



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Master's Thesis of Biological Sciences

Investigating the origins of the
eastern Eurasian sheep using
Xiongnu-period ancient sheep
genomes from Mongolia

흉노 시기 몽골 양 유전체를 통한
동아시아 양의 유전적 기원 연구

August 2022

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Jina Yena Min

Investigating the origins of the eastern Eurasian sheep using Xiongnu–period ancient sheep genomes from Mongolia

Advised by Professor Choongwon Jeong

Submitting a master's thesis of
Biological Sciences

August 2022

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Jina Yena Min

Confirming the master's thesis written by
Jina Yena Min
August 2022

Chair	_____	(Seal)
Vice Chair	_____	(Seal)
Examiner	_____	(Seal)

Abstract

Sheep (*Ovis aries*) is one of the earliest animals domesticated by humans about 10,000 to 12,000 years ago in Southwestern Asia when human subsistence patterns drastically transformed from a hunter–gatherer to a sedentary lifestyle. Sheep have dispersed through human activities and have been well adapted around the world. Especially, domestic sheep have become essential livestock in the Mongolian grasslands as a stable source of nutrition and hydration, where crop cultivation faces challenges and nomadic pastoralism is commonly practiced. However, the evolutionary history of domestic sheep in East Asia and its associated pastoralist societies are poorly understood. In this study, to explore the origins of present–day sheep in eastern Eurasia (i.e., East Asia), the first whole–genome data of ancient sheep from the Mongolian grasslands are presented. Duurlig Nars is one of the funeral sites of Xiongnu elite, nomadic pastoralist people who played an important role in cultural exchanges between the West and the East. Genome–wide data of six individuals from the satellite burials of Duurlig Nars tomb, along with previously published 11 ancient individuals, and 514 present–day individuals were generated using the 7497K SNPs identified in this study and population genetic analyses were conducted. Results indicate the strong genetic affinity between Xiongnu–period Duurlig Nars sheep and the present–day East Asian populations. Other ancient sheep

populations also show a high genetic affinity with the present-day populations geographically close to their excavation site. Also, a combined analysis of genome-wide data and mitochondrial DNA on ancient populations suggest ancestry continuity. In conclusion, this study reports the first whole-genome data of ancient East Asian sheep and confirms that the complex genetic structure in present-day sheep formed at least 2,000 years ago. Moreover, the worldwide sheep diversity panel of 7497K SNPs identified in this study will provide a high-resolution genomic analysis in further studies.

Keywords : Population genetics, Ancient DNA, Genomics, Evolutionary genetics, Sheep, Xiongnu

Student Number : 2020-20497

Table of Contents

Abstract	i
Table of Contents	iii
List of Tables.....	v
List of Figures	vi
1. Introduction.....	1
1.1. Domestication of sheep	1
1.2. Dispersal of domestic sheep	2
1.3. Duurlig Nars tomb site of Xiongnu-period.....	4
1.4. Aim of the study.....	6
2. Materials and Methods	10
2.1. Ancient DNA extraction and sequencing	10
2.2. Ancient DNA data collection and mapping	12
2.3. Ancient DNA damage patterns.....	13
2.4. Present-day DNA data collection and mapping	13
2.5. Read coverage and molecular sex determination	14
2.6. Variant calling and filtering	14
2.7. Genotype refinement	15
2.8. Genotyping of low-coverage individuals	19
2.9. Genetic kinship analysis	19

2.10. Data compilation and the master dataset	1 9
2.11. Principal component analysis (PCA)	2 0
2.12. Outgroup- f_3 analysis.....	2 0
2.13. TreeMix analysis	2 0
2.14. F_4 -statistics analysis	2 1
2.15. Mitochondrial DNA haplogroup analysis	2 1
3. Results	2 5
3.1. Ancient genome-wide data production.....	2 5
3.2. Worldwide sheep diversity panel and genotyping	2 6
3.3. Principal Component Analysis (PCA)	3 0
3.4. Outgroup- f_3 statistics analysis	3 4
3.5. TreeMix analysis	3 7
3.6. F_4 -statistics analysis	4 5
3.7. Mitochondrial DNA haplogroup analysis	4 9
4. Discussion.....	5 4
References	5 8
국문 초록.....	6 8
Appendices.....	7 0

List of Tables

Table 1. Number of SNPs per chromosome in each step of the worldwide sheep diversity panel generating process	17
Table 2. Software and Algorithms used in this study.....	23
Table 3. Summary statistics of six Duurlig Nars individuals.	29

List of Figures

Figure 1. Geographic distribution of wild sheep in Eurasia....	7
Figure 2. Geographic distribution of ancient sheep populations used in this study	8
Figure 3. A satellite image of Duurlig Nars tomb number 160 site from Google Map	9
Figure 4. A drawing of Duurlig Nars tomb number 160	11
Figure 5. Brief pipeline layout of the worldwide sheep diversity panel generating process	18
Figure 6. Geographic distribution of worldwide sheep populations used in this study	24
Figure 7. The mapDamage misincorporation plot of sequencing reads of six Duurlig Nars individuals mapped to the reference genome.....	28
Figure 8. PCA plot of top two components (PC1 & PC2) calculated with high-coverage present-day domestic and wild individuals	32
Figure 9. PCA plot of top two components (PC1 & PC2) calculated with high-coverage present-day domestic individuals.....	33
Figure 10. Outgroup- f_3 analysis plot of the form $f_3(\text{Urial};$ Duurlig Nars, worldwide)	35

Figure 11. Outgroup– f_3 scatter plot of the form $f_3(\text{Urial};$ Duurlig Nars, worldwide)	36
Figure 12. TreeMix tree constructed with nine representative populations, including Duurlig Nars	39
Figure 13. TreeMix tree constructed with nine representative populations, including Obishir5	41
Figure 14. TreeMix tree constructed with ten representative populations, including Duurlig Nars and Obishir5.....	43
Figure 15. F_4 –statistics plot of the form $f_4(\text{Urial}, \text{Waggir};$ Duurlig Nars, worldwide)	46
Figure 16. F_4 –statistics plot of the form $f_4(\text{Urial}, \text{Duurlig}$ Nars; Tibetan, worldwide)	47
Figure 17. F_4 –statistics plot of the form $f_4(\text{Urial},$ worldwide; Duurlig Nars, Tibetan)	48
Figure 18. Maximum Parsimony Tree reconstructed using MEGA analysis	51
Figure 19. Maximum Likelihood Tree reconstructed using MEGA analysis	52
Figure 20. Geographic distribution of mitochondrial DNA haplogroup of each sheep population used in this study	53

1. Introduction

1.1. Domestication of sheep

Sheep (*Ovis aries*) is one of the earliest animal species domesticated by humans. Its history of domestication goes back between 12,000 and 10,000 years during ‘the Neolithic transition’ when human subsistence patterns drastically transformed from a hunter–gatherer to a sedentary agricultural lifestyle in the Fertile Crescent region of Southwestern Asia (Alberto et al., 2018; Conolly et al., 2011; Sherratt, 1983; Zeder, 2008). Both archeological and genetic evidence indicate that domestic sheep (*O. aries*) is descended from wild Asiatic mouflon (*O. orientalis*), which is still found in Southwestern Asia (Alberto et al., 2018; Hiendleder et al., 1998; Nadler and RS, 1973; Rezaei et al., 2010; Stiner et al., 2014). Other closely related wild sheep species in Eurasia are argali (*O. ammon*) in the mountains of Central Asia, urial (*O. vignei*) widely distributed in Central Asia, and snow sheep (*O. nivicola*) distributed in northeast Siberia (Figure 1) (Korobitsyna et al., 1974; Mallon, 1985; Shackleton, 1997). Sheep are part of the so–called ‘Neolithic package’ (e.g., wheat, lentil, sheep, goat, and cattle) and have rapidly dispersed across the old world out from the Middle East (Cilingiroglu, 2005). As a stable source of nutrition and hydration with additional supplies of leather and wool, sheep have accompanied humans and have dispersed worldwide through human–mediated migrations, coping with different ecological settings beyond hot and arid native

habitat of their wild progenitor (Auffret et al., 2014; Hatziminaoglou and Boyazoglu, 2011; Zeder, 2008). Today, sheep are found in a diverse range of geographical locations around the world with diverse morphologies and over 1,400 breeds of sheep are recognized (Scherf, 2000).

Among the various geographical locations in which sheep inhabit, sheep is one of the main domestic livestock in the nomadic pastoralist societies in the Mongolian grasslands (MacHugh et al., 2017). While challenging for crop cultivation due to continental climate and low precipitation, the Mongolian grasslands (i.e., Mongolian Steppe or eastern Eurasian Steppe) is suitable for mobile herdings of domestic livestock, such as sheep, goats, cattle, and horses (Dong, 2016; Erdős et al., 2018; Neupert, 1999; Taylor et al., 2020). Investigating the evolutionary history of essential domestic animals can give insight into relationships between humans and their domestic animals and provide a foundation for a deeper understanding of human history (McHugo et al., 2019). Nevertheless, not much is known about the dispersal of domestic sheep, originally West Asian livestock, to inner Asia (Taylor et al., 2021).

1.2. Dispersal of domestic sheep

The domestication and dispersal processes that shaped the population dynamics of sheep we observe today traditionally have been studied through two complementary approaches: archaeological and genetic approaches (Frantz et al., 2020). The first approach

based on archeological data studies animal remains based on morpho-anatomical changes of domestication (i.e., zooarchaeology) and traces of animal husbandry. For example, zooarchaeological evidence suggest that domestic sheep have reached North Africa by 5,000 BCE and Iberia, Europe by 5,400 BCE (Klein and Scott, 1986; Zilhão, 2001). In Asia, the domestic sheep remains found in the Yellow River valley show that the domestic sheep have reached Central China by 2,100–1,800 BCE (Cai et al., 2007). However, it is challenging to distinguish between sheep/goats and wild/domestic sheep at the early domesticated stages solely on osteological evidence (Frantz et al., 2020; Salvagno and Albarella, 2017).

The second approach based on genetic data retrieved from ancient remains studies ancient DNA and protein, complementing archaeological information. Previously, Information achieved from mitochondrial DNA and protein from ancient remains has been used to imply the presence of domestic sheep. For example, mitochondrial DNA analysis of ancient sheep indicates the presence of domestic sheep in Inner Mongolia dates to 4,700 BCE through phylogenetic analysis (Cai et al., 2011). Sheep remain from the Afanasievo culture site in Altai region, a possible vector of ruminant livestock to East Asia, from 3,300–2,900 BCE had mitochondrial DNA haplogroup ‘D’ that only exists in domestic sheep, complementing the proteomic evidence of consumption of caprinae (sheep/goats) milk of Afanasievo culture in 3,000–2,500 BCE (Hermes et al., 2020; Wilkin et al., 2020). However, mitochondrial DNA and protein cannot provide sufficient information to test the ancestry, in local context, as

mitochondrial DNA is uniparental and broadly distributed and proteomic results cannot directly provide evidence for which animal species (i.e., sheep or goat) were exploited for dairy (Rüther et al., 2022). Thus, more profound understanding of the origins of domestic sheep can be achieved using genome-wide data. With the development of sheep reference genome and recent advances in sequencing methodology, genome-wide data of ancient genomes is used to demonstrate the evolutionary history of domestic sheep (McHugo et al., 2019). Only recently, paleogenomic studies on genome-wide data of ancient domestic sheep in Central Asia, Anatolia, and Iran have been reported (Rossi et al., 2021; Taylor et al., 2021; Yurtman et al., 2021). However, whole-genome data of ancient sheep in East Asia has not yet been reported.

1.3. Duurlig Nars tomb site of Xiongnu-period

Xiongnu was a confederation of nomadic pastoralists and founded the first nomadic empire in Eurasia, dominating the Eurasian steppe from 200 BCE to 100 CE (Miller, 2014; Pilipenko et al., 2018). Only in recent years have their complex origins been revealed using ancient genome data from Eurasian nomads (Jeong et al., 2020). Duurlig Nars is one of the funeral sites of Xiongnu elite located at Bayan-Adarga sum in the north of Khenti Aimag, 500 km northeast of the capital of Mongolia, Ulaanbaatar (Figure 2; Figure 3). Although Duurlig Nars site was first discovered in 1974, only a few expeditions were conducted and little is known about the site (Kim et al., 2010). According to a joint Japanese-Mongolian expedition in

1991, there are estimated to be more than 200 tombs in the Duurlig Nars site (Kim et al., 2010). Duurlig Nars tomb number 160 is a square-shaped grave with an adjoining entrance passage to the south. It is the biggest sized ‘凸’ shaped tomb in Duurlig Nars site with the length of the tomb of 75 m, width of 33 m, and an area of 1,573.8 m² (Figure 3; Figure 4) (National Museum of Korea, 2021). This unique ‘凸’ shape and the massive size are the characteristics of Xiongnu elite tomb, distinct from the average Xiongnu lower-class tombs (Eregzen, 2010). These ‘凸’ shaped Xiongnu elite tombs often, but not always, have circular satellite burials (Eregzen, 2010). Six satellite burials of Xiongnu-period Duurlig Nars tomb number 160 were discovered during the Korean–Mongolian joint expedition in 2019 (Figure 4) (National Museum of Korea, 2021). Each burial was occupied with one human individual in a wooden coffin along with domestic animal remains, including sheep, goats, horses, and cattle. These main domestic livestock will provide a deeper understanding of subsistence patterns of nomadic pastoralists of Xiongnu-period that roamed the Eurasian steppe.

1.4. Aim of the study

Sheep is one of the animal species first domesticated by humans and now play an essential role in the nomadic pastoralists societies in the Mongolian grasslands. However, the evolutionary history of domestic sheep in East Asia and its associated pastoralists societies are poorly understood. Ancient genomic study of sheep can provide insight into the origins of domestic sheep. Nevertheless, ancient whole-genome data of East Asian sheep has not been reported so far, especially in the Mongolian grasslands, which is the core of nomadic pastoralists societies (Neupert, 1999). Duurlig Nars is one of the funeral sites of Xiongnu elite, located at the northerneast of Mongolia. Interestingly, domestic animal remains were excavated along with humans, including sheep, in the satellite burials of Duurlig Nars tomb number 160. These remains are valuable resources as the Xiongnu empire expanded their power through cultural exchanges of the West and the East and were often accompanied by sheep (Houle and Broderick, 2011). Here in this study, to explore the genomic profile and the origins of East Asian sheep, I aimed to (1) generate whole-genome sequenced data of ancient sheep from the Xiongnu-period, (2) generate a worldwide sheep diversity panel to perform genomic analysis on a high-resolution scale, (3) and conduct population genomic analysis on ancient and present-day sheep populations.

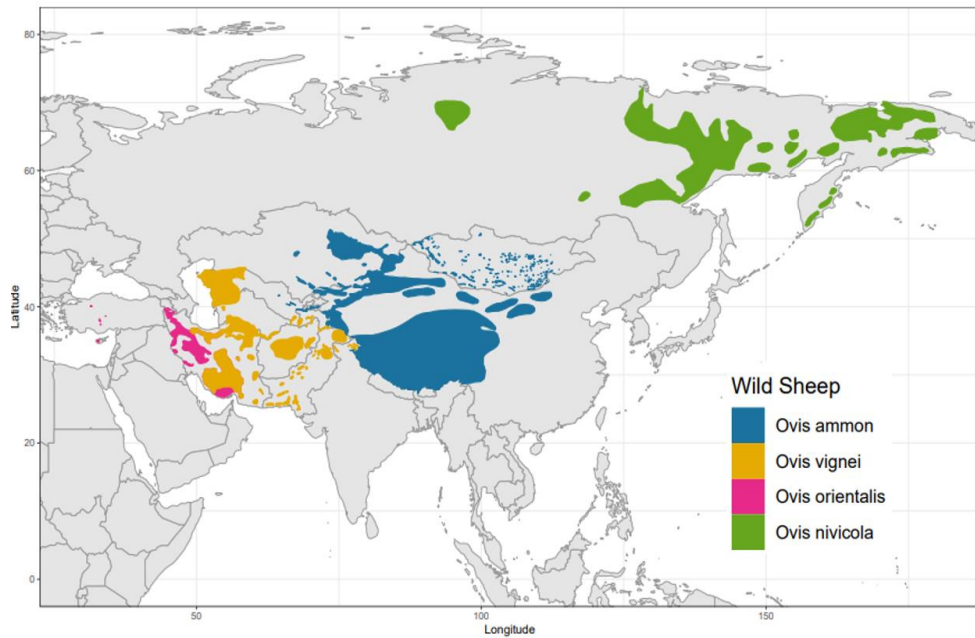


Figure 1. Geographic distribution of wild sheep in Eurasia. Geographic locations of four wild sheep species (i.e., *O. ammon*, *O. vignei*, *O. orientalis*, and *O. nivicola*) were based on the IUCN Red list (downloaded from <https://www.iucnredlist.org>, accessed on July 18, 2022).

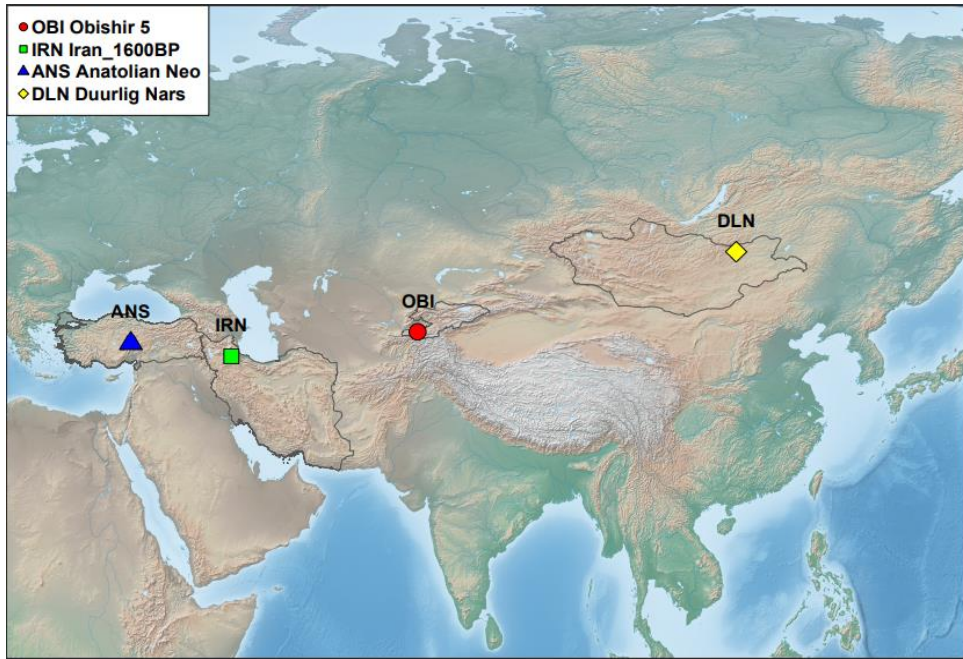


Figure 2. Geographic distribution of ancient sheep populations used in this study. Country border lines are represented on the map (From left to right: Turkey, Iran, Kyrgyzstan, and Mongolia). The map was produced using the Natural Earth public domain map dataset (source: <https://www.naturalearthdata.com/downloads/10m-raster-data/10m-cross-blend-hypso/>). The country border lines were produced using the GADM data (source: <https://gadm.org/data.html>).



Figure 3. A satellite image of Duurlig Nars tomb number 160 site from Google Map. Coordinates are $48^{\circ} 32'49.0''\text{N}$, $111^{\circ} 04'56.0''\text{E}$.

2. Materials and Methods

2.1. Ancient DNA extraction and sequencing

Six sheep individual remains for whole-genome sequencing were excavated at the satellite burials of Duurlig Nars tomb number 160 in Mongolia by a joint Korean–Mongolian expedition between 2017 and 2019. Two sheep samples DLN017.A0101 and DLN018.A0101 were excavated from a satellite burial named 160–E1 (Figure 4). Sample DLN019.A0101 was excavated from 160–E3, and the three samples DLN021.A0101, DLN023.A0101, and DLN024.A0101 came from 160–W1 (Figure 4). Metagenomics DNA of lower tooth, one from each individual, were used to generate a single-stranded paired-end Illumina sequencing library. DNA extraction and library preparation step were performed following previously published protocols (Dabney et al., 2013). All of the laboratory processes were performed in an ancient DNA clean room facility of the Max Plank Institute for the Science of Human History (HPI–SHH), Jena, Germany.

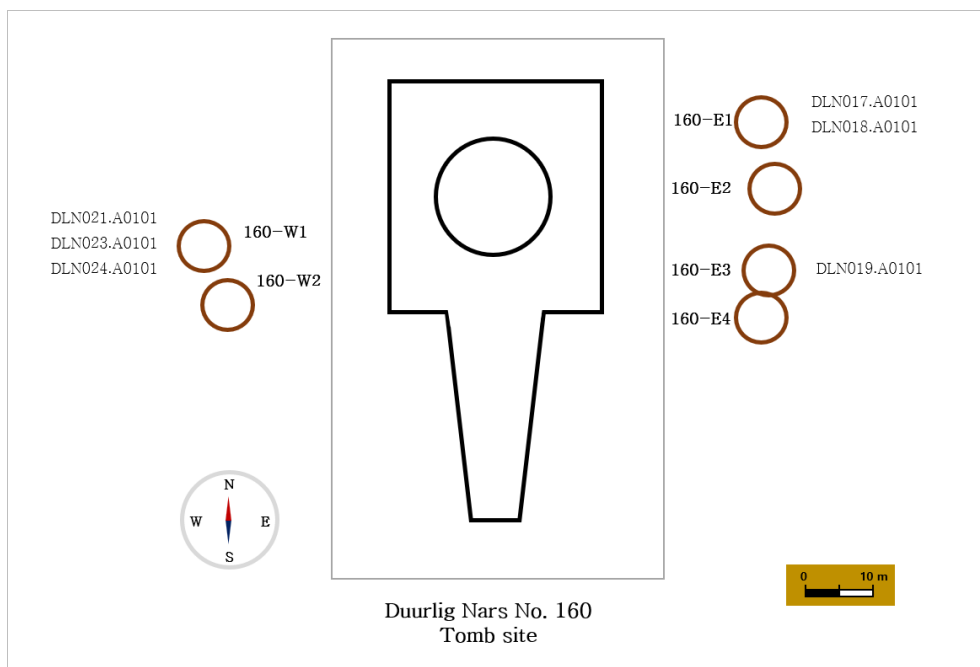


Figure 4. A drawing of Duurlig Nars tomb number 160. Six circular satellite burials were discovered surrounding the main tomb. 160-W1 and 160-W2 were located at the west side of the main tomb. 160-E1, 160-E2, 160-E3, and 160-E4 were located at the east of the main tomb (National Museum of Korea, 2021).

2.2. Ancient DNA data collection and mapping

Previously published whole-genome sequencing data of 11 ancient sheep individuals were downloaded from the European Nucleotide Archive (ENA) under the accession number PRJEB41594, PRJEB42261, and PRJEB36540 (Rossi et al., 2021; Taylor et al., 2021; Yurtman et al., 2021). Whole-genome sequences of previously published ancient sheep and Duurlig Nars sheep were used for downstream analyses.

Illumina adapter sequences were first trimmed from raw reads of ancient DNA using AdapterRemoval v2.3.0 (Schubert et al., 2016). Adapter-trimmed reads of 35 bp or longer were aligned to the *O. aries* reference genome (oviAri4, International Sheep Genome Consortium build.) using BWA-aln/samse v0.1.17 (Li and Durbin, 2009) with “-n 0.01” option. For non-uracil-DNA-glycosylase (non-UDG) enzyme treated ancient DNA libraries, seeding was disabled with “-l 9999” option in BWA aln program. This was to avoid high frequency of postmortem chemical damage in 5’ end bases. Reads that were properly aligned to the reference genome were kept by applying filters “-f 0x0003 -F 0x0004” in samtools v1.9 (Li et al., 2009) and merged into BAM files. PCR duplicates were removed using DeDup v0.12.5 (Peltzer et al., 2016), assuming that both ends of the reads were known. Uniquely mapped reads with Phred-scaled mapping quality score 30 or higher were kept by applying “-q30” filter in samtools v1.9 (Li et al., 2009).

2.3. Ancient DNA damage patterns

mapDamage v2.0.9 (Jónsson et al., 2013) was performed to visualize and tabulate postmortem chemical damage patterns of ancient DNA. mapDamage program calculates read length distribution and misincorporation patterns on the 3' and 5' ends of aligned sequencing read fragments in BAM files.

