation graph between the distance (km) between the domestication site (Nevali Çori) and sampling locations of European and African breeds in relation to their expected heterozygosities.

Description: We report plots representing Pearson correlations (with their P-values) between expected heterozygosity and distance between Nevali Çori and sampling location of European breeds not including (a) and including (b) insular breeds; and of African populations not including (c) and including (d) insular populations. Breeds acronyms can be found in Tables 1 and 2, and their country of origin is indicated with coloured points.

3.2 PAPER 2

The African introgression of Murciano Granadina goats has a Moroccan origin and displays remarkable levels of inter-individual variability.

Petretto E [1,2], Luigi-Sierra MG, Vacca GM[2], Martínez A[4], Delgado JV, Fernández Álvarez J, Castelló A[1], Pazzola M[2], Jordana J, Dettori ML[2], Amills M[1,3*]

[1] Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [2] Department of Veterinary Medicine, University of Sassari,07100, Sassari, Italy. [3] Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [4] Departamento de Genética, Universidad de Córdoba, Campus Universitario de Rabanales, Córdoba, Spain.

*Corresponding author: marcel.amills@uab.cat

E-mail addresses:

EP, elena.petretto@cragenomica.es

MGLS, marigls.ml@gmail.com

GMV, gmvacca@uniss.it

AMA, ib2mamaa@uco.es

JVD, juanviagr218@gmail.com

JFA, j.fernandez@caprigran.com

AC, anna.castello@uab.cat

MP, pazzola@uniss.it

JJ, jordi.jordana@uab.cat

MLD, mldettori@uniss.it

MA, marcel.amills@uab.cat

Summary

There is evidence that Murciano-Granadina (MG), the most important caprine dairy breed in Spain, has been introgressed by African goats, but the precise geographic origin of such introgression has not been identified yet. Moreover, an accurate estimate of the magnitude of this African introgression is lacking, since current estimates are based on small numbers of sampled individuals. The aim of our work was to tackle these two issues by genotyping 500 MG goats with the Goat SNP50 BeadChip and comparing their genotypes with those of reference populations from Spain (*Bermeya*), Switzerland (*Saanen*), Morocco (*Barcha, Draa, Ghazalia, Noire de Atlas, Nord, Moroccan*), Egypt (*Barki, Oasis, Saidi*), Algeria (*Arabia, Makatia, M'Zabite, Kabyle*) and Sudan (*Desert, Nilotic, Taggar*). The population of 500 MG goats was subdivided into 10 datasets of 50 individuals to ensure that sample sizes of the target (MG) and reference populations are balanced. By performing a supervised ADMIXTURE analysis, we found that the Moroccan genetic component reached a proportion of $4.2 \pm 3.8\%$ in MG goats, while the Algerian ($0.01 \pm 0.01\%$), Egyptian ($0.2 \pm 0.2\%$) and Sudanese ($0.13 \pm 0.1\%$) components were present in extremely small proportions. The historical circumstances of this introgression event are currently unknown, but several plausible scenarios are outlined. Moreover, our results show considerable inter-individual heterogeneity regarding the magnitude of the Moroccan introgression of MG goats (0% to 12% depending on the MG data set under analysis). This result implies that reliable estimates about the exotic introgression of livestock breeds can only be obtained by extensively sampling target populations.

Keywords: Murciano Granadina; goat; admixture; introgression; high-density SNP arrays.

Main text

There is convincing evidence of the introgression of European goat breeds by African populations, being especially significant in Italy and Spain (Manunza et al. 2016; Martínez et al. 2016; Colli et al. 2018). One of the most important Spanish goat dairy breeds is Murciano-Granadina (MG), which is mostly distributed in Southern Spain. MG goats have black or mahogany coats, generally do not have horns and they display eumetric proportions, being well adapted to harsh climatic conditions with hot and dry summers and cold winters (https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas). Several studies have contributed evidence about the African introgression of MG goats (Manunza et al. 2016), but the precise geographic origin of such introgression has not been defined yet. The goal of our work was to answer this question as well as to estimate the amount of African ancestry in MG goats.

To achieve this goal, we have extracted blood samples from 500 MG goats, mainly distributed in 15 farms located in the autonomous region of Andalusia (Southern Spain), and we have purified genomic DNA by using a previously published salting-out protocol (Luigi-Sierra et al. 2020). Subsequently, these 500 goats have been genotyped with the Goat SNP50 BeadChip (Illumina Inc., San Diego, CA, USA), which encompasses 53,347 single nucleotide polymorphism (SNP) markers, following the instructions of the manufacturer. In addition, we have retrieved Goat SNP50 BeadChip data from Moroccan breeds (N= 20; *Barcha, Draa, Ghazalia, Noire de Atlas, Nord, Moroccan*), Egyptian (N=20; *Barki, Oasis, Saidi*), Spanish (N=20; *Bermeya*) and French (N= 40; *Saanen*) breeds generated in the AdaptMap project (Stella et al. 2018). We have also gathered Goat SNP50 BeadChip data from Algerian (N= 48; *Arabia, Makatia, M'Zabite, Kabyle)* and Sudanese (N= 50; *Desert, Nilotic, Taggar*) breeds that were available at the Dryad Digital repository (Ouchene-K et al. 2018; Rahmatalla et al. 2017). The genomic location of the SNPs was obtained using the goat ARS1 genome (Bickhart et al. 2017) as reference, and the position and the name of each SNP was updated using the PLINK v 1.9 software (Chang et al. 2015).

We obtained 50,729 SNPs that were subsequently filtered with PLINK v 1.9 (Chang et al. 2015) to eliminate those meeting any of the following conditions: (1) SNPs with minor allele frequencies (MAF) lower than 0.05, (2) SNPs with missing call rates higher than 0.05, (3) SNPs that did not adjust to the Hardy–Weinberg expectation (*P*-value ≤0.001) and (4) unmapped SNPs. After filtering and quality control, we retained 38,428 SNPs for further analyses. To investigate population structure, we performed principal component analysis (PCA) using the "--pca" command line in PLINK v1.9 (Chang et al., 2015). The PCA results were plotted with

73

the '*ggplot2*' package of the RStudio program (RStudio Team 2020). To avoid the overcrowding of the PCA with the 500 MG goats, we randomly extracted 50 MG goats with the Python "*randint()*" method (https://docs.python.org/3/library/random.html#functions-for-integers). The 50 MG goats selected with this approach were merged with Spanish, African, and Saanen goats with PLINK v1.9 (Chang et al. 2015).

In the PCA shown in **Figure 1**, the principal component 1 (PC1), which accounts for 32.34% of the total variance, separated the Saanen (SAA), Spanish (BEY and MG) and African (MOR, ALG, EGY, SUD) populations. Principal component 2 (PC2), which accounts for 14.03% of the total variance, provided similar results. The degree of separation between the MG and BEY breeds from the African ones was similar to that between SAA and African breeds, suggesting that the introgression of MG goats with African blood was likely quite low.

The ADMIXTURE v1.3.0 software (Alexander et al. 2009; Alexander and Lange, 2011) was used to determine the ancestry of MG goats. Since the MG population (N=500) is very large and unbalanced sample size can distort the results of ADMIXTURE, we divided the data set of 500 MG goats into 10 datasets of 50 individuals, following the order of the total of 500 MG sample list (e.g. dataset 1 = from the first MG in the list to the 50th, dataset 2 = from the 51st to the 100th in the list etc.) The Moroccan (MOR), Algerian (ALG), Egyptian (EGY,), Sudanese (SUD), Bermeya (BEY), and Saanen (SAA) goat populations were set as reference populations, and each one of the 10 MG data sets was set as the target population. In other words, we carried out 10 independent ADMIXTURE analyses comprising one group of 50 MG goats plus the full set of reference populations. We performed a supervised analysis because we were interested in estimating the potentially admixed ancestries of MG goats by using as a reference European and African populations of known ancestry. According to Alexander et al. (2011), the supervised analysis provides more precise estimates of ancestry than the unsupervised one because there is less uncertainty in allele frequencies. Moreover, the display of results is simplified and run times are shorter because there are fewer parameters to estimate (Alexander et al. 2011). To carry out the supervised test of admixture, we used the flag '--*supervised*' in the command line and we set the number of reference populations (K=6).

Maximum likelihood estimation of individual ancestries from Goat SNP50 BeadChip genotypes for each one of the datasets are shown in **Figure 2**. The Spanish genetic background, represented by the BEY breed, was overwhelmingly predominant in MG goats (average proportion 95.1 ± 4%). Moreover, the Moroccan genetic component reached a proportion of 4.2 ± 3.8%, while the Algerian (0.01 ± 0.01%), Egyptian (0.2 ± 0.2%) and Sudanese (0.13 ± 0.1%) components were present in extremely small proportions. Noteworthy, the percentage of

Moroccan ancestry in MG goats was highly variable and dependent on the selected MG data set e.g., data set 3 did not show any sign of African introgression, while in data sets 4 and 9 the Moroccan ancestry component reached values of 10% and 12%, respectively.

Signatures of African introgression in European breeds have been detected in cattle (Decker et al. 2014), sheep (Ben Jemaa et al. 2019) and goats (Manunza et al. 2016; Martínez et al. 2016; Colli et al. 2018). Manunza et al. (2016) investigated the diversity and demographic history of seven Spanish goat breeds and detected moderate levels of African introgression (~ 25%) in the Andalusian Malagueña and MG breeds. Our results confirm the African ancestry of MG goats although of a much lower magnitude (~ 4%). More importantly, we also demonstrate the existence of a significant inter-individual variability regarding the magnitude of such African introgression. This means that large sample sizes are needed to accurately estimate the degree of exotic introgression into livestock breeds. The African component identified in the genomes of MG goats had fundamentally a Moroccan origin, while breeds from Algeria, Sudan, or Egypt did not make any significant contribution. This result can be explained by the geographical proximity between southern Spain and Morocco (Strait of Gibraltar = 13 km), which has facilitated the existence of a bidirectional gene flow between African and Spanish goat breeds (Martinez et al. 2016).

About the historical circumstances that led to the introgression of MG goats with North African blood, several non-mutually exclusive scenarios are possible. The genetic analysis of Iberian Bronze Age cattle from the Atapuerca archaeological site has evidenced the existence of prehistoric contacts over the Straits of Gibraltar which involved the exchange of domestic animals (Anderung et al. 2005). The Roman occupation of the Iberian Peninsula, which started in 218 B.C. and finalized in the 5th century, might have also favored the transportation of livestock from North Africa to Hispania. Indeed, the Roman army was supplied with animals for riding, cargo and/or transport by military-logistics centers distributed throughout the Roman empire (Colominas et al. 2022). Interestingly, Colominas and Edwards (2016) reported the identification of the African mitochondrial T1 haplotype in archaeological cattle remains from the Roman trading post of Empúries (Catalonia). A third plausible historical scenario is the invasion of Spain, in the 8th century, by a Muslim army mostly composed by Amazigh troops from the Al-Butr group, native to a geographic area between Tunisia and the Atlantic coast of Morocco, that fought under the command of the Omeya califate (Franco Moreno 2005). Because of the Muslim occupation of Spanish territories, several migratory waves resulted in the establishment of 40,000-50,000 Arabs, from Northeast Africa and Middle East, and hundreds of thousands of Amazigh settlers in Spain (Franco Moreno 2005). Muslim settlers

introduced many technical innovations to develop agriculture (Collins 1989) and new crops (O'Callaghan 1983, Fletcher 2006). Moreover, studies based on archaeofaunal assemblages coming from Andalusi contexts consistently point out the predominant presence of sheep and goat remains, evidencing that these two species were the most consumed ones (Carvajal López 2016; Moreno 2013). Our finding that the African genetic background detected in MG goats has a likely Moroccan origin would be consistent with the notion that most Muslims who settled in Spain after 711 had an Amazigh origin. Historical accounts of farming practices in Al Andalus are very scarce, so we do not know for sure whether African sheep and goats were transported to Spain. Our data might support such hypothesis, but the genetic analysis of dated archaeological remains will be needed to confirm it.

Conflict of interest statement

The authors declare that they have no competing interests.

Availability of data

Murciano Granadina goat SNP50 BeadChip genotypes are accessible at https://doi.org/10.6084/m9.figshare.18095825
The AdaptMap genotypic datasets can be accessed at: https://datadryad.org/stash/dataset/doi:10.5061/dryad.v8g21pt

Algerian goat genotype data are available:

https://doi.org/10.5061/dryad.5pt8nt8

Raw genotyping data of Sudanese goat breeds are available in: (Rahmatalla et al., 2017 – Additional file 10).

References

Alexander D.H., Novembre J., Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–64.

Alexander D.H. & Lange K., (2011) Enhancements to the admixture algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246.

Anderung C., Bouwman A., Persson P., Carretero J. M., Ortega A. I., Elburg R., Smith, C., Arsuaga J. L., Ellegren H., & Götherström, A. (2005) Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8431–5.

Ben Jemaa S., Kdidi S., Gdura A. M., Dayhum A. S., Eldaghayes I. M., Boussaha M., Rebours E., & Yahyaoui M. H. (2019) Inferring the population structure of the Maghreb sheep breeds using a medium-density SNP chip. *Animal Genetics* **50**, 526–533.

Bickhart D. M., Rosen B. D., Koren S., Sayre B. L., Hastie A. R., Chan S., Lee J., Lam E. T., Liachko I., Sullivan S. T., Burton J. N., Huson H. J., Nystrom J. C., Kelley C. M., Hutchison J. L., Zhou Y., Sun J., Crisà A., Ponce De León, F. A., … Smith T. P. L. (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics* **49**, 643–650.

Carvajal López J.C. (2016) Al-Andalus: Archaeology, History and Memory. In: *UCL Qatar Series in Archaeology and Cultural Heritage*, Akkadia Press, Edinburgh, UK .

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S. Purcell S.M., Lee J.J. (2015) Second generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**,7.

Colli L., Milanesi M., Talenti A., Bertolini F., Chen M., Crisà A., Daly K., G., del Corvo M., Guldbrandtsen B., Lenstra J., A., Rosen B., D., Vajana E., Catillo G., Joost S., Nicolazzi E., L., Rochat E., Rothschild M., F., Servin B., Sonstegard T., S., … Stella A., (2018) Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning

of diversity and highlights post-domestication migration routes. *Genetics Selection Evolution* **50**, 58.

Collins R. (1989) The Arab Conquest of Spain 710-797, Oxford UK: Blackwell.

Colominas L., & Edwards C. J. (2017) Livestock Trade during the Early Roman Period: First Clues from the Trading Post of Empúries (Catalonia). *International Journal of Osteoarchaeology* **27**, 167–179.

Colominas L., Olesti O., Guàrdia J., & Harzbecher K. (2022) Supplying the Roman army? Animal production and consumption in the countryside of north-eastern Iberia (Can Rubió, la Garrotxa). *Journal of Archaeological Science: Reports* **45**, 103561.

Decker J. E., McKay S. D., Rolf M. M., Kim J.W., Molina Alcalá A., Sonstegard T. S., Hanotte O., Götherström A., Seabury C.M., Praharani L., Babar M. E., Correia de Almeida Regitano L., Yildiz M. A., Heaton M. P., Liu W. S., Lei C. Z., Reecy J. M., Saif-Ur-Rehman M., Schnabel R. D. & Taylor J.F. (2014) Worldwide Patterns of Ancestry Divergence and Admixture in Domesticated Cattle. *PLoS Genetics* **10**, e1004254. Y otra vez, pero que locura es esta??

Fletcher R. A. (2006) *Moorish Spain*. University of California Press, Berkeley, CA.

Franco Moreno B. (2005) Distribución y asentamientos de tribus bereberes (Imazighen) en el territorio emeritense en época emiral (S. VIII-X). *Arqueología Y Territorio Medieval* **12**, 39–50.

Luigi-Sierra M. G., Landi V., Guan D., Delgado J.V., Castelló A., Cabrera B., Mármol-Sánchez E., Alvarez J F., Gómez-Carpio M., Martínez A., Such X., Jordana J. & Amills M. (2020) A genome-wide association analysis for body udder and leg conformation traits recorded in Murciano-Granadina goats. *Journal of Dairy Science* **103**, 11605–11617.

Manunza A., Noce A., Serradilla J., M., Goyache F., Martínez A., Capote J., Delgado J. V., Jordana J., Muñoz E., Molina A., Landi V., Pons A., Balteanu V., Traoré A., Vidilla M., Sánchez-Rodríguez M., Sànchez A., Cardoso T. F. & Amills M. (2016) A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genetics Selection Evolution* **48**, 52.

Martínez A., Manunza A., Delgado J. V., Landi V., Adebambo A., Ismaila M., Capote J., el Ouni M., Elbeltagy A., Abushady A. M., Galal S., Ferrando A., Gómez M., Pons A., Badaoui B., Jordana J., Vidal O. & Amills M., (2016) Detecting the existence of gene flow between Spanish and North African goats through a coalescent approach. *Scientific Reports* **6**, 38935.

Moreno M. (2013) Gestión y aprovechamiento de cabañas ganaderas en al-Andalus: aportaciones desde la arqueozoología. *Debates de Arqueología Medieval* **3**, 75–98.

O'Callaghan J. F. (1983) *A History of Medieval Spain*. Cornell University Press, Ithaca, NY.

Ouchene-Khelifi N. A., Lafri M., Pompanon F., Ouhrouch A., Ouchene N., Blanquet V., Lenstra J. A., Benjelloun B. & Silva A. (2018) Genetic homogeneity of North-African goats. *PLoS ONE* **13**, e0202196.

