Paleogenomics unveil domesticated camelid herding and wild camelid hunting in early pastoralist societies in the Atacama Desert

Conor. O'Hare[1], Paloma Diaz-Maroto[1], Isabel Cartajena[2], Michael. V Westbury[1]

1. Globe Institute, University of Copenhagen, Copenhagen, Denmark
2. Department of Anthropology University of Chile, Santiago, Chile

*Corresponding author. Email: m.westbury@sund.ku.dk

Abstract

The domestication of camelids and their integration into early pastoralist economies played a significant role in the development of Andean societies. Archaeological evidence suggests South American camelid domestication was a complex process that began with hunting, herd protection, and finally captivity/selective breeding. Although South American camelids can be osteometrically classified as large (putatively *Lama*), or small (putatively *Vicugna*), it is difficult to designate an individual as wild or domesticated, limiting insights into the use of early domesticates. Here we utilise palaeogenomic data to reveal the sex, ancestry, and domestication status of camelids recovered from archaeological sites from the Tulán ravine, a key area for understanding the transition from hunter-gatherer to initial pastoral societies in the South-Central Andes. We were able to infer the ancestry of 49 individuals with mean genome-wide coverages >0.001x and found evidence for both genera, *Vicugna* and *Lama*, in our dataset. Further filtering to 26 individuals >0.01x showed that all individuals, except one male llama, represented wild species. Similar male-to-female sex ratios across sites and genera further supported the hypothesis that these camelids were hunted rather than herded. Overall, our findings suggest that lost early domesticate lineages and/or wild camelids, not the direct ancestors of contemporary domesticated species, were the primary source of camelids used in both domestic and ceremonial contexts in the Tulán ravine during the Early Formative period (3,360–2,370 cal. yr BP).

Keywords: paleogenomics, South American camelids, admixture, hybridization, ancient DNA

Introduction

One of the most significant cultural innovations of humanity was the domestication of plants and animals (1). However, despite its remarkable significance, domestication processes (e.g. wild ancestors, timing, location, and underlying genetic mechanisms) remain unsolved for many species (2). This is especially true for the only large herbivores domesticated in the Americas, South American camelids (SACs) (1). The domestication of SACs played a significant role in the development of Andean societies. Their lone domesticated status and ability to occupy broad domesticate roles (transport, fibre, food) exemplifies the pivotal role these animals played in shaping ancient South American societies. Pack llama (*Lama glama*) acted to connect distant

communities, while camelid fibres and meat served both practical and ritual purposes (3–5). Even today, SACs remain the most important resource for many communities (6).

SACs comprise four extant species; two domesticated (llama [*Lama glama*], alpaca [*Vicugna pacos*]), and two wild (guanaco [*L.guanaco*], vicuña [*V. vicugna*]). Within the wild species there are two subspecies each: *L.g.cacsilensis*, *L.g.guanicoe*, *V.v.mensalis*, and *V.v.vicugna*. While the most likely location of domestication is somewhere in the Central Andes, Southern Bolivia and/or the Circumpuna (Northern Chile and NW Argentina) (7–10), the timing and number of domestication centres remain debated (11). Archaeological evidence suggests SAC domestication was a complex process and not synchronous across the Andes that began with hunting, herd protection, and finally captivity/selective breeding (12). However, although SACs can be osteometrically classified as large (putatively *Lama*), or small (putatively *Vicugna*), it is difficult to designate an individual as wild or domesticated[15], limiting insights into the presence of early domesticated individuals. Genetics is a useful tool that can complement the archaeological record to provide further insights into domestication processes (13). However, despite attempts to address theories of SAC domestication, the Spanish conquest of South America beginning ~1500 A.D and uncontrolled post-conquest hybridisation between domesticated species makes it difficult to identify genetic affinities prior to ~1500 A.D from contemporary individuals (14). Ancient DNA (aDNA) from pre-conquest individuals is therefore the best approach to reconstructing the uncertain past of SAC domestication and husbandry.

Here we utilise palaeogenomic data to investigate the use of camelids by ancient South American communities. We analyse the sex, ancestry, and domestication status from pre-conquest SACs recovered from archaeological sites from the Tulán ravine (Tulán-52, Tulán-54, Tulán-85, and Tulán-94) located in the eastern border of the Salar de Atacama that descends into Northern Chile. The Tulán ravine is a key area for understanding the transition from hunter-gatherer to initial pastoral societies in the South-Central Andes (15). The sites studied here belong to the Late Archaic period (Tulán-52; 5290-4840/4430-4090 cal. yBP) and to the Early Formative period (Tulán-54, Tulán-85, and Tulán-94; 3,360–2,370 cal. yr BP) when domesticated camelids are thought to have been incorporated into the economic activities of humans of this area. Specifically, archaeological remains from offerings and domestic activities evidence an important place of cultural and economic transformation from hunter-gatherers to early pastoralist communities in the highlands of the Southern Andes (16).

Materials and Methods

Archaeological contexts

It is proposed that the location of Archaic and Formative settlements in the Tulán area can be explained by a set of favourable cultural, social and environmental conditions, which allowed the change from a hunter-gatherer society to a mainly pastoralist one.

The Late Archaic period (ca. 5290-4090 cal. yBP) is marked by large campsites like Tulán-52, the emergence of substantial architecture with clusters of circular structures, and a diversified, innovative lithic industry. This period also saw the appearance of rock art, long-distance interactions, and a subsistence strategy that combined hunting with early camelid domestication. Altogether, these features point to a growing sociocultural complexity (17).

The Early Formative period (ca. 3360 - 2370 cal. yBP) is characterised by the emergence of large settlements with a complex and planned ritual architecture along with the intensification and expansion of productive practices, an emphasis on ritual expressions, the appearance of new technologies such as pottery, and three large contemporary Formative settlements (Tu-54, Tu-122 and Tu- 85) have been documented for this period. These are located within a 15 km area between the border of the Salar de Atacama and the Tulán ravine, at an altitude ranging from 2300–3200m. These sites have different functions. At Tulán-122, located in the middle sector of the ravine, residential aspects were emphasised. In contrast, Tulán-85, is located on the border of the resource-rich salt flat. These settlements shared a common political and religious organisation centred around the Tulán 54 site, the only site that possesses monumental architecture (15).

