Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **accepted version** of the article:

Mármol-Sánchez, Emilio; Luigi Sierra, Maria Gracia; Amills i Eras, Marcel. «Comparing the diversity of the casein genes in the Asian mouflon and domestic sheep». Animal Genetics, Vol. 51, Issue 3 (June 2020), p. 470-475. DOI 10.1111/age.12937

This version is available at https://ddd.uab.cat/record/224401

under the terms of the IN COPYRIGHT license

RUNNING HEAD: Ovine casein polymorphism

Comparing the diversity of the casein genes in the Asian mouflon and domestic sheep

M.G. Luigi-Sierra[1], E. Mármol-Sánchez[1] and M. Amills[1, 2]*

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. [2]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.

Corresponding author: Marcel Amills. Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. Tel. 34 93 5636600. E-mail: marcel.amills@uab.cat.

Abstract

Herewith, we aimed to determine whether casein variants that are currently segregating in ovine populations existed before the domestication of sheep or, on the contrary, if their emergence is much more recent. To this end, we have retrieved whole genome sequences from Iranian and domestic sheep from Africa, Europe, South and East Asia and West Asia. Population structure analysis based on 55,352,935 single nucleotide polymorphisms (SNPs) revealed a clear separation between Iranian mouflons and domestic sheep. Moreover, we also observed a strong genetic differentiation between Iranian mouflons sampled in geographic areas close to Tehran and Tabriz. Based on sequence data, hundreds of SNPs mapping to the casein α_{S1} (*CSN1S1*, 248 SNPs), casein α_{S2} (*CSN1S2*, 268 SNPs), casein ß (*CSN2*, 146 SNPs) and casein κ (*CSN3*, 112 SNPs) genes were identified. Approximately 25 – 63.02% of the casein variation was shared between Iranian mouflons and domestic sheep, and the four domestic sheep populations also shared 44.2 – 57.4 % of the casein polymorphic sites. These findings suggest that an important fraction of the casein variation present in domestic sheep was already segregating in the mouflon prior to its domestication. Genomic studies performed in horses and dogs are consistent with this view, suggesting that much of the diversity that we currently detect in domestic animals comes from standing variation already segregating in their wild ancestors.

Keywords: casein genes, sheep, mouflon, domestication.

Caseins represent the major milk protein fraction and they constitute an essential source of amino acids and calcium for the newborn. They can be classified into two main categories: calcium-sensitive caseins *i.e.* casein α_{S1} (CSN1S1), casein α_{S2} (CSN1S2) and casein ß (CSN2), which bind calcium and phosphorus; and the calcium insensitive casein κ (CSN3), which stabilizes the micelle and prevents milk coagulation (Martin 1993). In sheep, a number of casein variants have been reported and several of them have been associated with dairy traits (Selvaggi et al. 2014).

In a previous work (Guan et al. 2019), we compared the variability of the casein genes in domestic goats (*Capra hircus*) and their ancestor, the bezoar (*Capra aegagrus*). By doing so, we demonstrated that a relevant proportion of casein variants are shared by bezoars and domestic goats, suggesting that these variants emerged before goat domestication (Guan et al. 2019). The goal of the current work was to extend these studies to sheep. More specifically, we aimed to test the hypothesis that a significant fraction of casein polymorphisms that are currently segregating in domestic sheep (*Ovis aries*) come from standing variation already present in their wild ancestor, the mouflon (*Ovis orientalis*), prior to its domestication in the Fertile Crescent 10,000 YBP (Chessa et al. 2009; Stiner et al. 2014).