Rahmatalla S. A., Arends D., Reissmann M., Said Ahmed A., Wimmers K., Reyer H. & Brockmann G. A. (2017) Whole genome population genetics analysis of Sudanese goats identifies regions harboring genes associated with major traits. *BMC Genetics* **18**, 92.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Stella A., Nicolazzi E. L., van Tassell C. P., Rothschild M. F., Colli L., Rosen B. D., Sonstegard T. S., Crepaldi P., Tosser-Klopp G., Joost S., Amills M., Ajmone-Marsan P., Bertolini F., Boettcher P., Boyle Onzima R., Bradley D., Buja D., Cano Pereira M. E., Carta A., … Zhang, W. (2018) AdaptMap: Exploring goat diversity and adaptation. *Genetics Selection Evolution* **50**, 61.
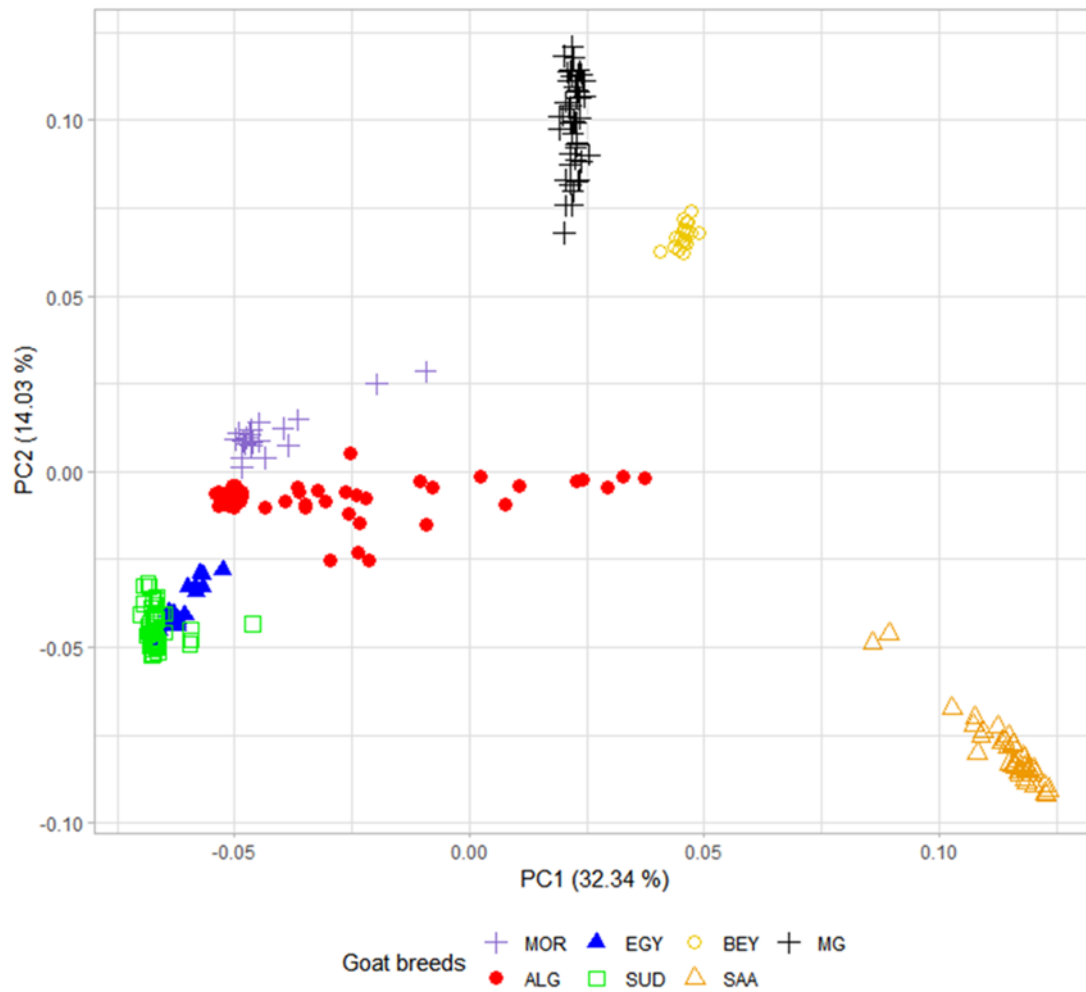
Paper 2-Figure 1. Principal component analysis (PCA) of 50 goats from the Murciano
Granadina (MG) breed randomly extracted from a population of 500 MG individuals plus
reference populations from Morocco (MOR), Algeria (ALG), Sudan (SUD), France (SAA,
Saanen) and Northern Spain (BEY, Bermeya).
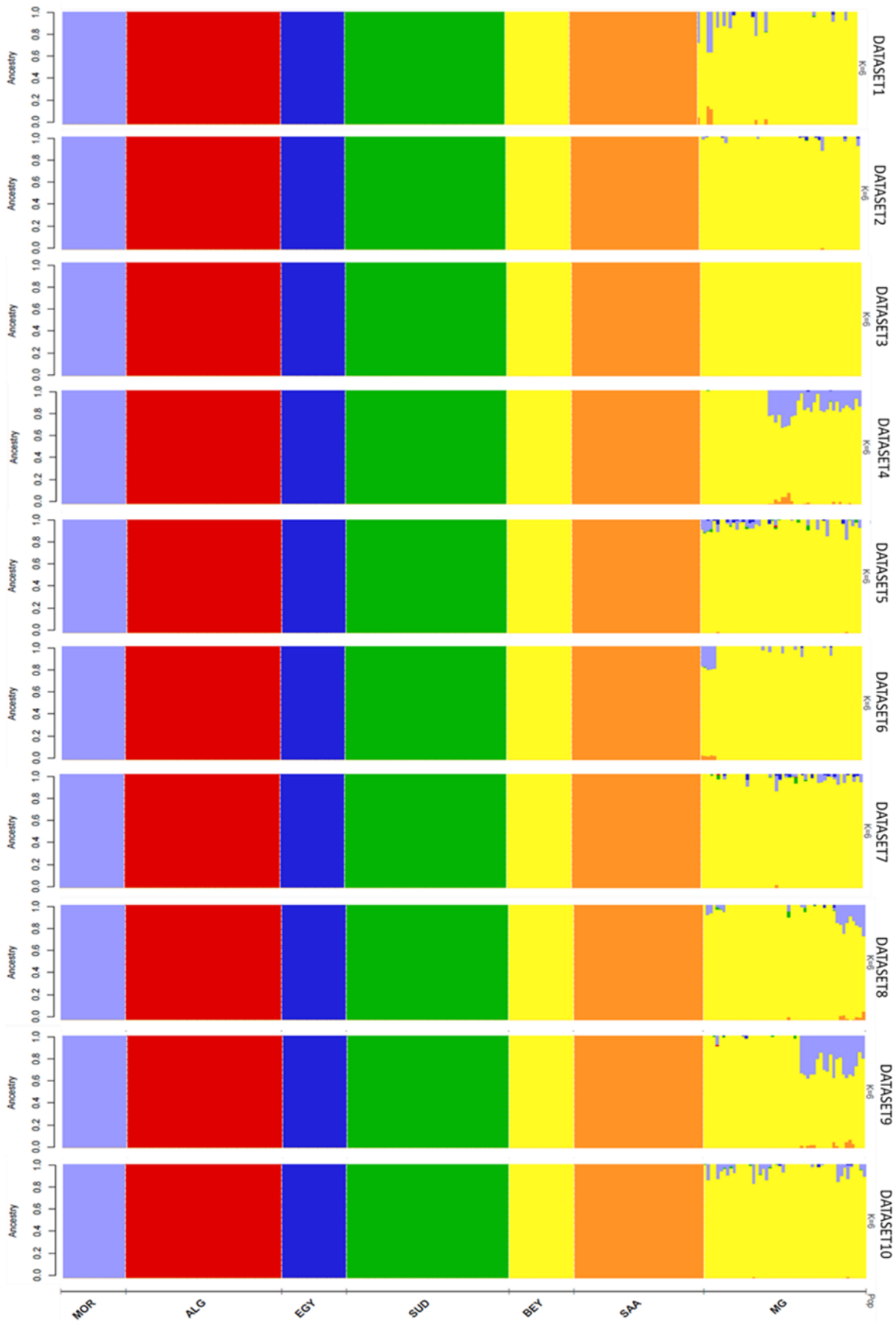
Paper 2-Figure 2. Supervised ADMIXTURE analysis of 10 groups of Murciano Granadina goats (MG) with 50 individuals each (datasets 1 to 10) plus reference populations from Morocco (MOR), Algeria (ALG), Sudan (SUD), Central Europe (SAA, Saanen) and Northern Spain (BEY, Bermeya). It can be seen that the levels of African (essentially Moroccan) introgression in MG goats are highly variable across datasets 1 to 10.

3.3 PAPER 3

scientific reports

Mitochondrial DNA diversity of the Sardinian local cattle stock

Elena Petretto[1,2], Maria Luisa Dettori[1, *], Michele Pazzola[1], Fabio Manca[1], Marcel Amills[2,3], and Giuseppe Massimo Vacca[1]

[1] University of Sassari, Department of Veterinary Medicine, via Vienna 2, 07100, Sassari, Italy. [2] Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193, Bellaterra, Spain. [3] Department of Animal Genetics, Center for Research in Agricultural Genomics (CRAG), CSICIRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain.

*mldettori@uniss.it

ABSTRACT

The aim of this research was to characterize the genetic diversity of the Sarda (Sa, n = 131), Sardo Bruna (SB, n = 44) and Sardo Modicana (SM, n = 26) cattle breeds, reared in the island of Sardinia (Italy). A portion of the mitochondrial DNA hypervariable region was sequenced, in order to identify a potential signature of African introgression. The FST coefficients among populations ranged between 0.056 for Sa vs SB and 0.167 for SB vs SM. AMOVA analysis indicated there was a significant differentiation of the three breeds, although most of diversity was gathered at the within-breed level. The Median Joining Network of the Sardinian sequences showed a potential founder effect signature. A MJ network including Sardinian cattle plus African, Italian, Iberian and Asian sequences, revealed the presence of haplogroup T3, already detected in Sa cattle, and the presence of Hg T1 and Hg T1'2'3, in Sa and SB. The presence of a private haplotype belonging to haplogroup T1, which is characteristic of African taurine breeds, may be due to the introgression of Sardinian breeds with African cattle, either directly (most probable source: North African cattle) or indirectly (through a Mediterranean intermediary already introgressed with African blood).

Introduction

Livestock breeding systems have experienced substantial changes during the twentieth century, mainly driven by mechanization, industrialization, and intensive selection. This process, which resulted in the adoption and diffusion throughout the world of highly selected cosmopolitan breeds [1], led to an impressive improvement of productions and to a genetic homogenization of farmed animals caused by the progressive replacement of rustic local breeds by their cosmopolitan counterparts [2]. Local breeds are an important cultural legacy and they play a fundamental role in landscape maintenance, being a key insurance against unknown forthcomings such as climate change and disease outbreaks [3,4]. Local and autochthonous breeds have undergone natural selection during millenia resulting in an optimal adaptation to a specific milieu [5]. For all these reasons, local breeds should be preserved as an essential asset for sustainable farming in the future [6]. Part of these conservation efforts have been devoted to the genetic characterization of these irreplaceable animal resources [6].

Sardinia (Italy) is a large and ancient island in the western Mediterranean Sea. Traditionally, sheep and goat farming have had an important impact in the rural economy of Sardinia [7,8]. In addition, three local cattle breeds are currently reared in Sardinia: the Sarda, the Sardo Bruna and the Sardo Modicana [9]. Sarda cattle are small sized, with high hardiness and resistance. They are perfectly adapted to the mountainous areas with arid soils in which they are typically raised. Historically, Sarda cattle provided milk, meat and labor to farmers, but during the 1880s, and for about fifty years, this breed was extensively crossed with bulls from the Brown breed, originally from Switzerland, with the aim of improving dairy traits. Moreover, Sarda cattle were also crossed with bulls from the Modicana breed, native to Sicily, with the goal of improving their work aptitude. Absorption crosses led to the transformation, in some areas of Sardinia, of the original Sarda cattle into two different populations i.e. Sardo Bruna, virtually equivalent to the Brown Swiss cattle, and Sardo Modicana, morphologically similar to the Sicilian Modicana. According to FAO [10], the number of Sarda cattle reared in Sardinia lies close to 21,800 individuals, while the Sardo Bruna and the Sardo Modicana breeds are represented by 27,670 heads and 2.200 heads, respectively. The government of Sardinia has established a herd register for each breed and herd books are managed by the Italian Breeders Association [11].

Decker et al. [12] investigated the patterns of ancestry, divergence and admixture of cattle by genotyping 43,043 single nucleotide polymorphisms (SNP) in 1,543 bovines from 134 breeds with a worldwide distribution. One of the main conclusions of this work was that Iberian and Italian cattle had been introgressed with African blood [12]. Although the genome-wide

diversity of the Sarda, Sardo Bruna and Sardo Modicana cattle has been characterized in previous studies [13,14] and the complete mitochondrial genome of one Sarda cattle (GenBank EU177832) has been sequenced [15], the potential African introgression of bovine breeds from Sardinia has never been explored in depth. In this regard, the analysis of mitochondrial data could be really useful because the T3 and T1 haplogroups are vastly predominant in Europe and Africa, respectively [16]. In the current work, we aimed to characterize the genetic diversity of the Sarda (Sa), Sardo Bruna (SB) and Sardo Modicana (SM) breeds through the partial sequencing of the mitochondrial DNA hypervariable region in order to identify a potential signature of African introgression.

Parameters	Sarda	Sardo Bruna	Sardo Modicana	Overall	Italian Brown*	Modicana*	Maremmana*	Italian Podolian*
Sample size	131	43	26	200	34	33	62	91
Number of polymorphic sites	26	22	6	34	–	–	–	–
Tajima's D	− 0.677 Ns	− 1.307 Ns	− 0.069	− 1.180 Ns	–	–	–	–
Nucleotide diversity	0.006	0.005	0.002	0.005	0.005	0.005	0.006	0.005
Number of haplotypes	22	15	5	32	–	–	–	–
Haplotype diversity	0.875	0.827	0.662	0.879	0.929	0.864	0.973	0.872
Detected haplogroups	T1 (22), T3 (104)	T1 (1), T3 (40)	T3 (26)	T1, T3	T1, T3	T1, T3	T1, T2	T1, T2
	T1'2'3 (5)	T1'2'3 (2)		T1'2'3	T5		T3	T3

Table 1. Distribution of mtDNA haplotypes in three local cattle breeds from Sardinia (616 bp mapping to the hypervariable region, range 15792 – 69 of acc. no. V00654). *Di Lorenzo et al. [19]. The range of the mtDNA hypervariable region was 15823- 215. Ns, non-significant.

Results

About 616 bp of the mtDNA hypervariable region (GenBank V00654 was the reference sequence) were successfully sequenced in 201 female cattle from the island of Sardinia (**supplementary Table S1**). Alignment of 200 sequences corresponding to Sa, SB and SM cattle revealed the occurrence of 34 polymorphic sites and 32 haplotypes (**supplementary Table S2**), while overall haplotype diversity was 0.878 (**Table 1**). The highest haplotype number was observed in the Sa breed, with 22 haplotypes out of 131 sampled animals, and it was similar to the SB breed, which had 15 haplotypes out of 43 sampled animals (**Table 1**). In SM cattle we found only 5 haplotypes out of 26 sequences, but it should be kept in mind that all individuals came from the same sampling site (Milis).

Geographic distribution of haplotypes in the island of Sardinia is shown in Figure 1. Different colours have been given to each haplotype. Moreover, each haplotype has been represented only once for each sampling site where it occurred, in order to avoid the overlapping of clusters. Sampling sites for the Sa breed are represented in Figure 1A (eleven sites), while sampling sites for the SB (five sites) and SM (only one site, Milis) breeds are shown in Figure 1B. The inspection of Figure 1 evidences that there was not any geographic structure associated with the distribution of mtDNA haplotypes in Sardinia.

The FST coefficients among populations ranged between 0.056 for Sa vs SB and 0.167 for SB vs SM (**Table 2**). AMOVA analysis indicated there was a significant differentiation of the three breeds (between-populations component of variation of 7.99 %) although most of diversity was gathered at the within-breed level (**Table 3**).

The MJ network only including the set of 200 mtDNA sequences generated by us plus the bovine reference sequence V00654 is shown in Figure 2A. The MJ network showed that most haplotypes were connected to each other in a star like fashion, with a central haplotype (H1) corresponding to the BRS (Acc. No. V00654). Eight haplotypes differed by one mutational event, while the remaining ones differed by two or more mutational events. Each breed showed private haplotypes, sometimes connected to the network through missing intermediate haplotypes (H13, H22, H18, H6). Haplotypes H19 and H4, as well as H6, were the most distant from the central haplotype.

The MJ network including Sardinian, European, Asian and African cattle (**Figure 2B**) revealed that about 80 % of the Sa cattle sequences belonged to the T3 haplogroup, and 15,8 % of sequences shared a specific haplotype belonging to Hg T1. In addition, two haplotypes belonged to Hg T1'2'3. All the SB cattle sequences belonged to Hg T3, except for one haplotype belonging to Hg T1, and one haplotype belonging to Hg T1'2'3, while all SM haplotypes belonged to Hg T3.

The private haplotype belonging to haplogroup T1 has been described in the current work for the first time. This haplotype was characterized by variations at nucleotide positions (np) 16050, 16113 and 16255 typical of Hg T1, and one variation at np 16022, which characterizes the sub-clade termed T1b1 [17]. In addition, two other variations (np 15948 and 16136) private to Sardinian native cattle (H6, supplementary Table S2) were detected.

