259. Deciphering the origins of Neolithic sheep from northern Iberian Peninsula

M. Serrano[1*], A. Fernández[1], C. González[1], J. Martínez-Blanch[2], M. Moreno-García[3], M. Rojo-Guerra[4] and J.H. Calvo[5]

[1]INIA-CSIC, Ctra. De la Coruña km 7,5, 28040 Madrid, Spain; [2]ADM Biopolis, Parque Científico Universitat de Agustín Escardino, 9, Carrer del Catedràtic Agustín Escardino Benlloch, 46980 Paterna, Valencia, Spain; [3]IH-CSIC, C. de Albasanz, 26, 28037 Madrid, Spain; [4]Universidad de Valladolid, Plaza del Campus s/n, 47011 Valladolid, Spain; [5]CITA-ARAID, Av. Montañana, 930, 50059 Zaragoza, Spain; malena@inia.es

Abstract

Domestication events in livestock can be studied by comparing modern DNA with ancient DNA (aDNA) from archaeological faunal remains. Investigating the phylogeny history of sheep remains from the Neolithic Spanish site of Cova de Els Trocs, provides an opportunity to gain some insights about their origins. In this work, aDNA recovered from 15 ovine specimens dated between 6000 to 4000 years BP were sequenced by whole genome sequencing (WGS), and mapped against the *O. aries* modern mitogenome. Complete mitogenomes were restored for 12 ancient samples. Bayesian Inference (BI) and Maximum Likelihood (ML) phylogenetic trees fitting the HKY+G+I model were performed including the mitogenomes of 12 ancient and 74 modern sheep and 4 wild species. All Neolithic samples belong to the mitochondrial haplotype B. Four ancient sheep constituted a clade (89%) with the Viena sheep breed (Karelia), suggesting a complex demographic history of their origins.

Introduction

Archaeological evidence points to the Neolithic as the moment in time of sheep introduction in Europe along the Mediterranean and Danube routes (Zeder 2017). Although mtDNA studies of worldwide sheep breeds have provided some clues about the history of sheep domestication and dispersion, questions such as population colonization, demographic history and patterns of gene flow to Europe, Asia and Africa (Lv *et al.* 2015) are yet unsolved. Current advances in paleo-genetics and aDNA recovering techniques have allowed deepening into the genetic origins and dispersion of sheep breeds by comparing DNA sequences of ancient specimens with those of modern breeds. In this work, aDNA from 15 *Ovis* recovered in the site Cova dels Trocs (6000 to 4000 years BP) were sequenced and mapped against the *O. aries* modern mitogenome. A phylogenetic analysis including complete mitogenomes from 12 Neolithic and 74 modern sheep, and 4 wild species were performed to elucidate the origin of ancient specimens.

Materials & methods

Ovine fossil material, samples processing and DNA extraction and purification. Cova de Els Trocs, situated at above 1,500 m.a.s.l., on the Iberian central Pyrenees, is one of the few sites with a fully documented Neolithic stratigraphic sequence spanning from the 6th to the last third of the 4th millennium cal BC (calibrated years before Christ). In this site, a high number sheep remains have been recovered, revealing the great involvement in herding activities by the human groups from this territory (Tejedor-Rodríguez *et al.* 2021). Fifteen sheep jaws and maxillary from Cova dels Trocs (Huesca Spain), which came from three different occupation phases: 5 from T1 (5800-6100 years BP), 5 from T2 (5500-5800 years BP) and 5 from T3 (3900-4600 years BP years), was the available material. To minimize contamination with modern DNA, two DNA extraction replicates were performed by using new equipment and reagents in two independent laboratories. One of the replicates was used for WGS, while the other one for Sanger sequencing. All steps such as bone/tooth cutting and surface sanding, grinding and solutions preparation, and DNA extraction in sterile conditions, were carried out in separate places and times. Teeth roots were

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_259, © M. Serrano *et al.* 2022

crushed for 90 seconds at maximum frequency within a ball mill with liquid nitrogen in a mill MM400 (Retsch, Germany). DNA extraction were performed following a modified protocol of Huel (Huel *et al.* 2012). DNA purification was carried out with QIAquick PCR purification kit (Qiagen Inc. USA).

DNA libraries and WGS. Libraries for next-generation sequencing (NGS) were built with the aDNA extracts using the Accel-NGS Methyl-Seq DNA Library Kit and Uniq Dual Indexing kit (Swift Biosciences) following manufacturer's protocol. This kit is specially indicated for sequencing aDNA when retention of fragments containing uracil nucleotides, resulting from damage, is not desired. PrePCR and PostPCR labours were conducted in different spaces to avoid contamination. In order to validate methodology, 4 ancient samples were repeated by using Meyer libraries (Meyer 2010). Sequencing was performed on the Illumina NextSeq 550 platform to obtain between 11 and 62 million 75nt paired-end reads per library.