The sunken temple structure identified in Tulán-54 at the centre of a refuse mound contained 27 human infant graves accompanied by offerings, rock art, structured fireplaces, and offering pits, among other elements, suggesting that the structure had a ceremonial function. The main structure identified at the site is a large, sunken semi-oval ceremonial structure surrounded by a perimeter wall made of large, vertically positioned stone blocks. The inner space is divided into six precincts, with a single oval structure in the middle (15, 16). The recovered camelid anatomical units display a very similar temporal and spatial pattern across all occupations and among the different precincts of the ceremonial structure, which was filled intentionally with the remains of butchering, processing, consumption and feasting activities. The waste was likely generated outside and then ritually incorporated (18). On the border of the Salar de Atacama, Tulán 85 corresponds to an extensive occupation characterised by dense domestic monticulated deposits. The site also includes marginal structures and a sector with burials of newborns (19). Finally, Tulan-94 site is composed of groups of circular and sub-circular structures, with a total of 19 living enclosures. The evidence suggests the transitional character of the site, since it has material components corresponding to the Late Archaic and Early Formative phases (15).

Samples

Our study consists of 80 modern individuals and 75 ancient individuals (Supplementary tables S1 and S2). The modern individuals comprise all extant *Lama* and *Vicugna* subspecies (Figure 1). The ancient individuals were sourced from four archaeological sites: Tulán-52, Tulán-54, Tulán-85, and Tulán-94. Sequencing data were generated from these individuals following (20) where the mitochondrial DNA from these specimens has been analysed. Morphological information on the size of the ancient individuals - big or small - were calculated during a previous study (20) and used here.

Bioinformatic processing

The raw sequencing data for the ancient Tulán individuals was generated in a previous study and obtained from the authors (20) while the raw sequencing reads for the modern individuals were downloaded from the European Nucleotide Archive (ENA) (Supplementary tables S1 and S2). For all individuals, we removed Illumina adapter sequences, low-quality reads (mean q <25), short reads (<30bp), and merged overlapping read pairs with Fastp v0.23.2 (21). We mapped the processed reads (merged only for the ancient individuals) to the alpaca reference genome (Genbank accession: GCF_000164845.3) using Burrows-wheeler-aligner (BWA) v0.7.15 (22) and either utilising the aln algorithm, with the seed disabled (-l 999) (otherwise default parameters) for the ancient individuals or the mem algorithm with default parameters for the modern individuals. We parsed the alignment files and removed duplicates and reads of mapping quality score <30 using SAMtools v1.6 (23). We checked for ancient DNA damage patterns of the ancient individuals using mapdamage2 (24). As the ancient individuals had also undergone targeted enrichment for mitogenomes and some selected genes, any regions of the alpaca genome aligning to the probes were removed from the bam files using bedtools v2.29.1 (25).

Genetic sex determination

We identified scaffolds putatively originating from the sex chromosomes in the alpaca assembly by aligning the assembly to the Cow X (Genbank accession: CM008168.2) and Human Y (Genbank accession: NC_000024.10) chromosomes using satsuma synteny v2.1 (26) with default parameters. We followed the SeXY pipeline (27) to identify the sex of each of the ancient individuals. We set a minimum threshold of 5,000 mapped reads to ensure the reliability of the results. If an individual had an X:A ratio of <0.7 it was designated as a male. If an individual had an X:A ratio of >0.8 it was designated a female. Individuals with ratios between 0.7 and 0.8 were deemed undetermined.

Ancient DNA Simulation

To assess the impact of ancient DNA damage on our analyses we simulated the damage patterns and read lengths of individual C21 onto a single representative per modern subspecies (six in total) using gargammel v1.1.4 (28). We chose C21 due to it having the highest coverage of the ancient samples. To do this, we generated consensus fasta files from the mapped bam files of the modern individuals using ANGSDv0.921(29) and the following parameters -dofasta 2 -setmindepthind 5 -minq 30 -minmapq 30. The modern individuals included *L.glama* (Llama2), *V.pacos* (Alpaca3), *L.g.cacsilensis* (Cacsilensis3), *L.g.guanicoe* (Guanicoe3), *V.v.mensalis* (Mensalis2), and *V.v.vicugna* (Vicugna1). We mapped the simulated ancient DNA reads to the alpaca reference genome following the same approach as for the ancient Tulán individuals. We additionally downsampled them to ~0.01x using SAMtools.

Principal Component Analysis

We evaluated the relationships between the modern and ancient camelids included in this study by performing PCAs using a pseudohaploid base call with ANGSD. We initially did this for

two different sample sets, either for all modern and ancient individuals >0.01x or all modern and ancient individuals >0.001x. We computed consensus pseudohaploid base calls specifying the parameters: minimum mapping and base qualities of 20 and 30 respectively (-minmapQ 20 -minQ 30), calculate genotype likelihoods using the GATK algorithm (-GL 2), calculate major and minor alleles based on genotype likelihoods (-doMajorMinor 1), remove transitions (-rmtrans 1), only consider autosomal chromosomes > 10Mb (-rf), skip triallelic sites (-skiptriallelic 1), only consider reads mapping to one region uniquely (-uniqueonly 1), make a consensus identity by state base calls (-doIBS 2), output a covariance matrix (-doCov 1) and only consider variable positions where the minor allele occurs in at least 2 individuals (-minminor 2). We also repeated this analysis but with only individuals belonging to the *Lama* genus. We set the minimum individual threshold to the number of modern individuals +1, i.e. 81 for the complete dataset and 57 for the *Lama* only.

Phylogenetic Analysis

To build a neighbour joining tree, we first constructed a distance matrix, which we output from ANGSD when performing PCA (setting the parameter -makematrix 1). We converted the distance matrix into a neighbour joining phylogenetic tree using fastME v2.1.6.1 (30) and default parameters

Ancestry proportions

We calculated the ancestry proportions of eight ancient individuals using admixfrog v0.7.2 (31). We included four big size camelids (C21, C38, C72, C75) and four small size camelids C20, C39, C50, C69. We selected these based on their higher coverage or unexpected placement within the PCA. Individuals C69 and C38 did not correspond to what was expected based on their size and C75 clustered differently compared to the other big sized ancient individuals. We also ran this analysis on the lone individual from Tulán-52 (C46) as it comes from an earlier time period which may be closer to the onset of domestication. Initially we only considered the wild species as the ancestry states, each represented by a single individual; *L.g.cacsilensis* - Cacsilensis4, *L.g.guanicoe* - Guanicoe1, *V.v.mensalis* - Mensalis4, and *V.v.vicugna* - Vicugna2. We repeated the analysis for the ancient *Lama* individuals using the same two *Lama sp.* subspecies as above plus Llama9 as ancestry states. As input for the reference panel we created a multi-individual variant call file with using BCFtools v1.15 (32), specifying autosomal scaffolds >10Mb (-f) and minimum mapping and base qualities of 20 (-q 20 -Q 20). We ran admixfrog specifying a minimum read length of 30bp, a bin size of 50kb, and otherwise default parameters. We only considered a bin in interpretations if it contained more than five SNPs. We pooled regions showing mixed ancestry between *Lama sp.* and *Vicugna sp.* as ancestral camelid in the first run and pooled regions showing any mixed ancestry in the *Lama sp.* only run as ancestral. Furthermore, to assess the accuracy and suitability of this method on our dataset, we also performed the same analysis (all wild (sub)species reference panel) on the modern individuals with simulated ancient DNA damage.

