

Goat ancient DNA analysis unveils a new lineage that may have hybridized with domestic goats

Laura Botigué based on reviews by *Torsten Günther* and *1 anonymous reviewer*

Open Access

A recommendation of:

A novel lineage of the *Capra* genus discovered in the Taurus Mountains of Turkey using ancient genomics

Kevin G. Daly, Benjamin S. Arbuckle, Conor Rossi, Valeria Mattiangeli, Phoebe A. Lawlor, Marjan Mashkour, Eberhard Sauer, Joséphine Lesur, Levent Atici, Cevdet Merih Ereğ, Daniel G. Bradley (2022), *bioRxiv*, 2022.04.08.487619, ver. 5 peer-reviewed and recommended by Peer Community in Genomics <https://doi.org/10.1101/2022.04.08.487619>

Submitted: 15 April 2022, Recommended: 02 August 2022

Recommendation

The genomic analysis of ancient remains has revolutionized the study of the past over the last decade. On top of the discoveries related to human evolution, plant and animal archaeogenomics has been used to gain new insights into the domestication process and the dispersal of domestic forms.

In this study, Daly and colleagues analyse the genomic data from seven goat specimens from the Epipalaeolithic recovered from the Direkli Cave in the Taurus Mountains in southern Turkey. They also generate new genomic data from *Capra* lineages across the phylogeny, contributing to the availability of genomic resources for this genus. Analysis of the ancient remains is compared to modern genomic variability and sheds light on the complexity of the Tur wild *Capra* lineages and their relationship with domestic goats and their wild ancestors.

Authors find that during the Late Pleistocene in the Taurus Mountains wild goats from the Tur lineage, today restricted to the Caucasus region, were not rare and cohabited with Bezoar, the wild goats that are the ancestors of domestic goats. They identify the ancient specimens as a

Published: 23 August 2022

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

separate lineage of the West and East Caucasus Tur modern lineages. Also, analysis of the genomic data and mitochondrial haplotypes evidences hybridization between the Tur and the Bezoar wild lineages. Interestingly, authors also find an uneven amount of Tur ancestry among Neolithic domestic goats, with European domestic goats showing evidence of this ancient Tur ancestry, whereas Neolithic Iranian domestic goats do not, a pattern that is also observed in some modern European domestic goats.

A modified D statistic, Dex, is developed to examine the contribution of the ancient Tur lineage in domestic goats through time and space. Dex measures the relative degree of allele sharing, derived specifically in a selected genome or group of genomes, and may have some utility in genera with complex admixture histories or admixture from ghost lineages. Results confirm that Neolithic European goat had an excess of allele sharing with this ancient Tur lineage, something that is absent in contemporary goats eastwards or in modern goats.

Interspecific gene flow is not uncommon among mammals, but the case of *Capra* has the additional motivation of understanding the origins of the domestic species. This work uncovers an ancient Tur lineage that is different from the modern ones and is additionally found in another geographic area. Furthermore, evidence shows that this ancient lineage exhibits substantial amounts of allele sharing with the wild ancestor of the domestic goat, but also with the Neolithic Eurasian domestic goats, highlighting the complexity of the domestication process.

This work has also important implications in understanding the effect of over-hunting and habitat disruption during the Anthropocene on the evolution of the *Capra* genus. The availability of more ancient specimens and better coverage of the modern genomic variability can help quantifying the lineages that went lost and identify the causes of their extinction.

This work is limited by the current availability of whole genomes from modern *Capra* specimens, but pieces of evidence as well that an effort is needed to obtain more genomic data from ancient goats from different geographic ranges to determine to what extent these lineages contributed to goat domestication.