Whole genome sequences corresponding to 40 domestic sheep from Africa (N= 10), Europe (N= 10), South and East Asia (N= 8) and West Asia (N= 12); and 17 mouflons from Iran (Tehran, N= 7; Tabriz, N= 10), were retrieved from the Sequence Read Archive (SRA) of the NCBI database (**Tables S1** and **S2**). Genome sequences were aligned to the ovine reference genome Oar_v3.1 (Archibald et al. 2010) using the BWA MEM algorithm (Li & Durbin 2010). The resulting Sequence Alignment Map (SAM) files were sorted and converted into binary files to remove the PCR duplicates with the Picard tool (https://broadinstitute.github.io/picard/). A pipeline for calling

single nucleotide polymorphisms (SNPs) was carried out using the Haplotype-Caller function of the Genome Analysis Toolkit (McKenna et al. 2010) with default parameters in order to create VCF files containing SNPs. After performing the SNP-calling process a total of 43,293,496 SNPs were found in domestic sheep and 42,388,204 SNPs in mouflons, making a total of 58,626,591 SNPs. We used Beagle 4.1 (Browning & Browning 2016) with default settings to impute missing genotypes in sheep and mouflons separately. We kept exclusively autosomal SNPs with a minor allele frequency higher than 0.01, and without strong departures (*P*-value > 1.10^{-6}) from the Hardy Weinberg equilibrium. The PLINK v.1.90 software (Chang et al. 2015) was used to prune the data. Once the SNPs were pruned, 39,931,839 and 40,241,476 SNPs were found in sheep and mouflons, respectively. In **Fig. S1**, we show how the amount of discovered SNPs depends on the number of analysed individuals in each population. The total data set (sheep and mouflons) comprised 55,352,935 SNPs and 50% of them displayed frequencies lower than 0.10 (**Fig. S2**). We used PLINK v.1.90 (Chang et al. 2015) to perform a principal component analysis (PCA) based on the variance-standardized relationship matrix. The first two components, explaining the majority of variation, were used to carry out the graphical visualization of the data with R (R core Team 2018). The PCA plot revealed that the mouflon population is subdivided into two highly differentiated groups (**Fig. 1A**). The analysis of the SNP data with Admixture v.1.3.0 (Alexander et al. 2009) confirmed such finding (**Fig. 1C**), and showed that the most significant K-value is 2 (**Fig. S3**). Furthermore, the software VCFtools (Danecek et al. 2011) was used to estimate the whole-genome nucleotide diversity (π) on a *per site* basis (--site-pi) and by considering windows of 1000 bp (--window-pi). In the *per site* analysis, the π –values estimated for domestic sheep, mouflons from Tehran and mouflons from Tabriz were $0.284 \pm 2.25x10^{-05}$, $0.163 \pm 3.34x10^{-05}$ and $0.250 \pm$

$2.59x10^{-05}$, respectively. In contrast, we obtained π –values of $0.0030 \pm 1.69x10^{-06}$, $0.0029 \pm 3.45x10^{-06}$ and $0.0042 \pm 1.8x10^{-06}$ when performing the *per window* analysis in domestic sheep, mouflons from Tehran and mouflons from Tabriz, respectively. Discrepancies between the π estimates obtained in the *per site* and *per window* analyses are probably due to the fact that the *per window* analysis takes into account invariant sites (Guo et al. 2019), so it yields π estimates that are more consistent with what has been published in the literature. Chen et al (2018) described significantly higher π levels in mouflon populations when comparing them with domestic sheep. Higher diversity in wild vs domestic species has been described in dogs (Gray et al., 2009) and pigs (Bosse et al., 2002), being the consequence of bottlenecks experienced during the domestication process. The existence of two highly differentiated mouflon populations would be compatible with a marked habitat fragmentation that strongly enhanced the effects of genetic drift on allele frequencies. Indeed, the two mouflon populations were sampled in geographic areas close to Tehran and Tabriz, that are 600 km apart (http://nextgen.epfl.ch/). Although the demographic history of Iranian mouflons is mostly unknown, the IUCN currently lists *Ovis orientalis* as a vulnerable species (https://www.iucnredlist.org/species/ 15739/5076068), indicating a strong reduction of population size.