SANGER sequencing. Twelve overlapping amplicons were amplified, assembling a total length of 4000 bp of the mtDNA after Sanger sequencing (positions 1 to 1,079, and 13,696 to 16,616 of the NC_001941.1 sheep mitochondrial genome). Amplicon sizes varied between 395 and 679 bp. As for WGS, PrePCR and PostPCR were performed in different labs.

Ancient DNA authentication; reads processing, mapping and variant calling. Sample reads were trimmed of adapter sequences and low quality bases by AdapterRemoval v.2.3.1 (Schubert *et al.* 2016) and mapped to the sheep mitochondrial genome (NC_001941.1) using BWA v. 0.7.17 (Li and Durbin 2009). Duplicates were removed with Picard Tools, and the genome coverage was estimated by Samtools v.1.10 (Li 2011). MapDamage 2.0 (Jónsson *et al.* 2013) was used to evaluate fragment lengths and nucleotide misincorporation patterns. Two different algorithms were used to produce the VCF files for mitochondrial positions: bcftools mpileup and GATK UnifiedGenotyper (DePristo *et al.* 2011). All the mutations reported were visually inspected in the Biomatters IGV software v2.3.66 (Robinson *et al.* 2011) to rule out artefacts due to misincorporations in low coverage regions. Finally, Samtools v1.10 was used to create the consensus sequence for all the samples.

Phylogenetic Analysis. Complete mtDNA sequences from Neolithic, modern and wild sheep were aligned using the Clustal Ω program. Evolutionary models were tested with MEGA-X. BI phylogenetic tree was constructed with MrBayes v3.1.2 using the Hasegawa-Kishino-Yano HKY+G+I model, running 4 simultaneous Markov chains for 20 million generations starting from a random tree. Sampling frequency was set to 1/500 discarding as burning the first 25% trees. Bootstrap values for the ML tree were generated with 1000 replicates using MEGA-X.

Results

Twelve of the 15 samples available yielded enough aDNA for sequencing. Table 1 shows average results of aDNA sequencing. MapDamage shows classical patterns of post-mortem damage in Neolithic samples, such as 5' C*T deamination and T missincorporations. However, 3' G*A deamination could not be detected due to the Adaptase technology used in the Accel-NGS Library Kit, which adds a low complexity polynucleotide tail to the 3' end of each fragment. The final mitochondrial fragment length of the Neolithic samples after correction and trimming was 16,616 bp (positions 1 to 16,616 of the sheep mitogenome).

Alignments of nucleotide sequences obtained by Sanger and WGS showed 100% identity, indicating that no contamination occurred in the aDNA extraction and amplification processes. Phylogenetic analysis (Figure 1) assigned the 12 ovine ancient samples to the mitochondrial haplotype B. No clustering was found between the aDNA samples based on their chronology. The distance matrix of Hamming dissimilarity showed that Neolithic sheep from Trocs were very close to Finnsheep, Lacaune, Rasa Aragonesa and Finn-

Table 1. Results of ancient DNA sequencing and mapping.

Parameter	No. paired-end sequences	Average Length	Average Quality	MT properly mapped	MT mean depth	% aMT recovered
Mean	16,569,276	149.9	36.1	2,660	33.6	0.017
SD	3,766,339	0.4	0.1	1,847	24.1	0.012
Min	11,360,788	148.8	35.9	550	6.3	0.002
Max	23,089,914	150.3	36.2	6,966	89.2	0.043