D-statistics

D-statistics calculates shared derived alleles between non-sister branches of a given topology [[[H1, H2], H3], Outgroup]. An allele is considered as derived if it is different to the outgroup allele. Although it is most commonly implemented to investigate differential levels of gene flow between non-sister taxon, it can also be used to investigate shared ancestral polymorphisms and therefore population structure (33, 34).

We subsampled the dataset to include four modern individuals per subspecies and only the eight ancient individuals also used in the ancestry proportions analysis above. We calculated the D-statistics using a random base call approach in ANGSD (-doabbababa 1), specifying only autosomal scaffolds >10Mb and the following parameters: -minmapQ 30 -minQ 30 -blocksize 1000000 -rmtrans 1 -uniqueonly 1. To infer the ancestral sequence (-anc) we used a wild bactrian camel (*Camelus ferus*; NCBI biosample SRR1947250) which we also mapped to the alpaca genome following the approach mentioned above. We summarised the results using a block jackknife approach with the Rscript available in the ANGSD toolsuite (ANGSD_jackknife.R).

We filtered the output into three different sets. First to investigate which genus each ancient individual was most closely related to, we filtered for H1=*Lama sp.* H2=*Vicugna sp.* H3=Ancient individuals. Based on these results we then looked into which subspecies the ancient individuals were most closely related to with either H1=*Lama sp.* H2=*Lama sp.* H3=Ancient *Lama sp.* individuals or H1=*Vicugna sp.* H2=*Vicugna sp.* H3=Ancient *Vicugna sp.* individuals.

To investigate whether the ancient DNA damage could cause biases in the results, we compared results when placing the individuals with simulated ancient DNA damage in H3 with the results generated using the high quality version of the same individual.

Results

Mapping results

After mapping to the alpaca reference genome, all modern samples had genome-wide coverages ranging from 8.34x to 41.11x (Supplementary table S1). As for the ancient individuals, genome-wide coverages ranged between ~0 and 0.28x or 4.5 k and 587M mapped base pairs (Supplementary table S2). The ancient samples showed typical ancient DNA damage patterns with elevated levels of A-G and C-T transitions towards the read ends (Supplementary Fig S1).

Genetic sex determination

Using the SeXY pipeline, we were able to genetically determine the sex of 31 males and 28 females (Supplementary table S2). We were unable to determine the sex of 16 individuals due to either low number of mapping reads or ambiguous results. Further separating into each site, we found: Tulán-52 (1 F), Tulán-54 (23 F and 25 M), Tulán-85 (6 M and 3 F), and Tulán-94 (1 F). Further splitting by taxonomic status, we found identical proportions of male and female *Lama* at Tulán-54, slightly more females than male *Lama* at Tulán-54, and more males than females at Tulán-85 for both genera (Table 1). The one individual that showed the highest proportion of llama ancestry (CL72) was a male.

Table 1: Genetic sexes of the individuals with known taxonomic status. Taxonomic status was based on the palaeogenomic results unless <0.001x then it was based on the osteological measurements - small = *Vicugna*, big = *Lama*.

Site	Tulán-54		Tulán-85	
Genus	*Lama sp.*	*Vicugna sp.*	*Lama sp.*	*Vicugna sp.*
# Males	13	8	4	2
# Females	13	10	2	1
# Total individuals	26	18	6	3
Proportion Males	0.5	0.44	0.67	0.67
Proportion Females	0.5	0.56	0.33	0.33

Principal Component Analysis

When analysing the entire dataset (*Lama sp.* and *Vicugna sp.*), the modern individuals fell into five distinct clusters, in most cases pertaining to their subspecies classification (Fig 2). However, we found *L.g.cacsilensis* and llamas clustering together, apart from two *L.g.cacsilensis* which clustered with *L.g.guanicoe* (Cacsilensis1 and Cacsilensis2). These two individuals are from the outside of the known range of *L.g.cacsilensis* and were therefore likely misidentified (14).

When considering the ancient samples >0.01x, we see three groupings; One closest to *V.v.vicugna* and mostly containing osteometrically small size individuals, one mostly containing osteometrically large size individuals somewhat in the middle of the two modern *Lama sp.* clusters, and a single ancient individual (C72) within the llama/*L.g.cacsilensis* cluster (Fig 2A). When reducing the dataset to only include *Lama sp.* individuals, we again see C72 within the llama/*L.g.cacsilensis* cluster, with the remaining individuals still forming their own cluster, but sitting more closely to the llama/*L.g.cacsilensis* cluster (Fig 2B). Results are similar when considering ancient individuals >0.001x; the cluster closest to *V.v.vicugna* remains but the structuring within *Lama sp.* becomes less clear (Supplementary Fig S2).

The individuals where the osteometric size did not match with their PCA placement included a single big individual (C38), and two small individuals (C69 and C70).

Phylogenetic Analysis

Our neighbour joining tree showed similar results as the PCA (Supplementary Fig S3). There is a split in the ancient individuals, with individuals being most closely related to *V.v.vicugna or* the llama/*L.g.cacsilensis* clade. Only individual C72 was found to be within the llama/*L.g.cacsilensis* clade.

Ancestry proportions

Investigations using the simulated aDNA damaged modern individuals showed a good ability to discern subspecies designation in the wild individuals; the highest proportion of single

subspecies ancestry in each individual was as expected (Supplementary Fig S4). In the llama we saw a relatively high proportion of ancestral camelid DNA (0.11 vs. 0.01 in the wild species) but negligible levels of *Vicugna* ancestry (0.01). However, in the alpaca, we saw a much higher proportion of ancestral camelid DNA (0.44) as well as relatively high levels of *Lama* ancestry (0.13).