With regard to domestic sheep, four different clusters are evident according to the geographical origin of the sampled individuals. European ovine breeds show a scattered distribution, whilst African, West Asian and South and East Asian breeds group more tightly (**Fig. 1B**), with the exception of two Barag sheep (Inner Mongolia, China) that are located far apart from other East Asian sheep. As shown in **Fig. 1C**, these clustering patterns were also detected with the Admixture v.1.3.0 software (Alexander et al. 2009). This result is consistent with population genetics data presented

by Kijas et al. (2012) showing a close relationship between the patterns of ovine genetic variation and geography, and it also agrees well with the eastwards and westwards dispersal routes of domestic sheep from their primary domestication center in the Fertile Crescent (Lv et al. 2015; Zeder 2008).

The genomic coordinates of the four casein genes (*CSN1S1*, *CSN1S2*, *CSN2* and *CSN3*) were obtained using as a reference the Oar_v3.1 ovine genome (Archibald et al. 2010). These coordinates were used to retrieve variants mapping to the casein genes in sheep and mouflon. Population structure analyses (**Figs, S4** to **S6**) were performed by using the whole set of 774 casein polymorphisms as well as for each casein gene independently. We did not observe an evident population structure when we used any of the aforementioned data sets, probably because they encompass a relatively low number of markers. Similar results were shown by Chessa et al., (2017), who reported weak population structure among domestic sheep breeds when considering 46 molecular markers mapping to the casein genes and other loci of economic interest. Additionally a neighbor-joining tree was constructed based on the casein sequences using the MEGA v7 software (Kumar et al. 2016). This analysis highlighted the existence of three different clusters: one containing exclusively domestic sheep and the other two a mixture of both sheep and mouflon individuals (**Fig. S7**).

The π-values of each casein gene were estimated with VCFtools (Danecek et al. 2011) as previously explained (**Table S7**). The π estimates obtained in the *per-site* analysis were consistent with those previously reported by Guan et al. (2019) in goats and bezoars. Moreover, we observed higher π-values in the sheep *CSN1S1, CSN1S2* and *CSN3* genes when compared to those of mouflons. In contrast, in the *per window* analysis mouflon populations exhibited higher variability than sheep in the four casein genes (**Fig. 2B, Table S7**). We consider that the *per window* analysis yields π-values

that are more comparable to the ones generated in other studies because it considers the existence of invariant sites. Overall, we can conclude that the variability of the casein genes is quite similar in sheep and mouflons. In principle, population size and geographic range of sheep are much larger than those of mouflons, but it is also true that domestication might have involved one or several bottlenecks decreasing genetic diversity (Alberto et al. 2018). Moreover, population subdivision in mouflons (**Fig. 1A**) might have contributed to create highly divergent gene pools, thus increasing overall diversity (**Fig. 2A**).

Hundreds of SNPs were identified into the casein genes in both sheep and mouflons (**Table S3**). We predicted the effects of the identified SNPs with the SNPeff v 4.3 software (Cingolani et al. 2012), while the potential effects of missense variation were inferred with the SIFT tool (Vaser et al. 2016). The broad majority of the SNPs identified in the four casein genes mapped to introns (**Tables S4** and **S5**), but we also detected a number of missense mutations that are reported in **Table S6**. Several of these missense mutations were described in previous publications (Bastos et al. 2001; Corral et al. 2010; Giambra & Erhardt 2012; Suárez-Vega et al. 2017), while others were novel (**Table S6**). Of course, part of the identified SNPs could have been produced by sequencing errors, so they should be further validated by using sequencing and typing techniques. We have also reconstructed haplotypes based on SNPs mapping to each casein gene by using fastPHASE 1.4 (Scheet & Stephens 2006). This software uses a clustering algorithm to infer the haplotype of each individual. Haplotype reconstruction has been performed for sheep and mouflons separately, and the numbers of casein haplotypes detected with this methodology are depicted in **Fig. S8**. It can be observed that the majority of casein haplotypes have very low frequencies (MAF < 0.05).