References

Daly KG, Arbuckle BS, Rossi C, Mattiangeli V, Lawlor PA, Mashkour M, Sauer E, Lesur J, Atici L, Cevdet CM and Bradley DG (2022) A novel lineage of the *Capra* genus discovered in the Taurus Mountains of Turkey using ancient genomics. bioRxiv, 2022.04.08.487619, ver. 5 peer-reviewed and recommended by Peer Community in Genomics. <https://doi.org/10.1101/2022.04.08.487619>

Cite this recommendation as:

Laura Botigué (2022) Goat ancient DNA analysis unveils a new lineage that may have hybridized with domestic goats. *Peer Community in Genomics*, 100020. <https://doi.org/10.24072/genomics.100020>

Reviews

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2022.04.08.487619>

Author's Reply, None

Download author's reply[Download tracked changes file](#)

Decision by *Laura Botigué*, 16 May 2022

The manuscript "A novel lineage of the Capra genus discovered in the Taurus mountains of Turkey using ancient genomics" analyses the genomes of different ~14,000 year-old Capra specimens from an archaeological site in southern Turkey. The time and location of these remains are of interest because they can shed more light into the details behind the domestication process of domestic goats. Both reviewers have found that the manuscript makes relevant contributions in several fields. Reviewer one is particularly enthusiastic about the Dext statistic and suggests that this statistic is exposed in a more relevant manner.

There are however, several concerns that the two reviewers share. The first of them, is that the manuscript revolves around the claim of the novel finding of a lost Capra lineage. Beyond the difficulties surrounding the very definition of species and subspecies that Reviewer 1 raises and the possibility raised by Reviewer 2 of these specimen being a hybrid, both reviewers point to the current gap of knowledge that exists on the population structure of the modern Capra genus. The Discussion should be more explicitly framed acknowledging this gap of knowledge. The second concern is about the processing of the low coverage genomes. Some additional analyses are proposed to avoid mapping biases and to quantify the mismapping effects associated with the low coverage genomes. I also think it would be necessary to provide more information on the number of SNPs used in the different analyses.

Something that the two reviewers do not comment on but I noticed is that in the legend of Figure 1B it is sated that the tree was done for genomes above 0.5x, but Tur2 is included and its genome is reported to be 0.02x.

I think that addressing these points would results in a significantly improved piece of work.

Reviewed by anonymous reviewer, 13 May 2022

In this article, the authors have sequenced multiple Capra ancient specimens from the Direkli and, in doing so, they have discovered one individual representing a tur-like novel lineage, Direkli 4, which might represent an extinct goat species (Taurasian tur) or subspecies highly related with *Capra caucasica*. Evidence of historical introgression of the Taurasian tur into ancient domestic goats is provided. It is a very interesting and relevant study, but several issues require further classification:

In the Introduction, the results obtained by Zheng et al. (Science Advances, DOI: 10.1126/sciadv.aaz5216) concerning the likely pre-domestication admixture between bezoars and tur-like species should be explained with more detail. In the light of these results, the placement of Direkli4 as sister to a clade of both Caucasian tur taxa would not be as unexpected as suggested in line 72. Indeed, Zheng et al indicate that among the 4 predomestication Anatolian bezoars 3 had a tur-like mitochondrial haplotype.

The authors used `bwa aln` to align ancient DNA reads to the goat reference genome. To mitigate the impact of reference bias (reads carrying the alternate allele are more difficult to map), which can have an adverse impact on the calculation of D-statistics, a variation-aware read aligner such as `vg` could be used instead.

It would be useful to assess the error rates associated with sequencing by contrasting the genetic distance of ancient samples of species A to an outgroup (distance 1) vs genetic distances of high quality modern samples of species A to an outgroup (distance 2). An inflation of genetic distance 1 with regard to distance 2 would provide a conservative estimate of such error rate. Another measure of interest would be to count differences between the sequenced DNA fragments and regions of the goat genome that are highly conserved across *Capra* species.

In line 120, I find paradoxical that the combination “tur-like genome, tur-like mitochondria” is never observed in any of the specimens. Parental tur and bezoar populations should be larger than the population of hybrids, right? Hybridization between *Capra* species can happen, and without a doubt, has happened, but probably is not a very frequent event.