2.4. Present-day DNA data collection and mapping

Previously published high-coverage whole-genome sequencing data of 411 present-day sheep individuals were downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI – SRA) under the accession number PRJNA24020 and PRJNA645671 (Deng et al., 2020; Li et al., 2020). Additional whole-genome sequencing data of 103 sheep individuals were downloaded from the ENA under the accession number PRJEB5463 and PRJEB23437 (Pan et al., 2018). Sequencing reads were first mapped to the reference genome (oviAri4) using BWA-mem v0.7.17-r1198-dirty (Li and Durbin, 2010). Reads that were properly aligned to the reference genome were kept by applying filters “-f 0x0003 -F 0x0004” in samtools v1.9 and merged into BAM files (Li et al., 2009). PCR duplicates were removed using Picard MarkDuplicate v2.20.0 (downloaded from <https://broadinstitute.github.io/picard/>). Uniquely mapped reads with Phred-scaled mapping quality score 30 or higher were kept by applying “-q30” filter in samtools v1.9 (Li et al., 2009).

2.5. Read coverage and molecular sex determination

Sequencing coverage of each sample was calculated using Qualimap v2.1 (García-Alcalde et al., 2012) from the analysis-ready BAM files. The number of mapped reads was normalized by the length of the chromosome. Since the sheep reference genome used in this paper does not include chromosome Y sequences, the individual molecular sex was estimated from the chromosome X to autosome sequencing coverage ratio. Females XX are expected to have a ratio of around 1.0 and males XY around 0.5. Number of reads aligned to the reference genome of each BAM file was also calculated using samtools v1.9 with an option “-c” (Li et al., 2009).

2.6. Variant calling and filtering

For the variant calling, the candidate variants were called on individual-level from each BAM file using GATK (Genome Analysis ToolKit) HaplotypeCaller v3.8.1.0 (download from <https://github.com/broadinstitute/gatk/>) with “-ERC GVCF” option and per chromosome GVCF files were created. GVCF files of high-coverage 415 individuals from studies PRJNA24020, PRJNA645671, and PRJEB5463 were merged using GATK-CombineGVCFs v3.8.1.0. The per chromosome raw VCF files were generated using GATK-GenotypeGVCFs v3.8.1.0 for downstream analyses.

Variant filtering was performed on the raw VCF files using

GATK–SelectVariants v3.8.1.0. INDELs were removed and SNPs with the QUAL 30 or higher were kept. Multi–allelic SNPs were removed using AWK and bi–allelic SNPs were extracted with “if (REF == "A" || REF == "C" || REF == "G" || REF == "T") && (ALT == "A" || ALT == "C" || ALT == "G" || ALT == "T")” option. To avoid possible false–positive calls, GATK–VariantFiltration v3.8.1.0 was further performed as recommended by the GATK best practice (McKenna et al., 2010): (1) SNPs with mean depth $< 1/5x$ and $> 5x$ (x , overall mean sequencing depth across all SNP sites); (2) quality by depth, QD < 2 ; (3) strand odds ratio, SOR > 3 ; (4) Fisher strand, FS > 60 ; (5) mapping quality, MQ ≤ 40 ; (6) mapping quality rank sum test, MQRankSum < -12.5 ; (7) read position rank sum test, ReadPosRankSum < -8 ; (8) and excess heterozygosity test, ExcessHet < 10 were filtered.

2.7. Genotype refinement

Imputation was done on the filtered high–quality bi–allelic SNPs to rescue as many missing genotype calls as possible using BEAGLE v4.1 (Browning and Browning, 2016) by comparison with the present–day sheep reference genome panel. Using a sliding window approach, imputation was performed on 170K SNPs in each window of 5Mb window length in the sheep reference genome, 12K SNPs overlapping between adjacent windows, the number of iterations 10, and with memory–saving option (option: window=170000;

overlap=12000; niterations=10; lowmem=true).

Genotype calling was performed on VCF files after BEAGLE imputation. The individual genotypes of GP \geq 0.99 were called in EIGENSTRAT format. SNPs were filtered with a missing genotype rate less than 2% and filtered rare alleles with minor allele count (mac) less than 25. To avoid ancient DNA chemical contamination bias, Transversion variants (Tv) were kept and Transition variants (Ts) were removed, leaving 7,497,624 SNPs. This 7497K SNPs were used to generate the worldwide sheep diversity panel (i.e., the 7497K SNP panel) for downstream population genetic analysis. Number of SNPs per chromosome in each step of generating the panel are listed in Table 1.

Table 1. Number of SNPs per chromosome in each step of the worldwide sheep diversity panel generating process. 7,497,624 SNPs spanning 26 autosomes were used to generate the genome-wide dataset.

Chr	Length in oviAri4 (bases)	# of SNPs					
		SelectVariants (SNP-only, QUAL >= 30)	AWK bi-allelic SNPs	VariantFiltration (hard filtering)	BEAGLE (GP >= 0.99, mac5)	mac25 missingness <= 2%	Tv-only 7497K
chr1	275,406,953	12,155,205	9,876,167	8,667,878	4,667,003	3,032,827	853,331
chr2	248,966,461	10,814,094	8,805,186	7,748,549	4,136,313	2,658,685	743,620
chr3	223,996,068	9,654,643	7,837,424	6,894,249	3,686,254	2,350,351	645,239
chr4	119,216,639	5,306,770	4,304,415	3,793,850	2,036,623	1,312,853	369,368
chr5	107,836,144	4,641,011	3,767,431	3,333,311	1,782,586	1,151,503	321,525
chr6	116,888,256	5,421,442	4,384,564	3,851,314	2,102,657	1,379,095	395,542
chr7	100,009,711	4,275,869	3,481,579	3,058,441	1,656,102	1,067,958	292,732
chr8	90,615,088	3,967,228	3,226,896	2,859,389	1,554,550	1,009,863	281,030
chr9	94,583,238	4,294,899	3,482,590	3,082,617	1,662,746	1,088,358	304,955
chr10	86,377,204	4,155,732	3,338,617	2,903,091	1,547,153	992,031	286,077
chr11	62,170,480	2,605,464	2,114,049	1,887,115	990,116	619,496	167,802
chr12	79,028,859	3,468,217	2,822,466	2,500,979	1,337,071	866,567	237,981
chr13	82,951,069	3,480,138	2,841,937	2,511,285	1,337,635	846,667	225,282
chr14	62,568,341	2,686,204	2,174,487	1,894,378	997,454	630,111	168,628
chr15	80,783,214	3,663,802	2,968,015	2,604,176	1,413,521	919,314	258,427
chr16	71,693,149	3,294,023	2,681,478	2,385,887	1,308,442	856,789	237,642
chr17	72,251,135	3,248,481	2,622,628	2,321,568	1,262,756	826,183	231,063
chr18	68,494,538	3,054,843	2,479,387	2,192,034	1,171,143	757,996	206,349
chr19	60,445,663	2,589,470	2,112,419	1,895,733	1,012,986	649,706	176,017
chr20	51,049,468	2,348,228	1,899,026	1,665,515	902,418	576,477	158,577
chr21	49,987,992	2,333,581	1,899,451	1,668,376	897,392	582,072	158,638
chr22	50,780,147	2,308,713	1,880,410	1,671,317	902,339	589,649	162,011
chr23	62,282,865	2,862,478	2,329,708	2,079,303	1,122,981	736,861	202,967
chr24	41,976,827	1,845,301	1,494,999	1,342,996	708,841	452,752	115,754
chr25	45,223,504	2,112,430	1,725,829	1,528,251	846,380	557,712	153,474
chr26	44,047,080	2,049,596	1,667,130	1,495,297	809,293	530,366	143,593
Total	2,449,630,093	108,637,862	88,218,288	77,836,899	41,852,755	27,042,242	7,497,624

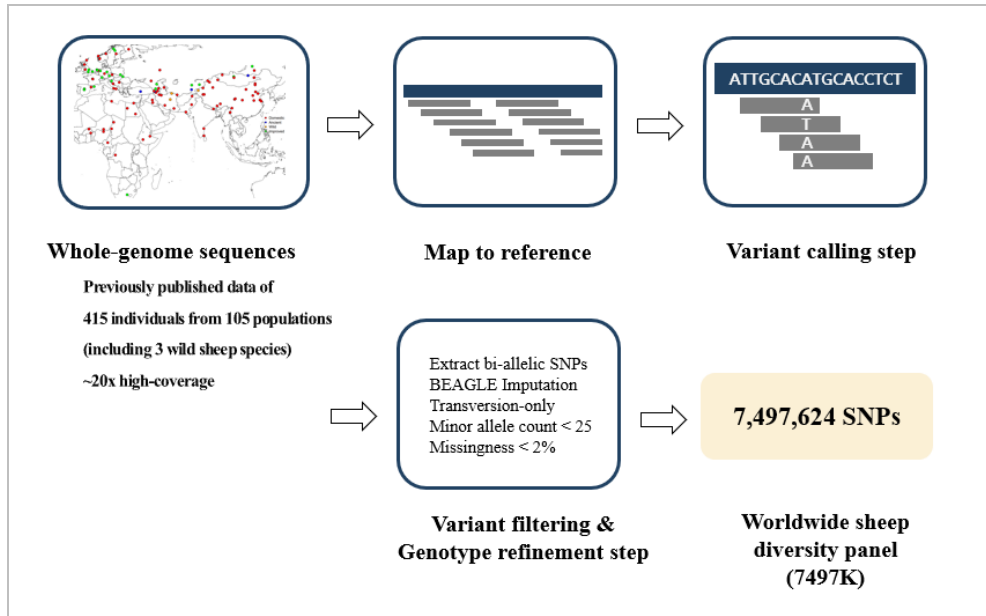


Figure 5. Brief pipeline layout of the worldwide sheep diversity panel generating process. Previously published high-coverage 415 individuals from 105 populations, including 3 wild sheep species, were used for the SNP identification.

2.8. Genotyping of low-coverage individuals

For ancient individuals and low-coverage present-day individuals, pseudo-haploid genotype call was performed using pileupCaller v1.4.0.5 (downloaded from <https://github.com/stschiff/sequenceTools>) with “—randomHaploid” option based on the 7497K panel. This was to reduce the reference bias by random allele sampling and genotype was called in EIGENSTRAT format.

2.9. Genetic kinship analysis

Pairwise mismatch rate (PMR) was calculated to detect close relative pairs in the dataset using pseudo-haploid genotype data (Kennett et al., 2017). From each pair of random haploid calls, kinship coefficients were called and one individual from each of the first-degree relatives and duplicates was removed for the downstream group-based analysis. For low-coverage individuals with multiple sequencing runs, duplicates were detected and the genotype data were merged based on PMR results.

2.10. Data compilation and the master dataset

Finally, genome-wide genotype data of 415 high-coverage present-day individuals, 99 low-coverage present-day individuals, and 17 ancient individuals typed on the 7497K SNP panel were compiled into a single master dataset. Individuals with low coverage were marked by adding ‘SG’ at the end of the sample name and

population name to distinguish them from the high-coverage individuals.

2.11. Principal component analysis (PCA)

Principal component analysis (PCA) was performed with the present-day sheep individuals used in generating the 7497K SNP panel with “lsqproject: YES; numchrom: 26” option, using smartpca v16000 (Patterson et al., 2006). Ancient individuals were projected on the PCA plots to understand the global genetic affinity.

2.12. Outgroup- f_3 analysis

Outgroup- f_3 statistics was calculated to quantify the genetic affinity between two populations from shared genetic drift. Shared genetic drift was calculated from the correlations of allele frequencies between Duurlig Nars and other sheep populations using wild sheep *Urrial* (*O. vignei*) as an outgroup. The outgroup- f_3 analysis was performed using qp3Pop v650 program in the AdmixTools v7.0 (Patterson et al., 2012) with “numchrom: 26” option.

2.13. TreeMix analysis

A PLINK format file of selected populations was generated and the missingness and allele frequencies were calculated using PLINK software v1.9 (Purcell et al., 2007). The output PLINK file was

converted into a TreeMix input file using a python script (downloaded from <https://github.com/ekirving/ctvt/>). TreeMix was performed using TreeMix software v1.13 (Pickrell and Pritchard, 2012) with SNPs that were present in all populations (grep -wv "0,0"), allowing up to 3 migration events with “-m 0-3” option. The inferred trees and residuals were visualized using author-provided R script (downloaded from <https://github.com/joepickrell/pophistory-tutorial/>).

2.14. F_4 -statistics analysis

F_4 -statistics analysis was performed using the qpDstat v970 program in the AdmixTools v7.0 (Patterson et al., 2012) with “f4mode: YES; numchrom: 26” option to test the phylogenetic relationships in the TreeMix tree using wild sheep Urial (*O. vignei*) as an outgroup.

2.15. Mitochondrial DNA haplogroup analysis

Mitochondrial consensus sequences were first retrieved from the BAM files aligned to the sheep reference genome (oviAri4) using ANGSD v0.929 program (Korneliussen et al., 2014) with “-r chrM -doFasta 2 -doCounts 1 -setMinDepth 3 -minQ 20 -minMapQ 30” option. Whole mitochondrial genome sequences of 10 present-day domestic sheep of mitochondrial DNA haplogroup A, B, C, D, and E (accession name: HM236174-HM236183) available

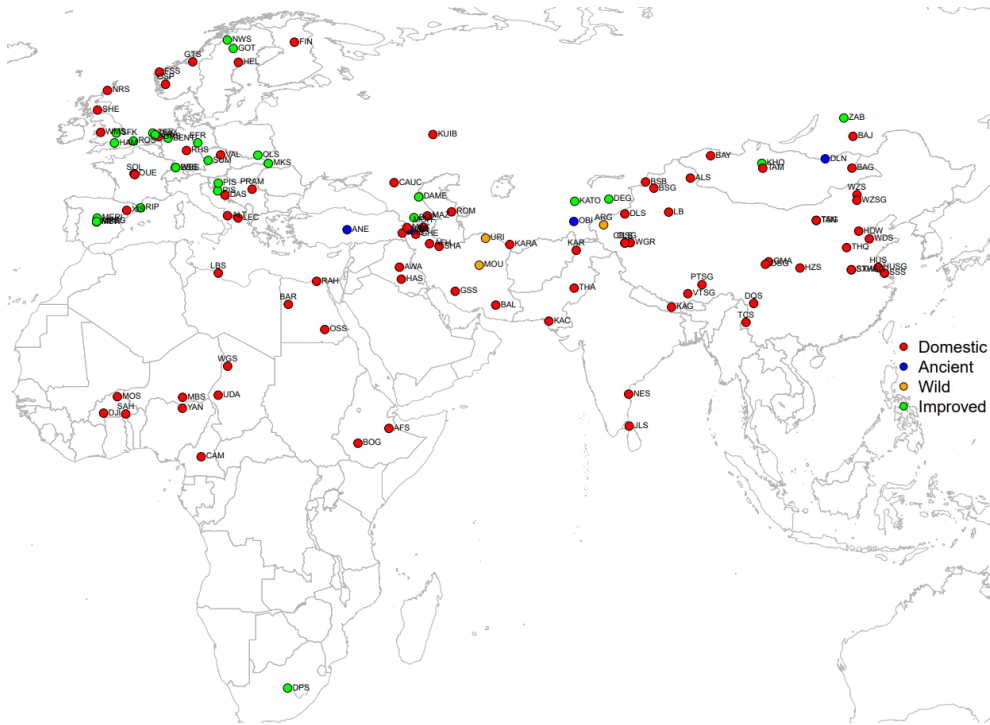
at NCBI GeneBank were downloaded (Meadows et al., 2011). From these sequences, 1,180 bp mitochondrial DNA control region (CR) sequences corresponded to positions 5437–16616 on the reference AF010406 (Hiendleder et al., 1998) were extracted and multiple sequence alignments were performed with MUSCLE v5.1 (Edgar, 2004).

First, the evolutionary history was inferred using the Maximum Parsimony (MP) method. Bootstrap test of 1,000 replicates was used in tree construction (Felsenstein, 1985). The MP tree was obtained using the Subtree–Pruning–Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). Evolutionary analyses were conducted in MEGA–CC 11 (Kumar et al., 2012; Tamura et al., 2021) and visualized with the Interactive Tree Of Life (iTOL) v5, an online tool (source: <https://itol.embl.de/>).

Lastly, the evolutionary history was inferred using the Maximum Likelihood (ML) method. Bootstrap test of 100 replicates was used in tree construction (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor–Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura–Nei model, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA 11 (Tamura et al., 2021) and visualized with the Interactive Tree Of Life (iTOL) v5, an online tool (source: <https://itol.embl.de/>).

Table 2. Software and Algorithms used in this study.

Software / Algorithms	References	Link
AdapterRemoval v2.3.0	(Schubert et al., 2016)	https://github.com/MikkelSchubert/adapterremoval/
AdmixTools v7.0	(Patterson et al., 2012)	https://github.com/DReichLab/AdmixTools/
ANGSD v0.929	(Korneliussen et al., 2014)	http://www.popgen.dk/angsd/index.php/ANGSD/
BCFtools v1.9	N/A	http://www.htslib.org/
BEAGLE v4.1	(Browning and Browning, 2016)	https://faculty.washington.edu/browning/beagle/
BWA-mem v0.7.17	(Li and Durbin, 2010)	http://bio-bwa.sourceforge.net/
BWA-aln v0.1.17	(Li and Durbin, 2009)	http://bio-bwa.sourceforge.net/
dedup v0.12.5	(Peltzer et al., 2016)	https://github.com/apeltzer/DeDup/
Eigensoft v7.2.1	(Patterson et al., 2006)	https://github.com/DReichLab/EIG/
Genome Analysis Toolkit (GATK) v3.8.1.0	(McKenna et al., 2010)	https://github.com/broadinstitute/gatk/
Integrative Genomics Viewer (IGV) v2.13.0	(Robinson et al., 2011)	https://software.broadinstitute.org/software/igv/
Interactive Tree Of Life (iTOL) v5	(Letunic and Bork, 2021)	https://itol.embl.de/
mapDamage v2.0.9	(Jónsson et al., 2013)	https://github.com/ginolhac/mapDamage/
MEGA-CC v11	(Kumar et al., 2012; Tamura et al., 2021)	https://www.megasoftware.net/
MUSCLE v5.1	(Edgar, 2004)	https://github.com/rcedgar/muscle/
Picard v2.20.0	N/A	https://broadinstitute.github.io/picard/
pileupCaller v1.4.0.5	N/A	https://github.com/stschiff/sequenceTools/
PLINK v1.9	(Purcell et al., 2007)	https://www.cog-genomics.org/plink/
Python v3.9.7	N/A	https://www.python.org/
Qualimap v2.1	(García-Alcalde et al., 2012)	http://qualimap.conesalab.org/
R v3.6.0	N/A	https://www.r-project.org/
samtools v1.9	(Li et al., 2009)	http://samtools.sourceforge.net/
TreeMix v1.13	(Pickrell and Pritchard, 2012)	https://github.com/carolinhms/TreeMix/
Python script plink2treemix.py	N/A	https://github.com/ekirving/cvtv/
R script TreeMix plotting_funcs.R	N/A	https://github.com/joepickrell/pophistory-tutorial/



AFS, Afar; AFH, Afshari; ALT, Altamurana; ALS, Altay; ANE, Anatolian_Neo; ARG, Argali; MOU, Asiatic_mouflon; AMSG, Australian_Merino.SG; AWA, Awassi; BAJ, Baidarak; BAL, Baluchi; BAG, Barga; BAR, Barki; BSB, Bashibai; BAY, Bayad; BSG, Bayinbuluke.SG; BENT, Bentheimer; BOG, Bonga; BOZ, Bozakh; BER, Brazilian_Bergamasca; SOM, Brazilian_Somali; BBS, Brown_Mountain; CAM, Cameroon; MKS, Carpathian_Mountain; CAUC, Caucasian; CLS, Cele_Black; CBSG, Cele_Black.SG; MFW, Chinese_Merino_finewool; MSF, Chinese_Merino_superfinewool; DAME, Dagestan_Mountain_Merino; DAS, Dalmatian; DEG, Degeres_Mutton-Wool; DQS, Diqing; DJI, Djallonke; DPS, Dorper; DRS, Drente_Heathen; DLS, Duolang; DLN, Duurlig_Nars; EFR, East_Friesian_Dairy; FSS, Feral; FIN, Finn; GHE, Ghezel; GOT, Gotland; GTS, Grey_Tronder; GSS, GreyShiraz; HAS, Hamdani; HAM, Hampshire; HZS, Hanzhong; HEL, Helsing; HUS, Hu; HUSG, Hu.SG; IRN, Iran_1600BP; JLS, Jaffna_Local; KAC, Kachi; KAG, Kage; KARA, Karakul; KAR, Karkul; KATO, Kazakh_Finewool; KHO, Khotont; KUIB, Kuibyshev; HDW, Large-tailed_Han; LEC, Leccese; LBS, Libyan_Barbar; LB, Lop; MAK, Makui; MAZ, Mazekh; MBS, Mbororo; MERI, Merino_Sheep; MOH, Moghani; MOS, Mossi; NES, Nellore; NRS, North_Ronaldsay; NWS, Norwegian_White; OBI, Obishir5; OSP, Old_Spael; OLS, Olkusa; OSS, Ossimi; OUE, Ouessant; OSG, Oula.SG; PIS, Pag_Island; PIS, PagIsland; PTSG, Prairie_Tibetan; PRAM, Pramenka; RAH, Rahmani; RHS, Rhoen; RIP, Ripollesa; ROM, Romanov; ROS, Romney; SAH, Sahelian; SIS, Santa_Ines; SHA, Shal; SHE, Shetland; SSS, Sishui_Fur; SXW, Small-tailed_Han; STHSG, Small-tailed_Han.SG; SOL, Solognote; SFK, Suffolk; SUM, Sumavska; THQ, Taihang_Fur; TAM, Tamir; TAN, Tan; TSG, Tan.SG; TCS, Tengchong; TEX, Texel; THA, Thalli; GMA, Tibetan; ZAB, Transbaikai_Finewool; UDA, Uda; WZS, Ujimqin; URI, Urial; VAL, Valachian; VTSG, Valley_Tibetan.SG; WDS, Wadi; WGR, Waggir; WMS, Welsh_Mountain; WGS, West_African_Dwarf; WBS, White_Mountain; WZSG, Wuzhumuqin.SG; XIS, Xisqueta; YAN, Yankasa; ZWA, Zwartbles.

Figure 6. Geographic distribution of worldwide sheep populations used in this study. Domestic, Ancient, Wild, and Improved sheep populations are represented by red, blue, orange, and green dots on the world map, respectively.

3. Results

3.1. Ancient genome-wide data production

Whole-genome data was generated from six ancient sheep remains excavated from the Duurlig Nars tomb number 160, Mongolia. Two remains were excavated from a satellite burial named 160-E1, one was from 160-E3, and the rest came from 160-W1 (Figure 4). Based on archeological contexts and radiocarbon dating of wooden coffin, these sheep are from Xiongnu-period which is about 2,000 BP (National Museum of Korea, 2021).

For each of the six individuals, a double-stranded Illumina sequencing library was built from metagenomics DNA extracted from a tooth sample. Sequenced genome data were mapped to the sheep reference genome (oviAri4), which consist of 26 autosomes, chrX, and chrM sequences. Then sequence coverage and postmortem damage pattern of mapped reads were measured to understand molecular properties of Duurlig Nars individuals. Six Duurlig Nars individuals had an average of 5.77% endogenous DNA preservation across the fragments, ranging from 1.77% to 11.51% and an average of 0.095x autosomal depth coverage spanning 0.025–0.169x (Table 3). These six libraries exhibited about 20% postmortem damage (PMD) induced C to T transitions at 5' ends of reads (Figure 7; Table 3). Molecular sex was inferred using the

chrX to autosomal read depth ratio. Half of the individuals were identified as female and the remaining half as male (Table 3).

3.2. Worldwide sheep diversity panel and genotyping

To understand the genetic diversity of sheep populations and their genetic profile, SNP identification process were performed to generate SNP list for further analysis. The publicly available high-coverage whole-genome sequencing data of a total of 415 individuals from 105 sheep populations were used in this study to produce the worldwide sheep diversity panel. This includes three wild sheep species: Urial, Asiatic Mouflon, and Argali. Sequence reads were aligned to the sheep reference genome and this 415 sheep data had read coverage of an average 19.11x, ranging from 10.00x to 26.92x. With these high-coverage genomes, variant discovery process was performed. Amongst the variants discovered, bi-allelic SNPs were selected and genotype refinement was done to lower the missing value. To avoid possible postmortem damage in ancient DNA, only transversion SNPs were kept for the downstream analysis. Finally, the worldwide sheep diversity panel of 7,497,624 SNPs (7497K SNP set) were produced.