The comparison of casein variation between sheep and mouflons revealed an extensive sharing of polymorphic sites (**Figs, 2C** and **S9**), a finding that agrees well with results reported by Guan et al. (2019) in goats and bezoars. Consistently, Chessa et al. (2017) showed the existence of a haplotype formed by 8 SNPs that segregates in 51% of the mouflons as well as in the majority of the domestic sheep breeds sampled in their study. One potential explanation for these results would be the introgression of mouflons with domestic sheep, but data depicted in **Fig. 1C** are not consistent with such hypothesis. More probably, an important fraction of the casein variation present in sheep was already segregating in the mouflon prior to its domestication. This conclusion is consistent with our observation that the four domestic sheep populations under study share an extensive amount of casein polymorphisms despite the fact that they were raised in different continents *i.e.* 44.21% of the SNPs shared in *CSN1S1*, 55.61% in CSN1S2, 57.49% in *CSN2* and 49.47% in *CSN3*.

Although large-effect polymorphisms present in domesticated species are usually recent mutations that were far easier to select on, the bulk of the variation present in livestock probably comes from the standing variation present in the wild progenitors (Wright 2015). According to our analyses, the ovine casein genes are no exception to this general trend, although we have also detected sheep casein polymorphisms that do not segregate in the mouflon (and vice versa). This result needs to be interpreted with caution, because sample sizes employed in our study are small, but we hypothesize that it could have been produced by the process of genetic differentiation that sheep and their wild progenitor underwent during the last 10,000 YBP.

Acknowledgments

This study makes use of data generated by the NextGen Consortium (https://www.epfl.ch/labs/nextgen/consortium). The European Union 7th Framework Programme (FP7/2010-2014) provided funding for the project under grant agreement no 244356-"NextGen". We acknowledge the support of the Spanish Ministry of Economy and Competitivity for the *Centro de Excelencia Severo Ochoa 2016–2019* award (SEV-2015-0533). The CERCA Programme of the Generalitat de Catalunya also supported our research. Maria Luigi-Sierra was funded with a fellowship Formación de Personal Investigador (BES-C-2017-079709) awarded by the Spanish Ministry of Economy and Competitivity. Emilio Mármol-Sánchez was funded with a PhD fellowship (FPU15/01733) granted by the Spanish Ministry of Education and Culture.

Availability of the Data.

The whole-genome SNP datasets used in this research are available in Figshare. The accession numbers are 10.6084/m9.figshare.11663598.v1, 10.6084/m9.figshare.11686809.v1, 10.6084/m9.figshare.11686962.v1, 10.6084/m9.figshare.11687118.v1, 10.6084/m9.figshare.11688276.v1 and 10.6084/m9.figshare.11688600.v2. Furthermore, VCF containing the SNPs mapping to the ovine casein genes can be found at 10.6084/m9.figshare.11663547.v1.

References

Alberto, F.J., Boyer, F., Orozco-Terwengel, P., Streeter, I., Servin, B., De Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H.R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M.W., Taberlet, P. & Pompanon, F. (2018) Convergent genomic signatures of domestication in sheep and goats. *Nature Communications* **9**, 813.

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

Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., McEwan, J.C., Hutton Oddy, V., Raadsma, H.W., Wade, C., Wang, J., Wang, W. & Xun, X. (2010) The sheep genome reference sequence: A work in progress. *Animal Genetics* **41**, 449-53.

Bastos, E., Cravador, A., Azevedo, J. & Guedes-Pinto, H. (2001) Single strand conformation polymorphism (SSCP) detection in six genes in Portuguese indigenous sheep breed "Churra da Terra Quente". *Biotechnology, Agronomy, Society and Environment* **5**, 7-15.

