

Following the Hidden Footsteps of Past Human Migrations via the Goat Genome

A. Manunza^{1*}, J. Ramirez Diaz^{1,2}, T. Bobbo^{1,3}, P. Cozzi¹, B. Lazzari¹, F. Biscarini¹, S. Biffani¹, A.M. Johansson⁴, L. Grøva⁵, A. Stella¹

¹ Institute of agricultural biology and biotechnology, National Research Council, via Edoardo Bassini, 15 - 20133 Milan, Italy; ² Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122, Piacenza, Italy; ³ University of Milan, Via Celoria 2, Milan, Italy; ⁴Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, SE-75007 Uppsala, Sweden; ⁵Norwegian Institute of Bioeconomy Research, Gunnars vei 6, NO-6630 Tingvoll, Norway. * arianna.manunza@ibba.cnr.it

Abstract

The aim of this study was to assess the genetic diversity and the degree of introgression of foreign alleles into autochthonous goat genomes following putative land and maritime routes of human migrations. Genotypic data from sixteen populations were analysed for diversity statistics and admixture. Results show genetic fingerprints of old gene flow that could be related to past human expansions for explorations, raids or demographic reasons. In addition, this study could add missing pieces to indirectly recreate historically and still uncertain human land and maritime routes.

Introduction

Archaeological and historical sources have driven the reconstruction of human migrations. Recently, cutting-edge genomics and bioinformatics analysis has contributed to this field. Since their domestication, small ruminants constituted an unreplaceable source of food and ecosystem services for humans and their small size made them ideal for carrying them by land and sea, following human migratory routes (Amills *et al.*, 2017). A deeper analysis of literature (Colli *et al.* 2018, Cardoso *et al.*, 2018), investigating the genetic diversity in several breeds from the five continents, led us to catch subtle but intriguing similarities among geographically distant populations. In this study, we aimed to infer signs of gene flow following putative routes of human migrations by analysing the genomes of native modern goat populations. In particular, we focused on two prolonged migration events that occurred between the 500 BCE and the 900 CE: the Celtic expansion from Continental Europe and the Viking diaspora from Scandinavia (Viking Age, VA).

Materials & Methods

Our dataset includes data from Stella *et al.*, (2018): (ARR, BLB, OIG) = Ireland, (BAG, OEG) = UK, DNK = Denmark, FIN = Finland, ICL = Iceland, NLD = The Netherlands, NRW = Norwegian landrace from Tanzania; from Berg *et al.*, (2020) (SKO, SEL) = Norway; from Cortellari *et al.*, (2021) (ORO, NVE, VLS) = northern Italy; from Smarter project, <https://www.smarterproject.eu/> SWE = Sweden. Genotype data were from Illumina GoatSNP50 BeadChip (Tosser-Klopp *et al.*, 2014). Data were remapped against ARS1.2 assembly. Quality control was performed using PLINK v1.9 (Chang *et al.* 2015). We used Arlequin v.3.5 (Excoffier *et al.*, 2010) to calculate the pairwise *F*_{st} values and the SNPrelete R package (Zheng *et al.*, 2012) to perform the Principal Component Analysis (PCA). We inferred the estimation of individual ancestries using ADMIXTURE v.1.3 with K ranging from 2 to 30 (Alexander *et al.*, 2009). To make a comparative dataset considering the maternal contributions, we selected from GenBank the HV1 control regions (392 bp) of 63

domestic goats (Scandinavia, British Isles and northern Italy). A Maximum-Likelihood (ML) phylogeny analyses using Tamura-Nei and 1000 bootstrap was carried out with MEGA v.11 (Tamura *et al.*, 2021) and the resulting tree was drawn using iTOL v.6.4.3 program (Letunic and Bork, 2021).