Additional low-coverage genome data of 99 modern sheep individuals and previously published 11 ancient sheep individuals were download. A total of 531 individuals were genotyped on the 7497K SNPs and compiled in EIGENSTRAT file. With this master

dataset, population analyses were conducted. The average genotyped rate on the 7497K panel of high-coverage 415 sheep was 99.93%, low-coverage 99 individuals was 97.64%, and Duurlig Nars individuals was 6.34%.

Using genotype data in the master dataset, Pairwise Mismatch Rate (PMR) were calculated to detect close relativeness in sample set. One Individual with relativeness of first-degree or duplicate were marked and excluded in the further analysis. No genetic relationships were detected in ancient sheep.

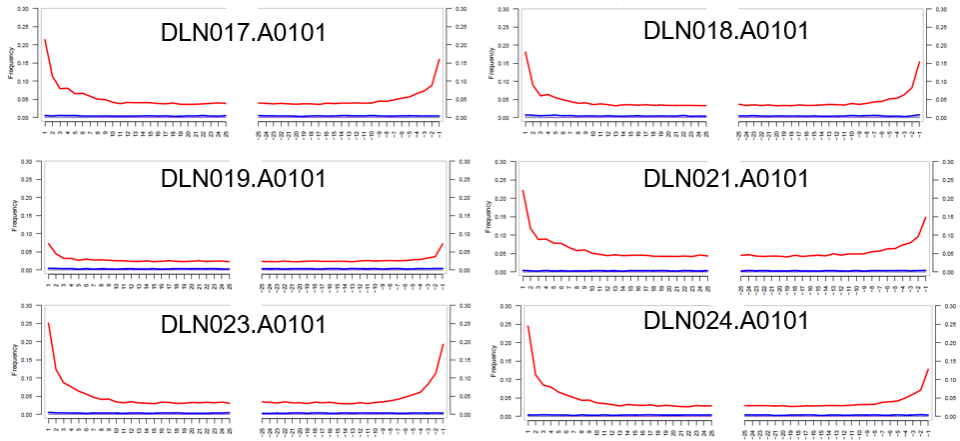


Figure 7. The mapDamage misincorporation plot of sequencing reads of six Duurlig Nars individuals mapped to the reference genome. The left plots and the right plot represent substitutions from the 5' and 3' ends, respectively. Red line represents C to T substitutions and blue line represents G to A substitutions.

Table 3. Summary statistics of six Duurlig Nars individuals.

No.	Sample ID	Tomb	Specimen	Sex	MT Hap	Covered in 7497K	
						# of SNPs	% of SNPs
1	DLN017.A0101	160-E1	Tooth (molar)	F	B	143,323	1.91
2	DLN018.A0101	160-E1	Tooth (molar)	F	A	147,800	1.97
3	DLN019.A0101	160-E3	Tooth (premolar)	M	A	799,841	10.67
4	DLN021.A0101	160-W1	Tooth (molar)	F	B	806,823	10.76
5	DLN023.A0101	160-W1	Tooth	M	B	275,632	3.68
6	DLN024.A0101	160-W1	Tooth	M	A	682,429	9.10

No.	Sample ID	# of reads				Endogenous DNA %
		sequenced	mapped	uniquely mapped	uniquely mapped with > q30	
1	DLN017.A0101	323,426,447	2,362,245	2,126,590	1,514,444	1.77
2	DLN018.A0101	314,837,499	2,016,698	1,828,153	1,431,708	1.48
3	DLN019.A0101	250,492,403	13,439,251	11,430,807	8,045,112	11.51
4	DLN021.A0101	317,228,525	14,351,132	12,568,016	9,021,638	9.43
5	DLN023.A0101	316,703,909	4,504,902	4,014,793	2,882,250	3.00
6	DLN024.A0101	319,637,449	10,985,529	9,772,366	7,189,886	7.44

No.	Sample ID	Coverage			Damage Pattern			
		Auto	X	MT	First base 5'	Second base 5'	First base 3'	Second base 3'
1	DLN017.A0101	0.0267	0.0273	1.465	0.2148	0.1139	0.0045	0.0045
2	DLN018.A0101	0.0248	0.0244	2.808	0.1819	0.0903	0.0081	0.0050
3	DLN019.A0101	0.1665	0.0857	5.431	0.0732	0.0448	0.0046	0.0044
4	DLN021.A0101	0.1687	0.1695	8.870	0.2222	0.1185	0.0047	0.0040
5	DLN023.A0101	0.0515	0.0272	4.074	0.2515	0.1239	0.0036	0.0043
6	DLN024.A0101	0.1331	0.0705	14.958	0.2459	0.1129	0.0037	0.0049

3.3. Principal Component Analysis (PCA)

PCA analysis was conducted to study the genetic structure of sheep populations. Present-day individuals used in SNP identification were used in the calculation. Firstly, the 7497K SNP was evaluated whether it represents the diverse genetic profile of present-day sheep and its resolution with PCA analysis. Clusters of individuals based on their similarity were shown in the PCA plots. With this clustering pattern, sheep were classified into eight categories according to their taxa and geographical locations: Wild, European, African, South Asian, West Asian, Central Asian, East Asian, and Improved Breeds. Outlier individuals were identified regarding overall population differentiation. With this global structure, ancient individuals were projected on the PCs to understand their global genetic affinity.

Secondly, PCA analysis were ran on all modern sheep and ancient individuals were projected on this PCs (n=17). PC1 vs PC2 plot was used to infer whether ancient sheep individuals are genetically domestic or wild. In this all present-day PCA, PC1 differentiates the domestic breeds (*Ovis aries*) from wild sheep species (Figure 8). All the ancient individuals, including Duurlig Nars, are located on top of the present-day domestic populations, indicating that the ancient individuals used in this study are genetically domestic (Figure 8).

Lastly, PCA analysis on present-day domestic individuals

was performed, excluding wild individuals in the calculation. Ancient sheep individuals were also projected on this PCs (n=17). Interestingly, all the ancient populations fell on top of the present-day populations which are geographically close to the excavation sites (Figure 9). For example, Duurlig Nars sheep (DLN) discovered in Mongolia are located on top of present-day East Asian populations (Figure 9). Anatolian Neolithic sheep (ANS) discovered in Turkey are located on top of present-day European populations (Figure 9). Obishir5 population (OBI) discovered in Kyrgyzstan are located on top of present-day Central Asian populations (Figure 9). 1,600-year-old sheep discovered in Iran are located on top of present-day West Asian populations (Figure 9).

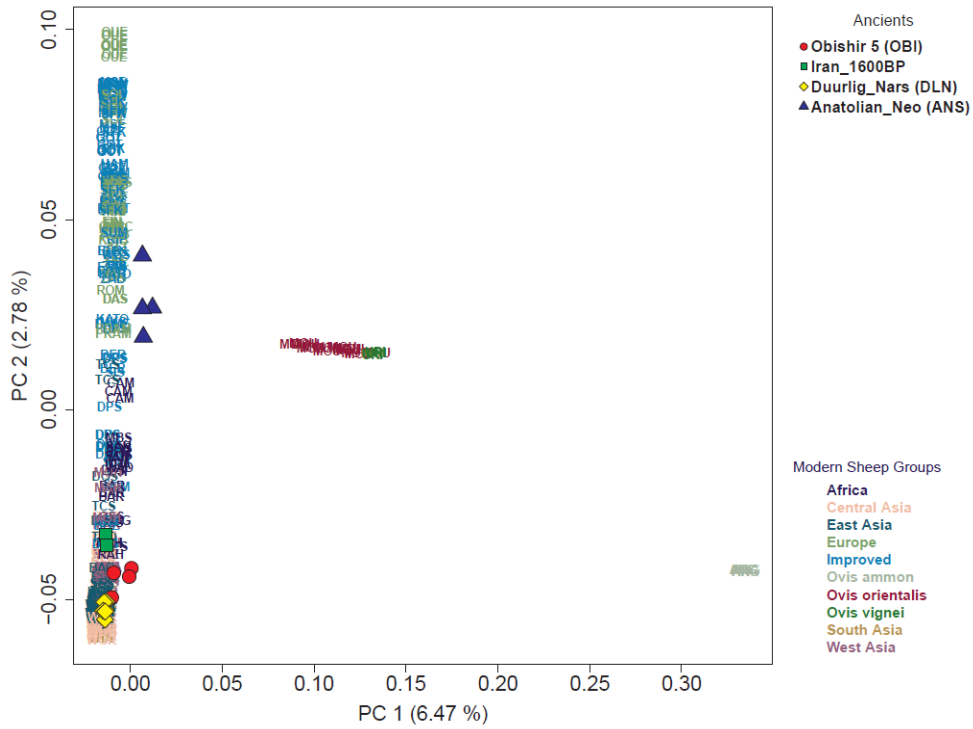


Figure 8. PCA plot of top two components (PC1 & PC2) calculated with high-coverage present-day domestic and wild individuals. Ancient individuals are projected on top of present-day sheep individuals. Present-day individuals used in the analysis are color coded in the PCs.

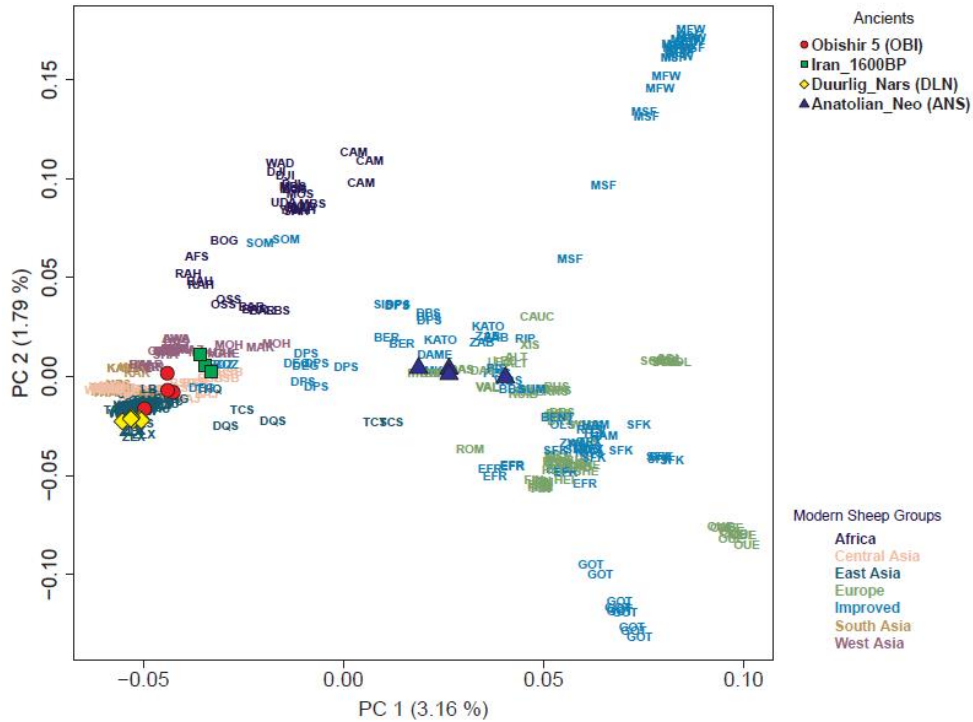


Figure 9. PCA plot of top two components (PC1 & PC2) calculated with high-coverage present-day domestic individuals. Ancient individuals are projected on top of present-day sheep individuals. Present-day individuals used in the analysis are color coded in the PCs.

3.4. Outgroup- f_3 statistics analysis

outgroup- f_3 statistics analysis was calculated to measure the shared genetic drift between Duurlig Nars and other sheep populations, using wild sheep Urial as an outgroup. The outgroup- f_3 result of the form $f_3(\text{Urial}; \text{Duurlig Nars}, \text{worldwide})$ showed that among present-day sheep populations, Duurlig Nars Xiongnu sheep share the highest genetic affinity with present-day East Asian sheep as indicated by the distribution on the PCA components (Figure 9; Figure 10). Above all, Duurlig Nars shared the most genetic affinity with Obishir5, an ancient Central Asian sheep population. Interestingly, among the East Asian populations, Duurlig Nars sheep had the highest genetic affinity with the present-day Tibetan breeds in the Tibetan Plateau region (Tibet, Kham, and Ando province of China) than the Northern Chinese sheep populations which are geographically closer (Figure 10; Figure 11. Bottom).

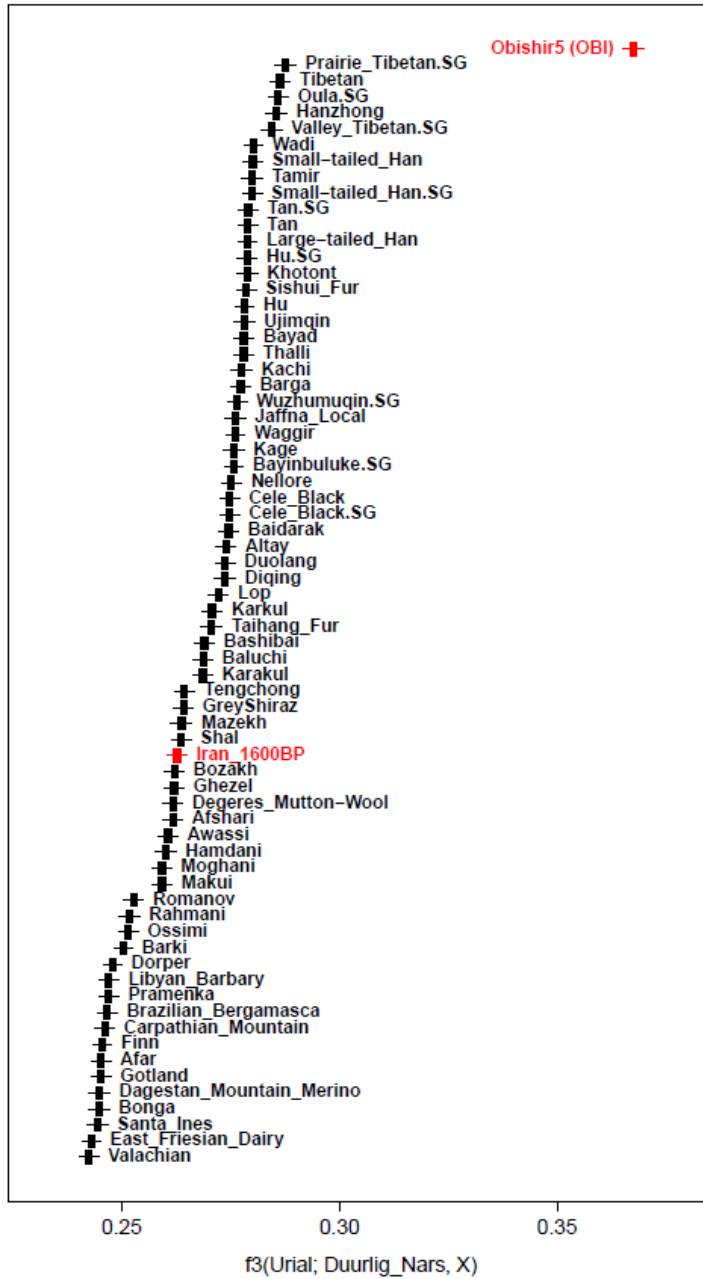


Figure 10. Outgroup- f_3 analysis plot of the form $f_3(\text{Urial}; \text{Duurlig_Nars}, \text{worldwide})$. This analysis quantifies the shared genetic drift between Duurlig Nars sheep and worldwide sheep, including ancient sheep using Urial as an outgroup. Ancient populations are highlighted in red.

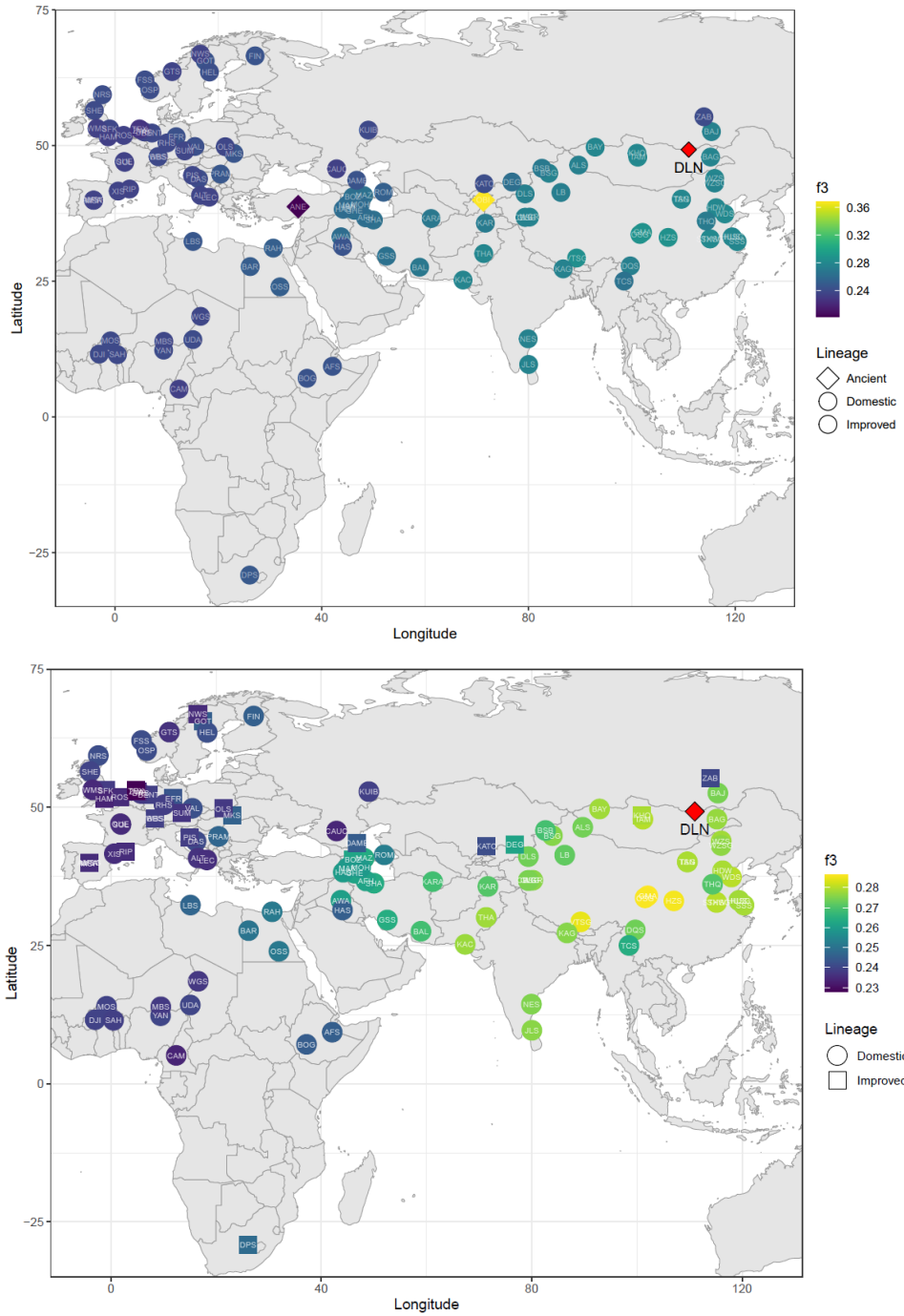


Figure 11. Outgroup- f_3 scatter plot of the form $f_3(\text{Urial}; \text{Duurlig Nars, worldwide})$. (Top) outgroup- f_3 scatter plot of worldwide populations, including ancient populations. (Bottom) outgroup- f_3 scatter plot of all present-day populations.

3.5. TreeMix analysis

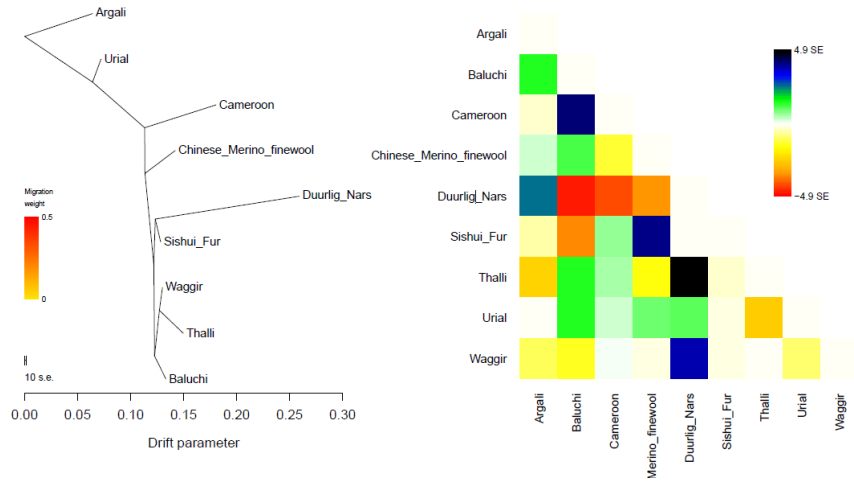
TreeMix analyses were conducted to confirm the population structure and possible admixture events based on the covariance matrix of allele frequencies. Firstly, TreeMix tree was constructed allowing 0–3 migration using 9 representative populations (Duurlig Nars, ancient; Urial, wild; Cameroon, African; Chinese_Merino_finewool, European; Baluchi, West Asian; Thalli, South Asian; Sishui_Fur, East Asian; Waggir, Central Asian; and Argali as an outgroup) (Figure 12). 2,228,031 SNPs were used in this analysis. With these TreeMix results, phylogenetic relationship of present-day sheep and Duurlig Nars and possible migration events of Duurlig Nars and its admixture proportion α were inferred. The TreeMix tree showed a phylogenetic relationship of sheep which Eurasian sheep forming a clade (Figure 12). Duurlig Nars sheep fell into a sister group with the present-day East Asian sheep, Sishui_Fur. The internal branch for Eurasian populations were shorter compared to internal branch of European and African populations (Figure 12). A migration event between Duurlig Nars and the branch node of South Asian and Central Asian populations was detected at the migration edge of $m=1$ (Figure 12). A putative wild introgression of argali sheep to Duurlig Nars was detected at the migration edge of $m=3$ (Figure 12).

Secondly, TreeMix tree was constructed allowing 0–3 migration using 9 representative populations (Obishir5, ancient; Urial, wild; Cameroon, African; Chinese_Merino_finewool, European;

Baluchi, West Asian; Thalli, South Asian; Sishui_Fur, East Asian; Waggir, Central Asian; and Argali as an outgroup) (Figure 13). 294,805 SNPs were used in this analysis. With these TreeMix results, phylogenetic relationship of present-day sheep and Obishir5 and possible migration events of Obishir5 and its admixture proportion α were inferred. Interestingly, a putative wild introgression to Obishir5 was detected from the first migration edge of $m=1$ (Figure 13). Obishir5 fell into a sister group with the present-day East Asian sheep, Sishui_Fur (Figure 13).

Lastly, TreeMix tree was constructed allowing 0–3 migration using 10 representative populations (Obishir5 and Duurlig Nars, ancient; Urial, wild; Cameroon, African; Chinese_Merino_finewool, European; Baluchi, West Asian; Thalli, South Asian; Sishui_Fur, East Asian; Waggir, Central Asian; and Argali as an outgroup) (Figure 14). 126,847 SNPs were used in this analysis. With these TreeMix results, phylogenetic relationship of present-day sheep and ancient sheep and possible migration events of ancient sheep and its admixture proportion α were inferred. Interestingly, Obishir5 and Duurlig Nars sheep formed a sister group on all trees (Figure 14).