MT = mitochondrial; aMT = ancien mitochondrial

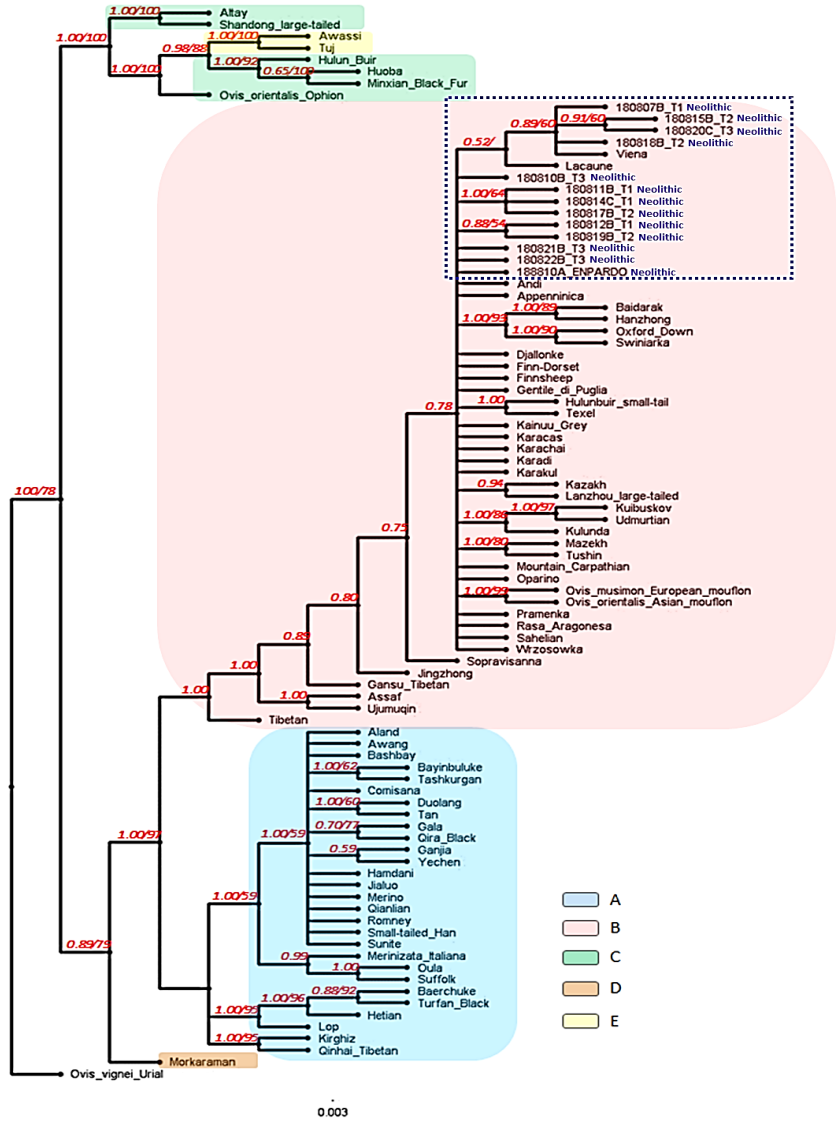

Figure 1. Phylogeny of Neolithic, modern and wild sheep inferred from 90 complete mitogenomes using BI and ML methods. On nodes: posterior probability/bootstrap values

Dorset breeds. Phylogenetic tree shows that 4 samples from Trocs formed a clade with the Viena sheep breed (Karelia, Russia).

Discussion

Accel-NGS Library Kit is suitable to work with aDNA, however its particular chemistry does not allow detecting in an optimal way the post-mortem damage patterns. However, by using Meyer libraries, both deamination patterns, 5' C*T and 3' G*A, could be detected. Ancient fragment lengths were longer than the average ~55 bp expected (Skoglund *et al.* 2012), probably due to the good preservation conditions of the Trocs cave. The small distance of all Neolithic samples from Trocs to Finnish breeds suggest a North European origin. Four Neolithic samples constituted a clade with an 89% of posterior probability with the Viena sheep, and ancient breed from Karelia, Russia (Meadows *et al.* 2005). In 2006, Tapio *et al.* found the haplotype A in the old native breeds of Nordic Countries and Russian Karelia, suggesting a previously unrecognized migration of sheep to northern Europe through Russia. This fact could indicate that some of the specimens found in Trocs, could have an origin in sheep populations that migrated to northern Europe, despite they exhibit the B haplotype.

References

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R. *et al.* (2011) Nat Genet. 43(5): 491-498. https://doi.org/10.1038/ng.806

Huel R., Amory S., Bilić A., Vidović S., Jasaragić E. *et al.* (2012) Methods Mol. Biol. 830: 185-198. https://doi.org/10.1007/978-1-61779-461-2_13

Jónsson H., Ginolhac A., Schubert M., Johnson P.L., Orlando L. (2013) Bioinformatics 29(13):1682-1684. https://doi-org/10.1093/bioinformatics/btt193

Li H. (2011) Bioinformatics 27(21) 2987-93. https://doi.org/10.1093/bioinformatics/btr509

Li H. and Durbin R. (2009) Bioinformatics, 25, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Lv F.H., Peng W.F., Yang J., Zhao Y.X., Li W.R. *et al.* (2015) Mol. Biol. Evol. 32(10): 2515-2533. https://doi.org/10.1093/molbev/msv139

Meadows J.R.S., Li K., Kantanen J., Tapio M., Sipos W. *et al.* (2005). Journal of Heredity 96(5): 494-501. https://doi.org/10.1093/jhered/esi100

Meyer M, Kircher M. (2010) Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot5448

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S. *et al.* (2011). Nature biotechnology, 29(1), 24–26. https://doi.org/10.1038/nbt.1754

Schubert M., Lindgreen S., Orlando L. (2016). BMC Research Notes, 9: 88. https://doi.org/10.1186/s13104-016-1900-2

Skoglund P., Malmström H., Raghavan M., Storå J., Hall P. *et al.* (2012) Science. 336 (6080): 466-9. https://doi.org/10.1126/science.1216304

Tapio M, Marzanov N, Ozerov M, Cinkulov M, Gonzarenko G *et al.* (2006) Mol. Biol. Evol. 23: 1776–1783. https://doi.org/10.1093/molbev/msl043

Tejedor-Rodríguez C., Moreno-García M., Tornero C., Hoffmann A., García-Martínez de Lagrán Í. *et al.* (2021) PlosOne 16, e0244139. https://doi.org/10.1371/journal.pone.0244139

Zeder M.A. (2017) Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In Boivin N., Crassard R. and Petraglia M. (Eds) Dispersal and Species Movement. Cambridge University Press. https://doi.org/10.1017/9781316686942.012