Bosse, M., Megens, H.J., Madsen, O., Paudel, Y., Frantz, L.A.F., Schook, L.B., Crooijmans, R.P.M.A. & Groenen, M.A.M. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics* **8**, e1003100.

Browning, B.L. & Browning, S.R. (2016) Genotype imputation with millions of reference samples. *American Journal Human Genetics* **98**, 116-26.

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

Chen, Z.H., Zhang, M., Lv, F.H., Ren, X., Li, W.R., Liu, M.J., Nam, K., Bruford, M.W. & Li, M.H. (2018) Contrasting patterns of genomic diversity reveal accelerated genetic drift but reduced directional selection on X-chromosome in wild and domestic sheep species. *Genome Biology and Evolution* **10**, 1282-97.

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R.R., Pemberton, J.M., Beraldi, D., Stear, M.J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M.H., Kazwala, R.R., Zhang, Y.P., Arranz, J.J., Ali, B.A., Wang, Z., Uzun, M., Dione, M.M., Olsaker, I., Holm, L.E., Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M.J., Paolo, A.M., Bruford, M.W., Kantanen, J., Spencer, T.E. & Palmarini, M. (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science* **324***,* 532-6.

Chessa, S., Giambra, I.J., Brandt, H., Caroli, A.M., Gootwine, E. & Erhardt, G. (2017) Genetic diversity within economically important loci in European, Middle Eastern, and African sheep breeds: An insight into their development. *Small Ruminant Research* **155**, 72-80.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 1-13.

Corral, J.M., Padilla, J.A., & Izquierdo, M. (2010) Associations between milk protein genetic polymorphisms and milk production traits in Merino sheep breed. *Livestock Science* **129**, 73-9.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G. & Durbin, R. (2011) The variant call format and VCFtools. *Bioinformatics.* **27**, 2156-8.

Giambra, I.J., & Erhardt, G. (2012) Molecular genetic characterization of ovine *CSN1S2* variants C and D reveal further important variability within *CSN1S2*. *Animal Genetics* **43**, 642–5.

Gray, M.M., Granka, J.M., Bustamante, C.D., Sutter, N.B., Boyko, A.R., Zhu, L., Ostrander, E.A. & Wayne, R.K. (2009). Linkage disequilibrium and demographic history of wild and domestic canids. *Genetic* **181**, 1493-1505.

Guan, D., Mármol-Sánchez, E., Cardoso, T. F. .,Such, X., Landi, V., Tawari , N. R. & Amills, M. (2019) A genomic analysis about the origins of extant casein variation in goats. *Journal of Dairy Science* **102**, 5230-41.

Guo, L., Yao, H., Shepherd, B., Sepulveda-Villet, O.J., Zhang, D.C. & Wang, H.P. (2019) Development of a genomic resource and identification of nucleotide diversity of yellow perch by RAD sequencing. *Frontiers in Genetics* **10**, 992.

Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R., Cristobal, M.S., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J. & Dalrymple, B. (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10,** e1001258.

Kumar, S., G. Stecher, & K. Tamura. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and evolution* **33**, 1870-4.

Li, H. & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95.

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., Li, J.Q., Liu, Y.G., Shen, Z.Q., Zhao, S.G., Hehua, Ee., Gorkhali, N.A., Farhad Vahidi, S.M., Muladno, M., Naqvi, A.N., Tabell, J., Iso-Touru, T., Bruford, M.W., Kantanen, J., Han, J.L. & Li, M.H. (2015) Mitogenomic meta-analysis identifies two phases of migration in the history of Eastern Eurasian sheep. *Molecular Biology and evolution* **32**, 2515-33.

Martin P. (1993) Polymorphisme génétique des lactoprotéines caprines. *Le Lait* **73**, 511-532.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-303.

R Core Team. (2018) R: A language and environment for statistical computing. *R Foundation for Statistical Computing,Vienna, Austria.* URL http//www. R-project. org.

Scheet, P. & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* **78,** 629-44.