Results

F-statistics highlighted a surprisingly low genetic diversity that characterizes some populations, considering a European continental context. F_{st} ranged from 0.071 to 0.621. The NVE breed shows the lowest F_{st} values in all pairwise comparisons and a close similarity with all populations from Scandinavia (above all DNK, 0.096) and secondarily with the British Isles (0.143-0.193). The ICL breed differentiates in a well-defined cluster (ADMIXTURE and PCA analysis). It presents F_{st} values a little lower of 0.3 with the Scandinavian breeds, definitely a lower value in comparison with British breeds (0.6) and slightly higher with Italian breeds (0.4). The breeds from the British Isles are quite similar to each other but show a lower genetic distance with DNK and NRW (0.159-0.208) followed by the Italian populations (0.184-0.217). The cross-validation method in the ADMIXTURE (Figure 1) analysis indicates $K=19$ as the most probable number of ancestries. Figure 1A shows a significant level of admixture in most of breeds, except for ICL, ARR, SKO and NLD. A small introgression level is detected among some populations from Scandinavia, British Isles and Italy. The genetic clustering analysis by PCA (Figure 2) groups the populations in three clusters: Icelandic breed is closer to Scandinavian breeds; British Isles forms a unique cluster with Italian breed near to NLD, whereas DNK is a bit far from the rest making a single cluster. Also, we can see a certain level of overlapping between NVE with BLB and among SKO, SEL and SWE (dotted circle in the Figure 2).

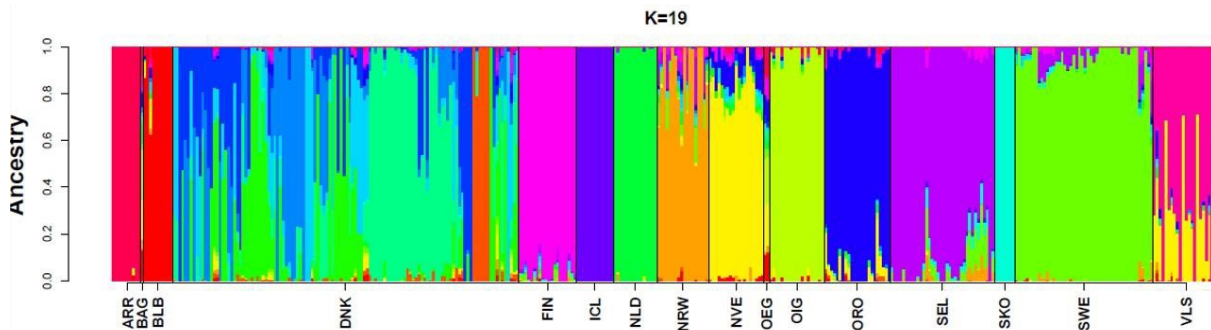


Figure 1. Individual ancestries inferred by ADMIXTURE analysis: different colours refer to different ancestries.

The control region tree depicts an interesting link across populations. Individuals are coloured depending on their geographical origin and the groups of interest are highlighted in different colours. Six well-established groups are identified, whose nodes are supported by a bootstrap value around 50% or higher. mtDNA analysis confirms a certain degree of past introgression between NVE and DNK, Norwegian with British Isles individuals (with historical samples in the same branches) and Icelandic with NVE. The ancient samples from the British Isles are also clustering together. H= historical and M=modern prefix ID of samples as in Cassidy *et al.* (2017).