(A)



(B)

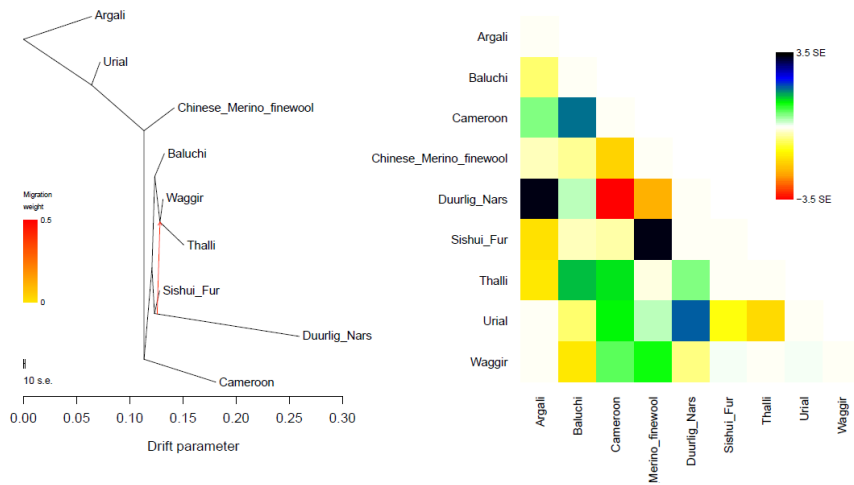
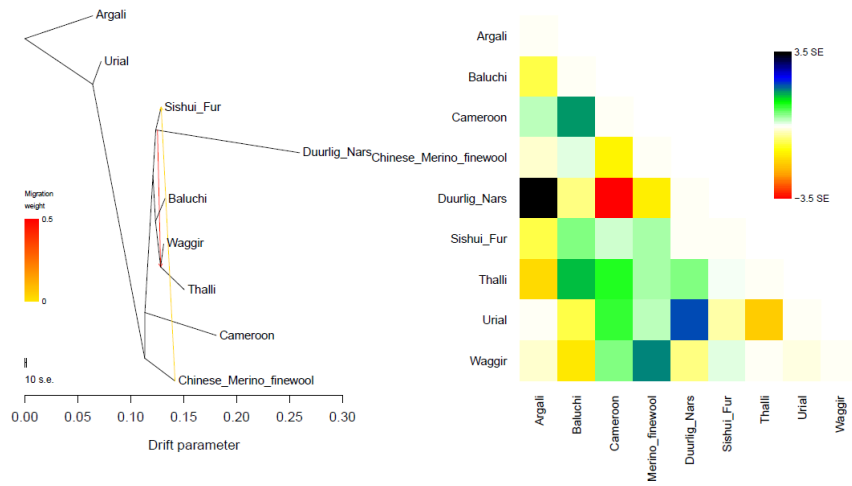


Figure 12. TreeMix tree constructed with nine representative populations of Urial, Cameroon, Chinese_Merino_finewool, Baluchi, Thalli, Sishui_Fur, Waggir, Argali (outgroup), and Duurlig Nars. Tree (A)–(D) represents a tree with migration edge from $m=0$ to $m=3$, respectively.

(C)



(D)

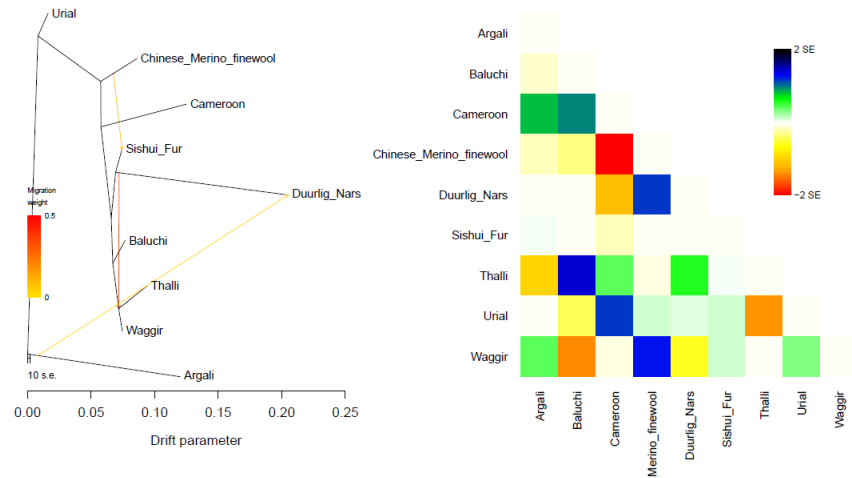
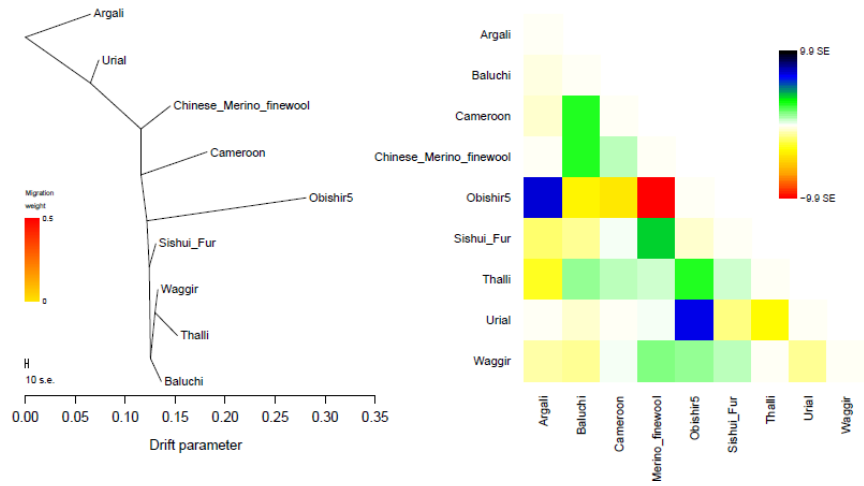


Figure 12. (continued)

(A)



(B)

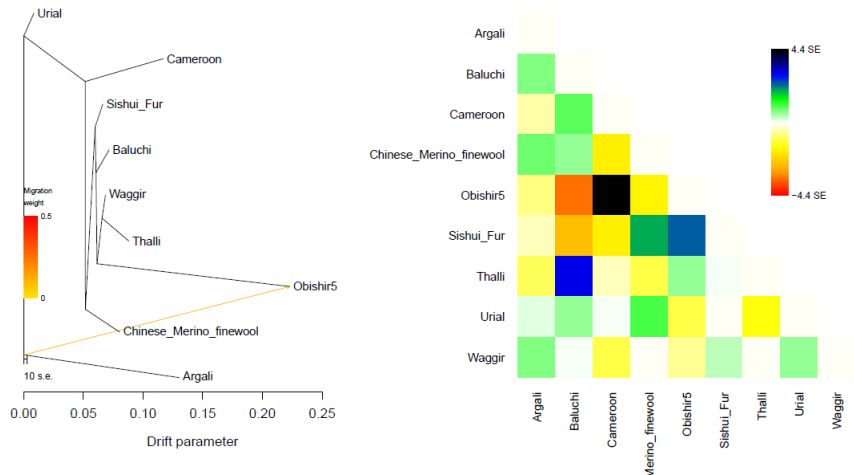
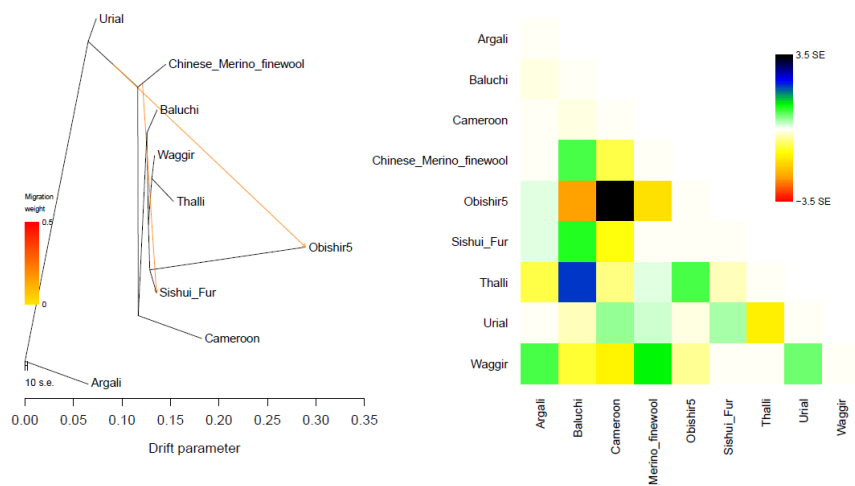


Figure 13. TreeMix tree constructed with nine representative populations of Urial, Cameroon, Chinese_Merino_finewool, Baluchi, Thalli, Sishui_Fur, Waggir, Argali (outgroup), and Obishir5. Tree (A) – (D) represents a tree with migration edge from $m=0$ to $m=3$, respectively.

(C)



(D)

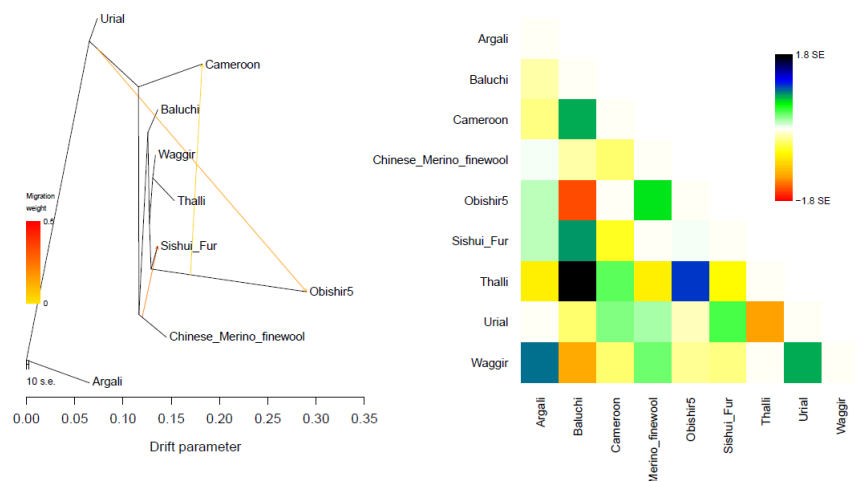
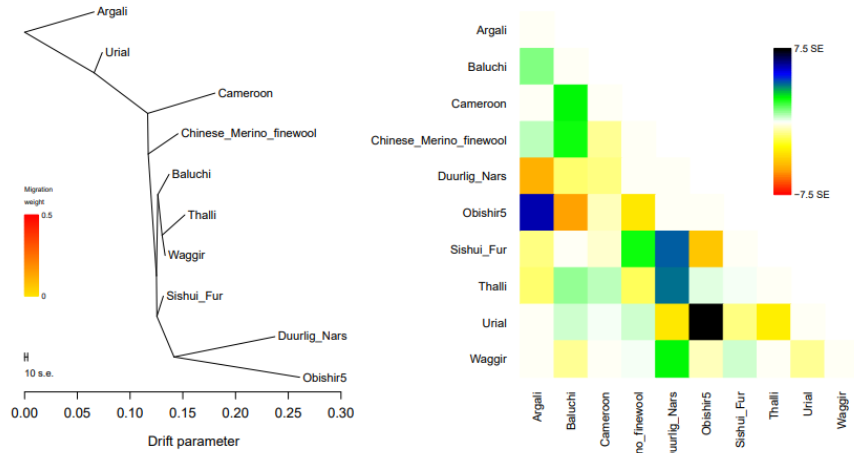


Figure 13. (continued)

(A)



(B)

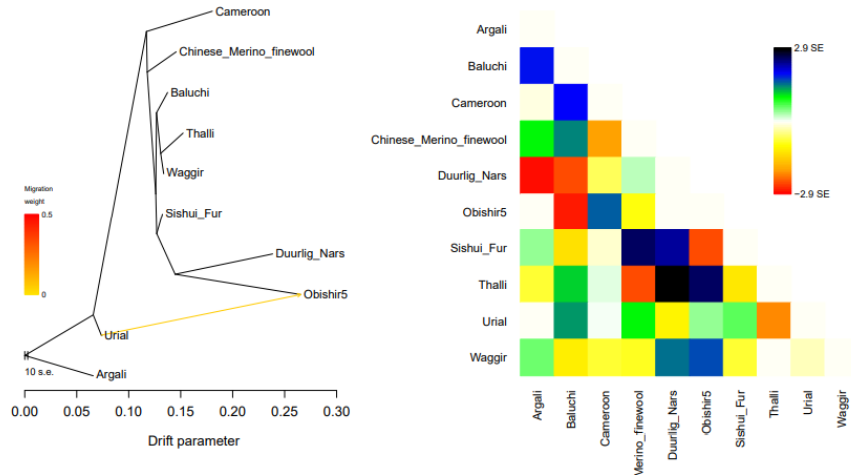
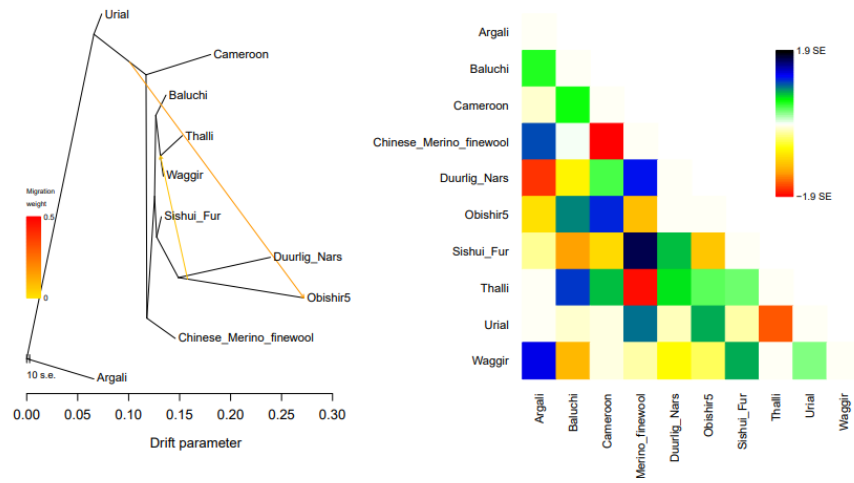


Figure 14. TreeMix tree constructed with ten representative populations of Urial, Cameroon, Chinese_Merino_finewool, Baluchi, Thalli, Sishui_Fur, Waggir, Argali (outgroup), Obishir5, and Duurlig Nars. Tree (A)–(D) represents a tree with migration edge from $m=0$ to $m=3$, respectively.

(C)



(D)

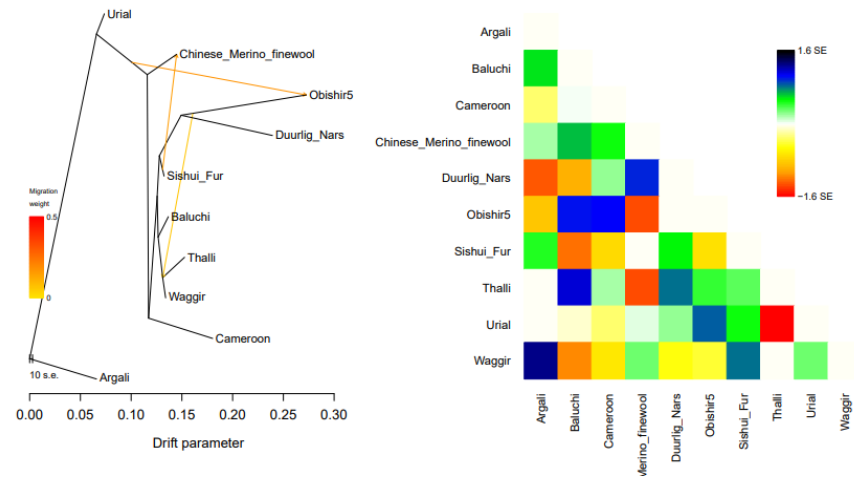


Figure 14. (Continued)

3.6. F_4 -statistics analysis

F_4 -statistics analysis was performed to test the phylogenetic relationships in the TreeMix tree. Firstly, f_4 -statistics analysis of the form $f_4(\text{Urial}, \text{Waggir}; \text{Duurlig Nars}, \text{East Asian})$ was performed to check if Duurlig Nars population have the strongest genetic affinity for the modern East Asian groups. This f_4 -statistics quantifies the symmetry between the Urial/Waggir and Duurlig Nars/East Asian. Waggir represents Central Asian population and most of the East Asian populations did not break the symmetry ($|Z| < 3$) (Figure 15).

Secondly, f_4 -statistics analysis of the form $f_4(\text{Urial}, \text{Duurlig_Nars}; \text{Tibetan}, \text{worldwide})$ was performed to confirm the highest shared genetic drift between Duurlig Nars and Tibetan breeds in present-day populations (Figure 10). All present-day populations except for Prairie_Tibetan.SG and Valley_Tibetan.SG broke the symmetry ($|Z| > 3$), confirming the previous result (Figure 16).

Lastly, f_4 -statistics analysis of the form $f_4(\text{Urial}, \text{worldwide}; \text{Duurlig Nars}, \text{Tibetan})$ was performed to confirm the highest shared genetic drift between Duurlig Nars and Obishir5 ancient sheep (Figure 10). Obishir5 showed significant negative value ($|Z| > 3$), consistent with the previous results. For most of the present-day and ancient populations, f_4 -statistics were non-significant ($|Z| < 3$) (Figure 17).

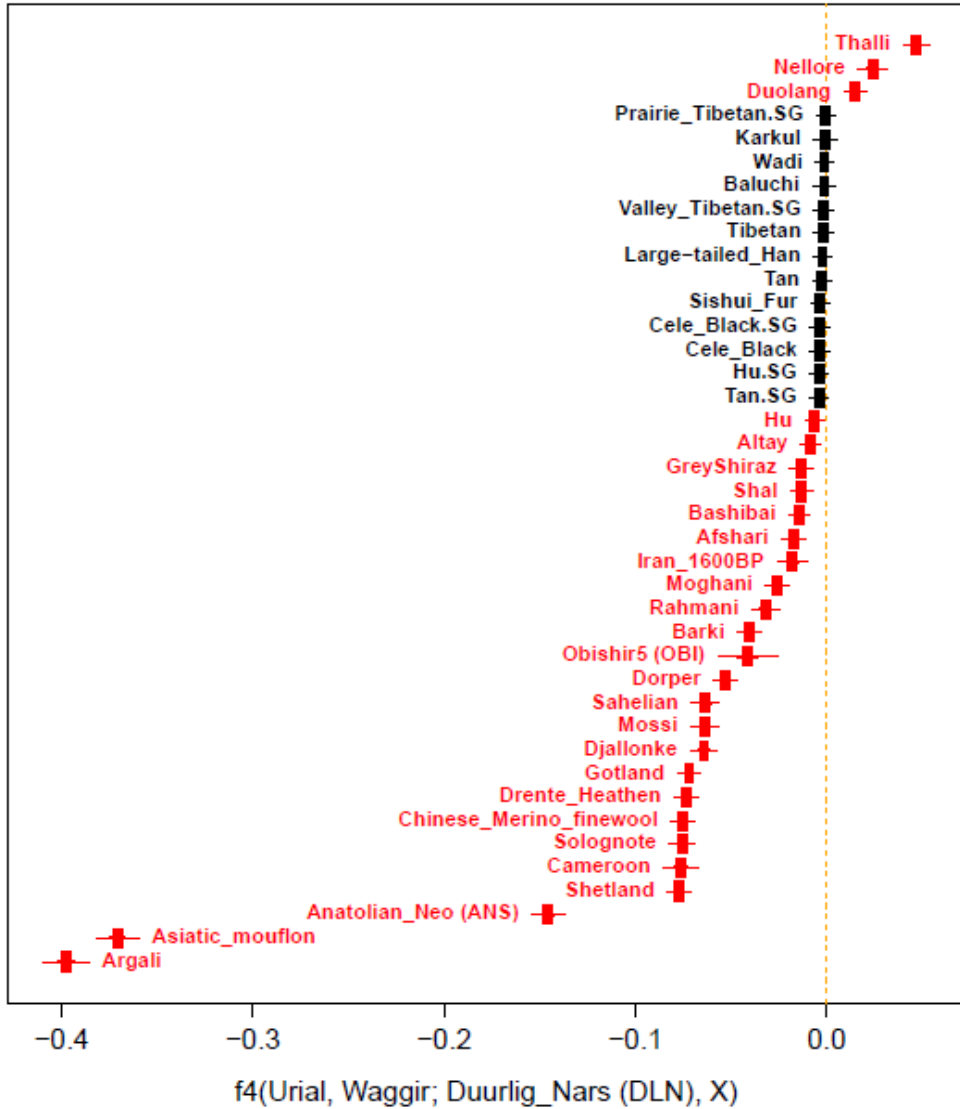


Figure 15. F_4 -statistics plot of the form $f_4(\text{Urial, Waggir; Duurlig_Nars, worldwide})$. F_4 -statistics with $|Z| > 3$ are highlighted in red. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure, respectively.

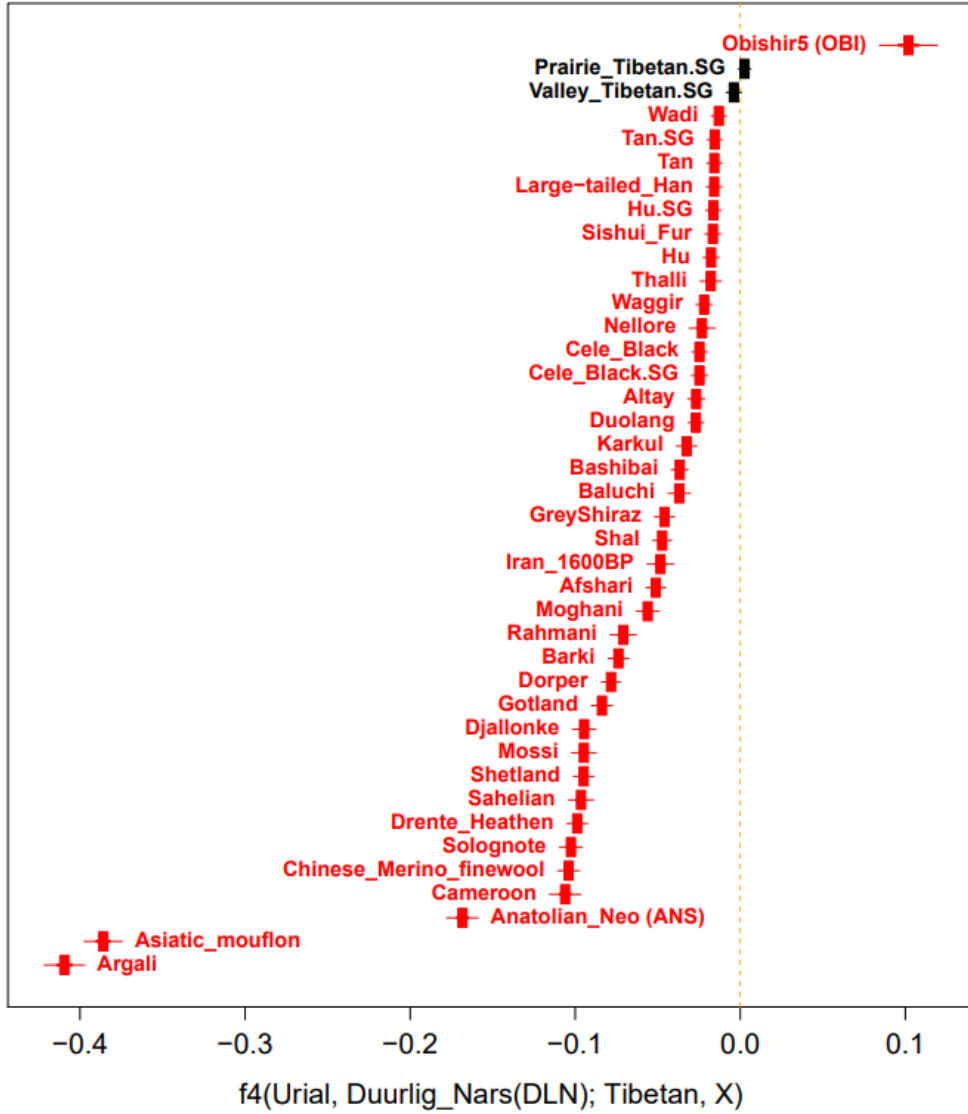


Figure 16. F_4 -statistics plot of the form $f_4(\text{Urrial, Duurlig Nars; Tibetan, worldwide})$. F_4 -statistics with $|Z| > 3$ are highlighted in red. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure, respectively.