Both haplotypes H4 and H19 (supplementary Table S2) showed a cytosine at np 16255 (typical of Hg T1) but at np 16050 and 16113 they harboured C and T, respectively, as in Hg T3. The nucleotide combination at positions 16050, 16113 and 16255 displayed by haplotypes H4 and H19 was the same one found in haplotype T1'2'3 (Acc. No. EU177840), which is considered a

common ancestor of the three T1, T2 and T3 haplogroups [15]. In addition to these nucleotide positions characteristic of the T1'2'3 haplotype, H19 shared with H4 the G > A variation at np 15825, which had been previously reported in only one subject belonging to an unidentified breed [18], while H4 had an additional rare variation at np 15915, reported only for a subject of the Chianina breed, but within a different haplotype [18]. We detected haplotype H4 in four Sa and one SB cattle distributed in three sampling sites from North East Sardinia (Oschiri and Trinità d'Agultu), while H19 was identified in the Sa breed in just one sampling site (Orotelli) located in the mountains of Central Sardinia.
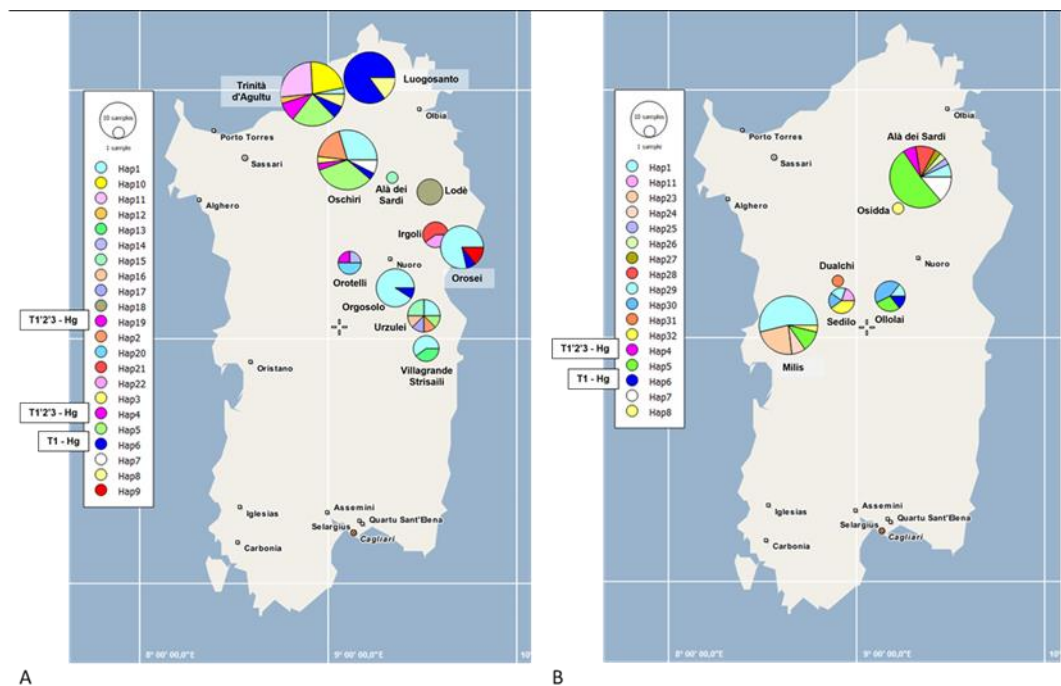
Paper 3-Figure 1. Geographic distribution of mtDNA haplotypes. The map was created with the PopArt software (Leigh and Bryant, 2015). Pie chart slices are proportional to the frequency of the haplotypes. A, Sampling sites and haplotypes found in the Sarda breed cattle. B, Sampling sites and haplotypes segregating in the Sardo Bruna breed cattle, except for the Milis sampling site, where all animals were of the Sardo Modicana breed. Haplotypes belonging to the T1 haplogroup are indicated in blue (T1 – Hg), Haplotypes of the T1'2'3 haplogroup are indicated in pink (T1'2'3 – Hg), all the other haplotypes were represented with pastel colours, and belong to the T3 haplogroup.

	Sarda	Sardo Modicana
Sardo Modicana	0.136	
Sardo Bruna	0.056	0.167

Table 2. F_{ST} values calculated with DnaSP based on mtDNA data from three Sardinian local cattle breeds.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	2	17.659	0.14172 Va	7.99
Within populations	197	321.536	1.63216 Vb	92.01
Total	199	339.195	1.77388	
Fixation Index F_{ST}: 0.07989		P-value = 0.00010		

Table 3. Analysis of molecular variance (AMOVA) based on mtDNA data from Sardinian cattle breeds.
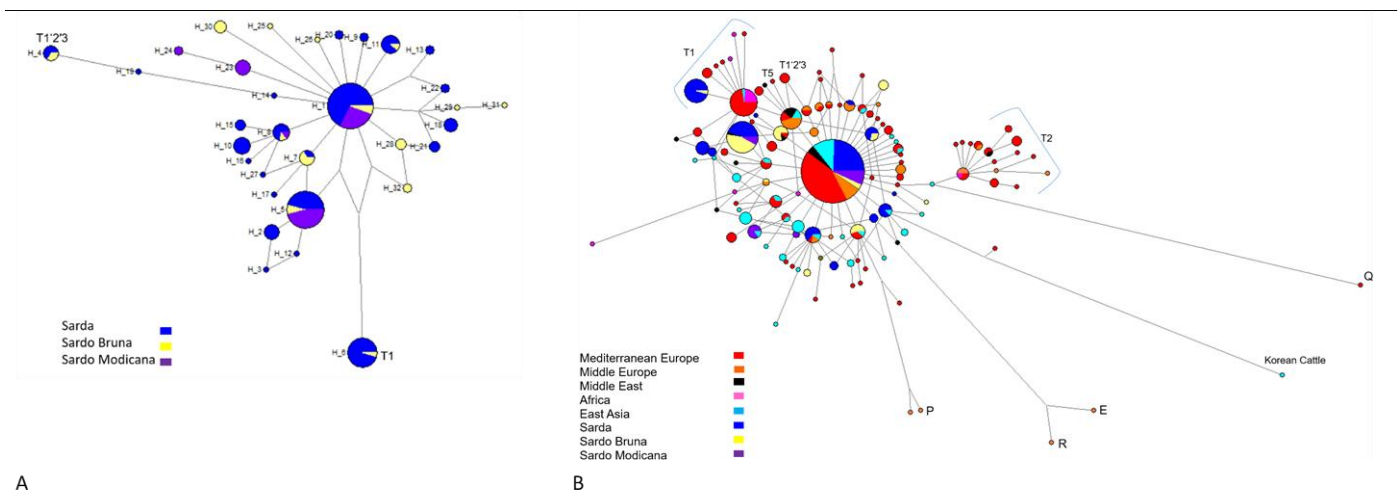
Paper 3-Figure 2. A, Median Joining Network based on mtDNA hypervariable sequence haplotypes of 157 Sardinian cattle, including sequences from Sarda, Sardo Bruna and Sardo Modicana cattle breeds. T1and T1,2,3 indicate the Haplogroup assignment. Aligned sequences were 616 bp long. B, Median joining network based on mtDNA hypervariable sequence haplotypes of 586 sequences from 157 Sardinian cattle and European, Asian and African bovines retrieved from public databases and representative of all known mitochondrial haplogroups. Sequences have been trimmed to obtain an alignment of 487 bp. Mediterranean Europe included sequences from Italy (North, South, Center), Spain, Portugal, and France. Central Europe encompassed sequences from Germany, Denmark, Romania, Bosnia, Croatia, Bulgaria. Middle East comprised sequences from Turkey, Iraq, Iran, Syria, Israel. Africa included sequences from Sudan, Kenya, Morocco, Libya, Mozambique, Egypt, Algeria, Tunisia. East Asia sequences were from Korea, South Korea, Japan, China, Nepal, Mongolia.

Discussion

The hypervariable region of mtDNA was analysed to obtain information about genetic diversity of three local cattle breeds, namely Sarda (Sa), Sardo Bruna (SB) and Sardo Modicana (SM), reared in the island of Sardinia. The FST analysis revealed a remarkable degree of differentiation between SM and SB. Besides, AMOVA was highly significant, revealing a differentiation between the three breeds. Such genetic differentiation between Sardinian breeds has been also observed by Cesarani et al. [13] and Mastrangelo et al. [14]. Cesarani et al. [13] genotyped 19 Sarda, 10 Sardo Bruna and 12 Sardo Modicana cattle with a medium density SNP chip and they revealed that Sardo Modicana individuals cluster close to the Modicana specimens and far apart from Sarda cattle, while Sardo Bruna individuals are placed at an intermediate location between the Brown Swiss and Sarda populations. In an additional study, the genetic diversity of 30 Sarda, 10 Sardo Bruna and 28 Sardo Modicana cattle was investigated with the BovineSNP50 BeadChip [14]. This latter study showed that Sarda animals cluster with Northern and Northern Central Italian breeds [14]. In addition, the lowest FST value corresponded to the Sarda vs Sardo Bruna pairwise comparison (FST = 0.016), while Sardo Modicana cattle were more similar to Modicana cattle than to the Sarda ones [14].

The overall haplotype diversity of Sardinian local cattle (Hd 0.879) was low, when compared to some continental Italian cattle breeds [19], especially for the SM breed (Hd 0.66). Negative Tajima's D values were calculated for all three breeds (although they were not significant), which might support the hypothesis of a founder effect or a bottleneck [20]. Indeed, the MJ network describing the genetic relationships between the three Sardinian local cattle (Figure 2A) has a star shaped topology consistent with the occurrence of a single founder effect. This kind of haplotype distribution has already been observed in goats from insular territories [21]. A geographical distribution of major taurine mtDNA haplogroups is reported in supplementary Figure S1.

The MJ network depicted in **Figure 2B** illustrates the relationships between the three local breeds of the current study and mtDNA sequences retrieved from public databases which represent, North Africa, Middle East, Middle Europe and Mediterranean Europe. The MJ network showed that most of Sardinian samples belonged to Hg T3, as already published for one Sardinian cattle by Achilli et al. [15]. Haplogroup T3 has been reported to be the most widespread in South West Europe and originates from the Near east to Europe migration of cattle herds which took place in the Neolithic [15].

We also detected the presence of one haplotype belonging to Hg T1, and two haplotypes belonging to Hg T1'2'3. The T1 haplotype segregated in both Sa and SB bovines, and it might

be private to Sardinian cattle, as a Blast search did not reveal its presence in any other bovine breed. The presence of a private T1 haplotype is consistent with the African introgression of Sardinian cattle breeds, as Hg T1 is representative of African taurine cattle, although Hg T1 has been also identified at low frequencies in continents other than Africa [22]. According to Decker et al. [12], both Iberian and Italian cattle display introgression from African taurine genomes, which probably occurred in two separate events. The Iberian breeds show signatures of a potential introgression from Western African taurine breeds, while several Italian breeds were likely introgressed by East African taurine breeds in which indicine introgression had already occurred [12].

The presence of Bos taurus in Sardinia has been verified in archaeological sites of both Neolithic and Chalcolithic ages, although no zoo-archaeological remains attributable to Bos primigenius have been found [23]. It has been reported that in the Neolithic age, maritime routes across the Mediterranean Sea already connected North Africa with Southern Europe [24]. The introduction of African haplotypes into Sardinia might have occurred at that time or later. On the other hand, Sardinia has been historically connected with other territories facing the Mediterranean Sea, from Spain to North Africa, up to present-day Lebanon (Phoenicians), so an indirect African introgression of Sardinian cattle (e.g. through an Iberian intermediary) is also feasible. For instance, zoo-archaeological and molecular studies (mtDNA) conducted in the Sus genus, revealed that pigs were traded between the Italian Peninsula and Sardinia by the end of the second millennium BC (late Bronze age and Iron age) and this gene flow left a genetic signature still detectable in Sardinian feral pigs [25]. During the Bronze Age, the inhabitants of Sardinia were part of the Sea People, who migrated to the Levant at that time, with routes to Sicily and Crete [23,25].

In conclusion, two hundred and one mtDNA sequences of three Sardinian cattle breeds (Sarda, Sardo Bruna and Sardo Modicana) were analysed in the present investigation. We found a moderate level of haplotype diversity in the Sa and the SB breeds, and low haplotype diversity in the SM. Most haplotypes belonged to haplogroup T3, which is widespread in Europe. In addition, we detected one haplotype belonging to haplogroup T1, and two haplotypes belonging to haplogroup T1'2'3. This T1 haplotype might derive from the African introgression of Sardinian cattle, which might have occurred directly or indirectly.

Methods

Sampling and DNA purification. A total of 201 blood samples were collected from Sarda (Sa, n = 131), Sardo Bruna (SB, n = 44), and Sardo Modicana (SM, n = 26) cattle, reared in 16

different areas of Sardinia (**supplementary Table S1**). In each farm, one to thirty-one female cattle were randomly chosen. Cows were managed under extensive farming systems based on mountainous territories with low agricultural productivity and typically associated with goat farming [9]. DNA was extracted from leukocytes using the Puregene DNA isolation kit (Gentra, Qiagen).

Mitochondrial DNA analysis. Based on the Bovine Reference Sequence (BRS) GenBank acc. no. V00654 [26], the primer pair MTF: 5'-GACTCAAGGAAGAAACTGC-3' and MTR: 5'-GACTCATCTAGGCATTTTCA-3' [27] was used to amplify a 1029 bp long segment of the mitochondrial DNA control region, in the nucleotide position range 15792 – 69 (V00654). Amplification conditions were as follows: 100 ng genomic DNA, 1.5 mM MgCl2, 0·2 mM dNTPs, 1X reaction buffer, 0·2 μM of each primer, and 1-unit Taq DNA polymerase (Platinum, Invitrogen, Life Technologies) in a 25 μl final volume. Thermal protocol was set for an initial denaturation at 94 °C for 2.30 min, and then 35 cycles of 94 °C for 20 sec, 56 °C for 30 sec, 72 °C for 1.20 min, followed by 72 °C for 5 min were carried out. Amplicons were purified with the ChargeSwitch PCR Clean-Up Kit (Invitrogen, Carlsbad, CA, USA) and then used to perform Sanger sequencing reactions with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Sequencing reactions were run in an Applied Biosystems 3730 DNA Analyser (Applied Biosystems, Foster City, CA, USA). Sequencing reactions yielded lower then expected length in many samples, then to make sure that the same fragment is analysed in all individuals, sequences were trimmed to 616 bp. All sequences were submitted to GenBank and given accession numbers KX923119 to KX923319.

Population Genetics Analyses. Sequence KX923305 was excluded from the dataset due to an 11 bp deletion. The MEGA version 7.0 software [28] (https://www.megasoftware.net/) was used to align mtDNA sequences and the DnaSP v.5.10.01 software [29] (http://www.ub.edu/dnasp/) was employed to estimate nucleotide and haplotype diversities as well as to calculate the FST coefficients of differentiation according to Hudson et al. [30]. The blastn suite of BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to screen the GenBank nucleotide collection database. We limited our search to Bos taurus (taxid:9913). Median-Joining (MJ) networks based on mtDNA data were built with the Network v.10 tool [31] (https://www.fluxus-engineering.com/sharenet.htm). A MJ network encompassing 586 sequences, including the 200 Sardinian mtDNA sequences generated by us, plus 386 European, Asian and African cattle mtDNA sequences retrieved from the public databases and representative of all known haplogroups (Hg) (**supplementary Table S3**) [32-53], is shown in

Figure 2B. Sequences have been trimmed to obtain an alignment of 487 bp. Polymorphic sites were weighted inversely to the number of mutational events according to Martínez et al. [54]. Transversions and transitions were given weights of 3 and 1, respectively. The analysis of molecular variance (AMOVA) was carried out with the Arlequin 3.5 software [55] (http://cmpg.unibe.ch/software/arlequin35/) and default parameters, while mtDNA haplotype frequencies relative to each sampling location were displayed with the POPART v.1.7 software [56] (http://popart.otago.ac.nz/index.shtml).

Ethics Statement. The DNA samples used for the present study were extracted from blood samples collected in the context of livestock sanitary programs featured by official veterinarians at local health institutions (Azienda per la Tutela della Salute, ATS) of the Regional Government of Sardinia (Italy), in accordance with relevant guidelines and regulations. All the procedures were approved by the Ethical Animal Care and Experimental Use Committee (Organismo Preposto al Benessere e alla Sperimentazione Animale, OPBSA) of the University of Sassari (protocol number 0122890, approved on 28 September 2021). None of the authors were involved in the collection of the blood samples previously, and just previously collected blood samples were used in this study.

Data Availability Statement

The original contributions presented in the manuscript are included in the article and Supplementary Material, further inquiries can be directed to the corresponding author.

References

1. Marsoner, T., Egarter V., L., Manck, F., Jaritz, G., Tappeiner, U. & Tasser, E. Indigenous livestock breeds as indicators for cultural ecosystem services: A spatial analysis within the Alpine Space. Ecological Indicators 94, 55-63 (2018).

2. FAO. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture, edited by B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome (available at http://www.fao.org/3/a-i4787e/index.html). (2015).

3. Zander, K. K., Signorello, G., De Salvo, M., Gandini, G. & Drucker, A. G. Assessing the total economic value of threatened livestock breeds in Italy: Implications for conservation policy. Ecological Economics 93, 219-229 (2013).

4. Pazzola, M., Dettori, M. L., Pira, E., Noce, A., Paschino, P. & Vacca, G. M. Effect of polymorphisms at the casein gene cluster on milk renneting properties of the Sarda goat. Small Rum. Res. 117, 124-130 (2014).

5. Stronen, A. V., Pertoldi, C., Iacolina, L., Kadarmideen, H. N. & Kristensen, T. N. Genomic analyses suggest adaptive differentiation of northern European native cattle breeds. Evolutionary Applications 12, 1096–1113 (2019).

6. FAO. World Livestock: Transforming the livestock sector through the Sustainable Development Goals. Rome. 222 pp. Licence: CC BY-NC-SA 3.0 IGO. (2018).