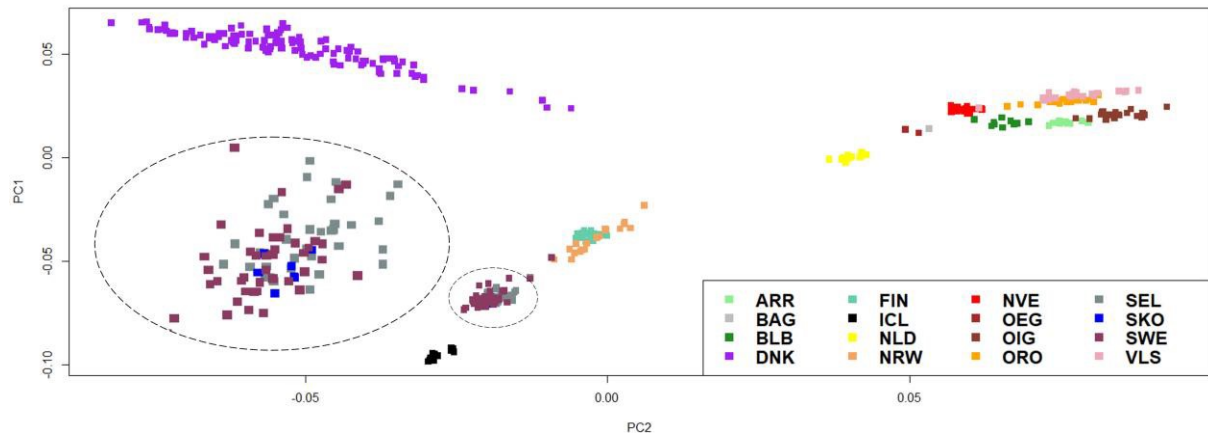


Figure 2. PCA analysis: the first two PCs explain the 10.3% of variance.

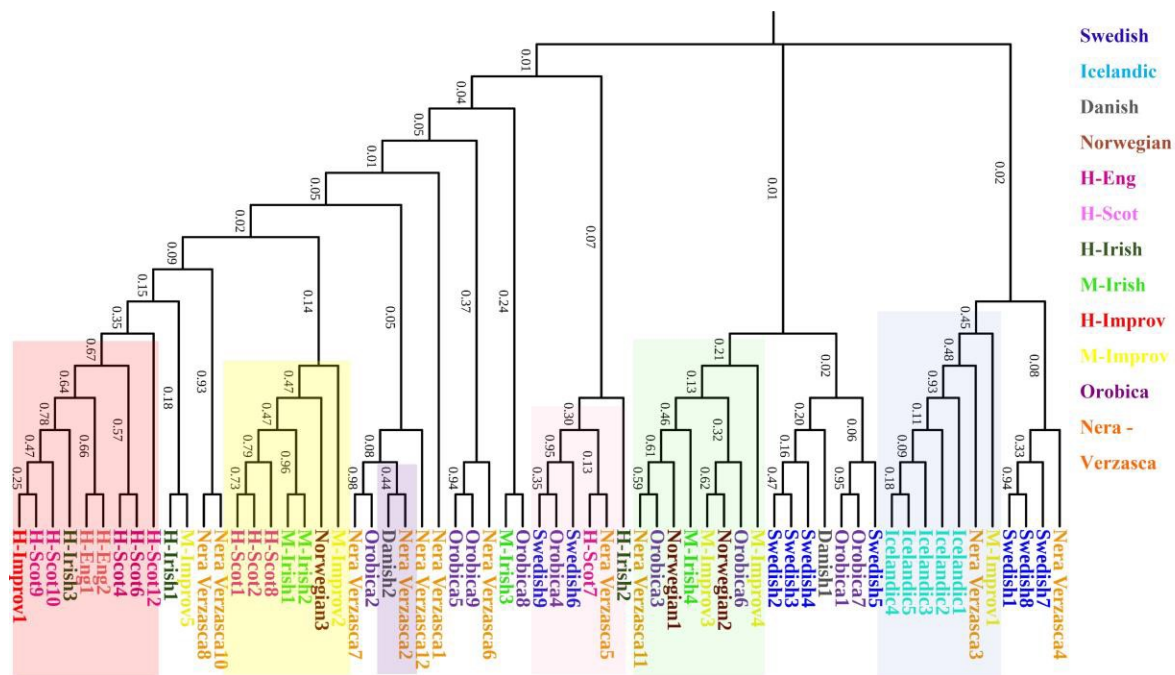


Figure 3. Phylogeographic analysis of 63 control regions of native goat breeds.