The empirical ancestry proportion results were in line with the PCA and phylogenetic tree results (Fig 2C and D). C72 had more *L.g.cacsilensis* ancestry relative to *L.g.guanicoe* compared to the other ancient *Lama sp.* individuals (Fig 2C). When looking into *Lama* sp. specific ancestry, C72 contained relatively higher levels of llama ancestry than the other three investigated individuals (Fig 2D). Overall, the genetically identified *Lama sp.* individuals had more *L.g.cacsilensis* ancestry than *L.g.guanicoe* and *Vicugna sp* (Fig 2C). Putative *Vicugna* individuals have more *V.v.vicugna* ancestry than *V.v.mensalis* (Fig 2C). The small sized C69 contains mostly *L.guanicoe* ancestry, and the big size C38 contains mostly *V.vicugna* ancestry. Only a very small proportion of each individual contained ancestral SAC ancestry <0.04 (Fig 2C).

When using the wild species reference panel, the vast majority of individual C46 from Tulán-52's ancestry could not be determined beyond ancestral *Lama* with the highest proportion of single subspecies ancestry (2%) coming from *L.g.cacsilensis*, the next closest being and only 0.1% from *L.g.guanicoe.* When using the *Lama* only reference panel we found 48.9% of its ancestry as ancestral *Lama*, 50.1% as ancestral *L.g.cacsilensis*/llama, 0.58% as *L.g.cacsilensis* only, 0.25% as llama only, and 0.13% as *L.g.guanicoe* only.

D-statistics

Placing an individual with ancient DNA damage into the H3 position produced comparable results to when placing a high quality version of the same individual in the H3 position (Supplementary Figure S5). Therefore, we could proceed with confidence knowing ancient DNA damage would not bias our results.

D-statistics results using the topology H1=*Lama sp.* H2=*Vicugna sp.* H3=Ancient individuals returned results consistent with the PCA, phylogenetic tree, and ancestry proportion analysis; individuals C75, C72, C89, C21 were more closely related to *Lama sp.* and individuals C50, C39, C38, and C20 were more closely related to *Vicugna sp.* (Fig 3A). Further investigations into the *Lama sp.* individuals showed that most had an overall closer relation to *L.g.cacsilensis* relative to llama or *L.g.guanicoe* (Fig 3B). However, C72 had similar levels of *L.g.cacsilensis and L.glama* ancestry, as seen by few significant results when calculating D-scores using *L.g.cacsilensis and L.glama* in the H1 and H2 position respectively. Further investigations into the *Vicugna sp.* individuals showed an overall closer relation to *V.v.vicugna* relative to alpaca or *V.v.mensalis* (Fig 3C).

Discussion

Through the analysis of palaeogenomic data we were able to infer the ancestry of 49 individuals with mean genome-wide coverages >0.001x from the ceremonial sites of Tulán-52 (Late

Archaic; n=1) and Tulán-54 (Early Formative; n=40) as well as the residential site of Tulán-85 (Early Formative; n=8). We found evidence for both genera, *Vicugna* and *Lama*, in our dataset. Further filtering to 26 individuals >0.01x showed that all individuals, except one, represented wild individuals.

Vicugna individuals belonging to the subspecies *V.v.vicugna* is not unexpected given that the local subspecies of *Vicugna* is *V.v.vicugna* (35). However, *Lama* individuals were more difficult to place. All individuals had a tendency towards *L.g.cacsilensis* and all but one had higher levels of *L.g.guanicoe* ancestry relative to our reference panel. The current distribution indicates that *L.g.cascilensis* are generally found further north but may slightly overlap with *L.g.guanicoe* in Northern Chile (36, 37). Therefore, the ancient Tulán individuals may represent a locally extinct population that hybridised with the local *L.g.guanicoe* individuals. As the one individual placed as most closely related to *L.g.cacsilensis* and llama showed relatively high levels of llama ancestry, we therefore suggest this was an early ancestor of contemporary domesticated llama. Being an early domesticated llama may have also been the case for our one sample from the Late Archaic Tulán-52 individual (C46) which contained mostly *L.g.cacsilensis*/llama ancestry with relatively little presence of *L.g.guanicoe*. There is a cultural continuity between the Late Archaic and the Early Formative, so it is to be expected that these populations continued to develop the domestication / specialisation process. However, the very low coverage (0.008x) nature of this sample prevented us from delving deeper. Taken together, our results suggest individuals that were not the direct ancestors of contemporary domesticated species were the main source of SACs used in both domestic and ceremonial activities in the Tulán ravine during the Early Formative period (3,360–2,370 cal. yr BP).

All but one of the excavated samples of Tulán-54 were previously osteometrically identified as wild individuals (16). However, the large degree of overlap between domestic and wild individuals makes clear results difficult to obtain (38). Taking these difficulties together with the fact that there are other lines of evidence (e.g. artefacts and fibres (39, 40)) supporting the notion that there were domesticated animals in the Tulan ravine since the Late Archaic period (41), suggest the large animals were utilised in a domesticated role despite sharing different ancestry to contemporary llama. As it has been hypothesised that llama domestication was a complex process and may have occurred in several different areas of the Andes (12), these individuals could represent a lineage of domesticates that were lost through time. Lost domesticated lineages that were replaced by contemporary lineages are also known from other domesticated mammals such as pigs and horses (42, 43) so could very well be the case here.

A previous study utilising mitochondrial genomes of the same individuals suggested that individuals whose size did not match their maternal ancestry (e.g. small size but *Lama* ancestry or big size but *Vicugna* ancestry) may have been hybrid individuals, which was interpreted as indicating domesticated individuals (20). Our palaeogenomic analyses find that the hybrid individuals do not contain clearly higher levels of shared ancestry between genera. Therefore, the

mitochondrial genome and osteometry may not be sufficient to determine the hybridisation status of an individual. Furthermore, although previous studies have utilised mitochondrial data from pre-conquest individuals (20, 44, 45), it only represent a single locus, and can be confounded by hybridisation and stochastic events (46) which may lead to difficulties in correctly assigning the overall ancestry of an individual. Instead these hybrid individuals may be due to mislabelling or misidentified juveniles/adults.

Wild vicuñas have one of the most valuable natural fibres in the world that has been appreciated since pre-Hispanic times (6), and is still greatly appreciated today. This may be the reason that most of the small animals in Tulán correspond to vicuñas. The use of camelid fibre is supported by microscopic camelid fibre analyses (47), and while the overwhelming majority have been attributed to light brown natural colours, a small quantity of pure black fleece suggests colour variation, hinting at early colour selection of pre-domesticated camelids (48) which could again suggest a lost domesticate lineage, but in this case for the alpaca. However this could also be the result of dye so should be considered with caution (49).