7. Vacca, G. M., Dettori, M. L., Carcangiu, V., Rocchigiani, A. M. & Pazzola, M. Relationships between milk characteristics and somatic cell score in milk from primiparous browsing goats. Animal Science Journal 81, 594-599 (2010).

8. Noce, A. et al. Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep. Animal Genetics 47, 717-726 (2016).

9. Pazzola, M. et al. Exploring the Genotype at CSN3 Gene, Milk Composition, Coagulation and Cheese-Yield Traits of the Sardo Modicana, an Autochthonous Cattle Breed from the Sardinia Region, Italy. Animals 10, 1995, 1-10 (2020).

10. Domestic Animal Diversity Information System (DAD-IS) link: http://www.fao.org/dad-is/data/en/ (Last accessed 23 June 2021).

11. AIA 2020. Associazione Italiana Allevatori [Italian breeders' association]. Web site. [accessed 2021 Jan 25] http://www.aia.it/aia-website/it/home.

12. Decker, J. E. et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genet. 10(3):e1004254. DOI: 10.1371/journal.pgen.1004254 (2014).

13. Cesarani, A. et al. Genome-wide variability and selection signatures in Italian island cattle breeds. Animal Genetics 49, 371-383 (2018).

14. Mastrangelo, S. et al. Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50, 35 (2018).

15. Achilli, A. et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. Current Biology 18, R157-R158 (2009). a

16. Lenstra, J. A., et al. Meta-Analysis of Mitochondrial DNA Reveals Several Population Bottlenecks during Worldwide Migrations of Cattle. Diversity 6, 178-187, DOI: https://doi.org/10.3390/d6010178 (2014).

17. Bonfiglio, S. et al. Origin and Spread of Bos taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1. PlosOne 7, e38601 (2012).

18. Bonfiglio, S. et al.. The Enigmatic Origin of Bovine mtDNA Haplogroup R: Sporadic Interbreeding or an Independent Event of Bos primigenius Domestication in Italy? PlosOne 5, e15760 (2010).

19. Di Lorenzo, P. et al. Mitochondrial DNA variants of Podolian cattle breeds testify for a dual maternal origin. PLoS ONE 13, e0192567, DOI: https://doi.org/10.1371/journal.pone.0192567 (2018).

20. Gattepaille, L. M., Jakobsson, M. & Blum, M. G. Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. Heredity 110, 409-19 (2013).

21. Ferrando, A. et al. A mitochondrial analysis reveals distinct founder effect signatures in Canarian and Balearic goats. Animal Genetics 46, 452-456 (2015).

22. Olivieri, A. et al. Mitogenomes from Egyptian Cattle Breeds: New Clues on the Origin of Haplogroup Q and the Early Spread of Bos taurus from the Near East. PLoS ONE 10, e0141170, DOI: 10.1371/journal.pone.0141170 (2015).

23. Wilkens, B. Archeozoologia. Il Mediterraneo, la storia, la Sardegna. Sassari, Italy. Editrice Democratica Sarda. (2012).

24. Zilhao, J. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. PNAS 98, 14180-14185 (2001).

25. Lega, C. et al. Like a pig out of water: seaborne spread of domestic pigs in Southern Italy and Sardinia during the Bronze and Iron Ages. Heredity 118, 154-159 DOI: 10.1038/hdy.2016.74 (2017).

26. Anderson, S., de Bruijn, M. H., Coulson, A. R., Eperon, I. C., Sanger, F., & Young, I. G. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156, 683–717 (1982).

27. Qu, K. X. et al. Mitochondrial D-loop Variation and Genetic Background of Brahman Cattle. Zoological Research 27, 615-620 (2006).

28. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

29. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452, DOI: 10.1093/bioinformatics/btp187 (2009).

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

31. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37-48 (1999).

32. Salim B., Taha K. M., Hanotte O., & Mwacharo J. M. Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle. Anim. Genet. 45, 782-90 (2014).

33. Takeda K., Satoh, M., Neopane, S. P., Kuwar, B.S., Joshi, H. D., Shrestha, N. P., Fujise, H., Tasai, M., Tagami, T., & Hanada, H. Mitochondrial DNA analysis of Nepalese domestic dwarf cattle Lulu. Animal Science Journal 75, 103-110 (2004).

34. Fujise, H., Murakami, M., Devkota, B., Dhakal, I. P., Takeda, K., Hanada, H., Fujitani, H. Sasaki, M., & Kobaiashi, K. Breeding distribution and maternal genetic lineages in Lulu, a dwarf cattle population in Nepal. Animal Science Journal 74, 1-5 (2003).

35. Mannen, H., Kohno, M., Nagata, Y., Tsuji, S., Bradley, D. G., Yeo, J. S., Nyamsamba, D., Zagdsuren, Y., Yokohama, M., Nomura, K., & Amano, T. Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. Mol. Phylogent. Evol. 32, 539-544 (2004).

36. Kim, K. I., Lee, J. H., Lee, S. S., & Yang, Y. H. Phylogenetic relationships of Northeast Asian cattle to other cattle populations determined using mitochondrial DNA D-loop sequence polymorphism. Biochem Genet. 41, 91-8 (2003).

37. Lai, S. J., Liu, Y.P., Liu, Y.X., Li, X.W., Yao, & Y.G. Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. Mol. Phylogenet. Evol. 38, 146-54 (2006).

38. Lei, C. Z., Chen, H., Zhang, H. C., Cai, X., Liu, R. Y., Luo, L. Y., Wang, C. F., Zhang, W., Ge, Q. L., Zhang, R. F., Lan, X. Y. & Sun, W. B. Origin and phylogeographical structure of Chinese cattle. Animal Genetics 37, 579-582 (2006).

39. Qu, K. X., Wu, G. S., Liao, X. L., Jin, X. D., Zhao, G., Yang, G. R., Yuan, X. P., Huang, B. Z., Wen, J. K., Zhang, Y. P. Zoological Research 27, 615-620 (2006).

40. Jia, S., Zhou, Y., Lei, C., Yao, R., Zhang, Z., Fang, X., & Chen H. A new insight into cattle's maternal origin in six Asian countries. J. Genet. Genomics. 37,173-180 (2010).

41. Dadi, H., Tibbo, M., Takahashi, Y., Nomura, K., Hanada, H., & Amano, T. Variation in mitochondrial DNA and maternal genetic ancestry of Ethiopian cattle populations. Anim Genet. 40, 556-9 (2009).

42. Berthouly, C., Maillard, J. C., Pham Doan, L., Nhu Van, T., Bed'Hom, B., Leroy, G., Hoang Thanh, H., Laloë, D., Bruneau, N., Vu Chi, C., Nguyen Dang, V., Verrier, E., & Rognon, X.

Revealing fine scale subpopulation structure in the Vietnamese H'mong cattle breed for conservation purposes. BMC Genet 11, 45 (2010).

43. Kim, J. H., Lee, S. S., Kim, S. C., Choi, S. B., Kim, S. H., Lee, C. W., Jung, K. S., Kim, E. S., Choi, Y. S., Kim, S. B., Kim, W. H., & Cho, C. Y. Haplogroup Classification of Korean Cattle Breeds Based on Sequence Variations of mtDNA Control Region. Asian-Australasian journal of animal sciences 29, 624-630 (2016).

44. Mannen, H., Yonesaka, R., Noda, A., Shimogiri, T., Oshima, I., Katahira, K., Kanemaki, M., Kunieda, T., Inayoshi, Y., Mukai, F., & Sasazaki, S. Low mitochondrial DNA diversity of Japanese Polled and Kuchinoshima feral cattle. Anim Sci J. 88, 739-744 (2017).

45. Achilli. A., Bonfiglio, S., Olivieri, A., Malusà, A., Pala, M., et al. (2009) The Multifaceted Origin of Taurine Cattle Reflected by the Mitochondrial Genome. PLoS ONE 4, e5753 (2009).

46. Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J. D., & Thomas, M. G. Modern Taurine Cattle Descended from Small Number of Near-Eastern Founders. Molecular Biology and Evolution 29, 2101-2104 (2012).

47. Scheu, A., Powell, A., Bollongino, R. et al. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet 16, 54 (2015).

48. Hristov, P., Spassov, N., Iliev, N., & Radoslavov, G. An independent event of Neolithic cattle domestication on the South-eastern Balkans: evidence from prehistoric aurochs and cattle populations. Mitochondrial DNA A DNA Mapp Seq Anal. 28, 383-391 (2017).

49. Ilie, D. E., Cean, A., Cziszter, L. T., Gavojdian, D., Ivan, A., & Kusza, S. Microsatellite and Mitochondrial DNA Study of Native Eastern European Cattle Populations: The Case of the Romanian Grey. PLoS ONE 10(9), e0138736 (2015).

50. Edwards, C. J., Magee, D. A., Park, S. D. E., McGettigan, P. A., Lohan, A. J., et al. A Complete Mitochondrial Genome Sequence from a Mesolithic Wild Aurochs (Bos primigenius). PLoS ONE 5, e9255 (2010).

51. Ginja, C., Penedo, M. C., Melucci, L., Quiroz, J., Martínez López, O. R., Revidatti, M. A., Martínez-Martínez, A., Delgado, J. V., & Gama, L. T. Origins and genetic diversity of New World Creole cattle: inferences from mitochondrial and Y chromosome polymorphisms. Anim. Genet. 41, 128-41 (2010).

52. Lopez-Oceja, A., Muro-Verde, A., Gamarra, D., Cardoso, S., & de Pancorbo M. M. New Q lineage found in bovine (Bos taurus) of Iberian Peninsula. Mitochondrial DNA A DNA Mapp Seq Anal. 27, 3597-601 (2016).

53. Fajardo, V., González, I., López-Calleja, I., Martín, I., Rojas, M., García, T., Hernández, P.E., & Martín, R. PCR identification of meats from chamois (Rupicapra rupicapra), pyrenean ibex (Capra pyrenaica), and mouflon (Ovis ammon) targeting specific sequences from the mitochondrial D-loop region. Meat Sci. 76, 644-52 (2007).

54. Martinez, A. et al. Inferring the demographic history of a highly endangered goat breed through the analysis of nuclear and mitochondrial genetic signatures. Small Rum. Res. 104, 78-84 (2012).

55. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10, 564–567 (2010).

56. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116 (2015).

Acknowledgments

Author Contributions

Conceptualization, G.M.V. and F.M.; methodology, M.L.D., M.A. and E.P.; formal analysis, E.P. and F.M.; bioinformatic analysis, E.P. and M.L.D.; resources, G.M.V. and M.P.; data curation, M.L.D. and E.P.; data interpretation, M.L.D, E.P. and M.A., writing—original draft preparation, M.L.D. G.M.V. and M.P.; writing—review and editing, M.A. and M.L.D.; funding acquisition, G.M.V. All authors have read and agreed to the published version of the manuscript.

Additional Information

Competing Interests Statement. The authors declare no conflict of interest.

4 General discussion

4.1 The post-domestication dispersal of livestock in Europe and Africa

4.1.1 Major clusters of genetic differentiation in European goat breeds and routes of dispersal

In the **study 1** of this thesis, we have investigated whether the post-domestication dispersal of goats left a genetic signature that can be detected nowadays by taking advantage of the wide availability of genotyping data derived from the AdaptMap project (https://www.biomedcentral.com/collections/ADAPTmap). More specifically, we aimed to find out whether the amount of diversity of current modern goat populations correlates with distance to the domestication centre in the Fertile Crescent, thus generating a gradient of declining genetic variation from it. By analysing genotyping data from 1,148 goats from 38 European breeds, we were able to detect a genetic differentiation between breeds from Northern Europe (Finland, Denmark, the Netherlands) and the British-Irish islands when compared to the remaining European populations (**see study 1 – Supplementary Figure S2**). Such clusters of genetic differentiation strongly correlating with geography were also found in a study about the genome-wide SNPs diversity of Balkan and Southwest Asian, Central and Northern European, and Mediterranean ovine breeds (Ciani et al., 2020). Through spatial models of supervised PCA, the presence one east-west genetic cline was observed, i.e. from the Balkans towards the Mediterranean coast. In addition, a south-north cline was also highlighted, underlining the contrast between northern European and south-western European breeds (Ciani et al., 2020). On the other hand, Ciani et al. (2020) proposed that Greek sheep breeds might have constituted a barrier to gene flow between Asian and European breeds.

The differentiation of Northern vs Southern European breeds could be partly explained by the existence of two different post-domestication dispersal routes (Mediterranean and Danubian) by which Neolithic peoples managed to reach different parts of Europe. The first evidence of livestock expansion from the Near East was found in Cyprus, with the transportation of animals from the mainland to the island, and later onto Crete 9,000 YBP (Efstratiou et al., 2004). In this early expansion there was a specific livestock profile, i.e. there was an abundance of goats (mainly sheep) and a lower number of cattle and pigs. Such livestock proportions were typical of Neolithic colonisers from southern and western Anatolia (Perlès et al., 2013). This characteristic composition

of transported livestock was also found in mainland Greece (8500 BC), Dalmatia (8,000 BC) and further northwards at the entrance to the Adriatic (7,700 BC) (Zeder, 2017). While maritime diffusion played a key role in the dispersal of livestock across the Mediterranean basin, the entry of livestock in Northern Europe probably took place through the Danubian corridor. In the Rhine valley and eastern Switzerland there was an abundance of cattle, whereas in the Rhone valley there was a dominance of sheep (Manning et al., 2013a). This difference could be due to the fact that in eastern Switzerland, Neolithic colonizers arrived via the Danube corridor, whereas in eastern France and southern Switzerland they came via the Mediterranean route. The arrival of Neolithic settlers in the British Isles happened much later (5,800 BC), probably from mainland European populations as evidenced by the presence of traditional carinated bowls typical of these peoples in Scotland, Ireland, and southern Britain (Schulting, 2013). Also confirming the arrival of Neolithic peoples in Britain from the continental route is the presence of large numbers of cattle in comparison to other livestock species (Schulting, 2013).

So, part of the genetic differentiation that we detect between Southern and Northern European goat breeds could be related to differences in the dispersal routes of the Neolithic stocks that originated such populations. Studies on the Y chromosome confirm such ancestral genetic differentiation: in fact haplogroup Y1B is the most widespread in central and northern Europe, which can be traced back to expansions and bottlenecks that occurred during the migration of Neolithic farmers via the Danube route (Cymbron et al., 2005); whereas, haplogroup Y1AA has been identified in Spain, sub-Saharan Africa and Anatolia in ancient samples, but has a low frequency in Europe today, further confirming such ancient bottlenecks and expansions (Nijman et al., 2022).

4.1.2 Absence of a significant gradient of diversity associated with distance to the domestication centre in European goats.

We investigated whether the heterozygosity of European goat populations correlates with distance to Nevali Çori, an archaeological site representative of the ancestral centre of domestication in Southern Anatolia. This analysis was carried out with two data sets with and without insular populations. By performing correlation analyses between heterozygosities (observed and expected) and the distance of the sampling site of these breeds from the domestication centre, we obtained a negative and significant correlation for both heterozygosities (H_o: r = -0.44, P-value = 0.005; H_e: r = -0.4, P-value = 0.011) in the complete dataset. In contrast, removing the insular breeds from the analysis resulted in non-significant correlations (H_o: r = -0.23, P-value = 0.21; H_e: r = -0.16, P-

value = 0.37). A gradient from west to east is not observed in Europe. Most European breeds show high to moderate heterozygosity values, e.g., the diversity of Spanish goats is not lower than that of the Italian ones despite the greater distance separating Spain from Nevali Çori.

Our analysis shows more significant results with the dataset including British-Irish breeds likely because of the impact of geographic isolation on the genetic diversity of insular populations. It is important to point out that the effects of isolation depend on several factors such as the history of breeds and their geographical distribution. Cardoso et al. (2018) found that goats from Madagascar, Iceland, Ireland, and La Palma had higher levels of homozygosity than those from Mediterranean island breeds. As a matter of fact, Mediterranean insular goats (such as the Sardinian breed) showed similar levels of homozygosity than those detected in continental breeds. Clearly, Sardinia, as well as Corsica, Sicily and Majorca are in the heart of the Mediterranean basin, becoming transit points in the maritime routes of the Phoenicians, through the Greeks, the Romans, and the Arabs over the centuries (Abulafia et al., 2011). Such migrations and transitions of different cultures have led to the preservation of a certain level of genetic diversity in Mediterranean island goat breeds, in contrast to northern European insular breeds located in less accessible areas.

In **study 1**, having a significant gradient when considering the British and Irish breeds, was clearly due to their low diversity. Such limited variation is not necessarily due to the footprint of the post-domestication dispersal. In fact, Old English goats and English goats, are at a critical risk of extinction (DAD-is database, https://www.fao.org/dad-is/en/). In Ireland, Old Irish goats are also in a very critical situation, being exclusively found in remote mountain ranges roaming in feral herds. The main cause of these strong demographic declines is the replacement of local goat breeds by more productive Swiss breeds, such as Toggenburg and Saanen. However, we cannot rule out the possibility that the initial dispersal of domestic goats in Europe involved a strong founder effect in the United Kingdom and Ireland. Indeed, a study using a series of simulation models on the Neolithic introduction of cattle into Britain revealed that probably only a limited number of cattle arrived to the British islands (Cummings and Morris, 2022). Given the intrinsic difficulties of transport by sea, Neolithic populations from continental Europe might have introduced a reduced number of animals (30 cattle) in the United Kingdom (Cummings and Morris, 2022). The low diversity values of British breeds have also been confirmed in cattle through microsatellite studies (Cymbron et al., 2005), and this seems to be a general trend in Northern Europe. However, it is very difficult to ascertain whether this is the result of serial founder effects associated with the post-domestication dispersal of livestock in Northern Europe or to the demographic decline of

many Northern European goat breeds due to highly intensive breeding practices carried out in septentrional Europe as well as to the replacement of local stocks by transboundary breeds.