Discussion

All our findings suggest that crossing and introgression events among these native breeds occurred in the past, possibly following human migrations. The BLB breed for example, seems to have contributed to the genetic composition of both Italian and Scandinavian breeds, during two different and temporally separated contacts or through a slow but continued genetic exchange. We can speculate that the foreign fingerprint found in these populations is traceable back to the Celtic migration from central Europe towards Ireland and Britain and later with the Viking diaspora, spanning from the 500 BCE to the 900 CE. A previous study on the historical gene flow based on mtDNA of British Isles breeds (Cassidy *et al.*, 2017) found that ancient samples from the islands of Scotland clustered together with modern Irish goats and a Norwegian landrace goat. Another interesting grouping contains a modern British goat from central-east Ireland alongside Welsh, English and Icelandic goats. Both results suggest a genetic link with native goats from Scandinavian countries and settlements. In our

mtDNA analyses, we included more individuals from Scandinavian countries, Iceland and northern Italy. It is noteworthy that a deeper inspection of the inferred ancestries and the goat haplotypes distributions reveals additional similarity between Scandinavian breeds and the rest of breeds. This suggests a bidirectional introgression and that a common situation has emerged in the genetic composition of modern Scandinavian and British Isles human populations. In fact, two studies aiming to explore the genomic history of the Celtic insular origin and the VA (Byrne *et al.*, 2018; Margaryan *et al.*, 2020), showed that the genetic composition of modern individuals from Scandinavia and Ireland exhibited different degrees of admixture that mirror the most famous expansion routes (Norwegian-like haplotypes into peoples from British Isles and Iceland, Danish-like haplotypes into Britain). In addition, several individuals with large amounts of South European ancestry in Denmark and southwest Sweden were retrieved during the VA, supporting our findings. Considering these results in the context of historical recorded raids and human demographic movements, we can postulate that the genome of modern indigenous goats could retain clues of old foreign genetic contribution as a consequence of an extensive trade and transport of goats in their history.

References

- Alexander D.H., Novembre J., Lange K. (2009) *Genome Res.* 19:1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Amills M., Capote J., Tosser-Klopp. (2017) *Anim. Genet.* 48(6):631-644. <https://doi.org/10.1111/age.12598>.
- Berg P., Groeneveld L.F., Brekke C., Våge D.I., Sørheim K.M. *et al.* (2020) *Acta Agric. Scand.* 69:47-52. <https://doi.org/10.1080/09064702.2020.1729852>.
- Byrne R.P., G., *et al.* (2018) *PLoS Genet.* 14(1): e1007152. <https://doi.org/10.1371/journal.pgen.1007152>.
- Cardoso T.F., *et al.* (2018) *Genet. Sel. Evol.* 50, 56. <https://doi.org/10.1186/s12711-018-0425-7>.
- Cassidy L.M., *et al.* (2017) *Biol. Lett.* 13(3): 20160876. <https://doi.org/10.1098/rsbl.2016.0876>.
- Chang C. C., Chow C. C., Tellier LCAM, Vattikuti S., Purcell S.M., *et al.* (2015) *GigaScience* 4(1): s13742-015-0047-8. <https://doi.org/10.1186/s13742-015-0047-8>.
- Colli L., Milanesi M., Talenti A., Bertolini F., Chen M. *et al.* (2018) *Genet Sel Evol* 50, 58 <https://doi.org/10.1186/s12711-018-0422-x>
- Cortellari M., Barbato M., Talenti A., Bionda A., Carta A., *et al.* (2021) *Sci. Rep.* 11, 19042. <https://doi.org/10.1038/s41598-021-89900-2>.
- Excoffier L., Lischer H.E.L. (2010) *Mol. Ecol. Resour.* 10: 564–567.
- Letunic I., Bork P. (2021) *Nucleic Acids Res.* 49(W1):W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- Margaryan A., *et al.* (2020) *Nature* 585, 390–396. <https://doi.org/10.1038/s41586-020-2688-8>.
- Stella A., *et al.* (2018) *Genet. Sel. Evol.* 50, 61. <https://doi.org/10.1186/s12711-018-0427-5>.
- Tamura K., Stecher G., Kumar S. (2021) *Mol. Biol. Evol.* 38(7):3022–3027. <https://doi.org/10.1093/molbev/msab120>.
- Tosser-Klopp G., Bardou P., Bouchez O., Cabau C., Crooijmans R. *et al.* (2014) *PLoS One* 9(3): e86227. <https://doi.org/10.1371/journal.pone.0086227>.
- Zheng, X., Levine D., Shen J., Gogarten S.M., Laurie C., *et al.* (2012) *Bioinformatics* 28(24):3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>.

This project was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 772787 (SMARTER).