SACs are sexually monomorphic (50) showing sexual dimorphism when specific bones are available (e.g. pelvis and canine teeth) (45). While it is possible to sex SACs, this can rarely be done due to the fragmentary nature of most assemblages. Therefore, the palaeogenomic data presented a unique opportunity to make inferences on sex. Phalanxes and Astragali have been used in this study, due to their high frequency, completeness, and in the case of the former, as their fusion stage indicates the presence of adult animals (over 24 months) (51). Thus, our results examine only the adult sex ratio in archaeological assemblages. Sex distribution is the result of both hunting and breeding strategies over the time. A typical assemblage resulting from hunting is dominated by the remains of adult animals, with the sexes evenly represented or tilted toward males. A typical assemblage resulting from herding is dominated by the remains of immature animals, with the adult sex ratio heavily skewed toward females (52). Therefore, the similar proportions of females and males when comparing by site and by genus (Table 1) suggests hunting was the main source of the camelids found at these sites, although the sample size at the Tulán-85 site is low and should be interpreted with caution. This result is in line with our other palaeogenomic results showing that the vast majority of individuals at these sites were wild. However, as male llama were also used as pack animals, our finding of more males than expected could suggest transport between sites was important. In fact the individual that showed the highest proportion of llama ancestry (CL72) was a male and therefore may have been involved in transport to some degree.

Acknowledgements:
We would like to thank Anders Hansen for support generating the sequencing data and Alba Rey-Iglesia for support in curating the data.

Data availability

Raw sequencing data will be available upon acceptance. Coordinates of the capture probes in the alpaca reference genome can be downloaded from https://sid.erda.dk/share_redirect/BzxbWhZqWB

References:

1. G. Larson, D. Q. Fuller, The Evolution of Animal Domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).

2. L. A. F. Frantz, D. G. Bradley, G. Larson, L. Orlando, Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* **21**, 449–460 (2020).

3. J. C. Wheeler, A. J. F. Russel, H. Redden, Llamas and Alpacas: Pre-conquest breeds and post-conquest hybrids. *J. Archaeol. Sci.* **22**, 833–840 (1995).

4. B. Hesse, Archaeological evidence for camelid exploitation in the Chilean Andes. *Säugetierkundliche Mitteilungen* **30**, 201–211 (1982).

5. K. L. Broughton, South American camelids in Central Andean religious practices. (2010).

6. B. Vilá, Y. Arzamendia, South American Camelids: their values and contributions to people. *Sustainability Sci.* 1–18 (2020).

7. G. L. M. Goñalons, H. D. Yacobaccio, The domestication of South American camelids. *Documenting Domestication--New Genetic and Archaeological Paradigms* 228–244 (2006).

8. I. Cartajena Fasting, Explorando la variabilidad morfométrica del conjunto de camélidos pequeños durante el Arcaico Tardío y el Formativo Temprano en Quebrada Tulán, norte de Chile. *Revista del Museo de Antropología* **2**, 199–212 (2009).

9. K. M. Moore, “9. Grace Under Pressure: Responses to Changing Environments by Herders and Fishers in the Formative Lake Titicaca Basin, Bolivia” in *Sustainable Lifeways*, (University of Pennsylvania Press, 2012), pp. 244–272.

10. K. M. Moore, Early domesticated camelids in the Andes. *The archaeology of Andean pastoralism* (2016).

11. G. L. Mengoni Goñalons, Camelids in ancient Andean societies: A review of the zooarchaeological evidence. *Quat. Int.* **185**, 59–68 (2008).

12. H. D. Yacobaccio, The domestication of South American camelids: a review. *Anim Front* **11**, 43–51 (2021).

13. M. A. Zeder, E. Emshwiller, B. D. Smith, D. G. Bradley, Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* **22**, 139–155 (2006).

14. R. Fan, *et al.*, Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. *Genome Biol.* **21**, 159 (2020).

15. L. Nuñez A, I. Cartajena F, C. Carrasco G, P. de Souza H, M. Grosjean, Emergencia de comunidades pastoralistas formativas en el sureste de la Puna de Atacama. *Estud. atacameños* 93–117 (2006).

16. L. Núñez, *et al.*, The temple of Tulán-54: Early Formative ceremonial architecture in the Atacama Desert. *Antiquity* **91**, 901–915 (2017).

17. S. P. De, I. Cartajena, L. Núñez, C. Carrasco, Cazadores-recolectores del Arcaico Tardío y desarrollo de complejidad social en la puna de Atacama: las evidencias del sitio Tulan-52 (norte árido de Chile). *Revista Werkén* **13**, 91–118 (2010).

18. I. Cartajena, P. L. Mendoza, B. Santander, L. Núñez, "Discarding practices among early herders on the western watershed of the Puna de Atacama, Chile (ca. 3000-2300 cal BP)" in *Animals: Cultural Identifiers in Ancient Societies?*, J. Peters, G. McGlynn, V. Goebel, Eds. (Documenta Archaeobiologiae 15 VML Verlag Marie Leidorf, 2019), pp. 65–84.

19. P. Lopez, I. Cartajena, L. Nunez, Analyses of stable isotopes in camelids collagen bones from Tulan Ravine, Atacama Puna, early formative period (CA 3, 1000-2,400 BP). *Revista de antropologia chilena* **45**, 237–247 (2013).

20. P. Diaz-Maroto, *et al.*, Ancient DNA reveals the lost domestication history of South American camelids in Northern Chile and across the Andes. *eLife* **10**, e63390 (2021).

21. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

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

23. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

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

25. A. R. Quinlan, BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12.1–34 (2014).

26. M. G. Grabherr, *et al.*, Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**, 1145–1151 (2010).

27. A. A. Cabrera, *et al.*, How low can you go? Introducing SeXY: sex identification from low-quantity sequencing data despite lacking assembled sex chromosomes. *Ecol. Evol.* **12**, e9185 (2022).

28. G. Renaud, K. Hanghøj, E. Willerslev, L. Orlando, gargammel: a sequence simulator for ancient DNA. *Bioinformatics* **33**, 577–579 (2017).

29. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

30. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).

31. B. M. Peter, 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. *bioRxiv* (2020).

32. V. Narasimhan, *et al.*, BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).

33. M. V. Westbury, *et al.*, Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. *Mol. Biol. Evol.* **35**, 1225–1237 (2018).