On the other hand, we also considered that one of the reasons that explain the absence of a significant gradient of diversity in continental European breeds is the process of dispersal itself. In principle, progressive overland dispersal is more expected to generate a cline of diversity that long-distance maritime voyaging, which would be more associated to a leapfrog pattern of variation. In a recent work, Isern et al. (2017) demonstrated that the spread of the Neolithic stock along the Mediterranean Basin involved navigation through considerable distances (300-450 km), with multiple points of entry along the coast, and with a speed (8 km/year) considerably higher than that achieved through land (1 km/year). Since pre-neolithic times, there is evidence of sea travel in the eastern Mediterranean (Ammerman, 2010). These navigation abilities were refined, following the arrival of Neolithic colonizers to Cyprus and Crete, in order to settle in new areas by transporting cargo and groups of humans (Broodbank and Strasser, 1991; Vigne and Cucchi, 2005). The art of navigation was also developed in the western Mediterranean, based on cabotage in coastal waters with at most short crossings between islands or between islands and the mainland but rarely in open sea areas (Zilhão, 2014).

Furthermore, many historical events that took place well after domestication might have also diluted the initial genetic signature of post-domestication dispersal in Europe. For instance, the Iberian Peninsula was a land of conquest and transit for many ancient civilizations from Europe such as the Romans and Visigoths (Colominas et al., 2012), and from Africa such as the Phoenicians and Muslims (this theme is discussed in **study 2** of the thesis). Colonizing populations probably transported livestock with them during their migrations, contributing to erase of the gradient left by the post-domestication dispersal of livestock.

The presence of a diversity gradient in European goat breeds has been investigated previously. In a study on the genetic diversity of European and Middle Eastern goat breeds, a cline of decreasing genetic diversity from south-east to north-west Europe was observed (Cañon et al., 2006). These findings do not match ours (e.g., Spanish breeds are as diverse as the Italian ones), and a potential explanation for this discrepancy would be that Cañon et al. (2006) based their analysis on a reduced number of microsatellite markers (30 microsatellites). Another explanation could be that our sampling of Northern European breeds is quite limited, while East Europe was poorly sampled (the only breed representative of this region was Carpathian). As a matter of fact, Cañon and co-authors used a dataset in which European breeds were more homogeneously distributed in geographical

terms than in ours. In addition, a decrease in genetic diversity was observed with increasing distance from the domestication site (in south-west Asia) in a study comparing the diversity of goat breeds in northern Europe with those in central Europe, the Mediterranean and south-west Asia (Lenstra et al., 2016). In that study, the number of animals and the geographical distribution of breeds were homogeneous and balanced, and they detected not only a south-east to north-west cline but also an east to west Mediterranean cline, although less steep than the former. Therefore, the low representation of Eastern European breeds and the limited number of Northern European breeds in our dataset may have influenced our results.

In sheep, the existence of a diversity gradient associated with distance from the domestication centre through whole genome analysis has been carried out in breeds distributed worldwide (Kijas et al., 2012). This study showed high levels of heterozygosity in Mediterranean and southern European sheep breeds. In addition, a moderate correlation (r = −0.40) between heterozygosity and distance from the domestication centre was found, consistent with the results that we have reported in **study 1**. Kijas et al. (2012) attributed such moderate correlation between genetic diversity and distance from the domestication centre to the extensive gene flow between different ovine populations after domestication. Further confirming this conclusion, Kijas et al. (2012) reported a high haplotypic sharing between sheep breeds from different geographical areas and a weak population structure due to gene flow and introgression events between breeds.

4.1.3 Major clusters of genetic differentiation in African goat breeds and detection of a significant gradient of diversity

Regarding African goat breeds, in **study 1** we observed different genetic clusters corresponding to the geographical origins of the investigated breeds: south, north, east, west Africa and Madagascar (**see study 1 – Figure S3 in additional file 1**). Such clear differentiation may be due to the complex spread of domestic animals in Africa compared to Europe, considering the vast size and different habitats and temperatures along the African continent (Zeder, 2017). In Africa, the presence of geographic barriers is probably of considerable importance to explain the differentiation of goat breeds. The vast extension of the Sahara Desert, with more than 9 million km2 (about one third of Africa), has limited to a considerable extent the expansion of pastoralist societies throughout the continent. This hypothesis is supported by a study of the non-metric cranial traits of late Holocene populations in North Africa (Nikita et al., 2012), which demonstrated that the Garamantians, who populated the Sahara Desert around 3000 years ago and controlled trade in

this area, had distant biological affinities to the Egyptian, Algerian, Sudanese, and Tunisian nearby populations (Nikita et al., 2012). This suggests that although there is archaeological evidence of contact between these populations (Liverani, 2000), the Sahara constituted a barrier influencing gene flow between the southwestern Libyan populations and the other North African populations under study (Nikita et al., 2012). Therefore, the Sahara Desert permitted mainly only short-distance contacts (Nikita et al., 2012), contributing to differentiate Northern African populations from those residing in the Sahel and southwards.

The differentiation of African goat breeds could be also explained by the routes of dispersal of the first domesticated animals. Such spread has been hypothesised to have occurred via two routes: one by land from southern Israel passing through the Sinai Peninsula, reaching the Red Sea coast around 7,000 YBP (Smith, 1992); and another one by sea along the Mediterranean coast (Pereira et al., 2009). Despite the existence of a maritime diffusion in the northern coastline of the continent, the gradient of diversity observed by us is more consistent with a pattern of overland dispersal of the Neolithic stock from north to south. It has been shown that the presence of crops and domesticated animals was initially confined to North Africa, in the coastal areas and the Nile Valley where the Mediterranean climate and annual floodings, respectively, made possible cereal cultivation and animal farming (Blench, 2000). A number of sites in Libya, northern Algeria and the Mediterranean coast of Morocco dated between 7,000 and 6,500 YBP have yielded archaeological evidence compatible with the Neolithic package typical of the farming populations scattered in the Mediterranean basin at that time (impressed ceramics and domesticated animals mostly goats and cattle) (Higgs, 1967; Gilman, 1975; Roubet and Carter, 1984; Klein and Scott 1986). Subsequently, pastoral communities that initially lived in the niche of the Sahara grasslands expanded to the west of the Niger (6,000 YBP), south of Sudan and into the western Sahel (5,400 YBP) (Gautier, 1987; Peters, 1986; MacDonald and MacDonald 2000) in conjunction with the increasing aridity of the Sahara. The spread and arrival of pastoral populations to East and West Africa is dated between 4,000 and 3,500 YBP (Marshall and Weissbrod, 2011). A late expansion of domestic animals into Sub-Saharan Africa may be due to factors related to the low immunity of introduced domestic livestock to diseases typical of the tropical rainforest. However, the opening of grazing corridors traversing infested areas and the acquisition of immunity by livestock allowed the spread of domestic animals to Central, East and West Africa (Gifford-Gonzalez and Hanotte, 2011). Alternatively, Chritz et al. (2015) have proposed that the slow expansion of livestock southwards the Sahel is probably related to social factors rather than to resistance to tropical

diseases. The arrival of sheep and goats to southern Africa is dated between 2,000 and 1,900 YBP (Smith, 1992; Sealy and Yates, 1994) and that of cattle later at 1,600 YBP (Orton, 2012). It is hypothesised that this gradual spread of early domesticated animals into South Africa was due to the initial adoption of elements of Neolithic culture by indigenous peoples engaged in long-distance trade (Orton, 2015); while, later, the arrival of additional domesticated animals (cattle) occurred with the arrival of Neolithic farming populations.

Limited gene flow may have caused a decrease of diversity in breeds from Central and East Africa, as evidenced in our results (**study 1**). Importantly, this region coincides with the Tsetse fly belt, which constitutes a formidable biological barrier. The Tsetse fly is the vector of a protozoan organism called Trypanosoma that upon infection leads to a disease-causing fever, weight loss, and anaemia affecting goats as well as other livestock (Ebhodaghe et al., 2018). The distribution area of the Tsetse fly is largely confined to sub-Saharan Africa (Dorn et al., 2017) and, since the 20th century, it has moved further southwards (Courtin et al., 2010). This may have contributed to a limited spread and reduced gene flow between goats susceptible to the parasite in this large geographic area. In this regard, Traoré et al. (2012) demonstrated that the presence of the Tsetse fly, and hence of trypanosomiasis, in a certain area of Burkina Faso influenced the genetic variability of goats. In this way, Traorè and co-authors showed that this protozoan disease may have constituted a barrier for the spread of trypanosomiasis sensitive goats (typically Sahelian goats) but, at the same time, it also led to a strong selection of goats immune to the disease (e.g., Djallonkè breed) in infested areas.

As with the European breeds, we investigated whether the diversity of two data sets of African breeds (with and without insular populations) correlates with distance to the domestication centre in Southern Anatolia. Interestingly, significant negative correlations between the diversity of African goat populations and distance to Nevali Çori were observed in both datasets, resulting in coefficients of -0.35 (H_o, P-value = 0.033) and -0.39 (H_e, P-value = 0.018) in the dataset without islands, and of -0.47 (H_o, P-value = 0.0022) and -0.49 (H_e, P-value = 0.0012) in the dataset with islands, respectively. North African breeds from Egypt and Algeria show a high level of heterozygosity, which decreases when considering a south-eastern direction (Malawi and Mozambique). In South African breeds, genetic diversity is again high, with observed and expected heterozygosity values of 0.409 and 0.43, respectively (see more details about these results in **Papers section - study 1**).

The significant diversity gradient that we have detected in African goat populations may be explained by the gradual terrestrial dispersal across the continent, facilitating the creation of a gradient of genetic variation. This could be due both to the fact that progression by land is slower than by sea (Isern et al., 2017), and to the presence of geographical barriers, such as the Sahara and the Kalahari deserts, which could have slowed or stopped the passage of humans and animals from one place to another. The presence of the Tsetse fly belt and thus the occurrence of Trypanosomiasis may also have initially blocked or retarded the passage of early pastoralists with their disease-susceptible animals to tropical areas (Dorn et al., 2017). As explained before, there are also differences in the dispersal patterns of the different species that were part of the Neolithic pack (cattle, sheep, goats, and pigs) both between Europe and Africa but also within the African continent itself (Zeder, 2017). The climatic conditions and the presence of the Mediterranean Sea and the Nile River as water sources in North Africa allowed not only the cultivation of cereals typical of the Near East, but also the farming of all domesticated Neolithic species (Zeder, 2017). In contrast, in the more extensive and dispersed areas of Sub-Saharan Africa, indigenous populations preferred to integrate goats and sheep into their agricultural practices over cattle, being easier to transport but also a good source of milk and meat (Zeder, 2017). Hence, the presence of goats throughout Africa and their dispersion may also have depended on the easiness of their integration into pre-existing local economies and the environmental conditions important for the survival of both crops and animals.

The South African goat breeds analysed in **study 1** mainly included two groups: Boer, Kahalari Red and Savanna (commercial goats) on the one hand, and local indigenous goats from Limpopo, Freestate, Gauteng and Northwest on the other hand. We decided to exclude the Boer breed from the correlation analysis because of its highly admixed origin, which we do not consider representative of the South African indigenous stock. Indeed, the Boer breed originated when the first European colonists arrived at the Eastern Cape in South Africa and they decided to cross goats with Indian ancestry, such as Anglo-Nubian, with local African dappled breeds (Porter et al., 2016). Such indigenous breeds might have arrived to South Africa through the migrations of African tribes along the east coast (Smith, 1992; Porter et al., 2016). The mixture of goat populations with Asian and African origins to establish the Boer breed might explain its high levels of observed and expected heterozygosities, which in our study took values of 0.397 (H_o) and 0.405 (H_e). As a consequence of these crossbreeding practices, Boer goats and their derivatives increased in size, which is highly advantageous in meat production. Selection for meat conformation began in 1920s

but the breed society for the 'improved Boer' was established in 1959 (Porter et al., 2016). This admixed origin is supported by phylogenetic studies in which the Boer breed tends to be separated from other African breeds and shows evidence of having a mixed African and Southwest Asian background (Colli et al., 2018). We have also excluded from the gradient analysis Boer derivatives i.e. Kalahari Red and Savanna breeds. Kalahari Red goats have red coats and originate from a crossing between red-headed Boer goats and 'unimproved' indigenous South African and Namibian brown goats (Campbell, 2003). The Savanna goat breed has a white coat developed from indigenous South African breeds, also known as the white Boer goat (Campbell, 2003). This breed was highly successful because of its white colour, which was important for religious events, and its high-quality carcass (Campbell, 2003). Population structure studies have highlighted that Kalahari Red and Savanna goats display a high level of introgression from the Boer breed (Chokoe et al., 2020). As the Boer breed originated from a crossbreed and was in turn used as a source of introgression for the improvement of other breeds from South Africa (e.g., Kalahari Red), we decided to exclude Boer and its derivatives from the gradient of diversity analysis.

Finally, one potential limitation of **study 1** is that it did not include ancient samples. Their inclusion would have been very helpful to identify with more confidence a gradient of genetic diversity since in our study the genetic signature left by the post domestication dispersal goats is superseded by many subsequent demographic and migratory events that took place well after. For example, Daly et al. (2018), through the use of genomic data from ancient goat samples (from the Palaeolithic to the Middle Ages), demonstrated the existence of multiple goat domestication centres across the Fertile Crescent.

4.2 The African introgression of Southern European cattle and goat breeds

4.2.1 Tracing the origins of the African introgression of Murciano-Granadina Goats

The geographic origin of the introgression of Murciano-Granadina goats by their African counterparts is the subject of **study 2** of this thesis. In addition, we aimed to characterize at a large scale the magnitude of such introgression. Our results show the presence of an African component in the genetic background of Murciano-Granadina goats. This result is consistent with the study of Colli et al. (2018), based on whole-genome SNPs genotyping data of goat breeds with a worldwide distribution, which revealed the presence of African alleles in European goat breeds. This study detected signs of introgression by West African breeds into Spanish breeds and to a lesser extent in French and Italian breeds. These events, which probably occurred after the post-domestication

expansion process, may have contributed to increase the genetic diversity of Southern European goat populations. Several lines of evidence exist about the occurrence of gene flow between cattle, sheep and goat populations from North Africa and the Iberian Peninsula (Pereira et al., 2006; Decker et al., 2016, Ben Jemaa et al., 2019), further supporting our result. A bidirectional genetic flow between North African and southern Spanish breeds has been detected in goats (Martinez et al., 2016), consistent with the discovery of genetic connections between Maghreb and Spanish goat breeds, as previously described by Pereira et al. (2006). The presence of gene flow between a Moroccan sheep breed (D'man) and a Spanish sheep breed (Castellana) was highlighted in a population structure study of Maghreb sheep, with the Castellana sheep positioned close to the North African clade (Ben Jemaa et al., 2019). In cattle, there is proof of the African introgression in Mediterranean breeds especially in Italian and Spanish cattle breeds (Decker et al., 2014, Vidal et al., 2017). This suggest that introgression events could be due to trading activities which led to the admixture of breeds from different parts of the Mediterranean Sea (Decker et al., 2014). These findings are also supported by studies of genetic diversity in humans, where the North African ancestry present in Spanish and Canarian populations is attributable to populations originating from the Maghreb (Moroccans, Tunisian Berbers and West Saharans) rather than from sub-Saharan Africa (Botiguè et al., 2013). Another result highlighted by Botiguè and co-authors is the clear difference in the distribution of the African component in southern European populations: for example, Spain has a higher level of African introgression than France (Botiguè et al., 2013).

The African introgression of Murciano-Granadina goats may be explained by the intertwined histories of the Iberian Peninsula and Northern Africa as well as by the close distance separating both territories i.e. the Strait of Gibraltar narrows to 8 miles (13 km) in width between Point Marroquí (Spain) and Point Cires (Morocco), constituting a corridor for the exchange of livestock and goods. This proximity and flow between the two territories (South Spain and Morocco) may explain the most probable Moroccan origin of African introgression in Murciano-Granadina goats that we highlighted in **study 2**, with an average percentage of around 4%. This result is consistent with a genetic study involving seven Spanish goat breeds, including the Murciano-Grandina breed. In this study, an important (25%) African component was found in the genetic background of two Andalusian breeds (Malagueña and Murciano-Granadina) (Manunza et al., 2016). In addition, in this work, it is evident that the African component is almost absent in the Bermeya and Blanca de Rasquera goat breeds, which originate from northern Spain (Manunza et al., 2016).

The presence of African blood in southern Spanish goat breeds is well established from our data and previous studies, although it is complicated to ascertain when these African introgression events occurred. As shown in a study of wild boar (*Sus scrofa*), contacts between breeds of southern Spain and North Africa (Morocco) occurred as early as the late Pleistocene (around 90,000 years ago) (Soria-Boix et al., 2017). Wild animals may have crossed the Gibraltar Strait during glacial periods in the Pleistocene because of a succession of sea-level increases and decreases which facilitated passage (Gibert et al., 2003). Even during the Neolithic, there is evidence of possible contacts and influences between North African and South Spanish populations. For example, Neolithic Andalusian human populations appear to be composed of several groups, including one of African origin (Caro, 2002). This can be deduced from archaeological evidence indicating the use of red ochre for the decoration of ceramic surfaces, a technique that seems to be typical of African cultures in the Maghreb (Camps, 1984; Escacena, 2000). In addition, the Egyptian Badaric culture dated to the 5th millennium B.C. and the Late Neolithic culture in western Andalusia shared several similarities. This evidence suggest that these two cultures may have originated from a common area, such as pastoral societies inhabiting the Sahara (Mederos, 1996).