Selvaggi, M., Laudadio, V., Dario, C. & Tufarelli, V. (2014) Investigating the genetic polymorphism of sheep milk proteins: A useful tool for dairy production. *Journal of the Science of Food and Agriculture* **94**, 3090-9.

Stiner, M.C., Buitenhuis, H., Duru, G., Kuhn, S.L., Mentzer, S.M., Munro, N.D., Pöllath, N., Quade, J., Tsartsidou, G. & 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 USA* **111**, 8404-

9.

Suárez-Vega, A., Gutiérrez-Gil, B., Klopp, C., Tosser-Klopp, G. & Arranz, J.J. (2017) Variant discovery in the sheep milk transcriptome using RNA sequencing. *BMC Genomics* **18**, 170.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. & Ng, P.C. (2016) SIFT missense predictions for genomes. *Nature Protocols* **11**, 1-9.

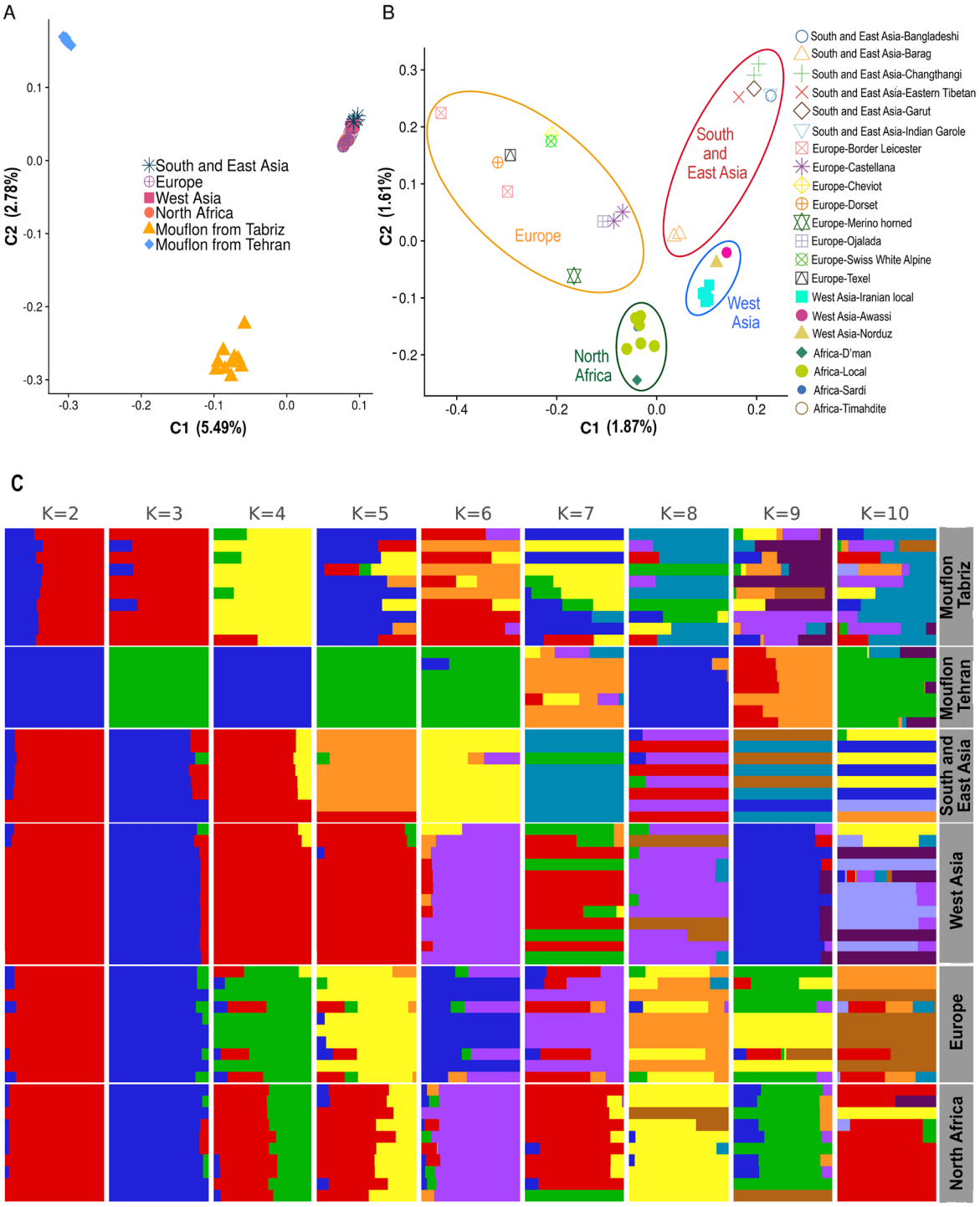
Wright, D. (2015) The genetic architecture of domestication in animals. *Bioinformatics and Biology Insights* **9**, 11-20.