34. T. Segawa, A. Rey-Iglesia, E. D. Lorenzen, M. V. Westbury, The origins and diversification of Holarctic brown bear populations inferred from genomes of past and present populations. *Proc. Biol. Sci.* **291** (2024).

35. B. A. González, *et al.*, Phylogeography and Population Genetics of Vicugna vicugna: Evolution in the Arid Andean High Plateau. *Front. Genet.* **10** (2019).

36. F. León, *et al.*, History of diversification and adaptation from north to south revealed by genomic data: Guanacos from the desert to sub-Antarctica. *Genome Biol. Evol.* **16**, evae085 (2024).

37. B. A. González, H. Samaniego, J. C. Marín, C. F. Estades, Unveiling current Guanaco distribution in chile based upon niche structure of phylogeographic lineages: Andean puna to subpolar forests. *PLoS One* **8**, e78894 (2013).

38. P. López M., I. Cartajena, R. Loyola, L. Núñez, C. Carrasco, The use of hunting and herding spaces: Stable isotope analysis of Late Archaic and Early Formative camelids in the Tulan transect (Puna de Atacama, Chile): Stable isotopes analyses on camelids in tulan transect (Chile). *Int. J. Osteoarchaeol.* **27**, 1059–1069 (2017).

39. G. Arias, A. Benavente, P. Gecele, Identificación y variabilidad del uso del animal a través de textiles arqueológicos: Contraste con patrones fanéreos actuales in *Actas Del Xii Congreso Nacional de Arqueología Chilena*, (1993), pp. 151–162.

40. I. Cartajena, M. A. Benavente, L. Núñez, C. Thomas, La utilización de los camélidos durante el Formativo Temprano: una comparación entre la cuenca del Loa medio y el Salar de Atacama. *Zooarqueología y Tafonomía en el Confín del Mundo* 45–58 (2009).

41. R. Loyola, *et al.*, Hunting and feasting in the pre-Columbian Andes: Exploring the nature and scale of early ceremonial aggregations in Tulan Ravine (5300 to 2400 yr cal. BP) through the circulation of obsidian artefacts. *J. Anthropol. Archaeol.* **64**, 101360 (2021).

42. L. A. F. Frantz, *et al.*, Ancient pigs reveal a near-complete genomic turnover following their

introduction to Europe. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 17231–17238 (2019).

43. P. Librado, *et al.*, Widespread horse-based mobility arose around 2200 BCE in Eurasia. *Nature* **631**, 819–825 (2024).

44. C. C. Abbona, *et al.*, Were domestic camelids present on the prehispanic South American agricultural frontier? An ancient DNA study. *PLoS One* **15**, e0240474 (2020).

45. M. Westbury, *et al.*, First complete mitochondrial genome data from ancient South American camelids - The mystery of the chilihueques from Isla Mocha (Chile). *Sci. Rep.* **6**, 38708 (2016).

46. M. V. Westbury, *et al.*, Hyena paleogenomes reveal a complex evolutionary history of cross-continental gene flow between spotted and cave hyena. *Science Advances* **6**, eaay0456 (2020).

47. M. A. Benavente, Análisis lanimétrico de fanéreos de los sitios Tulán-52 y 54. *Ms Informe Final Proyecto FONDECYT* **1020316** (2005).

48. P. Dransart, Llamas, herders and the exploitation of raw materials in the Atacama Desert. *World Archaeol.* **22**, 304–319 (1991).

49. P. Dransart, Camelid Fleece and the Hidden Histories of Colors in the Andes. (2020).

50. G. C. Miranda-de la Lama, M. Villarroel, Behavioural biology of South American domestic camelids: An overview from a welfare perspective. *Small Rumin. Res.* **220**, 106918 (2023).

51. J. D. Kent, The Domestication and exploitation of the South American camelids: methods of analysis and their application to circum-lacustrine archaeological sites in Bolivia and Peru. (1982).

52. N. Russell, L. Martin, H. Buitenhuis, Cattle domestication at Çatalhöyük revisited. *Curr. Anthropol.* **46**, S101–S108 (2005).

Figures:

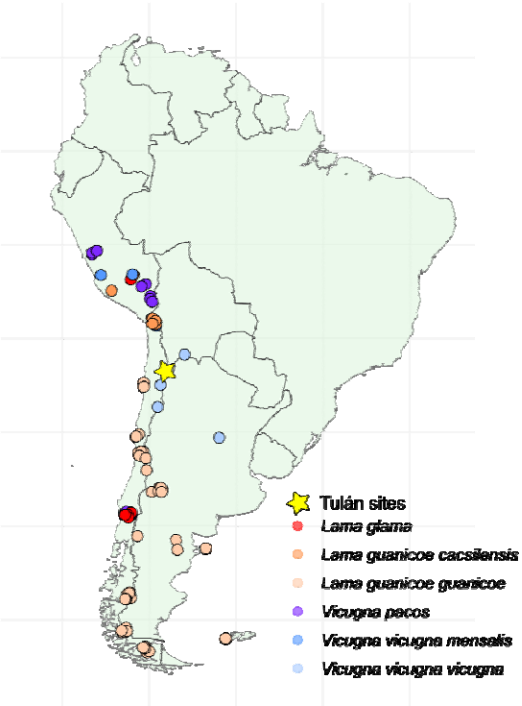


Figure 1: Map indicating the location of the Tulán archaeological sites and the sampling location of the modern South American camelid individuals used in this study.