Moving forward in historical events and through the study of genetic evidence in ancient samples of North African and Spanish breeds (Colominas et al., 2015), a further occasion of contact may have occurred during the period of conquest of the Iberian Peninsula by the Romans (218 BC). During the Roman Empire, animals were transported from North Africa to supply Roman troops (Colominas et al., 2012). The presence of a Moroccan genetic component in Murciano-Granadina goats may also be due to the Muslim conquest of the Iberian Peninsula by the Moors, which began in the 8th century and ended in the 15th century. The presence of the Moors for such a long period of time surely facilitated the importation and exchange of livestock by the Berbers into Spanish territories (Cymbron et al., 1999). Indeed, during the Arab-Berber invasion of the Iberian Peninsula, around 40,000 Arabs and Berbers settled in Spain, arriving from the Middle East as well as from north Africa, which had previously been invaded by Muslims (Cymbron et al., 1999; Franco Moreno, 2005). Thus, North African Berbers may have transported animals of African origin (specifically from the Maghreb) to Spain which subsequently introgressed local populations. Within the 10 datasets analyzed by us, it is important to note that the percentages of African (mainly Moroccan) components in the Murciano-Granadina goats vary considerably (range from 0% to 12%). This result suggests that the assessment of admixture and introgression in domestic animals requires the use of a large number of individuals. Most of the studies analysing population structure

and admixture events sample about 25 individuals per population (Manunza et al., 2016; Chokoe et al., 2020; Papachristou et al., 2020), which may be insufficient for estimating the proportion of ancestral components. It is important to emphasise that in our case, a supervised analysis was carried out with the software Admixture, which is more precise in calculating estimates than the classic unsupervised analysis (Alexander and Lange, 2011). As shown in our **study 2** depending on the data set selected, considerable discrepancies regarding the magnitude of the Moroccan introgression of Murciano-Granadina goats can be observed, which can lead to opposite conclusions depending on the data set of choice.

4.2.2   Analyzing the diversity and African introgression of Sardinian cattle

In the **study 3** of this thesis, we analysed a portion of the hypervariable region of mitochondrial DNA in the three local Sardinian cattle breeds: Sarda, Sardo-Bruna and Sardo Modicana breeds, officially recognised (FAO, 2023). We observed a high level of differentiation between the three breeds (approximately 8% variation between populations (AMOVA)). The diversity of Sardinian local breeds at the genomic level has already been studied, confirming a high genetic differentiation between the three Sardinian cattle breeds (Cesarani et al., 2018; Mastrangelo et al., 2018). Regarding the genetic relationship between these breeds, we carried out a MJ network (see **study 3 – Figure 2A**) that shows the distribution of haplotypes with a 'star-like' structure, a typical indication of a founder effect and of the isolation given by insularity (Ferrando et al., 2015; Cardoso et al., 2018). However, in a study of homozygosity levels in island and continental breeds, it was seen that breeds originating from Sardinia showed similar levels of homozygosity when compared to their continental counterparts (Cardoso et al., 2018). This can be explained by its geographical position, in the centre of the Mediterranean Sea, whereby Sardinia has long been at the crossroads of trade routes connecting distant coastal enclaves. Such populations transported their livestock through long-distance trips, facilitating the admixture of cattle with different genetic profiles and thus contrasting the loss of genetic diversity produced by the founder effect (Cardoso et al., 2018). We also integrated public mtDNA data obtained from cattle from Central and Southern Europe, North Africa, and the Middle East into our dataset to investigate the relationships between these breeds and Sardinian cattle. As previously reported by Achilli et al. (2009), almost all the haplotypes were attributed to haplogroup T3, the most widespread in Europe. In addition, a haplotype (H6) showing the mutational motif of the control region of haplogroup T1 was assigned

to sub-haplogroup T1b1, due to the presence of the control region transition (16022) (Bonfiglio et al., 2012). We highlighted two nucleotide variations (15948 and 16136) unique to Sardinian cattle. The high mitochondrial diversity of Sardinian cattle, as evidenced by us and others, could be due to the fact that Sardinia is one of the Italian islands at the centre of the Mediterranean Sea with a complex history, marked by the influence of various civilisations and cultures over the centuries: starting with the first Neolithic and Eneolithic populations (Wilkens, 2012), Phoenicians, Carthaginians, Romans, Saracens, Byzantines and ending with the reign of the Aragon Crown from Spain (Abulafia et al., 2011; Murgia, 2012). Maritime routes between North Africa and Southern Europe had been traversed since the Neolithic period (Zilhão, 2001), probably leading to the spread of the T1 haplotype directly from Africa or indirectly from Spain although not in recent times (Beja-Pereira et al., 2006).

The introduction of Bos taurus in Sardinia might have occurred during the Nuragic period (1,800 – 900 B.C.), from the Bronze Age to the Iron Age. The Nuragics, indigenous inhabitants of the island, were excellent sailors and many authors believe they were the 'people of the sea' mentioned in Egyptian texts (Cavillier, 2019). The Egyptians called them 'Shardana', and they traded with numerous merchant population in different regions of the Mediterranean (Cavillier, 2019). The presence of a Near Eastern haplotype on the islands of Sardinia and Sicily during the Bronze Age has been revealed through the analysis of the mitochondrial DNA of ancient Sus scrofa samples. This probably indicates the maritime transport of domestic pigs by humans, at least during 1,600-1,300 BC (Lega et al., 2016). All these civilisations have left their traces on the island, which are often found on a genetic level, both in human genetics (Di Gaetano et al., 2014) and in livestock genetics such as in cattle (Cesarani et al., 2018; Mastrangelo et al., 2018).

In **study 3**, we showed that Sardinian cattle do not share nucleotide variations with Spanish cattle, both for the T1 and T3 haplogroups. Spanish colonisation of the island has been relatively recent (1400-1700), however during the reign of the Crown of Aragon there is evidence of the existence of the so-called 'ruta de las islas', the main trade route (Murgia, 2012). This commercial route connected the western Mediterranean areas (Barcelona, the kingdom of Valencia and the Balearic Islands) with North Africa and the Near East (Manca et al., 1966; Anatra, 1981). During these exchanges, it is possible that not only cereals, salts and metals were transported but also livestock, leading to crossbreeding and gene flow between breeds with an African introgression already present from different areas of the Mediterranean trade routes.

5 Conclusions

1. We have observed that the heterozygosity of African goat populations decreases with distance to the domestication center in Southeastern Anatolia, while such trend is not evident in European populations. The post-domestication dispersal of the Neolithic stock in Africa probably took place fundamentally by land. This feature combined with the existence of important geographic and biological barriers might have contributed to preserve the gradient of diversity detected in our study. In Europe, in contrast, maritime transportation might have played a more relevant role in the diffusion of the Neolithic caprine stock across the continent and, moreover, the absence of important barriers to gene flow in Europe contributed to erase any gradient of diversity produced by the dispersal of domesticates.

2. The Murciano-Granadina breed has been introgressed by Moroccan goats, and such levels of introgression are quite variable with an average of 4% and a range of 0-12%. Gene flow between Moroccan and Murciano-Granadina breeds would have taken place as early as the Bronze Age, since there is archaeological evidence of contacts across the Strait of Gibraltar, or during the Roman or Muslim occupations of the Iberian Peninsula.

3. A moderate haplotype diversity was found in the Sarda and Sardo Bruna breeds, whereas a lower level of haplotypic variation was found in the Sarda Modicana cattle. The T3 haplogroup (the most widespread in Europe) was the most abundant one in Sardinian cattle, but we also identified haplotypes belonging to the T1 and T1′2′3 haplogroups. The presence of haplogroup T1 is compatible with the African introgression of Sardinian bovine breeds, an event which may have occurred directly or indirectly given the strategic location of Sardinia in the centre of the Mediterranean Sea.

6 References

Abulafia, D. (2011). The Great Sea: a human history of the Mediterranean. Oxford University Press.

Achilli, A. et al. (2009). The multifaceted origin of taurine cattle reflected by the mitochondrial genome. PLoS One. 4 (6), e5753.

Ahmad, H.I. et al. (2020). The Domestication Makeup: Evolution, Survival, and Challenges. Front. Ecol. Evol. 8, 103.

Aitken, N. et al. (2004). Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. Mol. Ecol. 13(6), 1423-1431.

Ajmone-Marsan, P. et al. (2023). Genomic characterization of animal genetic resources. FAO Animal Production and Health Guidelines. Rome, FAO. 32.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinform. 12, 246.

Alexander, D.H., et al. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Amills, M. et al. (2017). Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. Anim. Genet. 48(6), 631–644.

Ammerman, A.J. (2010). The First Argonauts: Towards the Study of the Earliest Seafaring in the Mediterranean. In: A. Anderson, J. Barrett, J., K. Boyle (eds.) The Global Origins and Development of Seafaring. The McDonald Institute for Archaeological Research, Cambridge, 81-92.

Anatra, B. (1989). Economia sarda e commercio mediterraneo nel basso medioevo e nell'età moderna. In: Storia dei sardi e della Sardegna, III. L'età moderna. Dagli Aragonesi alla fine della dominazione spagnola. M. Guidetti (ed.). Jaca Book, Milano, 109-132.

Anderung, C. et al. (2005). Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. Proc. Natl. Acad. Sci. 102(24), 8431–8435.

Bandelt, H.J. et al. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16, 37-48.

Beja-Pereira, A. et al. (2006). The origin of European cattle: Evidence from modern and ancient DNA. Proc Natl Acad Sci. 103, 8113-8118.

Ben Jemaa, S. et al. (2019). Inferring the population structure of the Maghreb sheep breeds using a medium-density SNP chip. Anim. Genet. 50, 526-533.

Biscarini, F. et al. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. Front. Genet. 6, 33.

Biscarini, F. et al. (2018). Lessons learnt on the analysis of large sequence data in animal genomics. Anim. Genet. 49, 147–158.

Blench, R.M. (2000). A survey of ethnographic and linguistic evidence for the history of livestock in Africa. In: The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics, and Ethnography, ed. Blench R.M. and MacDonald K. CA. London and New York: Routledge. pp. 18–27.

Boettcher, P. J. et al. (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. Anim. Genet. 41, 64-77.

Bonfiglio, S. et al. (2010). The enigmatic origin of bovine mtDNA haplogroup R: Sporadic interbreeding or an independent event of Bos primigenius domestication in Italy? PLoS One, 5(12), e15760.

Bonfiglio, S. et al. (2012). Origin and Spread of Bos taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1. PLoS One. 7(6), e38601.

Botigué, L. R. et al (2013). Gene flow from North Africa contributes to differential human genetic diversity in southern europe. Proc. Natl. Acad. Sci. U.S.A. 110(29), 11791-11796.

Bradić, M. et al. (2011). Genotyping with Sequenom. Methods mol. biol. 772, 193–210.

Broodbank, C. & Strasser, T.F. (1991). Migrant farmers and the Neolithic colonization of Crete. Antiquity. 65, 233-245.

Campbell, Q.P. (2003). The Origin and Description of Southern Africa' s Indigenous Goats. S. Afr. J. Anim. Sci. 4, 18–22.

Camps, G. (1984). In: Scripta Praehistorica: Francisco Jordá Oblata, ed. Fortea, J. University of Salamanca, Salamanca, Spain. 187-208.

Cañon, J. et al. (2006). Geographical partitioning of goat diversity in Europe and the Middle East. Anim. Genet. 37(4), 327–334.

Caramelli, D (2006). The origins of domesticated cattle. Hum. Evol. 21, 107–122.

Cardoso, T.F. et al. (2018). Patterns of homozygosity in insular and continental goat breeds. Genet. Sel. Evol. 50, 56.

Caro, A. (2002). Ensayo sobre cerámica en arqueología. Sevilla: Eds Agrija.

Cavalli-Sforza, L. L. et al. (1993). Demic Expansion and Human Evolution. Science. 259, 639-646.

Cavalli-Sforza, L.L. et al. (1994) The history and geography of human genes. Princeton University Press, Princeton.

Cavillier, G. (2002). Gli Shardana dell'Egitto o l'Egitto degli Shardana? La visione del mercenario nell'Egitto ramesside. Aegyptus, 82, 67–80.

Cesarani, A. et al. (2018). Genome-Wide Variability and Selection Signatures in Italian Island Cattle Breeds. Anim. Genet. 49, 371-383.

Cesarani, A. et al. (2022). Unravelling the effect of environment on the genome of Sarda breed ewes using Runs of Homozygosity. J. Anim. Breed. Genet. 139, 292-306.

Chang, C.C., et al. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4, 7.

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

Chen, S. Y. et al. (2005). Mitochondrial diversity and phylogeographic structure of Chinese domestic goats. Mol. Phylogenet. Evol. 37, 804–814.

Chokoe, T. C. et al. (2020). Genetic Diversity of South African Indigenous Goat Population from Four Provinces Using Genome-Wide SNP Data. Sustainability, 12, 10361.

Chritz, K.L. et al. (2015). Environments and trypanosomiasis risks for early herders in the later Holocene of the Lake Victoria basin, Kenya. Proc. Natl. Acad. Sci. 112, 3674–3679.

Ciani, E. et al. (2020). On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genet. Sel. Evol. 52, 25.

Cinar Kul, B. et al. (2015). Y-chromosomal variation of local goat breeds of Turkey close to the domestication centre. J. Anim. Breed. Genet. 132, 449-453.

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

Colli, L. et al. (2018). Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet. Sel. Evol. 50, 58.

Colominas, L. & Edwards, C. J. (2017). Livestock Trade during the Early Roman Period: First Clues from the Trading Post of Empúries (Catalonia). Int. J. Osteoarchaeol. 27, 167–179.

Colominas, L. et al. (2015). Detecting the T1 cattle haplogroup in the Iberian Peninsula from Neolithic to medieval times: New clues to continuous cattle migration through time. J. Archaeol. Sci. 59, 110–117.

Colominas, L. et al. (2022). Supplying the Roman army? Animal production and consumption in the countryside of north-eastern Iberia (Can Rubió, la Garrotxa). J. Archaeol. Sci.: Reports. 45, 103561.

Courtin, F. et al. (2010). Updating the Northern Tsetse Limit in Burkina Faso (1949–2009): Impact of Global Change. Int. J. Environ. Res. Public Health. 7, 1708–1719.

Cox, C. et al. (2016). Biogeography: An Ecological and Evolutionary Approach, 9th Edition.

Cummings, V. & Morris, J. (2022). Neolithic Explanations Revisited: Modelling the Arrival and Spread of Domesticated Cattle into Neolithic Britain. Environ. Archaeol. 27, 20-30.

Cymbron, T. et al. (1999). Mitochondrial Sequence Variation Suggests an African Influence in Portuguese Cattle. Proc. Biol. Sci. 266, 597–603.

Cymbron, T. et al. (2005). Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. Proc. R. Soc. B. 272, 1837–1843.

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

Dayo, G. K. et al. (2009). Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. Mol. Ecol. 18, 1801–1813.

Decker, J. E. et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet. 10, e1004254.

Delgado, J. V. et al. (2018). Murciano-Granadina goat: A Spanish local breed ready for the challenges of the twenty-first century. Sustainable Goat Production in Adverse Environments: Volume II: Local Goat Breeds, 205-219.

Dettori, M. L. et al. (2020). Assessing the diversity and population substructure of sarda breed bucks by using mtdna and y-chromosome markers. Animals. 10, 2194.

Di Gaetano, C. et al. (2014) Sardinians Genetic Background Explained by Runs of Homozygosity and Genomic Regions under Positive Selection. PLoS One. 9, e91237.

Di Lorenzo, P. et al. (2018). Mitochondrial DNA variants of Podolian cattle breeds testify for a dual maternal origin. PLoS One. 13, e0192567.

Dorn, P. L. et al. (2017). Genetics of Major Insect Vectors. In Genetics and Evolution of Infectious Diseases. Elsevier. pp 341–382.

Druet, T. & Gautier, M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. Mol. Ecol. 26, 5820–5841.

Dunbar, R. (1984). Scapegoat for a thousand deserts. New Sci. 104, 30 –33.

Ebhodaghe, F. et al. (2018). A systematic review and meta-analysis of small ruminant and porcine trypanosomiasis prevalence in sub-Saharan Africa (1986 to 2018). Acta Trop. l88, 118-131.

Edwards, C. J. et al. (2011). Dual origins of dairy cattle farming - Evidence from a comprehensive survey of european Y-chromosomal variation. PLoS One. 6, e15922.

Efstratiou, N. et al. (2004). The Neolithic settlement of Knossos: new light on an old picture. Annu. Br. Sch. at Athens.12, 39-49.

Escacena, J. L. (2000). In: Prehistoric Iberia: Genetics, Anthropology, and Linguistics, ed. Arnaiz-Villena, A. (Kluwer–Plenum, New York), pp. 125-162.

Eusebi, P. G. et al. (2020). Genomic tools for effective conservation of livestock breed diversity. Diversity. 12, 8.

Excoffier, L. & H.E. L. Lischer (2010). Arlequin suite v. 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564-567.

Excoffier, L. et al. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genet. 131, 479-491.

Felius M. et al. (2015). Conservation of cattle genetic resources: the role of breeds. J. Agric. Sci. 153, 152–62.

Felius, M. et al. (2014). On the history of cattle genetic resources. Diversity. 6, 705-50.

Ferrando, A. et al.(2015). A mitochondrial analysis reveals distinct founder effect signatures in Canarian and Balearic goats. Anim. Genet. 46, 452-456.

Francis, R. M. (2017). Pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17, 27-32.

Franco Moreno, B. (2005). Distribución y asentamientos de tribus bereberes (Imazighen) en el territorio emeritense en época emiral (S. VIII-X). Arqueología y territorio medieval.12, 39-50.

Frankham, R. et al. (2010). An Introduction to Conservation Genetics. University Press, New York, Cambridge, pp. 618.