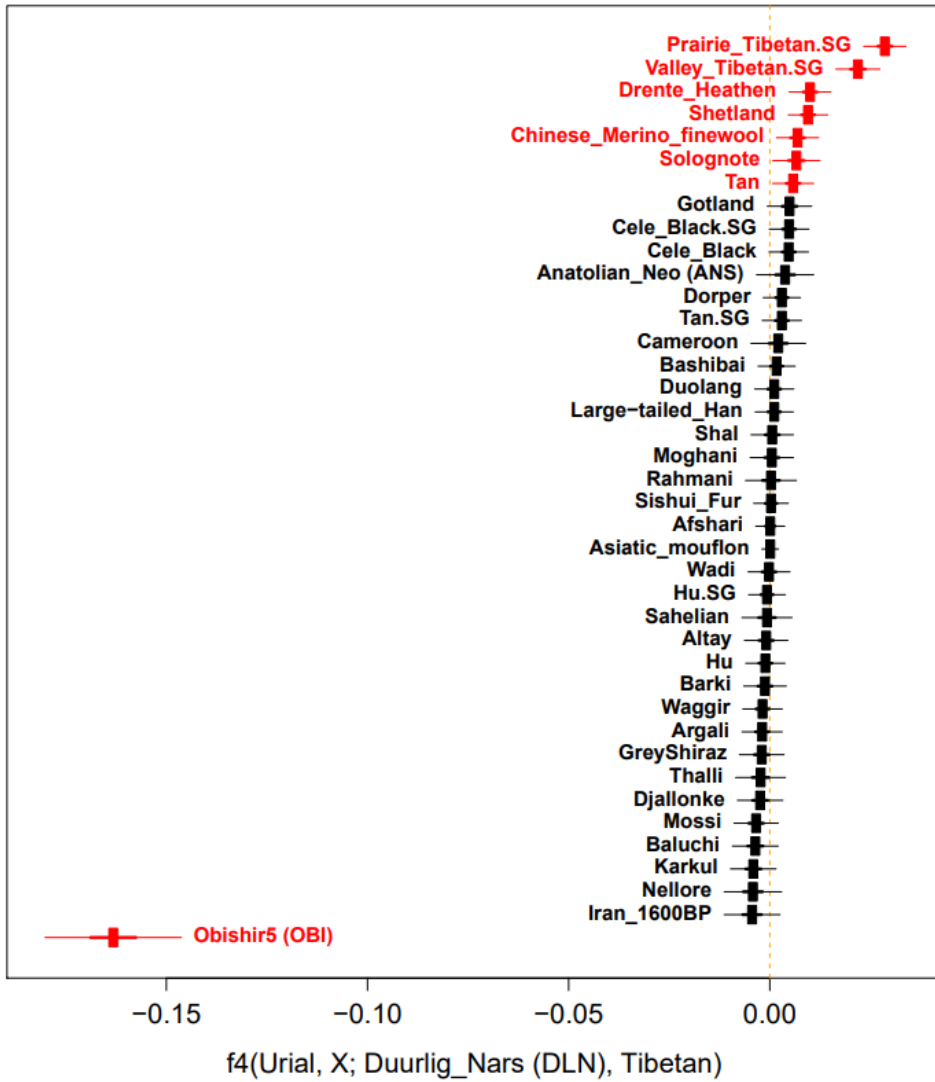


Figure 17. F_4 -statistics plot of the form $f_4(\text{Urial, worldwide; Duurlig Nars, Tibetan})$. F_4 -statistics with $|Z| > 3$ are highlighted in red. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure, respectively.

3.7. Mitochondrial DNA haplogroup analysis

Mitochondrial DNA (mtDNA) haplogroup was observed for comprehensive understanding of sheep used in this study. The mtDNA haplogroup information was inferred using the phylogenetic relationships with previously reported mtDNA haplogroup references downloaded from NCBI under the accession name HM236174–HM236183. Mitochondrial control region (CR) sequences of 1,180 bp corresponding to positions 5437–16616 on the reference AF010406 were extracted from the reads mapped to the sheep reference genome. Individuals with at least 3x sequence depth were used in the analysis. Then, multiple sequence alignment was performed on these sequences.

Maximum Parsimony (MP) tree analysis was conducted using MEGA command line. From phylogenetic relationships with the references, mtDNA haplogroups of sheep individuals was inferred (Figure 18). The most parsimonious tree with length=1,728 was generated. The consistency index was 0.225, the retention index was 0.884, and the composite index was 0.199 for all sites. There was a total of 1,186 positions in the final dataset. Among 17 ancient individuals, only 14 of them were able to infer the mtDNA haplogroup due to lack of mtDNA sequence information. Among six individuals excavated from Duurlig Nars tomb number 160, half of them were grouped with mtDNA haplogroup A and the other half were grouped with mtDNA haplogroup B (Figure 18; Figure 19; Table 3).

Maximum Likelihood (ML) tree analysis was conducted using MEGA. From phylogenetic relationships with the references, mtDNA haplogroups of sheep individuals was inferred (Figure 19). This analysis involved 538 nucleotide sequences. There was a total of 1,186 positions in the final dataset. The mtDNA haplogroup pattern of ancient individuals inferred from the ML tree was consistent with the mtDNA haplogroup pattern inferred from the MP tree (Figure 18; Figure 19).

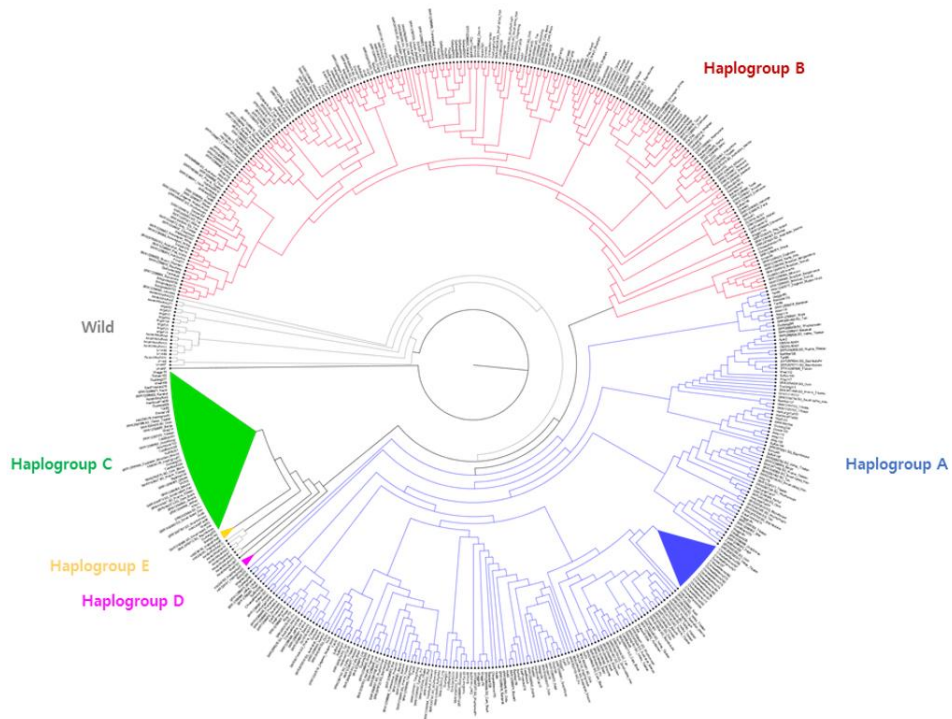


Figure 18. Maximum Parsimony Tree reconstructed using MEGA analysis. Ur191 was used as an outgroup in tree building. The tree shows the mitochondrial DNA relationships of sheep used in this study versus reference genomes of *O. aries* mtDNA haplogroup A–E. The MEGA output tree was visualized using the Interactive Tree Of Life (iTOL) v5, an online tool (source: <https://itol.embl.de/>).

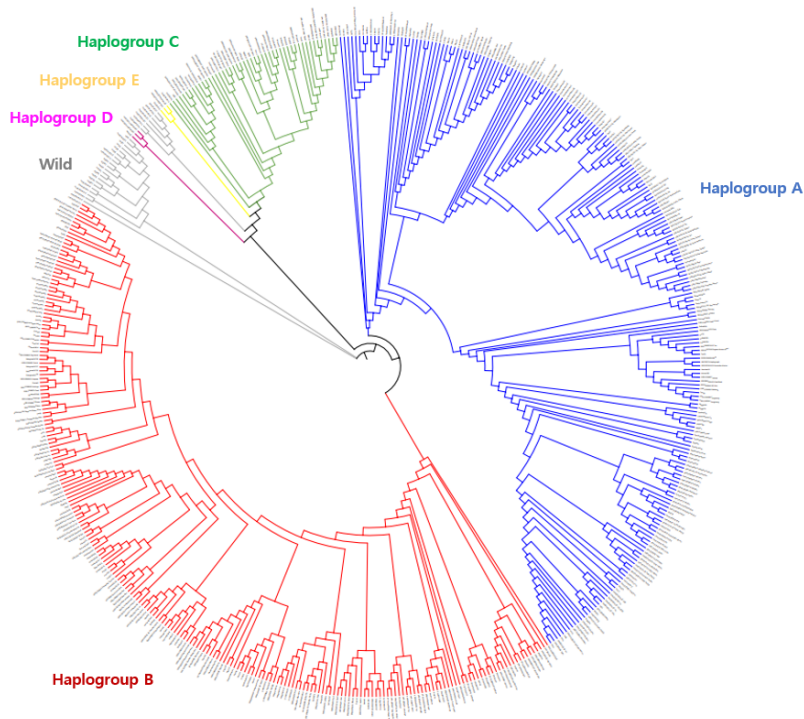
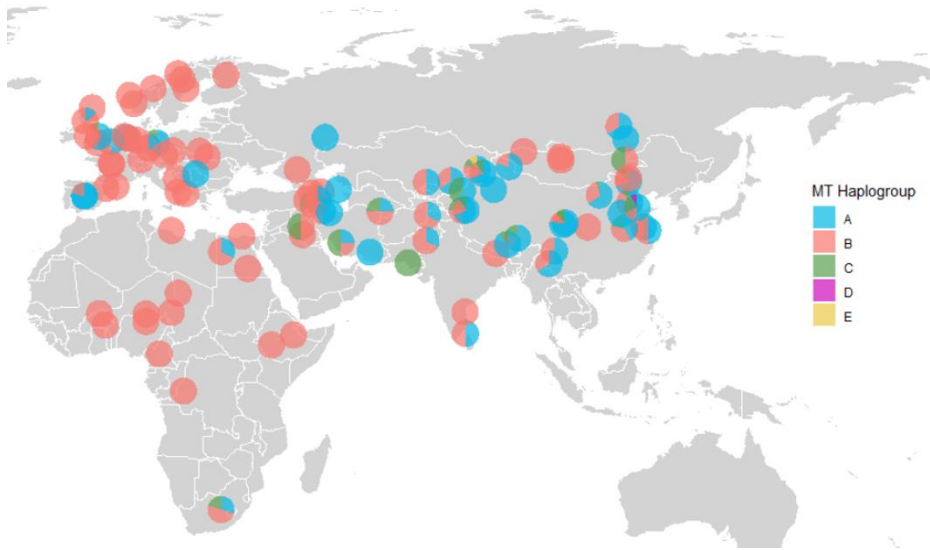


Figure 19 Maximum Likelihood Tree reconstructed using MEGA analysis. The tree shows the mitochondrial DNA relationships of sheep used in this study versus reference genomes of *O. aries* mtDNA haplogroup A–E. The MEGA output tree was visualized using the Interactive Tree Of Life (iTOL) v5, an online tool (source: <https://itol.embl.de/>).

(A)



(B)

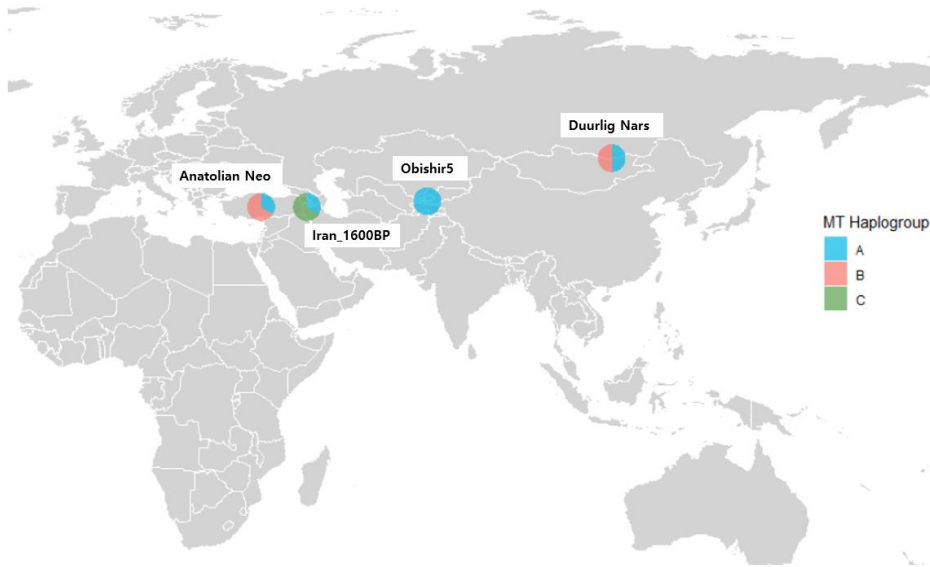


Figure 20. Geographic distribution of mitochondrial DNA haplogroup of each sheep population used in this study. The mtDNA haplogroup pattern was inferred from the Maximum Parsimony tree from Figure 17. (A) The mtDNA haplogroup pie chart of present-day populations (B) The mtDNA haplogroup pie chart of ancient populations.

4. Discussion

The comparative analyses of genome-wide data and mitochondrial DNA from Xiongnu-period sheep from Mongolia provide insight into the origins of East Asian sheep. In this study, genome data of six sheep individuals from satellite burials of Xiongnu-period Duurlig Nars tomb number 160 of Mongolia were sequenced. Genotype data of ancient sheep and present-day sheep populations were typed on the transversion-only 7,497,624 SNPs (7497K) identified in this study. With this genome-wide polymorphism data, population genomic analyses were performed. All PCA, outgroup- f_3 , TreeMix, and f_4 -statistics results indicate the shared lineage between present-day East Asian sheep and Xiongnu-period Duurlig Nars sheep, suggesting ancestry continuity in East Asia (Figure 9; Figure 10; Figure 12; Figure 15). Other ancient sheep populations also show high genetic affinities with present-day populations geographically close to their excavation sites (Figure 2; Figure 9). These results suggest that the complex genetic structure observed in present-day sheep populations was already constructed in the past, thus, supporting ancestry continuity.

The mitochondrial DNA analysis combined with genome-wide analysis of ancient sheep supports the genetic structure in present-day populations. Ovine (genus *Ovis*) mtDNA has five main haplogroups; A, B, C, D, and E. MtDNA haplogroups A and B are the most common and distributed worldwide (Lv et al., 2015). MtDNA

haplogroup A is frequent in the Eastern world and mtDNA haplogroup B is predominant in the Western world (Lv et al., 2015). MtDNA haplogroup C is found primarily in Southwestern Asia, Mongolia, and Northern China. Lastly, mtDNA haplogroup D and E are rare as they are found only in Southwestern Asia (Lv et al., 2015). This distinct distribution of maternal relationships between present-day sheep populations was also observed in worldwide sheep populations used in this study (Figure 20. A). For ancient sheep populations, Anatolian Neolithic sheep (ANS) showed a ratio of 1:2 and Neolithic Central Asian sheep (OBI) showed a ratio of 2:0 of mtDNA haplogroup A to B (Figure 20. B). This supports the distinct geographical distribution of the West-East of mtDNA haplogroup. MtDNA haplogroup C was found in ancient Iranian sheep (IRN), which is consistent with the distribution region of haplogroup C (Figure 20. B). Xiongnu-period Duurlig Nars sheep (DLN) were expected to have a higher percentage of mtDNA haplogroup A, which is predominant in present-day Mongolian sheep. Nevertheless, they were identified to have a ratio of 1:1 of mtDNA haplogroup A and B (Ganbold et al., 2019).

The high genetic affinity between Neolithic Central Asian sheep (OBI) and Xiongnu-period East Asian sheep (DLN) suggested in outgroup- f_3 and f_4 -statistics show discontinuity of geographical ancestry (Figure 10; Figure 15; Figure 16; Figure 17). One possible explanation for this result is that these two ancient Eurasian sheep populations share genetic similarity, this genetic

signal was later lost, and no longer detectable in present-day Eurasian populations. Other possible explanation can be a sequencing bias made in library preparation step as both sequencing libraries were made in the same laboratory in Max Plank Institute (HPI-SHH). However, this scenario is very unlikely as both sequenced reads passed the contamination test. Further analysis on additional ancient sheep individuals will provide more detailed information on this relationship between the ancient sheep in Eurasia.

Most of the present-day sheep show genetic affinity coinciding with the geographical locations except for the Tibetan breeds and Duurlig Nars sheep (Figure 10). Duurlig Nars is located at Northeastern Mongolia, which is geographically closer to Northern China than to the Tibetan region, Southern China. Interestingly, sheep in Northern China and Southern China are distinct in tail phenotype. Northern Chinese breeds tend to have a fat-tailed phenotype, which is also present in Mongolian sheep, and Southern Chinese breeds tend to have a thin-tailed phenotype (Kalds et al., 2021; Luo et al., 2021). Moreover, the distinct genetic pattern in Northern Chinese sheep and Southern Chinese sheep has been reported in previous genetic studies (Zhao et al., 2017). In this study, researchers detected a distinctive selection pattern of tail phenotype in Chinese sheep of the Northern region and the Southern region. Western introgression in Northern Chinese sheep after the divergence of Northern and Southern Chinese sheep was

inferred using the approximate Bayesian computation approach. Archeologically this could be the result of the southward migration of the Hui and the Mongols in the 12th and 13th centuries which could have formed a distinct genetic pattern between Tibetan breeds and Northern Chinese sheep (Christian, 2000).

Although this study confirms the ancestry continuity in East Asian sheep, suggesting that the origins of present-day East Asian sheep goes back at least 2,000 years from now, it cannot tell the whole evolutionary history of the initial arrival, expansion, and admixture of sheep in East Asia. To understand further on how the genetic profile of East Asian sheep that we observe today was formed, further analysis on additional ancient sheep individuals is crucial and this will provide more precise information on the relationship between sheep in Eurasia. Moreover, the sheep diversity panel of 7497K SNPs identified in this study will provide a high-resolution genomic analysis in further studies.

References

Alberto, F.J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., et al. (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications* *9*, 813. 10.1038/s41467-018-03206-y.

Auffret, A.G., Berg, J., and Cousins, S.A.O. (2014). The geography of human-mediated dispersal. *Diversity and Distributions* *20*, 1450–1456. 10.1111/ddi.12251.

Browning, Brian L., and Browning, Sharon R. (2016). Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics* *98*, 116–126. 10.1016/j.ajhg.2015.11.020.

Cai, D.-W., Han, L., Zhang, X.-L., Zhou, H., and Zhu, H. (2007). DNA analysis of archaeological sheep remains from China. *Journal of Archaeological Science* *34*, 1347–1355. 10.1016/j.jas.2006.10.020.

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

Christian, D. (2000). Silk Roads or Steppe Roads? The Silk Roads in World History. *Journal of World History* *11*, 1–26.

Cilingiroglu, C. (2005). The concept of 'Neolithic package': Considering its meaning and applicability. *Documenta Praehistorica* *32*, 1–13. 10.4312/dp.32.1.

Conolly, J., Colledge, S., Dobney, K., Vigne, J.-D., Peters, J., Stopp, B., Manning, K., and Shennan, S. (2011). Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. *Journal of*

Archaeological Science *38*, 538–545. 10.1016/j.jas.2010.10.008.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.–T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.–L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* *110*, 15758–15763. 10.1073/pnas.1314445110.

Deng, J., Xie, X.–L., Wang, D.–F., Zhao, C., Lv, F.–H., Li, X., Yang, J., Yu, J.–L., Shen, M., Gao, L., et al. (2020). Paternal Origins and Migratory Episodes of Domestic Sheep. *Current Biology* *30*, 4085–4095.e4086. 10.1016/j.cub.2020.07.077.

Dong, S. (2016). Overview: Pastoralism in the World. In *Building Resilience of Human–Natural Systems of Pastoralism in the Developing World: Interdisciplinary Perspectives*, S. Dong, K.–A.S. Kassam, J.F. Tourrand, and R.B. Boone, eds. (Springer International Publishing), pp. 1–37. 10.1007/978-3-319-30732-9_1.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* *32*, 1792–1797. 10.1093/nar/gkh340.

Erdős, L., Ambarlı, D., Anenkhonov, O.A., Bátori, Z., Cserhalmi, D., Kiss, M., Kröel–Dulay, G., Liu, H., Magnes, M., Molnár, Z., et al. (2018). The edge of two worlds: A new review and synthesis on Eurasian forest–steppes. *Applied Vegetation Science* *21*, 345–362. 10.1111/avsc.12382.

Eregzen, G. (2010). A study of Xiongnu tombs in Mongolia. *The Korean Association For The Central Asian Studies* *15*. 10.29174/cas.2010.15..003.

Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* *39*, 783–791. 10.2307/2408678

Frantz, L.A.F., Bradley, D.G., Larson, G., and Orlando, L. (2020).

Animal domestication in the era of ancient genomics. *Nature Reviews Genetics* *21*, 449–460. 10.1038/s41576-020-0225-0.

Ganbold, O., Lee, S.-H., Seo, D., Paek, W.K., Manjula, P., Munkhbayar, M., and Lee, J.H. (2019). Genetic diversity and the origin of Mongolian native sheep. *Livestock Science* *220*, 17–25. 10.1016/j.livsci.2018.12.007.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* *28*, 2678–2679. 10.1093/bioinformatics/bts503.

Hatziminaoglou, J., and Boyazoglu, J. (2011). Husbandry of Dairy Animals | Sheep: Multipurpose Management. In *Encyclopedia of Dairy Sciences* (Second Edition), J.W. Fuquay, ed. (Academic Press), pp. 875–881. 10.1016/B978-0-12-374407-4.00241-7.

Hermes, T.R., Tishkin, A.A., Kosintsev, P.A., Stepanova, N.F., Krause-Kyora, B., and Makarewicz, C.A. (2020). Mitochondrial DNA of domesticated sheep confirms pastoralist component of Afanasievo subsistence economy in the Altai Mountains (3300–2900 cal BC). *Archaeological Research in Asia* *24*, 100232. 10.1016/j.ara.2020.100232.

Hiendleder, S., Lewalski, H., Wassmuth, R., and Janke, A. (1998). The Complete Mitochondrial DNA Sequence of the Domestic Sheep (*Ovis aries*) and Comparison with the Other Major Ovine Haplotype. *Journal of Molecular Evolution* *47*, 441–448. 10.1007/PL00006401.

Houle, J.-L., and Broderick, L.G. (2011). Settlement patterns and domestic economy of the Xiongnu in Khanui Valley, Mongolia. *Xiongnu archaeology: Multidisciplinary perspectives of the first steppe empire in Inner Asia*, 137–152.

Jeong, C., Wang, K., Wilkin, S., Taylor, W.T.T., Miller, B.K., Bemmman, J.H., Stahl, R., Chioveli, C., Knolle, F., Ulziibayar, S., et al. (2020). A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* *183*, 890–904.e829.

10.1016/j.cell.2020.10.015.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* *29*, 1682–1684. 10.1093/bioinformatics/btt193.

Kalds, P., Luo, Q., Sun, K., Zhou, S., Chen, Y., and Wang, X. (2021). Trends towards revealing the genetic architecture of sheep tail patterning: Promising genes and investigatory pathways. *Animal Genetics* *52*, 799–812. 10.1111/age.13133.

Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nature Communications* *8*, 14115. 10.1038/ncomms14115.

Kim, K., Brenner, C.H., Mair, V.H., Lee, K.-H., Kim, J.-H., Gelegdorj, E., Batbold, N., Song, Y.-C., Yun, H.-W., Chang, E.-J., et al. (2010). A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *American Journal of Physical Anthropology* *142*, 429–440. 10.1002/ajpa.21242.

Klein, R., and Scott, K. (1986). Re-analysis of faunal assemblages from the Haua Fteah and other Late Quaternary archaeological sites in Cyrenaican Libya. *Journal of Archaeological Science* *13*, 515–542. 10.1016/0305-4403(86)90038-5.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* *15*, 356. 10.1186/s12859-014-0356-4.

Korobitsyna, K.V., Nadler, C.F., Vorontsov, N.N., and Hoffmann, R.S. (1974). Chromosomes of the Siberian Snow Sheep, *Ovis Nivicola*, and Implications Concerning the Origin of Amphiberingian wild Sheep (subgenus *Pachyceros*). *Quaternary Research* *4*, 235–245. 10.1016/0033-5894(74)90013-1.

Kumar, S., Stecher, G., Peterson, D., and Tamura, K. (2012).

MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* *28*, 2685–2686. 10.1093/bioinformatics/bts507.

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* *49*, W293–W296. 10.1093/nar/gkab301.

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

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* *26*, 589–595. 10.1093/bioinformatics/btp698.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079. 10.1093/bioinformatics/btp352.

Li, X., Yang, J., Shen, M., Xie, X.-L., Liu, G.-J., Xu, Y.-X., Lv, F.-H., Yang, H., Yang, Y.-L., Liu, C.-B., et al. (2020). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications* *11*, 2815. 10.1038/s41467-020-16485-1.