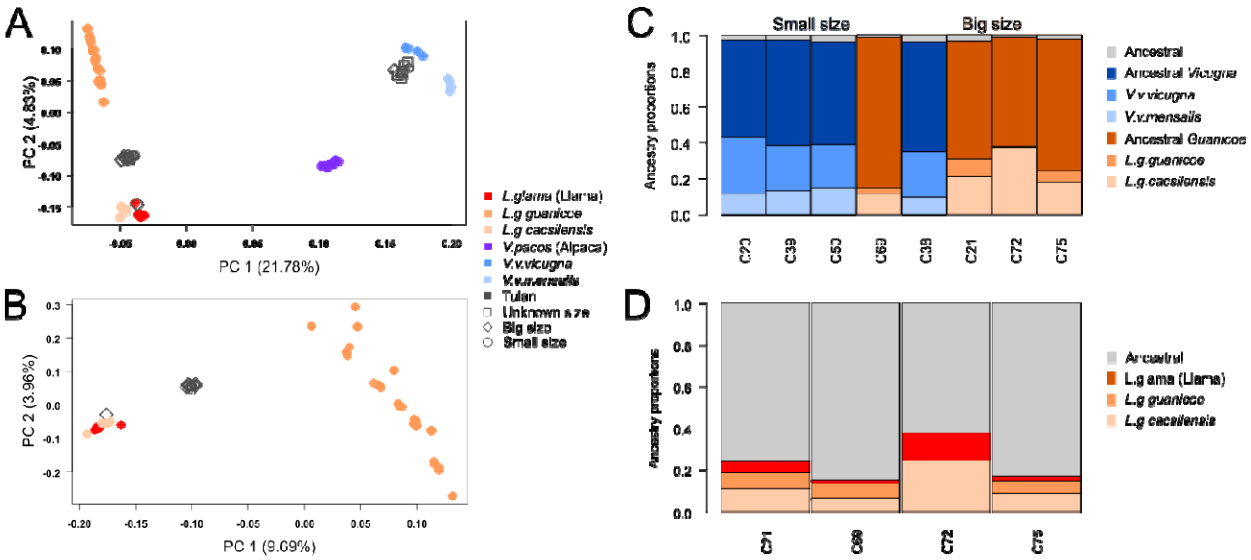


Figure 2: Relationships between modern and ancient individuals. A) PCA performed on all modern individuals and ancient individuals >0.01x. B) PCA performed on modern *Lama*

individuals and ancient *Lama* individuals > 0.01x. C) Ancestry states of eight ancient individuals. Ancestry states are derived from a single individual; *L.g.cacsilensis* - Cacsilensis4, *L.g.guanicoe* - Guanicoe1, *V.v.mensalis* - Mensalis4, and *V.v.vicugna.* Ancestral is defined as having combined ancestry between both *Lama* and *Vicugna* (C) or mixed ancestry between *Lama* (sub)species (D).