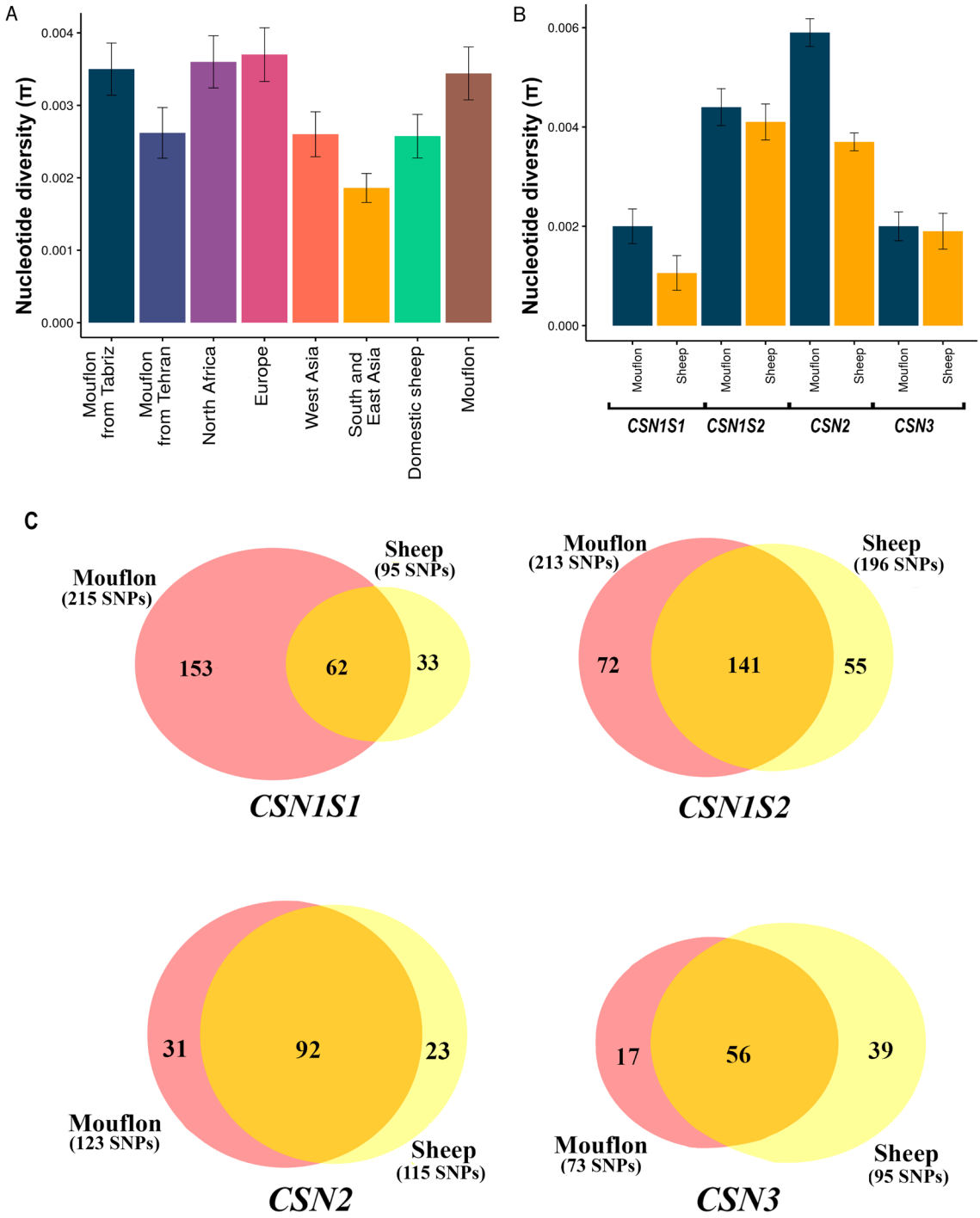
Zeder, M.A. (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the USA* **105**, 11597-604.