If I understand well, the Direkli 4 individual has a tur-like autosomal genome (Figure 1B) and a bezoar-like mitochondrial genome (Figure 2B), so could it be an hybrid between the two species, tur and bezoar, rather than a new Taurasian tur species (*Capra taurensis*) or subspecies (*Capra caucasica taurensis*)? Although mtDNA coalescent estimates suggest that this “Taurasian tur” may have diverged from the Caucasian lineages 130-200kya, this seems quite a short divergence time in terms of speciation (shorter than the *Bos taurus*/*Bos indicus* divergence time for instance). Indeed, in the paper authored by the same authors in *Science* (Daly et al. 2018), the predomestic radiation of the divergent bezoar/goat Levant population is assumed to have taken place 38,500 to 195,200 BP and it is not treated as a subspecies. There is even a more compelling example: the estimated time of divergence between Asian and Western pigs is 1 million years BP, and these two populations are not treated as different *suid* species. In summary, evidence that the new lineage discovered by the authors corresponds to a new species or subspecies is, in my view, not proved beyond a reasonable doubt.

Besides, how do we know that this divergent lineage is extinct in modern turs? I do not think that the variation of this species has been investigated in depth. It would be good to collect *Capra* genus modern sequences and construct a phylogenetic tree encompassing both modern and ancient *Capra* sequences to assess the variation of each species on a broader basis.

A multivariate analysis of the data via MDS (only ancient and ancient + modern) would be also advisable

The results obtained by Zheng should be incorporated more broadly into the Discussion because they are strongly connected with the main findings of Daly and coworkers.

In the Supplementary Materials (media 1-file), each Figure should have the corresponding caption to facilitate its interpretation.

I have found very few typos, but some additional revision is advised to eliminate all of them. For instance: Table 1 should be provenance or provenience, but not providence that has a different meaning (an interesting discussion can be found in: <https://www.merriam-webster.com/words-at-play/usage-of-province-providence-provenience-provenance>). In Table 1,

but also in other places, Pyrenica should be Pyrenaica (*Capra pyrenaica*). In page 8 of the suppl file, "were are" seems incorrect. In page 9, Caucasian is also incorrect.

Reviewed by Torsten Günther, 10 May 2022

This is a review of the preprint by Daly et al entitled "A novel lineage of the *Capra* genus discovered in the Taurus mountains of Turkey using ancient genomics". In Direkli Cave in the Taurus mountains of modern-day Turkey, they found ~14,000 year old remains that appear to belong to an unknown lineage of the genus *Capra*, closely related to Caucasian tur species. This lineage is specifically represented by the genome of Direkli4. The authors created a broad dataset of many different ancient *Capra* genomes. While this study had a pretty specific focus, we are sure that this data could become handy for many other projects in the future. All sequence data is already available through public databases.

Additionally, the authors present an extended D statistic, Dex, to infer gene flow from a specific source population (by conditioning on all other potential outgroups/sources are ancestral). In this particular study, the method is used to detect allele sharing between the tur lineage and ancient (but not modern) European domestic goats. Code for the method is available through GitHub. This method could be useful in similar scenarios. Something that comes to mind are anatomically modern humans and archaic hominins, where "ghost" admixture can affect D statistic results and, for instance, Neandertal admixture can influence tests where only Denisovans and modern populations are used in the test itself. The utility of Dex for other studies could actually be stressed further by explaining in what cases one can apply it.

We generally think that this study represents a very interesting example that highlights the power of archaeogenomics to gain surprising insights into past distributions of species and genera. We have some comments below but none of them should systemically affect the major finding of an unknown tur population in the Taurus Mountains.