Luo, R., Zhang, X., Wang, L., Zhang, L., Li, G., and Zheng, Z. (2021). GLIS1, a potential candidate gene affect fat deposition in sheep tail. *Molecular Biology Reports* *48*, 4925–4931. 10.1007/s11033-021-06468-w.

Lv, F.H., Peng, W.F., Yang, J., Zhao, Y.X., Li, W.R., Liu, M.J., Ma, Y.H., Zhao, Q.J., Yang, G.L., Wang, F., et al. (2015). Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep. *Mol Biol Evol* *32*, 2515–2533. 10.1093/molbev/msv139.

MacHugh, D.E., Larson, G., and Orlando, L. (2017). Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual*

Review of Animal Biosciences *5*, 329–351. 10.1146/annurev-animal-022516-022747.

Mallon, D. (1985). Wild sheep in Mongolia. Wild sheep distribution, abundance, management and conservation of wild sheep of the world and closely related ungulates. Northern Wild Sheep and Goat Council Special Report. Yukon Wildlife Branch, Whitehorse, Canada, 179–187.

McHugo, G.P., Dover, M.J., and MacHugh, D.E. (2019). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. BMC Biology *17*, 98. 10.1186/s12915-019-0724-7.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research *20*, 1297–1303. 10.1101/gr.107524.110.

Meadows, J.R.S., Hiendleder, S., and Kijas, J.W. (2011). Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. Heredity *106*, 700–706. 10.1038/hdy.2010.122.

Miller, B.K. (2014). Xiongnu “Kings” and the Political Order of the Steppe Empire. Journal of the Economic and Social History of the Orient *57*, 1–43. 10.1163/15685209-12341340.

Nadler, C., and RS, H. (1973). Cytogenetic differentiation, geographic distribution, and domestication in Palearctic sheep (*Ovis*).

National Museum of Korea. (2021). Research Report on Korean–Mongolian Expedition in Mongolia VIII (National Museum of Korea).

Neupert, R.F. (1999). Population, nomadic pastoralism and the environment in the Mongolian Plateau. Population and Environment *20*, 413–441.

Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., Miao, B., Hu, W.,

Wang, X., Hu, X., et al. (2018). Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *GigaScience* 7. 10.1093/gigascience/giy019.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in Human History. *Genetics* 192, 1065–1093. 10.1534/genetics.112.145037.

Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics* 2, e190. 10.1371/journal.pgen.0020190.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biology* 17, 60. 10.1186/s13059-016-0918-z.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics* 8, e1002967. 10.1371/journal.pgen.1002967.

Pilipenko, A.S., Cherdantsev, S.V., Trapezov, R.O., Zhuravlev, A.A., Babenko, V.N., Pozdnyakov, D.V., Konovalov, P.B., and Polosmak, N.V. (2018). Mitochondrial DNA diversity in a Transbaikalian Xiongnu population. *Archaeological and Anthropological Sciences* 10, 1557–1570. 10.1007/s12520-017-0481-x.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81, 559–575. 10.1086/519795.

Rezaei, H.R., Naderi, S., Chintauan-Marquier, I.C., Taberlet, P., Virk, A.T., Naghash, H.R., Rioux, D., Kaboli, M., and Pompanon, F. (2010). Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and*

Evolution 54, 315–326. 10.1016/j.ympev.2009.10.037.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nature Biotechnology* 29, 24–26. 10.1038/nbt.1754

Rossi, C., Ruß-Popa, G., Mattiangeli, V., McDaid, F., Hare, A.J., Davoudi, H., Laleh, H., Lorzadeh, Z., Khazaeli, R., Fathi, H., et al. (2021). Exceptional ancient DNA preservation and fibre remains of a Sasanian saltmine sheep mummy in Chehrābād, Iran. *Biology Letters* 17, 20210222. 10.1098/rsbl.2021.0222.

Rüther, P.L., Husic, I.M., Bangsgaard, P., Gregersen, K.M., Pantmann, P., Carvalho, M., Godinho, R.M., Friedl, L., Cascalheira, J., Taurozzi, A.J., et al. (2022). SPIN enables high throughput species identification of archaeological bone by proteomics. *Nature Communications* 13, 2458. 10.1038/s41467-022-30097-x.

Salvagno, L., and Albarella, U. (2017). A morphometric system to distinguish sheep and goat postcranial bones. *PloS one* 12, e0178543–e0178543. 10.1371/journal.pone.0178543.

Scherf, B.D. (2000). World watch list for domestic animal diversity (Food and Agriculture Organization (FAO)).

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* 9, 88. 10.1186/s13104-016-1900-2.

Shackleton, D.M. (1997). Wild sheep and goats and their relatives: status survey and conservation action plan for Caprinae.

Sherratt, A. (1983). The Secondary Exploitation of Animals in the Old World. *World Archaeology* 15, 90–104.

Stiner, M.C., Buitenhuis, H., Duru, G., Kuhn, S.L., Mentzer, S.M., Munro, N.D., Pöllath, N., Quade, J., Tsartsidou, G., and Özbaşaran, M.

(2014). A forager–herder trade–off, from broad–spectrum hunting to sheep management at Aşıklı Höyük, Turkey. *Proceedings of the National Academy of Sciences of the United States of America* *111*, 8404–8409. 10.1073/pnas.1322723111.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* *38*, 3022–3027. 10.1093/molbev/msab120.

Taylor, W.T.T., Clark, J., Bayarsaikhan, J., Tuvshinjargal, T., Jobe, J.T., Fitzhugh, W., Kortum, R., Spengler, R.N., Shnaider, S., Seersholm, F.V., et al. (2020). Early Pastoral Economies and Herding Transitions in Eastern Eurasia. *Scientific Reports* *10*, 1001. 10.1038/s41598-020-57735-y.

Taylor, W.T.T., Pruvost, M., Posth, C., Rendu, W., Krajcarz, M.T., Abdykanova, A., Brancaleoni, G., Spengler, R., Hermes, T., Schiavinato, S., et al. (2021). Evidence for early dispersal of domestic sheep into Central Asia. *Nature Human Behaviour* *5*, 1169–1179. 10.1038/s41562-021-01083-y.

Wilkin, S., Ventresca Miller, A., Taylor, W.T.T., Miller, B.K., Hagan, R.W., Bleasdale, M., Scott, A., Gankhuyg, S., Ramsøe, A., Uliziibayar, S., et al. (2020). Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years. *Nature Ecology & Evolution* *4*, 346–355. 10.1038/s41559-020-1120-y.

Yurtman, E., Özer, O., Yüncü, E., Dağtaş, N.D., Koptekin, D., Çakan, Y.G., Özkan, M., Akbaba, A., Kaptan, D., Atağ, G., et al. (2021). Archaeogenetic analysis of Neolithic sheep from Anatolia suggests a complex demographic history since domestication. *Communications Biology* *4*, 1279. 10.1038/s42003-021-02794-8.

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

Zhao, Y.-X., Yang, J., Lv, F.-H., Hu, X.-J., Xie, X.-L., Zhang, M., Li, W.-R., Liu, M.-J., Wang, Y.-T., Li, J.-Q., et al. (2017). Genomic Reconstruction of the History of Native Sheep Reveals the Peopling Patterns of Nomads and the Expansion of Early Pastoralism in East Asia. *Molecular Biology and Evolution* *34*, 2380–2395. 10.1093/molbev/msx181.

Zilhão, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proceedings of the National Academy of Sciences* *98*, 14180–14185. 10.1073/pnas.241522898.

국문 초록

홍노 시기 몽골 양 유전체를 통한 동아시아 양의 유전적 기원 연구

양(*Ovis aries*)은 지금으로부터 약 10,000–12,000년 전 서남아시아에서 가축화된 동물로 이 시기의 인간의 생활 방식은 수렵채집에서 농경과 정착으로 급격히 변했다. 가축화된 양은 인간의 활동에 의해 여러 지역으로 빠르게 퍼졌는데, 특히 양은 안정적인 수분과 영양 공급책으로서 곡식 농사가 힘든 동아시아 초원 지대의 유목 목축생활의 필수적인 가축이 되었다. 하지만 지금까지 서남아시아에서 처음 가축화된 양이 어떻게 동아시아까지 전파가 되었는지는 잘 알려지지 않았고, 동아시아의 동물 고유전체 자료에 의해 연구된 바가 없다. 본 연구에서는 최초의 유목 제국을 이룬 홍노 무덤인 몽골 북동쪽에 위치한 도르릭 나르스 유적의 배장묘에서 발굴된 양 6개체에서 총유전체 자료를 생산함으로써 현재 동아시아 양의 유전적 기원을 추적하고자 하였다. 홍노 고대양과 비교하기 위하여 총 415 개체의 현대양 총유전체 자료를 처리해 약 749만 개의 전 세계 양 다양성 염기서열 변이 패널을 구성하였다. 도르릭 나르스 홍노 양과 이미 출판된 고대양 11개체와 현대양 514개체의 유전체 정보를 이 변이 패널에 대해서 추출하여 집단유전학 분석을 실시하였다. 이를 통해 도르릭 나르스 홍노 양과 현대 동아시아 양 계통이 유전적으로 매우 유사함을 확인하였다. 또한 이와 같은 유전체 수준에서의 분석과 추가적인 미토콘드리아 DNA 분석에서 다른 고대양 집단에서도 고대양과 발굴된 지역의 현대양 집단 간의 유전적 유연성이 관측되었다.

이번 연구에서는 지금까지 보고되지 않았던 동아시아 양의 고유전체 자료를 생산하였고, 이를 통해 현대양에서 보이는 복잡한 유전적 구조가 상당히 오래전에 형성되었다는 것을 확인할 수 있었다. 이에 현대 동아시아의 양 계통의 역사는 적어도 약 2,000년 전까지 거슬러 올라가는 것으로 보인다. 또한 이번 연구에서 생산된 약 749만 개의 전 세계 양 다양성 염기서열 변이 패널은 앞으로 집단유전체학 연구를 하는데 있어 높은 해상도의 분석을 제공할 것으로 기대된다.

주요어: 집단유전학, 고유전체, 유전체학, 진화유전학, 양, 흉노

학번: 2020-20497

Appendices

Appendix 1. Summary statistics of sheep population used in this study.

No.	Population	Species Name	Origin	Code	PCA group	Latitude (° N)	Longitude (° E)	#
1	Afar	<i>O. aries</i>	Ethiopia	AFS	Africa	9.32	42.06	1
2	Afshari	<i>O. aries</i>	Iran	AFH	West Asia	36.67	48.48	4
3	Altamurana	<i>O. aries</i>	Italy	ALT	Europe	40.82	16.55	2
4	Altay	<i>O. aries</i>	Xinjiang	ALS	Central Asia	46.39	89.70	10
5	Anatolian_Neo	<i>O. aries</i>	N/A	ANS	N/A	38.74	35.45	4
6	Argali	<i>O. ammon</i>	Central Asia	ARG	Ovis ammon	39.47	75.99	8
7	Asiatic_mouflon	<i>O. orientalis</i>	West Asia	MOU	Ovis orientalis	33.48	56.33	16
8	Australian_Merino.SG	<i>O. aries</i>	Spain	AMSG	Improved	39.95	-4.14	10
9	Awassi	<i>O. aries</i>	Iraq	AWA	West Asia	33.19	43.72	2
10	Baidarak	<i>O. aries</i>	Russia	BAJ	Central Asia	52.57	115.40	3
11	Baluchi	<i>O. aries</i>	Iran	BAL	West Asia	27.58	58.94	3
12	Barga	<i>O. aries</i>	Mongolia	BAG	East Asia	47.87	115.23	2
13	Barki	<i>O. aries</i>	Egypt	BAR	Africa	27.69	26.16	3
14	Bashibai	<i>O. aries</i>	China	BSB	Central Asia	45.81	82.61	10
15	Bayad	<i>O. aries</i>	Mongolia	BAY	East Asia	49.70	92.90	1
16	Bayinbuluke.SG	<i>O. aries</i>	Xinjiang	BSG	East Asia	44.89	83.93	10
17	Bentheimer	<i>O. aries</i>	Germany	BENT	Improved	52.30	7.16	1
18	Bonga	<i>O. aries</i>	Ethiopia	BOG	Africa	7.12	37.17	1
19	Bozakh	<i>O. aries</i>	Azerbaijan	BOZ	Improved	40.53	46.02	2
20	Brazilian_Bergamasca	<i>O. aries</i>	Brazil	BER	Improved	-16.40	-49.26	2
21	Brazilian_Somali	<i>O. aries</i>	Brazil	SOM	Improved	-3.75	-40.34	2
22	Brown_Mountain	<i>O. aries</i>	Germany	BBS	Improved	47.95	8.36	1
23	Cameroon	<i>O. aries</i>	Cameroon	CAM	Africa	5.12	12.39	3
24	Carpathian_Mountain	<i>O. aries</i>	Ukraine	MKS	Improved	48.55	23.01	1
25	Caucasian	<i>O. aries</i>	Russia	CAUC	Europe	45.71	42.88	1
26	Cele_Black	<i>O. aries</i>	Xinjiang	CLS	East Asia	36.77	79.35	10
27	Cele_Black.SG	<i>O. aries</i>	Xinjiang	CBSG	East Asia	36.77	79.35	10
28	Chinese_Merino_fine wool	<i>O. aries</i>	Spain	MFW	Improved	39.93	-4.14	10

: Number of individuals in population

Appendix 1. Continued (2/4).

No.	Population	Species Name	Origin	Code	PCA group	Latitude (° N) Longitude (° E)		#
29	Chinese_Merino_supe rfinewool	<i>O. aries</i>	Spain	MSF	Improved	39.90	-4.14	10
30	Dagestan_Mountain_ Merino	<i>O. aries</i>	Russia	DAME	Improved	43.60	46.77	1
31	Dalmatian	<i>O. aries</i>	Croatia	DAS	Europe	43.86	16.16	3
32	Degeres_Mutton- Wool	<i>O. aries</i>	Kazakhstan	DEG	Improved	43.27	76.81	3
33	Diqing	<i>O. aries</i>	Yunnan	DQS	East Asia	27.82	99.70	2
34	Djallonke	<i>O. aries</i>	Burkina Faso	DJI	Africa	11.58	-3.03	4
35	Dorper	<i>O. aries</i>	South Africa	DPS	Improved	-29.14	26.06	10
36	Drente_Heathen	<i>O. aries</i>	Netherland s	DRS	Europe	52.51	5.66	5
37	Duolang	<i>O. aries</i>	Xinjiang	DLS	Central Asia	41.08	79.35	10
38	Duurlig_Nars	<i>O. aries</i>	N/A	DLN	N/A	48.55	111.08	6
39	East_Friesian_Dairy	<i>O. aries</i>	Germany	EFR	Improved	51.62	11.84	10
40	Feral	<i>O. aries</i>	Norway	FSS	Europe	62.09	5.79	3
41	Finn	<i>O. aries</i>	Finland	FIN	Europe	66.52	27.12	10
42	Ghezel	<i>O. aries</i>	Iran	GHE	West Asia	38.07	46.30	3
43	Gotland	<i>O. aries</i>	Sweden	GOT	Improved	65.59	17.46	10
44	Grey_Tronder	<i>O. aries</i>	Norway	GTS	Europe	63.61	11.06	1
45	GreyShiraz	<i>O. aries</i>	Iran	GSS	West Asia	29.62	52.53	4
46	Hamdani	<i>O. aries</i>	West Asia	HAS	West Asia	38.24	44.16	1
47	Hampshire	<i>O. aries</i>	UK	HAM	Improved	51.60	-1.32	2
48	Hanzhong	<i>O. aries</i>	Shanxi	HZS	East Asia	33.07	107.02	1
49	Helsing	<i>O. aries</i>	Sweden	HEL	Europe	63.54	18.28	3
50	Hu	<i>O. aries</i>	Jiangsu	HUS	East Asia	33.14	119.40	10
51	Hu.SG	<i>O. aries</i>	Jiangsu	HUSG	East Asia	33.14	119.40	10
52	Iran_1600BP	<i>O. aries</i>	N/A	IRN	N/A	37.33	48.02	3
53	Jaffna_Local	<i>O. aries</i>	Sri Lanka	JLS	South Asia	9.66	80.03	2
54	Kachi	<i>O. aries</i>	Iran	KAC	South Asia	25.21	67.34	1
55	Kage	<i>O. aries</i>	Nepal	KAG	South Asia	27.27	86.71	2
56	Karakul	<i>O. aries</i>	Iran	KARA	West Asia	36.55	61.16	4
57	Karkul	<i>O. aries</i>	Pakistan	KAR	South Asia	35.69	71.66	3
58	Kazakh_Finewool	<i>O. aries</i>	Kazakhstan	KATO	Improved	42.95	71.42	2
59	Khotont	<i>O. aries</i>	Mongolia	KHO	Improved	48.57	101.00	1
60	Kuibyshev	<i>O. aries</i>	Russia	KUIB	Europe	52.84	49.02	3

: Number of individuals in population

Appendix 1. Continued (3/4).

No.	Population	Species Name	Origin	Code	PCA group	Latitude (° N) Longitude (° E)		#
61	Large-tailed_Han	<i>O. aries</i>	China	HDW	East Asia	38.53	116.30	10
62	Leccese	<i>O. aries</i>	Italy	LEC	Europe	40.48	18.22	2
63	Libyan_Barbar	<i>O. aries</i>	Libya	LBS	Africa	32.32	15.10	1
64	Lop	<i>O. aries</i>	Xinjiang	LB	East Asia	41.35	86.27	3
65	Makui	<i>O. aries</i>	Iran	MAK	West Asia	39.05	44.93	2
66	Mazekh	<i>O. aries</i>	Azerbaijan	MAZ	West Asia	40.79	48.13	1
67	Mbororo	<i>O. aries</i>	Niger	MBS	Africa	13.92	9.46	3
68	Moghani	<i>O. aries</i>	Iran	MOH	West Asia	39.09	47.45	3
69	Mossi	<i>O. aries</i>	Burkina Faso	MOS	Africa	14.01	-0.86	3
70	Nellore	<i>O. aries</i>	India	NES	South Asia	14.39	79.94	1
71	North_Ronaldsay	<i>O. aries</i>	Scotland	NRS	Europe	59.37	-2.42	1
72	Norwegian_White	<i>O. aries</i>	Norway	NWS	Improved	66.84	16.52	3
73	Obishir5	<i>O. aries</i>	N/A	OBI	N/A	39.96	71.28	4
74	Old_Spael	<i>O. aries</i>	Norway	OSP	Europe	60.27	6.76	3
75	Olkuska	<i>O. aries</i>	Poland	OLS	Improved	49.78	21.34	1
76	Ossimi	<i>O. aries</i>	Egypt	OSS	Africa	23.96	31.92	2
77	Ouessant	<i>O. aries</i>	France	OUE	Europe	47.04	1.78	10
78	Oula.SG	<i>O. aries</i>	Croatia	OSG	East Asia	33.62	101.56	10
79	Pag_Island	<i>O. aries</i>	Tibet	PIS	Improved	44.50	15.00	3
80	Prairie_Tibetan.SG	<i>O. aries</i>	Tibet	PTSG	East Asia	30.61	91.54	10
81	Pramenka	<i>O. aries</i>	Serbia	PRAM	Europe	44.74	20.44	2
82	Rahmani	<i>O. aries</i>	Egypt	RAH	Africa	31.10	30.63	3
83	Rhoen	<i>O. aries</i>	Germany	RHS	Europe	50.47	10.07	3
84	Ripollesa	<i>O. aries</i>	Spain	RIP	Improved	41.98	2.82	1
85	Romanov	<i>O. aries</i>	Russia	ROM	Europe	41.39	51.95	1
86	Romney	<i>O. aries</i>	UK	ROS	Improved	51.87	1.68	2
87	Sahelian	<i>O. aries</i>	Burkina Faso	SAH	Africa	11.46	0.47	3
88	Santa_Ines	<i>O. aries</i>	Brazil	SIS	Improved	-3.68	-40.39	1
89	Shal	<i>O. aries</i>	Iran	SHA	West Asia	36.27	50.00	4
90	Shetland	<i>O. aries</i>	Scotland	SHE	Europe	56.47	-3.99	7
91	Sishui_Fur	<i>O. aries</i>	Shandong	SSS	East Asia	32.30	120.37	10
92	Small-tailed_Han	<i>O. aries</i>	Henan	SXW	East Asia	32.81	115.14	10

: Number of individuals in population

Appendix 1. Continued (4/4).

No.	Population	Species Name	Origin	Code	PCA group	Latitude (° N)	Longitude (° E)	#
93	Small-tailed_Han.SG	<i>O. aries</i>	Henan	STHS G	East Asia	32.81	115.14	9
94	Solognote	<i>O. aries</i>	France	SOL	Europe	46.90	1.89	10
95	Suffolk	<i>O. aries</i>	England	SFK	Improved	53.06	-1.02	10
96	Sumavska	<i>O. aries</i>	Czech Republic	SUM	Improved	49.00	13.49	1
97	Taihang_Fur	<i>O. aries</i>	Henan	THQ	East Asia	36.09	114.39	1
98	Tamir	<i>O. aries</i>	Mongolia	TAM	East Asia	47.85	101.14	1
99	Tan	<i>O. aries</i>	Ningxia	TAN	East Asia	40.12	109.59	10
100	Tan.SG	<i>O. aries</i>	Ningxia	TSG	East Asia	40.12	109.59	10
101	Tengchong	<i>O. aries</i>	Yunnan	TCS	East Asia	25.02	98.49	3
102	Texel	<i>O. aries</i>	Netherlands	TEX	Improved	53.06	4.75	3
103	Thalli	<i>O. aries</i>	Pakistan	THA	South Asia	30.09	71.34	3
104	Tibetan	<i>O. aries</i>	Tibet	GMA	East Asia	33.99	102.07	9
105	Transbaikals_Finewool	<i>O. aries</i>	Russia	ZAB	Improved	55.29	113.95	3
106	Uda	<i>O. aries</i>	Chad	UDA	Africa	14.24	15.07	1
107	Ujimqin	<i>O. aries</i>	Inner Mongolia	WZS	East Asia	43.92	116.03	2
108	Urial	<i>O. vignei</i>	Central Asia	URI	Ovis vignei	37.48	57.33	5
109	Valachian	<i>O. aries</i>	Czech Republic	VAL	Europe	49.82	15.47	2
110	Valley_Tibetan.SG	<i>O. aries</i>	Tibet	VTSG	East Asia	29.27	89.30	10
111	Wadi	<i>O. aries</i>	China	WDS	East Asia	37.39	117.98	13
112	Waggir	<i>O. aries</i>	Afghanistan	WGR	Central Asia	36.81	80.20	10
113	Welsh_Mountain	<i>O. aries</i>	UK	WMS	Europe	53.16	-3.49	1
114	West_African_Dwarf	<i>O. aries</i>	Chad	WGS	Africa	18.53	16.57	1
115	White_Mountain	<i>O. aries</i>	Germany	WBS	Improved	47.95	8.36	1
116	Wuzhumuqin.SG	<i>O. aries</i>	China	WZSG	East Asia	43.09	116.01	10
117	Xisqueta	<i>O. aries</i>	Spain	XIS	Europe	41.61	0.62	1
118	Yankasa	<i>O. aries</i>	Nigeria	YAN	Africa	12.30	9.42	1
119	Zwartbles	<i>O. aries</i>	Netherlands	ZWA	Improved	52.80	5.10	1

: Number of individuals in population

Appendix 2. Summary statistics of previously published present-day individuals used in this study.