Garcea, E. (2004). An alternative way towards food production: the perspectives from the Libyan Sahara. J. World Prehist. 18, 107-154.

Gautier, A. (1987). Prehistoric men and cattle in North Africa: a dearth of data and a surfeit of models. In: Prehistory of Arid North Africa, ed. A.E. Close. Dallas: Southern Methodist University Press. pp. 163–187.

Gibert, J. et al. (2003). The Gibraltar Strait: A Pleistocene door of Europe?. Hum. Evol. 18, 147-160 (2003).

Gifford-Gonzalez, D. & Hanotte, O. (2011). Domesticating animals in Africa: implications of genetic and archaeological findings. J. World Prehist. 24, 1-23.

Gilman, A. (1975). The Later Prehistory of Tangier, Morocco. Cambridge, Mass.: Peabody Museum. Bulletin of the American School of Prehistoric Research, 29.

Ginja, C. et al. (2019). The genetic ancestry of American Creole cattle inferred from uniparental and autosomal genetic markers. Sci. Rep. 9, 11486.

Groeneveld, L.F. et al. (2010). Genetic diversity in farm animals- a review. Anim. Genet. 41, 6-31.

Guintard C. (1999). On size of the ure-ox or aurochs (Bos primigenius Bijanus, 1827). In: Weniger G.C. (ed.). Archäologie und Biologie des Auerochesen: Mettmann: Neanderthal Museum. pp 7–21.

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, 68-76.

Higgs, E.S. (1967). Environment and chronology – the evidence from mammalian fauna. In: The Haua Fteah (Cyrenaica and the Stone Age of the South-East Mediterranean). ed. CA.B.M McBurney. Cambridge: Cambridge University. pp 149-164.

Holder, M. & Lewis, P. (2003). Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275-284.

Holland, P. M. et al. (1991). Detection of specific polymerase chain reaction product by utilizing the $5' \rightarrow 3'$ exonuclease activity of Thermus aquaticus DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 88, 7276-7280.

Isern, N. et al. (2017). Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc. Natl. Acad. Sci. U.S.A. 114, 897-902.

Jarcho, J. (2001). Restriction fragment length polymorphism analysis. Curr. Protoc. Hum. Genet. Chapter 2, Unit 2.7. 1-15.

Kijas, J. W. et al. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10, e1001258.

Klein, R.G. & Scott, K. (1986). Re-analysis of faunal assemblages from the Haua Fteah and other Late Quaternary archaeological sites in Cyrenaican Libya. J. Archaeol. Sci. 13, 515-542.

Kristensen, T.N. et al. (2015). What can livestock breeders learn from conservation genetics and vice versa? Front. Genet., 6, 38.

Lega, C. et al. (2016). Like a pig out of water: seaborne spread of domestic pigs in Southern Italy and Sardinia during the Bronze and Iron Ages. Heredity. 118, 154-159.

Lenstra J.A. & The ECONOGENE Consortium (2005). Evolutionary and demographic history of sheep and goats suggested by nuclear, mtDNA and Y-chromosomal markers. Proc. Int. Workshop 'The role of biotechnology for the characterization of crop, forestry, animal and fishery genetic resources'. Turin, March 5–7, 2005.

Lenstra, J. A. et al. (2017). Microsatellite diversity of the Nordic type of goats in relation to breed conservation: how relevant is pure ancestry? J. Anim. Breed. Genet. 134, 78-84.

Liverani, M. (2000). The Garamantes: a fresh approach. Libyan Studies. 31,17-28.

Lou, R.N. et al. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. Mol. Ecol. 30, 5966-5993.

Luigi-Sierra, M. G. et al. (2022). Genomic patterns of homozygosity and inbreeding depression in Murciano-Granadina goats. J. Anim. Sci. Biotechnol. 13, 35.

Luikart, G. et al. (2001). Multiple maternal origins and weak phylogeographic structure in domestic goats. Proc. Natl. Acad. Sci. 98, 5927-5932.

MacDonald, K.CA. & MacDonald, R.H. (2000). The origins and development of domesticated animals in arid West Africa. In: The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics, and Ethnography, ed. R.M. Blench.

Magee, D. A. et al. (2014). Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. Anim. Front. 4, 7-22.

Manca, C. (1966). Aspetti dell'espansione economica catalano-aragonese nel Mediterraneo occidentale. Il commercio internazionale del sale. Milano, Giuffrè. 8b, 32 -37.

Manen, C. & Sabatier, P. (2003). Chronique radiocarbone de la néolithisation en Méditerranée nord-occidentale. Bull. de la Soc. Prehist. Francaise. 100, 479-504.

Manning, K. et al. (2013a). Animal exploitation in the Early Neolithic of the Balkans and central Europe. In: The Origins and Spread of Domestic Animals in Southwest Asia and Europe. ed. S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan. Walnut Creek: Left Coast Press. pp 237-252.

Manunza, A. et al. (2016). A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. Genet. Sel. Evol. 48, 52.

Marshall, F. & Weissbrod, L. (2011). Domestication processes and morphological change: through the lens of the donkey and African pastoralism. Curr. Anthropol. 52, 397-413.

Martínez, A. et al. (2016). Detecting the existence of gene flow between Spanish and North African goats through a coalescent approach. Sci. Rep. 6, 38935.

Mastrangelo, S. et al. (2018). Conservation status and historical relatedness of Italian cattle breeds. Genet. Sel. Evol. 50, 35.

Mastrangelo, S. et al. (2020). Genome-wide detection of signatures of selection in three Valdostana cattle populations. J. Anim. Breed. Genet.137, 609-621.

McDonald, M.M.A. (1998). Early African pastoralism: a view from Dakhleh Oasis (South Central Egypt). J. Anthropol. Archaeol.17, 124-142.

McTavish, E. J. et al. (2013). New World cattle show ancestry from multiple independent domestication events. Proc. Natl. Acad. Sci. U.S.A. 110, e1398-e1406.

McVean, G. (2007). The structure of linkage disequilibrium around a selective sweep. Genet. 175, 1395-1406.

Meadow R.H. (1993). Animal domestication in the Middle East: a revised view from the eastern margin. In: Possehl GL, editor. Harappan Civilization. A Recent Perspective. 2nd ed. New Delhi: Oxford and IBH Publishing; pp 295–320.

Mederos Martín, A. (1996). La conexión Levantino-Chipriota. Indicios de comercio Atlántico con el Mediterráneo Oriental durante el Bronce Final (1150-950 a.C.). Trabajos De Prehistoria. 53, 95–115.

Meirmans, P. G. (2012). AMOVA-based clustering of population genetic data. J. Hered. 103, 744-750.

Menozzi, P. et al. (1978). Synthetic maps of human gene frequencies in Europeans. Science. 201, 786-792.

Michalakis, Y. & Excoffier, L. (1996). A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genet. 142, 1061-1064.

Murgia G. (2012). Un'isola, la sua storia. La Sardegna tra Aragona e Spagna (secoli XIV - XVII). Edizioni Grafica del Parteolla, Dolianova (CA). pp. 24.

Naderi, S. et al. (2007). Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. PLoS One. 2, e1012.

Naderi, S. et al. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals Proc. Natl. Acad. Sci. 105, 17659–17664.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U.S.A. 70, 3321-3323.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genet. 89, 583-590.

Nei, M. (1987). Molecular evolutionary genetics. Columbia university press.

Nicolazzi, E. L. et al. (2015). SNPchiMp v.3: Integrating and standardizing single nucleotide polymorphism data for livestock species. BMC Genom. 16, 283.

Nielsen, R. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One. 7, e37558.

Nijman, I. J. et al. (2022). Geographical contrasts of Y-chromosomal haplogroups from wild and domestic goats reveal ancient migrations and recent introgressions. Mol. Ecol. 31, 4364-4380.

Nikita, E. et al. (2012). Sahara: Barrier or corridor? Nonmetric cranial traits and biological affinities of North African late Holocene populations. Am. J. Phys. Anthropol. 147, 280-292.

Noce, A. et al. (2016). Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep. Anim. Genet. 47, 717–726.

Noda, A. et al. (2018) The mtDNA haplogroup P of modern Asian cattle: A genetic legacy of Asian aurochs? PLoS One. 13, e0190937.

Novembre, J. et al. (2008). Genes mirror geography within Europe. Nature. 456, 98-101.

Ojha, K. K. et al. (2022). Computational molecular phylogeny: concepts and applications. Bioinformatics, 67-89.

Olivieri, A. et al. (2015). Mitogenomes from Egyptian cattle breeds: New clues on the origin of haplogroup Q and the early spread of Bos taurus from the Near East. PLoS One. 10, e0141170.

Orton, J.D. (2012). Late Holocene archaeology in Namaqualand, South Africa: hunter-gatherers and herders in a semi-arid environment. DPhil. dissertation, University of Oxford.

Orton, J.D. (2015). The introduction of pastoralism to southernmost Africa: thoughts on new contributions to an ongoing debate. Azania: Archaeological Research in Africa. 50, 250-258.

Papachristou, D. et al. (2020). Genomic diversity and population structure of the indigenous Greek and Cypriot cattle populations. Genet. Sel. Evol. 52, 43.

Park, S.D. et al. (2015). Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. Genome Biol. 16, 234.

Pereira, F. et al. (2009). Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. Mol. Biol. and Evol. 26, 2765-2773.

Perlès, C. et al. (2013). Early seventh-millennium AMS dates from domestic seeds in the Initial Neolithic at Franchthi Cave (Argolid, Greece). Antiquity. 87, 1001-1015.

Peters, J. (1986). A revision of the faunal remains from two Central Sudanese sites: Khartoum Hospital and Esh Shaheinab. In: Archaeozoologica, 5th International Congress of Archaeozoologia, Bordeaux. pp 11-35.

Peterson, B. K. et al. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 7, e37135.

Philipsson, J. et al. (2011). Global perspectives on animal genetic resources for sustainable agriculture and food production in the tropics. In: Ojango, J.M., Malmfors, B. and Okeyo, A.M. (eds), Animal Genetics Training Resource, version 3. Nairobi, Kenya, ILRI and Uppsala, Sweden: Swedish University of Agricultural Sciences.

Piras, D. et al. (2012). Haplotype affinities resolve a major component of goat (Capra hircus) mtDNA D-loop diversity and reveal specific features of the Sardinian stock. PLoS One. 7, e30785.

Porter, V. et al. (2016) Goats. In: Mason's World Encyclopedia of Livestock Breeds and Breeding (volume 1). CABI, Oxfordshire, UK. pp 350–2.

Price, A.L. et al. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904-909.

Pritchard, J.K. et al. (2000). Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Purcell, S. et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81, 559-575.

Rauh, N. (2003). Merchants, Sailors and Pirates in the Roman World. Tempus Publishing, Stroud.

Rodero, A. et al. (1992). Primitive Andalusian livestock and their implications in the discovery of America. Arch. Zootec. 41, 383-400.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.1.

Schlumbaum, A. et al. (2006). Near East mtDNA haplotype variants in Roman cattle from Augusta Raurica, Switzerland, and in the Swiss Evolène breed. Anim. Genet. 37, 373-375.

Schulting, R.(2013). On the northwestern fringes: Earlier Neolithic subsistence in Britain and Ireland as seen through faunal remains and stable isotopes. In: The Origins and Spread of Domestic Animals in Southwest Asia and Europe. ed. S. Colledge, J. Conolloy, K. Dobney, K.Manning, and S. Shennan. Walnut Creek: Left Coast Press. pp 313-338.

Sealy, J. & Yates, R. (1994). The chronology of the introduction of pastoralism to the Cape, South Africa. Antiquity. 68, 58-67.

Smith, A. B. (1992). Origins and spread of pastoralism in Africa further annual reviews. Annu. Rev. Anthropol. 21, 25-14.

Soria-Boix, C. et al. (2017) Contacts in the last 90,000 years over the Strait of Gibraltar evidenced by genetic analysis of wild boar (Sus scrofa). PLoS One. 12, e0181929.

Tabata, R. et al. (2019). The Eurasian Steppe is an important goat propagation route: A phylogeographic analysis using mitochondrial DNA and Y-chromosome sequences of Kazakhstani goats. Anim. Sci. J. 90, 317-322.

Taberlet, P. et al. (2018). Environmental DNA: For biodiversity research and monitoring. Oxford University Press.

Tomich, P. Q. (1986). Mammals in Hawaii: A synopsis and notational bibliography. 2nd ed. Bishop Museum Press, Honolulu.

Tosser-Klopp, G. et al. (2014). Design and characterization of a 52K SNP chip for goats. PLoS One. 9, e0152632.

Traoré, A. et al. (2012). Ascertaining gene flow patterns in livestock populations of developing countries: A case study in Burkina Faso goat. BMC Genetics. 13,35.

Uerpmann, H-P. (1996): Animal domestication -accident or intention. In: The origins and spread of agriculture and pastoralism in Eurasia. D. R., ed. London: UCLPress. pp 227-237.

Upadhyay, M. et al. (2017). Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. Heredity. 118, 169-76.

Vacca, G. M. et al. (2010). Relationships between milk characteristics and somatic cell score in milk from primiparous browsing goats. Anim. Sci. J. 81, 594-599.

Verdugo M.P. et al. (2019) Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science. 365, 173-6.

Vidal, O. et al. (2017). Differential distribution of Y-chromosome haplotypes in Swiss and Southern European goat breeds. Sci. Rep. 7, 16161.

Vigne J. D. & Cucchi T. (2005). Premières navigations au Proche Orient: les informations indirectes de Chypre. Paléorient. 31, 186-194.

Waki, A. et al. (2015). Paternal phylogeography and genetic diversity of East Asian goats. Anim. Genet. 46, 337-339.

Whitlock, M. C. (2011). G'ST and D do not replace FST. Mol. Ecol. 20, 1083-1091.

Wilkens, B. (2012). Archeozoologia. Il Mediterraneo, la storia, la Sardegna. Sassari, Italy. Editrice Democratica Sarda.

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

Xia, X. et al. (2019). Genetic diversity of Chinese cattle revealed by Y-SNP and Y-STR markers. Anim. Genet. 50, 64-69.

Xiao, C. et al. (2021). The assembly of caprine Y chromosome sequence reveals a unique paternal phylogenetic pattern and improves our understanding of the origin of domestic goat. Ecol. Evol. 11, 7779-7795.

Yaro, M. et al. (2017). Molecular identification of livestock breeds: A tool for modern conservation biology. Biol. Rev. 92, 993-1010.

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

Zeder, M. A. (2017). Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In: N. Boivin, R. Crassard, & M. Petraglia (eds.), Human Dispersal and Species Movement: From Prehistory to the Present. Cambridge: Cambridge University Press. pp 261-303.

Zhang, K . et al. (2020). Evolution and domestication of the Bovini species. Anim. Genet. 51, 637-657.

Zheng, Z. et al. (2020). The origin of domestication genes in goats. Sci. Adv. 6, eaaz5216.

Zilhao, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proc. Natl. Acad. Sci. 98, 14180-14185.

Zilhão, J. (2000), The Ebro frontier: a model for the late extinction of Iberian Neanderthals. In: C.B. Stringer, R.N.E. Barton, J.C. Finlayson (eds), Neanderthals on the Edge. Oxbow Books, Oxford. pp 111-121.

Zilhao, J. (2014). Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. Eurasian Prehistory. 11, 185-200.

7   Annex

7.1 PAPER 1 – Supplementary material
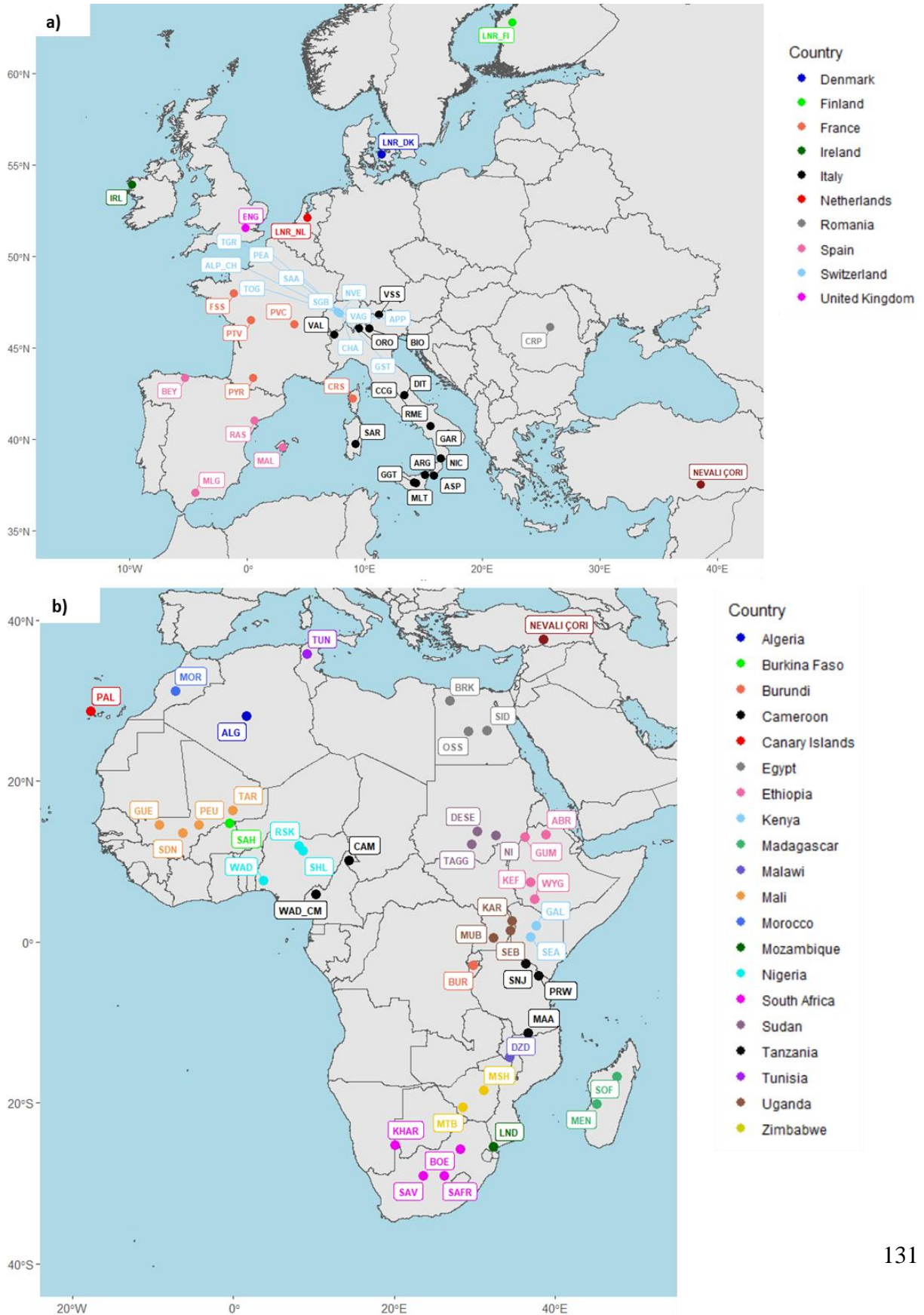
Supplementary Figures S1

Title: Geographical distribution of European (a) and African (b) breeds.