LEGENDS TO FIGURES

Fig. 1A. Principal component analysis (PCA) based on 55,352,935 SNPs and comprising two Iranian mouflon populations (Tabriz and Tehran) as well as four domestic sheep populations from North Africa, West Asia, Europe and South and East Asia. **1B** Principal component analysis based on 39,931,839 SNPs identified in domestic sheep and focused on the four North African, West Asian, European and South and East Asian populations mentioned before (breed of origin is also indicated). **1C**. Admixture analysis (K=2-10) encompassing the same number of markers and populations defined in **1A**.

Fig. 2. Nucleotide diversity (*per window* analysis) of the whole casein cluster (**2A**) and of each casein locus (**2B**) in two Iranian mouflon populations (Tabriz and Tehran) and in 4 domestic sheep populations from North Africa, West Asia, Europe and South and East Asia. Each bar represents the mean nucleotide diversity and its standard error. Venn diagrams (**2C**) depicting the proportion of casein SNPs shared by domestic sheep and mouflons.


A
B
C
South and East Asia
Europe
West Asia
North Africa
Mouflon from Tabriz
Mouflon from Tehran
C1 (5.49%)
C2 (2.78%)
C1 (1.87%)
C2 (1.61%)
South and East Asia-Bangladeshi
South and East Asia-Barag
South and East Asia-Changthangi
South and East Asia-Eastern Tibetan
South and East Asia-Garut
South and East Asia-Indian Garole
Europe-Border Leicester
Europe-Castellana
Europe-Cheviot
Europe-Dorset
Europe-Merino horned
Europe-Ojalada
Europe-Swiss White Alpine
Europe-Texel
West Asia-Iranian local
West Asia-Awassi
West Asia-Norduz
Africa-D'man
Africa-Local
Africa-Sardi
Africa-Timahdite
K=2
K=3
K=4
K=5
K=6
K=7
K=8
K=9
K=10
Mouflon Tabriz
Mouflon Tehran
South and East Asia
West Asia
Europe
North Africa


A
Nucleotide diversity (π)
0.004
0.003
0.002
0.001
0.000
Mouflon from Tabriz
Mouflon from Tehran
North Africa
Europe
West Asia
South and East Asia
Domestic sheep
Mouflon
B
Nucleotide diversity (π)
0.006
0.004
0.002
0.000
Mouflon
Sheep
CSN1S1
CSN1S2
CSN2
CSN3
C
Mouflon (215 SNPs)
Sheep (95 SNPs)
153
62
33
CSN1S1
Mouflon (213 SNPs)
Sheep (196 SNPs)
72
141
55
CSN1S2
31
92
23
Mouflon (123 SNPs)
Sheep (115 SNPs)
CSN2
17
56
39
Mouflon (73 SNPs)
Sheep (95 SNPs)
CSN3