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
1	Afar2746	Afar	F	B	17.25	17.17	287.73	[1]
2	Afshari202	Afshari	M	A	17.34	9.39	328.65	[1]
3	Altay109	Altay	F	B	23.72	22.95	1540.74	[1]
4	Altay111	Altay	F	A	22.50	22.21	760.25	[1]
5	Altay113	Altay	F	A	22.63	22.50	773.96	[1]
6	Altay114	Altay	F	A	22.70	22.36	930.57	[1]
7	Altay115	Altay	F	A	23.22	22.72	1071.48	[1]
8	Altay116	Altay	F	A	23.43	23.03	1035.57	[1]
9	Altay117	Altay	F	A	22.78	22.33	975.90	[1]
10	Altay118	Altay	F	B	26.08	26.10	910.10	[1]
11	Altay119	Altay	F	A	23.63	22.90	1095.08	[1]
12	Altay120	Altay	F	A	22.63	22.11	862.32	[1]
13	Argali104	Argali	M	N/A	15.64	8.42	317.22	[2]
14	Argali19	Argali	M	N/A	14.97	8.12	1295.32	[2]
15	Argali20	Argali	M	N/A	16.39	8.82	1332.82	[2]
16	Argali21	Argali	M	N/A	15.20	8.13	171.88	[2]
17	Argali23	Argali	M	N/A	18.95	10.26	1476.97	[2]
18	Argali31	Argali	M	N/A	16.77	9.04	305.82	[2]
19	Argali82	Argali	M	N/A	15.35	8.24	164.32	[2]
20	Argali93	Argali	M	N/A	17.63	9.51	99.45	[2]
21	AsiaticMouflon1	Asiatic_mouflon	M	N/A	30.31	16.21	236.48	[1]
22	AsiaticMouflon10	Asiatic_mouflon	M	N/A	21.74	11.64	213.52	[1]
23	AsiaticMouflon11	Asiatic_mouflon	F	N/A	32.81	32.21	246.28	[1]
24	AsiaticMouflon12	Asiatic_mouflon	F	N/A	25.38	24.83	396.42	[1]
25	AsiaticMouflon19	Asiatic_mouflon	F	N/A	21.84	21.82	330.80	[1]
26	AsiaticMouflon2	Asiatic_mouflon	M	N/A	21.65	11.70	231.11	[1]
27	AsiaticMouflon20	Asiatic_mouflon	F	N/A	22.17	21.59	615.13	[1]
28	AsiaticMouflon266	Asiatic_mouflon	M	N/A	26.08	14.07	338.39	[1]
29	AsiaticMouflon267	Asiatic_mouflon	M	N/A	21.25	11.44	207.74	[1]
30	AsiaticMouflon271	Asiatic_mouflon	F	N/A	23.73	23.74	398.34	[1]
31	AsiaticMouflon272	Asiatic_mouflon	F	N/A	25.94	25.52	260.44	[1]

Appendix 2. Continued (2/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
32	AsiaticMouflon273	Asiatic_mouflon	F	N/A	24.33	23.90	353.40	[1]
33	AsiaticMouflon4	Asiatic_mouflon	M	N/A	22.75	12.24	669.02	[1]
34	AsiaticMouflon6	Asiatic_mouflon	M	N/A	28.09	15.22	2548.36	[1]
35	AsiaticMouflon7	Asiatic_mouflon	M	N/A	23.09	12.46	502.20	[1]
36	AsiaticMouflon9	Asiatic_mouflon	M	N/A	22.09	11.93	141.94	[1]
37	Awassi021	Awassi	F	B	16.61	16.51	232.07	[1]
38	Bashbai126	Bashibai	F	A	22.81	22.60	701.23	[1]
39	Bashbai128	Bashibai	F	A	24.54	24.04	962.59	[1]
40	Bashbai129	Bashibai	F	C	23.81	23.53	935.15	[1]
41	Bashbai130	Bashibai	F	B	23.65	23.28	1060.82	[1]
42	Bashbai136	Bashibai	F	B	22.55	22.17	876.71	[1]
43	Bashbai137	Bashibai	F	A	24.34	23.78	981.42	[1]
44	Bashbai139	Bashibai	F	A	22.83	22.62	813.14	[1]
45	Bashbai140	Bashibai	F	E	23.65	23.36	925.81	[1]
46	Bashbai141	Bashibai	F	B	21.60	21.42	768.28	[1]
47	Bashibai123	Bashibai	F	A	22.72	22.24	800.35	[1]
48	Bonga2602	Bonga	F	B	18.26	18.12	491.32	[1]
49	Cameroon07	Cameroon	F	B	19.13	18.81	248.80	[1]
50	CeleBlack222	Cele_Black	F	A	22.18	21.98	631.19	[1]
51	CeleBlack226	Cele_Black	F	C	22.16	21.98	674.63	[1]
52	CeleBlack227	Cele_Black	F	C	22.32	22.13	455.98	[1]
53	CeleBlack228	Cele_Black	F	B	21.70	21.59	513.70	[1]
54	CeleBlack233	Cele_Black	F	A	22.45	22.28	645.82	[1]
55	CeleBlack234	Cele_Black	F	A	22.52	22.58	733.28	[1]
56	CeleBlack235	Cele_Black	F	A	25.15	24.75	593.28	[1]
57	CeleBlackA	Cele_Black	F	A	23.48	23.17	527.55	[1]
58	CeleBlackB	Cele_Black	F	A	23.43	23.05	591.87	[1]
59	CeleBlackD	Cele_Black	F	C	23.50	24.41	234.06	[1]
60	ChineseMerino157	Chinese_Merino_fin ewool	F	A	24.87	24.72	614.48	[1]
61	ChineseMerino159	Chinese_Merino_fin ewool	F	A	25.26	26.78	336.89	[1]
62	ChineseMerino160	Chinese_Merino_fin ewool	F	A	22.67	21.85	1092.37	[1]

Appendix 2. Continued (3/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
63	ChineseMerino161	Chinese_Merino_fin ewool	F	A	24.52	24.31	612.22	[1]
64	ChineseMerino163	Chinese_Merino_fin ewool	F	A	22.98	22.73	583.83	[1]
65	ChineseMerino164	Chinese_Merino_fin ewool	F	A	23.74	22.96	703.48	[1]
66	ChineseMerino165	Chinese_Merino_fin ewool	F	A	23.62	23.40	721.03	[1]
67	ChineseMerino167	Chinese_Merino_fin ewool	F	A	25.77	27.34	238.96	[1]
68	ChineseMerino169	Chinese_Merino_fin ewool	F	A	27.44	27.48	995.46	[1]
69	ChineseMerino170	Chinese_Merino_fin ewool	F	A	24.56	24.61	563.02	[1]
70	ChineseMerino173	Chinese_Merino_su perfinewool	F	A	24.64	24.16	585.11	[1]
71	ChineseMerino174	Chinese_Merino_su perfinewool	F	A	27.49	26.91	597.25	[1]
72	ChineseMerino175	Chinese_Merino_su perfinewool	F	A	23.68	23.20	585.40	[1]
73	ChineseMerino176	Chinese_Merino_su perfinewool	F	B	22.40	22.02	541.72	[1]
74	ChineseMerino177	Chinese_Merino_su perfinewool	F	A	23.31	22.94	602.16	[1]
75	ChineseMerino180	Chinese_Merino_su perfinewool	F	A	23.18	22.96	635.69	[1]
76	ChineseMerino181	Chinese_Merino_su perfinewool	F	B	29.22	29.77	678.72	[1]
77	ChineseMerino184	Chinese_Merino_su perfinewool	F	A	23.95	24.38	553.05	[1]
78	ChineseMerino186	Chinese_Merino_su perfinewool	F	A	23.75	23.99	691.85	[1]
79	ChineseMerino190	Chinese_Merino_su perfinewool	F	A	21.94	21.74	722.38	[1]
80	Djallonke180b	Djallonke	F	B	18.10	17.70	149.09	[1]
81	Dorper142	Dorper	F	C	22.61	22.16	863.53	[1]
82	Dorper147	Dorper	F	B	22.22	21.66	993.67	[1]
83	Dorper149	Dorper	F	B	22.22	21.79	1042.73	[1]
84	Dorper153	Dorper	F	C	23.35	22.86	1114.85	[1]
85	Dorper155	Dorper	F	A	22.37	21.96	1185.80	[1]
86	Dorper357	Dorper	F	A	23.49	23.27	810.15	[1]
87	Dorper358	Dorper	F	B	23.20	22.78	974.37	[1]
88	Dorper360	Dorper	F	A	22.49	22.01	662.47	[1]

Appendix 2. Continued (4/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
89	Dorper369	Dorper	F	B	22.68	22.24	737.97	[1]
90	Dorper370	Dorper	F	B	22.34	21.98	793.86	[1]
91	Drente1	Drente_Heathen	M	B	18.18	9.84	473.46	[1]
92	Drente2	Drente_Heathen	M	B	22.91	12.36	736.11	[1]
93	Drente3	Drente_Heathen	M	B	23.37	12.61	737.33	[1]
94	Drente4	Drente_Heathen	M	B	23.90	12.95	689.26	[1]
95	Drente5	Drente_Heathen	M	B	22.46	12.20	685.50	[1]
96	Duolang249	Duolang	F	A	24.89	24.54	1399.55	[1]
97	Duolang305	Duolang	M	C	23.34	12.59	939.30	[1]
98	Duolang309	Duolang	M	A	25.46	13.76	972.54	[1]
99	Duolang311	Duolang	M	A	25.36	13.47	1262.12	[1]
100	Duolang313	Duolang	M	A	27.90	15.02	998.20	[1]
101	Duolang317	Duolang	M	C	22.69	12.30	1070.31	[1]
102	Duolang325	Duolang	F	C	22.70	22.38	902.96	[1]
103	Duolang333	Duolang	F	C	22.28	21.96	932.87	[1]
104	Duolang334	Duolang	F	A	24.89	24.36	837.68	[1]
105	Duolang348	Duolang	F	A	23.24	23.32	599.38	[1]
106	EastFriesian02	East_Friesian_Dairy	M	A	22.43	12.45	188.00	[1]
107	EastFriesian06	East_Friesian_Dairy	F	B	21.92	21.61	222.82	[1]
108	EastFriesian10	East_Friesian_Dairy	F	A	25.13	24.53	175.25	[1]
109	EastFriesian12	East_Friesian_Dairy	F	B	22.86	22.29	135.55	[1]
110	EastFriesian208	East_Friesian_Dairy	F	B	25.41	26.51	617.78	[1]
111	EastFriesian212	East_Friesian_Dairy	F	B	22.62	22.18	971.81	[1]
112	EastFriesian213	East_Friesian_Dairy	F	A	24.29	24.28	833.49	[1]
113	EastFriesian215	East_Friesian_Dairy	F	C	22.33	22.52	596.76	[1]
114	EastFriesian216	East_Friesian_Dairy	F	A	23.73	23.23	743.97	[1]
115	EastFriesian219	East_Friesian_Dairy	F	A	21.73	21.60	671.30	[1]
116	Finn300	Finn	F	B	23.81	23.25	351.51	[1]
117	Finn303	Finn	F	B	23.63	23.24	246.15	[1]
118	Finn304	Finn	F	B	23.09	22.76	398.70	[1]
119	Finn305	Finn	F	B	27.12	26.80	482.11	[1]
120	Finn306	Finn	F	B	21.95	21.66	284.25	[1]

Appendix 2. Continued (5/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
121	Finn307	Finn	F	B	25.61	25.27	260.25	[1]
122	Finn313	Finn	F	B	25.65	25.32	288.26	[1]
123	Finn314	Finn	F	B	24.01	23.61	369.60	[1]
124	Finn315	Finn	F	B	22.66	22.31	410.68	[1]
125	Finn317	Finn	F	B	25.50	25.14	388.02	[1]
126	Ghezel152	Ghezel	F	A	16.42	16.26	140.81	[1]
127	Gotland24506	Gotland	F	B	22.41	21.91	519.12	[1]
128	Gotland24507	Gotland	F	B	23.04	22.96	188.43	[1]
129	Gotland24508	Gotland	F	B	22.69	22.52	490.00	[1]
130	Gotland24510	Gotland	M	B	22.58	12.09	206.67	[1]
131	Gotland24512	Gotland	F	B	22.60	22.12	158.12	[1]
132	Gotland24513	Gotland	F	B	22.09	21.57	266.95	[1]
133	Gotland24514	Gotland	M	B	22.91	12.44	296.54	[1]
134	Gotland24515	Gotland	M	B	21.92	11.81	173.08	[1]
135	Gotland24516	Gotland	M	B	22.34	11.96	160.31	[1]
136	Gotland24517	Gotland	F	B	22.27	21.76	286.44	[1]
137	GreyShiraz102	GreyShiraz	M	C	15.62	8.49	252.61	[1]
138	Hamdani18	Hamdani	F	A	17.29	16.50	179.20	[1]
139	HanLargeTail47	Large-tailed_Han	F	A	21.27	20.92	451.18	[1]
140	HanLargeTail48	Large-tailed_Han	F	C	21.72	21.31	676.52	[1]
141	HanLargeTail49	Large-tailed_Han	F	B	22.01	21.62	696.10	[1]
142	HanLargeTail50	Large-tailed_Han	F	A	21.89	21.67	612.65	[1]
143	HanLargeTail51	Large-tailed_Han	F	B	21.83	21.54	723.78	[1]
144	HanLargeTail52	Large-tailed_Han	F	B	22.29	21.71	965.55	[1]
145	HanLargeTail54	Large-tailed_Han	F	A	21.36	21.07	453.00	[1]
146	HanLargeTail56	Large-tailed_Han	F	A	21.36	20.92	589.67	[1]
147	HanLargeTail59	Large-tailed_Han	F	B	21.33	20.98	724.53	[1]
148	HanLargeTail60	Large-tailed_Han	F	B	21.31	20.79	606.41	[1]
149	HanSmallTail16	Small-tailed_Han	F	B	22.06	21.89	416.87	[1]
150	HanSmallTail17	Small-tailed_Han	F	A	25.13	24.04	366.45	[1]
151	HanSmallTail19	Small-tailed_Han	F	B	24.61	24.01	440.91	[1]
152	HanSmallTail20	Small-tailed_Han	F	B	24.57	23.69	497.26	[1]

Appendix 2. Continued (6/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
153	HanSmallTail23	Small-tailed_Han	F	A	24.56	24.89	342.98	[1]
154	HanSmallTail25	Small-tailed_Han	F	A	23.73	24.25	349.77	[1]
155	HanSmallTail26	Small-tailed_Han	F	A	25.90	25.43	520.76	[1]
156	HanSmallTail28	Small-tailed_Han	F	B	24.94	24.88	383.88	[1]
157	HanSmallTail29	Small-tailed_Han	F	B	22.74	22.28	307.66	[1]
158	HanSmallTail30	Small-tailed_Han	F	C	25.26	24.53	448.49	[1]
159	Hu62	Hu	F	A	21.71	21.18	487.73	[1]
160	Hu63	Hu	F	A	21.32	20.83	445.17	[1]
161	Hu64	Hu	F	A	21.07	20.55	456.97	[1]
162	Hu65	Hu	F	B	20.69	20.28	432.48	[1]
163	Hu66	Hu	F	A	21.37	21.03	542.01	[1]
164	Hu68	Hu	F	C	21.18	20.61	433.65	[1]
165	Hu69	Hu	F	A	21.24	20.80	468.62	[1]
166	Hu73	Hu	F	B	21.07	20.63	418.37	[1]
167	Hu75	Hu	F	A	21.80	21.23	567.92	[1]
168	Hu76	Hu	F	A	21.87	21.48	378.29	[1]
169	Karakul4265	Karakul	M	B	17.63	9.05	180.68	[1]
170	Makui333	Makui	F	B	15.58	15.40	277.29	[1]
171	Mazekh10	Mazekh	F	B	18.36	18.21	251.20	[1]
172	Mbororo12	Mbororo	M	B	17.57	9.30	373.90	[1]
173	Moghani326	Moghani	F	A	14.81	14.63	122.28	[1]
174	Mossi99	Mossi	F	B	17.25	17.18	87.04	[1]
175	Ouessant0331	Ouessant	F	B	23.73	23.32	953.80	[1]
176	Ouessant24802	Ouessant	F	B	22.16	21.76	228.08	[1]
177	Ouessant24946	Ouessant	F	B	21.19	20.88	406.39	[1]
178	Ouessant24947	Ouessant	F	B	23.66	23.29	436.77	[1]
179	Ouessant24948	Ouessant	F	B	21.73	21.33	451.75	[1]
180	Ouessant24949	Ouessant	F	B	21.98	21.84	341.96	[1]
181	Ouessant24951	Ouessant	F	B	21.86	21.45	480.78	[1]
182	Ouessant24953	Ouessant	F	B	22.49	22.54	505.69	[1]
183	Ouessant72647	Ouessant	M	B	23.49	12.72	644.75	[1]
184	Ouessant81037	Ouessant	F	B	22.05	21.61	402.67	[1]

Appendix 2. Continued (7/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
185	SRR12396859	Baluchi	M	A	16.02	8.76	131.61	[2]
186	SRR12396860	Pag_Island	M	B	17.23	9.32	191.69	[2]
187	SRR12396861	Mossi	M	B	19.52	10.56	266.72	[2]
188	SRR12396862	Ossimi	M	B	17.41	9.43	260.18	[2]
189	SRR12396863	Ossimi	M	B	18.26	9.87	342.46	[2]
190	SRR12396864	Pag_Island	M	B	16.94	9.14	155.87	[2]
191	SRR12396865	Helsing	M	B	16.95	9.19	371.60	[2]
192	SRR12396866	Norwegian_White	M	B	17.30	9.36	221.30	[2]
193	SRR12396867	Norwegian_White	M	B	15.89	8.60	192.96	[2]
194	SRR12396868	Norwegian_White	M	B	15.88	8.59	224.74	[2]
195	SRR12396869	Feral	M	B	16.31	8.84	272.03	[2]
196	SRR12396870	Feral	M	B	17.55	9.53	533.89	[2]
197	SRR12396871	Feral	M	B	15.46	8.41	313.12	[2]
198	SRR12396872	Grey_Tronder	M	B	15.93	8.65	367.78	[2]
199	SRR12396873	Old_Spael	M	B	16.62	9.06	310.78	[2]
200	SRR12396874	Mossi	M	B	17.90	9.74	305.55	[2]
201	SRR12396875	Old_Spael	M	B	17.89	9.73	348.88	[2]
202	SRR12396876	Old_Spael	M	B	15.28	8.31	301.14	[2]
203	SRR12396877	Carpathian_Mountain	M	B	18.78	10.08	225.74	[2]
204	SRR12396878	Olkuska	M	B	17.15	9.23	524.70	[2]
205	SRR12396879	Zwartbles	M	A	16.35	8.91	250.46	[2]
206	SRR12396884	Mbororo	M	B	16.71	9.01	341.39	[2]
207	SRR12396888	Santa_Ines	M	B	17.06	8.98	297.66	[2]
208	SRR12396889	Brazilian_Bergamasca	M	B	18.47	9.16	100.61	[2]
209	SRR12396890	Brazilian_Bergamasca	M	B	17.91	9.47	117.42	[2]
210	SRR12396891	Romanov	M	A	18.07	9.62	123.65	[2]
211	SRR12396892	Kuibyshev	M	A	16.69	8.99	333.10	[2]
212	SRR12396893	Kuibyshev	M	A	15.70	8.43	269.09	[2]
213	SRR12396894	Kuibyshev	M	A	16.34	8.83	366.26	[2]
214	SRR12396895	Sahelian	M	B	17.46	9.41	339.42	[2]
215	SRR12396896	Dagestan_Mountain_Merino	M	N/A	17.88	9.67	253.38	[2]
216	SRR12396897	Caucasian	M	B	15.69	8.47	277.46	[2]

Appendix 2. Continued (8/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
217	SRR12396898	Pramenka	M	A	17.39	9.29	228.62	[2]
218	SRR12396899	Pramenka	M	A	16.49	8.90	330.21	[2]
219	SRR12396900	Leccese	M	B	19.61	10.52	241.17	[2]
220	SRR12396901	Leccese	M	B	17.89	9.66	246.55	[2]
221	SRR12396902	Altamura	M	B	17.82	9.41	236.76	[2]
222	SRR12396903	Altamura	M	B	20.47	11.07	271.26	[2]
223	SRR12396904	White_Mountain	M	B	19.11	10.28	421.64	[2]
224	SRR12396905	Rhoen	M	A	19.01	10.23	242.72	[2]
225	SRR12396906	Sahelian	M	B	18.80	10.09	289.63	[2]
226	SRR12396907	Rhoen	M	B	18.75	10.14	304.88	[2]
227	SRR12396908	Rhoen	M	B	16.33	8.80	262.44	[2]
228	SRR12396909	Bentheimer	M	B	20.48	11.10	298.68	[2]
229	SRR12396910	Ripollesa	M	B	17.70	9.62	25.36	[2]
230	SRR12396911	North_Ronaldsay	M	B	16.78	8.96	136.28	[2]
231	SRR12396912	Dalmatian	M	B	17.93	9.62	193.22	[2]
232	SRR12396913	Dalmatian	M	B	17.88	9.60	196.29	[2]
233	SRR12396914	Dalmatian	M	B	17.32	9.31	175.46	[2]
234	SRR12396915	Pag_Island	M	B	19.38	10.50	210.35	[2]
235	SRR12396916	Baidarak	M	A	18.32	9.86	325.24	[2]
236	SRR12396917	Baidarak	M	A	17.95	9.62	317.62	[2]
237	SRR12396918	Baidarak	M	A	17.59	9.47	423.82	[2]
238	SRR12396919	Brazilian_Somali	F	B	14.70	14.52	136.04	[2]
239	SRR12396920	Brazilian_Somali	F	B	18.78	18.62	149.56	[2]
240	SRR12396921	Xisqueta	F	B	16.86	16.91	32.63	[2]
241	SRR12396922	Brown_Mountain	F	B	17.36	16.85	344.86	[2]
242	SRR12396923	Djallonke	M	B	17.81	9.81	331.87	[2]
243	SRR12396924	Hampshire	F	B	17.79	17.42	166.84	[2]
244	SRR12396925	Romney	F	B	17.49	17.12	205.89	[2]
245	SRR12396926	Welsh_Mountain	F	B	16.03	15.89	294.55	[2]
246	SRR12396927	Helsing	F	B	17.29	17.31	329.33	[2]
247	SRR12396928	Helsing	F	B	16.86	16.64	249.21	[2]
248	SRR12396929	Jaffna_Local	F	A	12.21	11.81	378.12	[2]

Appendix 2. Continued (9/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
249	SRR12396930	Thalli	F	A	18.11	17.30	273.92	[2]
250	SRR12396931	Karkul	F	A	18.01	17.86	490.16	[2]
251	SRR12396932	Karakul	F	C	17.34	17.24	145.05	[2]
252	SRR12396933	Karakul	F	B	16.34	16.33	156.43	[2]
253	SRR12396934	Djallonke	M	B	19.02	10.30	334.49	[2]
254	SRR12396935	Karakul	F	A	16.70	16.44	103.34	[2]
255	SRR12396936	GreyShiraz	F	A	16.88	16.83	257.82	[2]
256	SRR12396937	Baluchi	F	A	17.23	17.26	163.70	[2]
257	SRR12396939	Wadi	F	A	26.51	25.57	380.94	[2]
258	SRR12396940	Wadi	F	A	25.80	26.20	1063.11	[2]
259	SRR12396941	Wadi	F	A	21.83	22.12	1162.93	[2]
260	SRR12396942	Diqing	F	B	17.86	17.30	722.33	[2]
261	SRR12396943	Diqing	F	A	16.31	17.02	162.27	[2]
262	SRR12396944	Lop	F	A	18.12	18.05	238.70	[2]
263	SRR12396945	Mbororo	M	B	17.57	9.30	373.90	[2]
264	SRR12396946	Lop	F	A	13.32	12.71	196.01	[2]
265	SRR12396947	Rahmani	F	B	14.32	13.84	174.47	[2]
266	SRR12396948	Cameroon	F	B	19.51	19.02	261.63	[2]
267	SRR12396949	Cameroon	F	B	18.31	18.05	277.99	[2]
268	SRR12396952	Shal	M	A	14.15	7.70	395.04	[2]
269	SRR12396953	Shal	M	A	20.27	10.91	435.82	[2]
270	SRR12396954	Barki	M	A	20.40	11.06	466.34	[2]
271	SRR12396955	Shal	M	A	16.54	8.99	262.14	[2]
272	SRR12396956	Afshari	M	B	17.56	9.54	294.38	[2]
273	SRR12396957	Afshari	M	A	17.34	9.39	328.65	[2]
274	SRR12396958	Afshari	M	C	16.79	9.04	308.74	[2]
275	SRR12396959	Ghezel	M	B	16.72	9.09	171.63	[2]
276	SRR12396960	Ghezel	M	C	16.31	8.80	133.88	[2]
277	SRR12396961	GreyShiraz	M	B	16.42	8.92	331.98	[2]
278	SRR12396962	GreyShiraz	M	C	15.62	8.49	252.61	[2]
279	SRR12396963	Awassi	M	C	16.08	8.71	203.00	[2]
280	SRR12396964	Jaffna_Local	M	B	17.62	9.56	702.00	[2]