Definition of a new species:

The manuscript is strongly trying to assign some taxonomic status to the Direkli lineage, even potential species and subspecies names are proposed. As population geneticists, we usually do not have a strong opinion on what should be considered a (sub-)species. These populations were able to produce fertile offspring and interfertility seems to be common within the *Capra* genus. We simply believe that an attempt to define a mammalian species only based on a low coverage genome and without much morphological information has the potential to trigger some discussion and draw attention from the general quality of this research and the aspects of this finding that highlight the potential of ancient DNA research. The authors have found a tur population that seems to be extinct (within the limitations of the available data, see also below), so it represents a great example on how ancient DNA can help to uncover past distribution of species/taxa and how past climate change and other factors did influence the extinction of populations. Demographic reconstructions often require "ghost populations" to fit but we are struggling to place these in time and space. In this case, the Direkli lineage may actually represent one of these ghost populations that were suspected from previous results.

The fact that the “extinction” of this lineage must have taken place within the range of aDNA also highlights the possibility for follow up studies to specifically search for genetic traces of this group in modern *Capra* populations as well as archaeological material from Anatolia, the Caucasus and the Zagros mountain range. A discussion largely revolving around the taxonomic status of the lineage could distract from these findings and there are probably good arguments for different classifications that us geneticists may not even have in mind at this point. The mitochondrial divergence between the groups would fall into a range that is still considered one species in other systems. The exact status of this “Taurasian tur” could also have implications for the two Caucasian (sub-)species as well as local conservation efforts.

Writing of the text:

In general, the text is well written but there are some inconsistencies within the text that leave the feeling that it was at least partly written in a rushed way. For example, Introduction and Discussion seem quite strongly aiming at defining the new genome as a new (sub-) species while abstract and results seem more nuanced. There are also some seemingly contradictory statements, as e.g. (from the abstract): “West Eurasian domestic goats in the past, but not those today, appear enriched for Direkli4-specific alleles, and we further identify genomic regions introgressed in domestic goats with high affinity to Direkli4.” – this sentence refers to two completely different results/analyses which is clear after reading the main text, but maybe not at this point.

The supplementary material could also take some work, e.g. making sure all display items are described (figure legends) and referenced in the text, maybe even move the supplementary figures closer to the corresponding text bits.

The dataset and its limitations:

We believe that it should be made clearer that all conclusions are drawn with respect to the available data. The genus *Capra* seems to be understudied, so only a handful of full MT genomes are available and some species are only represented by relatively low quality aDNA nuclear genomes. This is especially true for the Caucasian tur species for which only 5 individuals are available. Furthermore, Direkli4 itself seems to be a hybrid as it carries a bezoar MT lineage. This means that a lot of the diversity of these species and within the genus itself is unknown. Adding that interbreeding seems to be quite common, this should be taken into account when discussing the species status of the new genome.

Ascertainment and reference bias:

The authors seem concerned that their results could be driven by certain biases. To avoid or mitigate them, they chose two approaches: mapping the reads to the sheep genome or using only sites that also segregate in sheep. While the first approach should reduce the effect of mapping/reference bias, the latter should avoid ascertainment bias. Most analyses are verified with the sheep ascertained variants but only the IBS tree is based on the sheep mapped data. We would be concerned that the goat mapped but sheep ascertained results of allele sharing analyses (e.g. Dex) could be influenced by mapping bias. Our recommendation would be to

repeat these analyses with the sheep mapped data and sheep ascertained data to avoid both biases.

Minor comments:

- Line 123: Considering that all these individuals carrying this T lineage so far were found in the same place, it may be difficult to conclude how many actual introgression events took place. There could be multiple closely related events or further distant in the past or other possibilities.
- Please add the species names to Figure 2B.
- Line 246: One could probably test those demographic scenarios with this dataset. That might be beyond the scope of this article but it would certainly add relevant information.
- Please list non default parameters instead of just writing "relaxing parameters (Meyer et al 2012).
- Was the `auto_only` option set to `FALSE` in `admixtools2`? By default, the package assumes human data, i.e. it ignores all chromosomes after `chr22`.

This review was written by Pedro Morell Miranda and Torsten Günther.