Description: It shows the points and the acronym name of the breed coloured according to the country of origin. The location of Nevali Çori is also indicated.

Principal component 2 (8.53 %)

Principal component 1 (10.58 %)

Subregions

● North Europe	● East Europe	● Central Europe	● West Europe	● South Europe
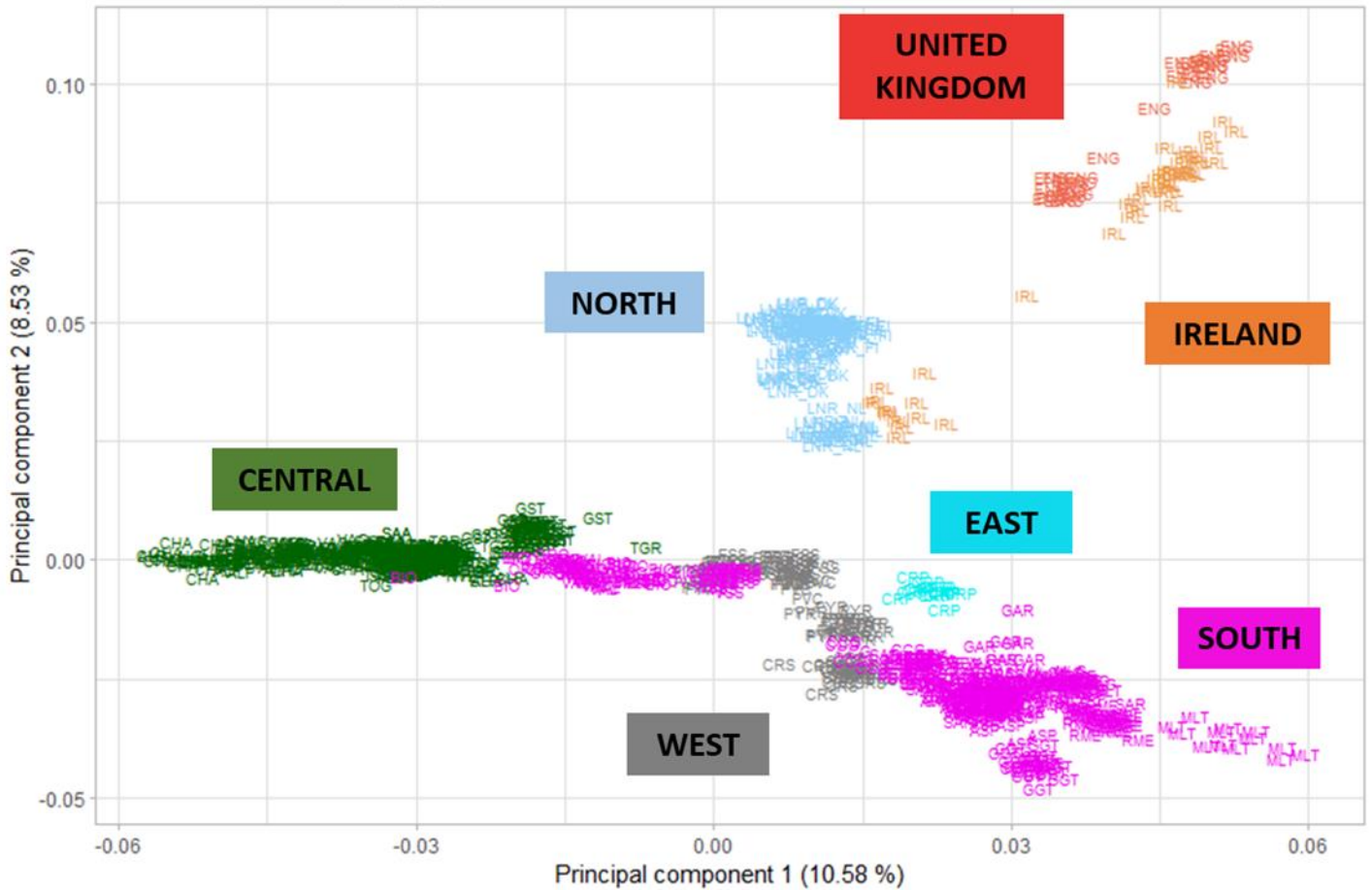
North Europe	East Europe	Central Europe		West Europe	South Europe	
LNR_DK	CRP	ALP_CH	SAA	CRS	ARG	RAS
LNR_FI		APP	SGB	FSS	ASP	RME
LNR_NL	● United Kingdom	CHA	TGR	PTV	BEY	MLG
	ENG	GST	TOG	PVC	BIO	MLT
		NVE	VAG	PYR	CCG	NIC
	● Ireland	PEA			DIT	ORO
	IRL				GAR	SAR
					GGT	VAL
					MAL	VSS

133

Supplementary Figures S2

Title: Principal Component Analysis (PCA) plot of the first two components for 38 European goat breeds.

Description: The percentage of variation explained by the two main components is shown in brackets. Individuals are coloured according to their subregion of sampling. Breed acronyms are as follows: ALP_CH= Alpine, APP=Appenzell, ARG= Argentata, ASP= Aspromontana, BEY= Bermeya, BIO= Bionda dell'Adamello, CCG= Ciociara Grigia, CHA= Swiss Chamois, CRP= Carpathian goat, CRS= Corse, MAL= Mallorquina, MLG= Malagueña, MLT=Maltese, NIC= Nicastrese, NVE= Nera Verzasca, ORO= Orobica, PEA= Peacock, PTV= Poitevine, PVC= Provençale, PYR= Pyrenean, DIT= Di Teramo, ENG= Old English Goat, FSS= Fosses, GAR= Garganica, GGT= Girgentana, GST= Grisons striped, IRL= Old Irish Goat, LNR_DK= Landrace Goat (Denmark), LNR_FI= Landrace Goat (Finland), LNR_NL= Landrace Goat (Netherlands), RAS= Blanca de Rasquera, RME= Rossa Mediterranea, SAA= Saanen, SAR= Sarda, SGB= Booted goat, TGR= Tessin grey goat, TOG= Toggenburg, VAG= Valais, VAL= Valdostana,VSS= Valpassiria.

Supplementary Figures S3

Title: Principal Component Analysis (PCA) plot of the first two components for the 43 African goat breeds.

Description: The respective percentage of variation explained by the two main components is shown in brackets. Individuals are coloured according to their subregions of sampling. Breed acronyms are as follows: ABR= Abergelle, ALG= Arabia,Makatia, and M'Zabite,Kabyle, BOE=Boer, BRK=Barki, BUR= Burundi goat, CAM= Cameroon Goat, DESE=Desert, DZD= Dedza, MAA= Maasai, MEN= Malagasy goat (Menabe), MOR= Barcha,Draa,Ghazalia, Moroccan goats, Noire de l'Atlas,Nord, MSH= Mashona, MTB= Matebele, MUB= Mubende, NI= Nilotic, OSS= Oasis, SDN= Soudanaise, SEA= Small East African, SEB= Sebei, SHL= Sahel, SID= Saidi, SNJ=Sonjo, SOF= Malagasy goat (Sofia), TAGG= Taggar, GAL= Galla, GUE= Guera, GUM= Gumez, KAR= Karamonja, KEF= Keffa, KHAR= Kalahari Red, LND= Landin, PAL= Palmera, PEU= Peulh, PRW= Pare White, RSK= Red Sokoto, SAFR= South Africa Local breeds (from Limpopo, Freestate,Gauteng,Nortwest), SAH= Sahel, SAV= Savanna. TAR= Targui, TUN= Tunisian, WAD_CM= West African Dwarf (Cameroon), WAD= West African Dwarf (Nigeria), WYG= Woyito Guji.

a) European dataset without islands
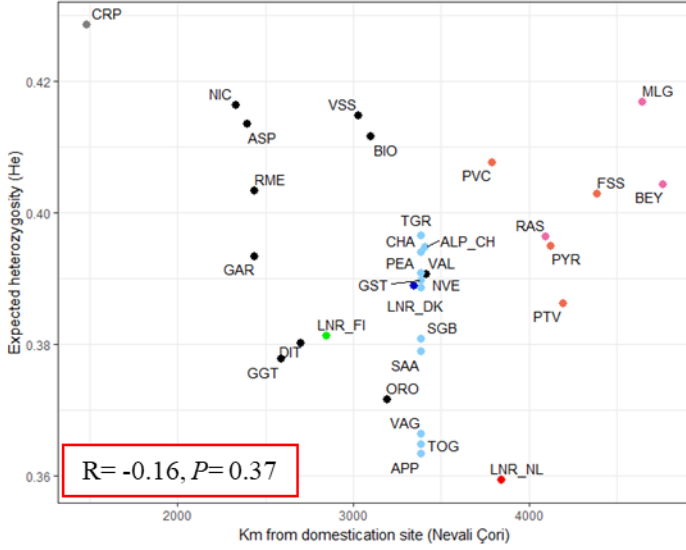
b) European dataset with islands
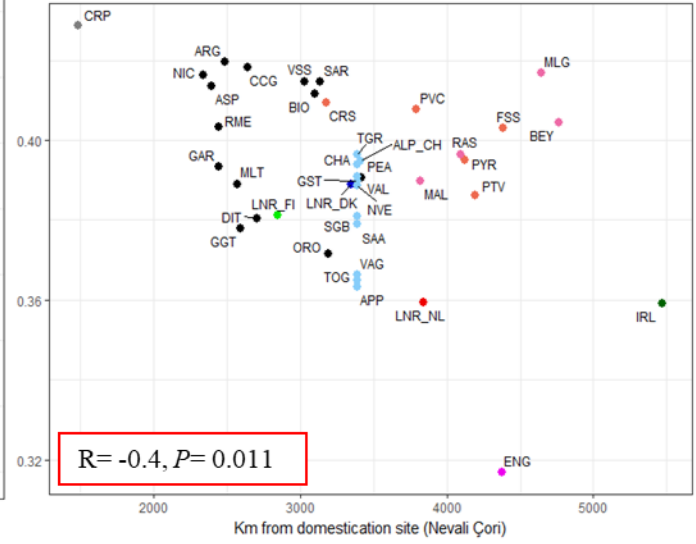
c) African dataset without islands
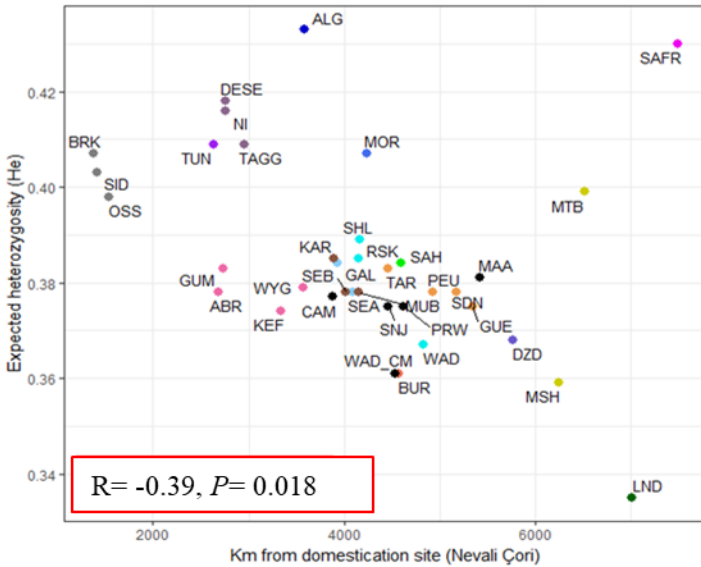
d) African dataset with islands
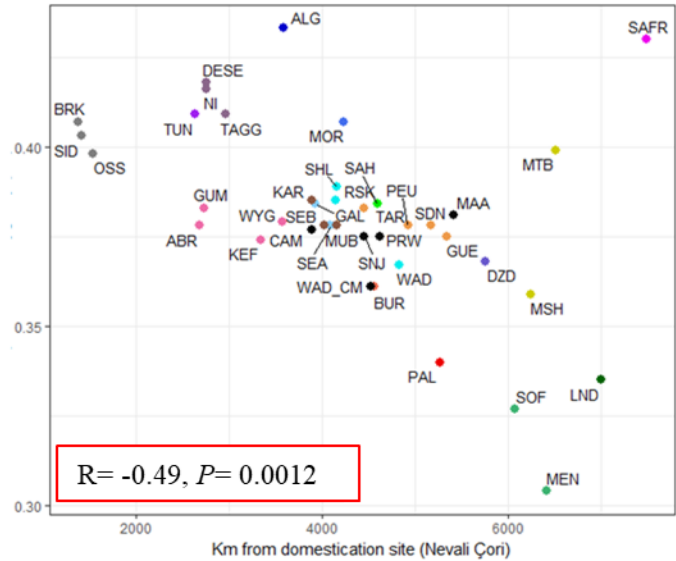
European sampling countries

- Denmark
- France
- Italy
- Romania
- Switzerland
- Finland
- Ireland
- Netherlands
- Spain
- United Kingdom

African sampling countries

- Algeria
- Canary Islands
- Madagascar
- Mozambique
- Tanzania
- Burkina Faso
- Egypt
- Malawi
- Nigeria
- Tunisia
- Burundi
- Ethiopia
- Mali
- South Africa
- Uganda
- Cameroon
- Kenya
- Morocco
- Sudan
- Zimbabwe

137

Supplementary Figures S4

Title: Correlation graph between the distance (km) from the domestication site (Nevali Çori) of sampling locations of European and African breeds and their expected heterozygosities.

Description: We report plots representing Pearson's correlations (with their P-values) between expected heterozygosity and distance between Nevali Çori and sampling location of European breeds not including (a) and including (b) island breeds and of African populations not including (c) and including (d) insular populations. Breeds acronyms can be found in Tables 1 and 2, and their country of origin is indicated with coloured points.

7.2 PAPER 3 – Supplementary material

All supplementary material are available at:

DOI: https://doi.org/10.1038/s41598-022-06420-3

8 Acknowledgements

Agradecimientos

Como conclusión de este trabajo, me gustaría mencionar a todas las personas que en mayor y en menor medida han contribuido a la realización de un trabajo de cuatro años que me ha llevado a un país maravilloso como España y me ha hecho crecer tanto profesional como humanamente.

Me gustaría empezar agradeciendo a la Universidad de Sassari (UNISS) y en particular al Departamento de Medicina Veterinaria y a la coordinadora del curso de doctorado en Ciencias Veterinarias, la profesora Fiammetta Berlinguer, que me ofrecieron la oportunidad de realizar parte de mis estudios de doctorado en la Universidad Autónoma de Barcelona (UAB), dándome la posibilidad de adquirir nuevos conocimientos, pero también de conocer diferentes culturas y personas increíbles. Gracias al profesor Vacca y a todo el grupo de investigación por haberme permitido realizar un doctorado.

Me gustaría dar las gracias a la UAB y en particular al coordinador del curso de doctorado en producción animal, el Dr. Joaquín Casellas por su disponibilidad y apoyo durante las diferentes etapas de mi periodo de estudios en Barcelona. Agradezco al Centro de Investigación en Agrigenómica (CRAG) por acogerme y ofrecerme los recursos e instalaciones necesarios para mi estancia y para la realización de mis estudios de doctorado en Barcelona.

Me gustaría agradecer a mis directores de tesis, la prof. Maria Luisa Dettori y el prof. Marcel Amills, que me guiaron con su experiencia y siempre me animaron a dar lo mejor de mí y a no rendirme. Atesoraré para siempre los valiosos consejos que recibí.

Gracias a todos los compañeros del CRAG 'viejos y nuevos': Cristina, Magi, Alice, Yron, Jesús, Joan, Ferran, Mingjing, KeWang, Manuel, Laura, Lino, etc... sin ellos no hubiera sido tan increíble. En particular, agradezco a María toda su ayuda y disponibilidad no sólo laboral sino también humanamente, y a Antonia que no sólo ha sido importante para mí en el trabajo, sino que ha sido mi familia en Barcelona.

Gracias a mis padres, a mis hermanos y cuñadas por apoyarme siempre, y a mis pequeños amores Albino y Grazia que siempre me han dado una energía infinita en los momentos un poco difíciles. Gracias a Valentino, que siempre me anima y me hace ver lo bueno de las nuevas oportunidades, apoyándome y haciéndome creer que todo es posible si lo deseas con todas tus fuerzas. También doy las gracias a mis cuñadas y cuñados, a mi ahijada y a toda la gran familia de Vale, que siempre cree en mí.