Appendix 2. Continued (10/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
281	SRR12396965	Barki	M	B	18.04	9.75	449.45	[2]
282	SRR12396966	Nellore	M	B	19.68	10.74	199.13	[2]
283	SRR12396967	Thalli	M	B	17.27	9.09	242.52	[2]
284	SRR12396968	Thalli	M	B	18.27	9.86	317.96	[2]
285	SRR12396969	Karkul	M	B	18.18	9.45	260.90	[2]
286	SRR12396970	Karkul	M	B	17.63	9.05	180.68	[2]
287	SRR12396971	Kachi	M	C	15.38	7.75	198.03	[2]
288	SRR12396972	Kage	M	B	17.30	9.37	169.90	[2]
289	SRR12396973	Kage	M	B	17.68	10.34	175.00	[2]
290	SRR12396974	Bozakh	M	B	15.99	8.61	196.63	[2]
291	SRR12396975	Bozakh	M	B	16.72	9.05	213.28	[2]
292	SRR12396976	Rahmani	M	B	17.51	9.49	416.52	[2]
293	SRR12396977	Khotont	M	B	15.02	8.16	187.25	[2]
294	SRR12396978	Tamir	M	B	17.12	9.27	180.39	[2]
295	SRR12396979	Barga	M	B	17.09	9.28	348.11	[2]
296	SRR12396980	Barga	M	C	16.09	8.75	295.10	[2]
297	SRR12396981	Bayad	M	B	15.06	8.15	197.33	[2]
298	SRR12396982	Ujimqin	M	B	21.41	11.72	254.69	[2]
299	SRR12396983	Ujimqin	M	B	21.45	11.92	220.92	[2]
300	SRR12396984	Transbaikal_Finewool	M	A	16.78	9.08	350.11	[2]
301	SRR12396985	Transbaikal_Finewool	M	B	18.36	9.92	272.88	[2]
302	SRR12396986	Transbaikal_Finewool	M	A	17.47	9.44	317.16	[2]
303	SRR12396987	Rahmani	M	B	17.04	9.24	321.09	[2]
304	SRR12396988	Texel	M	B	18.10	9.69	186.78	[2]
305	SRR12396989	Texel	M	B	19.17	10.39	387.88	[2]
306	SRR12396990	Texel	M	B	16.63	8.77	147.08	[2]
307	SRR12396991	Hampshire	M	B	12.50	6.63	146.46	[2]
308	SRR12396992	Romney	M	A	17.07	9.10	301.61	[2]
309	SRR12396993	Djallonke	M	B	17.27	9.30	197.25	[2]
310	SRR12396994	Sumavska	M	B	19.02	10.24	367.09	[2]
311	SRR12396995	Valachian	M	B	16.06	9.11	390.75	[2]
312	SRR12396996	Valachian	M	B	21.17	11.25	385.45	[2]

Appendix 2. Continued (11/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
313	SRR12396997	Taihang_Fur	M	A	17.78	9.63	224.43	[2]
314	SRR12396998	Hanzhong	M	B	15.45	8.73	128.36	[2]
315	SRR12396999	Tengchong	M	A	16.94	9.35	155.48	[2]
316	SRR12397000	Tengchong	M	B	17.49	9.61	178.71	[2]
317	SRR12397001	Tengchong	M	A	13.26	7.39	124.31	[2]
318	SRR12397002	Tibetan	M	A	17.82	9.75	189.79	[2]
319	SRR12397003	Tibetan	M	A	17.35	9.64	167.31	[2]
320	SRR12397004	Libyan_Barbar	M	B	17.66	9.74	561.33	[2]
321	SRR12397005	Tibetan	M	A	16.49	9.05	169.62	[2]
322	SRR12397006	Tibetan	M	B	17.52	9.76	110.54	[2]
323	SRR12397007	Tibetan	M	C	19.21	10.42	395.04	[2]
324	SRR12397008	Tibetan	M	A	18.81	10.21	245.44	[2]
325	SRR12397009	Tibetan	M	A	16.34	8.92	298.99	[2]
326	SRR12397010	Tibetan	M	A	17.81	9.66	255.47	[2]
327	SRR12397011	Tibetan	M	A	16.93	9.18	303.07	[2]
328	SRR12397012	Lop	M	A	18.14	10.04	158.87	[2]
329	SRR12397013	Kazakh_Finewool	M	A	18.20	9.66	507.19	[2]
330	SRR12397014	Kazakh_Finewool	M	B	17.87	9.46	419.48	[2]
331	SRR12397015	Barki	M	B	17.35	9.37	276.64	[2]
332	SRR12397016	Degeres_Mutton-Wool	M	B	18.76	9.81	476.96	[2]
333	SRR12397017	Degeres_Mutton-Wool	M	A	18.39	9.61	392.49	[2]
334	SRR12397018	Degeres_Mutton-Wool	M	A	17.14	9.23	118.11	[2]
335	SRR12397019	Moghani	M	B	16.54	8.94	224.43	[2]
336	SRR12397020	Moghani	M	B	16.66	8.99	191.67	[2]
337	SRR12397021	Makui	M	B	16.33	8.85	227.39	[2]
338	SRR12397022	Baluchi	M	A	13.73	7.49	97.56	[2]
339	Sahelian163	Sahelian	F	B	16.82	16.63	201.75	[1]
340	Shal229	Shal	M	A	20.27	10.91	435.82	[1]
341	Shetland23342	Shetland	F	B	21.63	21.15	209.12	[1]
342	Shetland23343	Shetland	F	B	22.13	21.93	266.46	[1]
343	Shetland23344	Shetland	F	B	22.19	21.75	156.87	[1]
344	Shetland23345	Shetland	F	B	21.50	21.16	533.81	[1]

Appendix 2. Continued (12/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
345	Shetland23346	Shetland	M	B	21.99	11.94	453.47	[1]
346	Shetland23347	Shetland	M	B	22.00	12.01	364.64	[1]
347	Shetland23354	Shetland	M	A	21.90	11.61	241.13	[1]
348	Sishui32	Sishui_Fur	M	A	21.37	11.51	481.66	[1]
349	Sishui35	Sishui_Fur	M	B	20.51	11.27	408.62	[1]
350	Sishui37	Sishui_Fur	M	A	21.31	11.50	219.48	[1]
351	Sishui39	Sishui_Fur	M	B	20.85	11.23	508.60	[1]
352	Sishui40	Sishui_Fur	M	A	20.91	11.29	346.37	[1]
353	Sishui41	Sishui_Fur	M	B	21.20	11.49	273.70	[1]
354	Sishui43	Sishui_Fur	M	B	21.28	11.45	425.30	[1]
355	Sishui44	Sishui_Fur	M	A	20.92	11.26	325.31	[1]
356	Sishui45	Sishui_Fur	M	B	21.46	11.59	304.91	[1]
357	Sishui46	Sishui_Fur	M	A	21.12	11.39	312.13	[1]
358	Solognote24727	Solognote	F	B	22.52	21.98	794.68	[1]
359	Solognote24728	Solognote	M	B	23.56	12.56	796.12	[1]
360	Solognote24729	Solognote	M	B	24.61	13.18	523.90	[1]
361	Solognote24730	Solognote	F	B	25.05	24.77	866.46	[1]
362	Solognote24731	Solognote	M	B	23.00	12.48	499.42	[1]
363	Solognote24732	Solognote	F	B	22.75	22.31	652.57	[1]
364	Solognote24733	Solognote	F	B	23.18	22.92	581.45	[1]
365	Solognote24734	Solognote	F	B	23.34	22.96	729.92	[1]
366	Solognote24735	Solognote	F	B	23.18	22.78	755.83	[1]
367	Solognote24736	Solognote	M	B	23.24	12.43	502.31	[1]
368	Suffolk193	Suffolk	F	A	22.64	21.35	2226.53	[1]
369	Suffolk194	Suffolk	F	A	24.64	24.03	871.23	[1]
370	Suffolk195	Suffolk	F	B	25.28	25.23	943.25	[1]
371	Suffolk197	Suffolk	F	A	21.87	22.50	500.65	[1]
372	Suffolk198	Suffolk	F	B	22.91	22.92	964.04	[1]
373	Suffolk201	Suffolk	F	C	25.13	25.14	838.26	[1]
374	Suffolk203	Suffolk	F	A	22.80	21.74	911.75	[1]
375	Suffolk204	Suffolk	F	A	22.35	22.92	336.68	[1]
376	Suffolk205	Suffolk	F	B	23.21	22.84	709.86	[1]

Appendix 2. Continued (13/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
377	Suffolk206	Suffolk	F	A	22.82	22.57	798.97	[1]
378	Tan78	Tan	F	A	22.42	22.16	489.55	[1]
379	Tan79	Tan	F	A	23.06	22.75	748.75	[1]
380	Tan80	Tan	F	A	22.72	22.59	635.54	[1]
381	Tan82	Tan	F	C	22.61	22.39	668.57	[1]
382	Tan86	Tan	F	B	23.76	23.66	853.29	[1]
383	Tan87	Tan	F	B	22.22	21.95	720.27	[1]
384	Tan88	Tan	F	A	23.83	23.73	851.88	[1]
385	Tan89	Tan	F	A	22.21	22.14	721.16	[1]
386	Tan90	Tan	F	A	22.75	22.57	729.56	[1]
387	Tan91	Tan	F	B	23.44	23.23	629.08	[1]
388	Uda3	Uda	F	B	18.56	18.32	259.60	[1]
389	Urial6	Urial	M	N/A	16.17	8.90	302.39	[2]
390	Urial87	Urial	F	N/A	10.00	9.22	312.18	[3]
391	Urial89	Urial	M	N/A	12.50	6.52	774.30	[3]
392	Urial90	Urial	M	N/A	12.92	6.63	783.65	[3]
393	Urial91	Urial	M	N/A	11.03	5.96	1165.03	[3]
394	Wadi102	Wadi	F	B	23.56	23.18	1016.50	[1]
395	Wadi109	Wadi	F	C	22.04	22.43	983.97	[1]
396	Wadi112	Wadi	F	A	18.53	18.61	903.58	[1]
397	Wadi123	Wadi	F	A	17.78	17.87	782.63	[1]
398	Wadi14	Wadi	F	C	16.37	16.39	390.29	[1]
399	Wadi148	Wadi	F	A	21.43	21.79	377.67	[1]
400	Wadi149	Wadi	F	B	24.43	25.33	578.07	[1]
401	Wadi163	Wadi	F	D	23.74	23.36	473.95	[1]
402	Wadi53	Wadi	F	C	20.00	19.80	690.67	[1]
403	Wadi60	Wadi	F	A	18.45	18.35	270.33	[1]
404	Waggir1	Waggir	M	A	24.36	13.36	290.89	[1]
405	Waggir10	Waggir	M	C	24.10	13.04	281.67	[1]
406	Waggir11N	Waggir	F	B	26.92	26.43	373.77	[1]
407	Waggir2N	Waggir	M	A	22.83	12.20	366.40	[1]
408	Waggir3N	Waggir	F	A	25.96	25.43	362.21	[1]

Appendix 2. Continued (14/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
409	Waggir5N	Waggir	F	B	26.26	25.68	264.50	[1]
410	Waggir6N	Waggir	F	A	21.97	21.50	369.26	[1]
411	Waggir7N	Waggir	F	A	24.51	23.87	365.08	[1]
412	Waggir8N	Waggir	F	A	25.40	24.88	344.13	[1]
413	Waggir9N	Waggir	F	A	26.43	25.96	416.04	[1]
414	WestAfricanDwarf16	West_African_Dwarf	F	B	19.37	19.22	294.07	[1]
415	Yankasa5	Yankasa	F	B	16.07	16.02	293.55	[1]
416	SRR3022928.SG	Valley_Tibetan.SG	M	A	5.17	5.09	51.19	[4]
417	SRR3023068.SG	Valley_Tibetan.SG	M	A	5.28	5.25	58.76	[4]
418	SRR3023075.SG	Hu.SG	M	B	5.29	2.90	86.72	[4]
419	SRR3023084.SG	Hu.SG	M	N/A	6.25	3.38	81.87	[4]
420	SRR3023153.SG	Hu.SG	M	A	6.03	3.26	74.31	[4]
421	SRR2973610.SG	Australian_Merino.SG	F	A	6.17	6.12	85.53	[4]
422	SRR2973635.SG	Australian_Merino.SG	F	B	5.19	5.23	131.76	[4]
423	SRR2976378.SG	Australian_Merino.SG	F	A	5.43	5.41	66.32	[4]
424	SRR2976380.SG	Australian_Merino.SG	F	B	5.43	5.48	99.61	[4]
425	SRR2976383.SG	Australian_Merino.SG	F	A	5.71	5.67	95.13	[4]
426	SRR2976535.SG	Bayinbuluke.SG	M	A	5.17	2.80	91.57	[4]
427	SRR2976538.SG	Bayinbuluke.SG	F	A	5.20	5.22	105.99	[4]
428	SRR2976541.SG	Bayinbuluke.SG	M	A	5.13	2.69	92.35	[4]
429	SRR2976544.SG	Bayinbuluke.SG	M	A	3.93	0.00	0.00	[4]
430	SRR2976545.SG	Bayinbuluke.SG	M	B	5.83	3.03	103.25	[4]
431	SRR2976708.SG	Bayinbuluke.SG	M	A	5.83	5.68	75.90	[4]
432	SRR2976711.SG	Bayinbuluke.SG	F	A	5.23	5.16	79.91	[4]
433	SRR2976713.SG	Bayinbuluke.SG	F	C	9.61	9.72	224.67	[4]
434	SRR2976722.SG	Bayinbuluke.SG	F	A	4.96	4.81	83.80	[4]
435	SRR3184728.SG	Small-tailed_Han.SG	F	B	5.74	5.93	171.73	[4]
436	SRR3184734.SG	Small-tailed_Han.SG	F	A	5.12	5.06	65.58	[4]
437	SRR3184746.SG	Small-tailed_Han.SG	M	A	5.36	5.36	67.89	[4]
438	SRR3184750.SG	Small-tailed_Han.SG	M	A	5.09	5.02	22.28	[4]
439	SRR2992575.SG	Cele_Black.SG	M	C	6.20	6.25	152.41	[4]

Appendix 2. Continued (15/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
440	SRR2992628.SG	Valley_Tibetan.SG	M	A	5.42	2.84	119.50	[4]
441	SRR2992761.SG	Valley_Tibetan.SG	F	A	5.72	3.00	90.49	[4]
442	SRR2992782.SG	Valley_Tibetan.SG	F	A	5.28	2.67	60.95	[4]
443	SRR2997388.SG	Valley_Tibetan.SG	F	C	5.32	2.72	108.21	[4]
444	SRR2997424.SG	Valley_Tibetan.SG	F	A	6.35	6.31	110.93	[4]
445	SRR2997425.SG	Valley_Tibetan.SG	F	A	5.56	5.36	44.01	[4]
446	SRR2980596.SG	Bayinbuluke.SG	F	B	5.70	5.60	117.85	[4]
447	SRR2980600.SG	Cele_Black.SG	M	A	5.33	2.90	96.20	[4]
448	SRR2980624.SG	Cele_Black.SG	M	B	6.04	3.30	81.94	[4]
449	SRR2982559.SG	Cele_Black.SG	M	A	5.98	3.23	83.88	[4]
450	SRR2982564.SG	Cele_Black.SG	M	B	5.97	3.22	101.09	[4]
451	SRR2982588.SG	Cele_Black.SG	M	B	5.52	2.97	88.92	[4]
452	SRR2984616.SG	Cele_Black.SG	F	B	5.94	5.90	138.74	[4]
453	SRR2984617.SG	Cele_Black.SG	F	A	5.96	5.99	162.43	[4]
454	SRR2984620.SG	Cele_Black.SG	F	A	5.67	5.64	122.23	[4]
455	SRR3193935.SG	Prairie_Tibetan.SG	M	A	5.89	3.21	126.05	[4]
456	SRR3193937.SG	Prairie_Tibetan.SG	M	C	5.37	2.74	82.38	[4]
457	SRR3193947.SG	Prairie_Tibetan.SG	M	C	5.95	3.16	104.68	[4]
458	SRR3193948.SG	Prairie_Tibetan.SG	F	A	5.26	2.75	68.26	[4]
459	SRR3193955.SG	Prairie_Tibetan.SG	F	B	5.08	5.07	76.56	[4]
460	SRR3087354.SG	Wuzhumuqin.SG	F	A	5.51	5.54	144.41	[4]
461	SRR3087367.SG	Wuzhumuqin.SG	M	C	8.68	8.77	205.71	[4]
462	SRR3087377.SG	Wuzhumuqin.SG	M	A	5.58	5.63	109.17	[4]
463	SRR3087381.SG	Wuzhumuqin.SG	M	A	4.99	4.70	117.28	[4]
464	SRR3087385.SG	Wuzhumuqin.SG	M	A	5.13	5.14	120.09	[4]
465	SRR2968869.SG	Australian_Merino.SG	F	B	5.05	2.74	90.05	[4]
466	SRR2970520.SG	Australian_Merino.SG	F	B	7.35	3.96	165.67	[4]
467	SRR2970524.SG	Australian_Merino.SG	F	A	5.23	2.83	106.04	[4]
468	SRR2970616.SG	Australian_Merino.SG	F	A	5.38	2.91	83.23	[4]
469	SRR2970619.SG	Australian_Merino.SG	F	B	5.67	3.08	97.90	[4]
470	SRR3471444.SG	Prairie_Tibetan.SG	F	A	5.47	5.38	100.55	[4]
471	SRR3471448.SG	Prairie_Tibetan.SG	F	A	6.44	6.36	107.98	[4]

Appendix 2. Continued (16/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
472	SRR3471450.SG	Prairie_Tibetan.SG	M	A	5.32	5.18	79.81	[4]
473	SRR3471451.SG	Prairie_Tibetan.SG	M	B	5.96	5.90	136.24	[4]
474	SRR3471453.SG	Prairie_Tibetan.SG	M	A	5.45	5.43	103.41	[4]
475	SRR3054904.SG	Oula.SG	M	A	5.48	2.98	68.23	[4]
476	SRR3054905.SG	Oula.SG	M	B	5.97	5.99	97.21	[4]
477	SRR3054976.SG	Oula.SG	M	C	5.50	5.22	98.58	[4]
478	SRR3055048.SG	Oula.SG	F	A	5.83	5.82	76.87	[4]
479	SRR3055050.SG	Oula.SG	F	A	5.70	5.71	67.26	[4]
480	SRR3055051.SG	Oula.SG	F	A	5.39	5.29	83.26	[4]
481	SRR3061456.SG	Tan.SG	M	A	5.30	3.02	61.30	[4]
482	SRR3061457.SG	Tan.SG	M	A	6.16	3.34	58.56	[4]
483	SRR3061462.SG	Tan.SG	M	A	5.90	3.12	61.20	[4]
484	SRR3061463.SG	Tan.SG	M	A	6.07	3.28	41.59	[4]
485	SRR3061464.SG	Tan.SG	F	A	5.03	2.80	50.81	[4]
486	SRR3036475.SG	Hu.SG	F	B	5.74	3.06	95.54	[4]
487	SRR3036476.SG	Hu.SG	F	B	5.43	2.88	62.41	[4]
488	SRR3036477.SG	Hu.SG	F	B	6.04	6.05	123.67	[4]
489	SRR3036494.SG	Hu.SG	F	A	5.26	5.19	78.48	[4]
490	SRR3036496.SG	Hu.SG	F	B	5.62	5.71	172.86	[4]
491	SRR3043125.SG	Hu.SG	F	A	6.02	5.79	54.37	[4]
492	SRR3043352.SG	Hu.SG	M	B	5.69	5.73	146.11	[4]
493	SRR3043353.SG	Oula.SG	M	A	6.08	3.31	110.76	[4]
494	SRR3043356.SG	Oula.SG	M	A	4.93	2.61	132.76	[4]
495	SRR3043358.SG	Oula.SG	M	A	5.22	2.85	18.81	[4]
496	SRR3043360.SG	Oula.SG	M	A	5.51	3.07	108.49	[4]
497	SRR3180855.SG	Small-tailed_Han.SG	F	A	5.49	2.97	41.21	[4]
498	SRR3180874.SG	Small-tailed_Han.SG	F	C	5.70	3.10	31.67	[4]
499	SRR3180876.SG	Small-tailed_Han.SG	F	B	5.48	2.95	96.58	[4]
500	SRR3180888.SG	Small-tailed_Han.SG	F	C	5.50	2.99	74.34	[4]
501	SRR3180890.SG	Small-tailed_Han.SG	M	C	5.57	3.01	37.07	[4]
502	SRR3065079.SG	Tan.SG	M	A	5.33	2.93	57.13	[4]
503	SRR3065080.SG	Tan.SG	M	A	4.98	2.73	26.33	[4]

Appendix 2. Continued (17/17).

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
504	SRR3065081.SG	Tan.SG	M	B	5.19	5.00	42.89	[4]
505	SRR3065082.SG	Tan.SG	M	B	5.58	5.58	60.95	[4]
506	SRR3065083.SG	Tan.SG	F	B	5.22	5.32	62.60	[4]
507	SRR3068438.SG	Wuzhumuqin.SG	F	A	5.83	3.17	97.59	[4]
508	SRR3068457.SG	Wuzhumuqin.SG	F	A	5.56	3.06	109.92	[4]
509	SRR3068467.SG	Wuzhumuqin.SG	F	B	5.42	2.96	119.74	[4]
510	SRR3068471.SG	Wuzhumuqin.SG	F	A	5.84	3.18	109.33	[4]
511	SRR3068478.SG	Wuzhumuqin.SG	F	B	5.65	3.08	89.70	[4]
512	SRR2992573.SG	Cele_Black.SG	F	C	5.13	5.11	177.13	[4]
513	SRR2997142.SG	Valley_Tibetan.SG	F	A	5.30	2.83	95.07	[4]
514	SRR2997390.SG	Valley_Tibetan.SG	F	A	5.86	5.49	66.04	[4]

Ref [1] : Study Accession (NCBI / PRJNA624020)

Ref [2] : Study Accession (NCBI / PRJNA645671)

Ref [3] : Study Accession (NCBI / PRJEB5463)

Ref [4] : Study Accession (NCBI / PRJNA304478)

Appendix 3. Summary statistics of previously published ancient individuals used in this study.

No.	Sample ID	Population	Sex	MT Hap	Coverage			Ref
					Auto	X	MT	
1	OBI013.A0101	Obishir5	F	A	0.0271	0.0255	0.49	[1]
2	OBI014.A0101	Obishir5	F	A	0.0168	0.0146	3.89	[1]
3	OBI017.A0101	Obishir5	F	N/A	0.0009	0.0007	0.08	[1]
4	OBI018.A0101	Obishir5	M	N/A	0.0237	0.0112	0.87	[1]
5	ERR5766949.SG	Iran_1600BP	M	C	0.0385	0.0195	5.49	[2]
6	ERR5766950.SG	Iran_1600BP	F	A	0.0686	0.0643	7.66	[2]
7	ERR5767136.SG*	Iran_1600BP	M	C	2.7674	1.5165	113.91	[2]
8	ERR3861592.SG	Anatolian_Neo	F	B	0.3560	0.3579	1.80	[3]
9	ERR3861593.SG	Anatolian_Neo	F	B	0.6387	0.6176	9.66	[3]
10	ERR3861594.SG	Anatolian_Neo	M	A	0.0631	0.0330	4.95	[3]
11	ERR3861595.SG	Anatolian_Neo	F	N/A	0.0466	0.0464	0.17	[3]

Ref [1] : Study Accession (NCBI / PRJEB41594)

Ref [2] : Study Accession (NCBI / PRJEB43881)

Ref [3] : Study Accession (NCBI / PRJEB36540)

No.	Sample ID	Covered in 7497K		Damage Pattern				Endogenous DNA %
		# of SNPs	% of SNPs	First base 5'	Second base 5'	First base 3'	Second base 3'	
1	OBI013.A0101	660,484	8.81	0.1569	0.0287	0.1147	0.0257	11.70
2	OBI014.A0101	423,168	5.64	0.0751	0.0138	0.0693	0.0165	10.44
3	OBI017.A0101	22,224	0.30	0.0698	0.0143	0.0680	0.0201	0.38
4	OBI018.A0101	612,799	8.17	0.1659	0.0178	0.1576	0.0238	10.12
5	ERR5766949.SG	294,736	3.93	0.0236	0.0187	0.0291	0.0200	86.52
6	ERR5766950.SG	500,795	6.68	0.0251	0.0198	0.0293	0.0202	45.99
7	ERR5767136.SG*	6,836,380	91.18	0.0100	0.0072	0.0099	0.0078	28.87
8	ERR3861592.SG	1,515,925	20.22	0.2780	0.1824	0.2556	0.1752	37.61
9	ERR3861593.SG	2,522,688	33.65	0.2618	0.1849	0.2519	0.1806	70.24
10	ERR3861594.SG	376,081	5.02	0.3405	0.2223	0.3366	0.2221	18.61
11	ERR3861595.SG	290,307	3.87	0.4712	0.2783	0.4903	0.2837	20.50

* Sample ERR5767136.SG had multiple sequencing runs and these data were merged in the data compilation step.