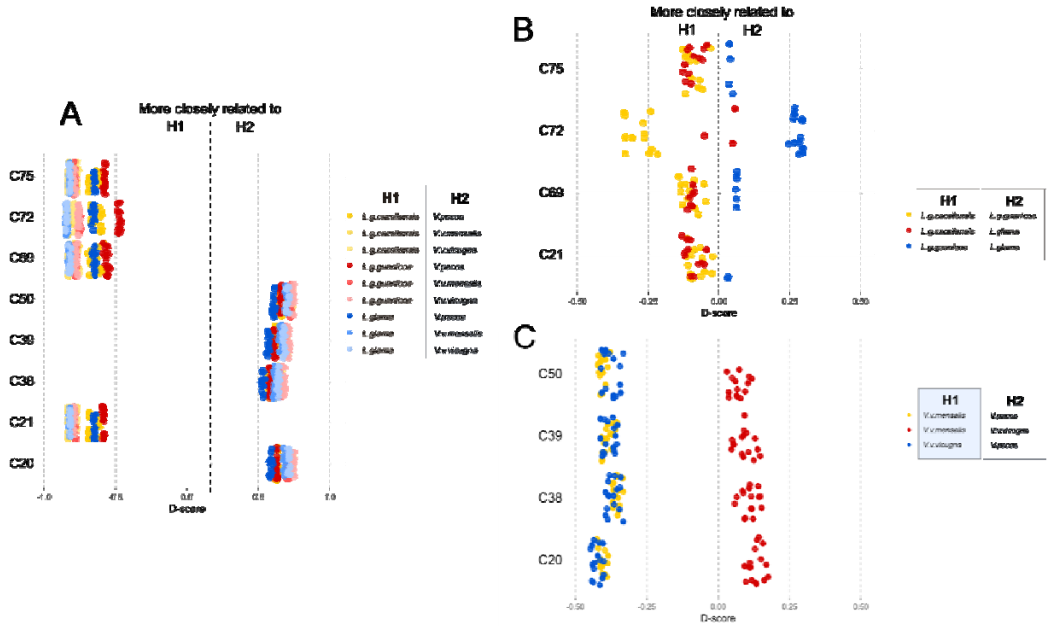
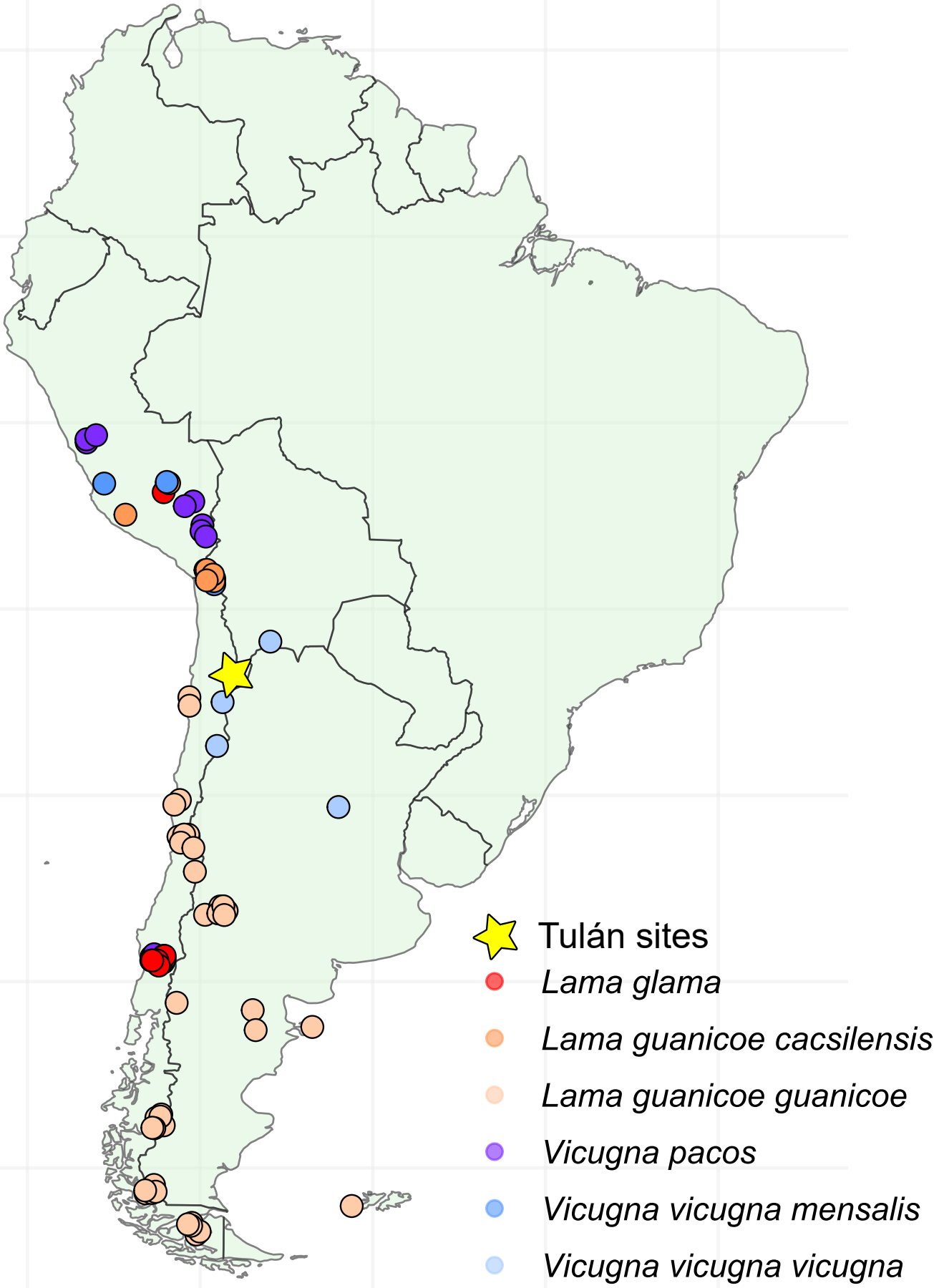
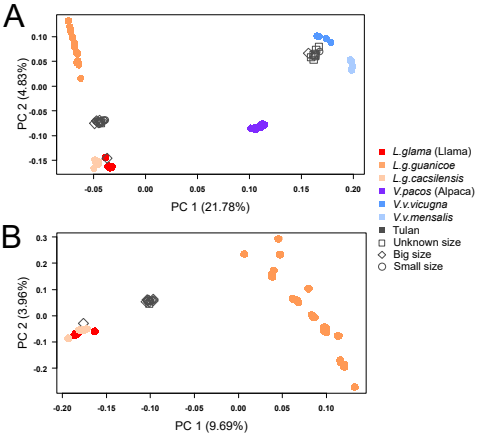
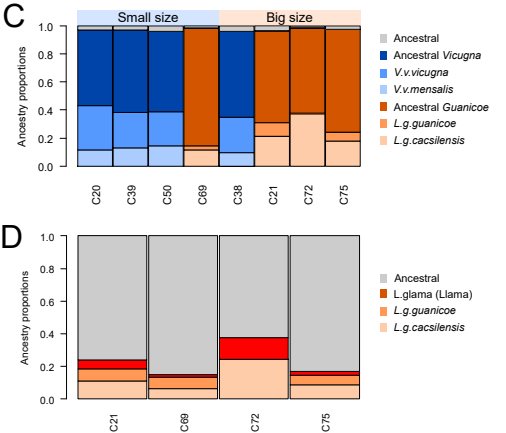
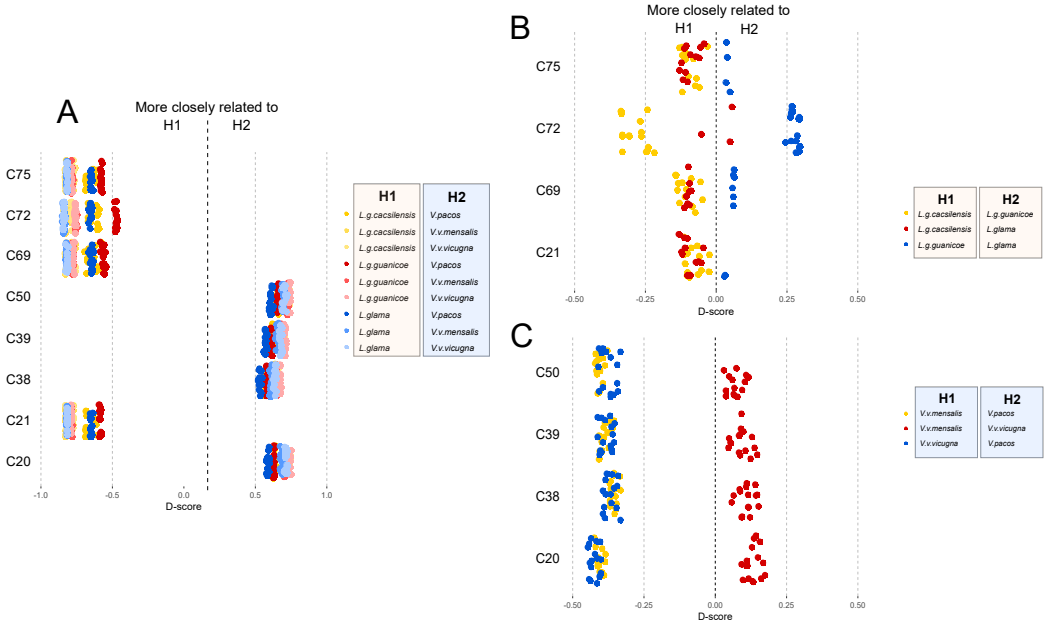


Figure 3: Test for closer evolutionary relationships between our ancient samples and the modern (sub)species using D-statistics and various input topologies. A) Comparisons between *Lama* and *Vicugna* B) Comparisons within *Lama* C) Comparisons within *Vicugna*. Non-significant values |Z|<3 not shown.


Tulán sites
Lama glama
Lama guanicoe cacsilensis
Lama guanicoe guanicoe
Vicugna pacos
Vicugna vicugna mensalis
Vicugna vicugna vicugna


A
PC 2 (4.83%)
PC 1 (21.78%)
B
PC 2 (3.96%)
PC 1 (9.69%)
L.glama (Llama)
L.g.guanicoe
L.g.cacsilensis
V.pacos (Alpaca)
V.v.vicugna
V.v.mensalis
Tulan
Unknown size
Big size
Small size


C
Small size
Big size
Ancestry proportions
C20
C39
C50
C69
C38
C21
C72
C75
Ancestral
Ancestral Vicugna
V.v.vicugna
V.v.mensalis
Ancestral Guanicoe
L.g.guanicoe
L.g.cacsilensis
D
Ancestry proportions
C21
C69
C72
C75
Ancestral
L.glama (Llama)
L.g.guanicoe
L.g.cacsilensis


A
More closely related to
H1
H2
C75
C72
C69
C50
C39
C38
C21
C20
-1.0
-0.5
0.0
0.5
1.0
D-score
H1
H2
L.g.cacsilensis
L.g.cacsilensis
L.g.cacsilensis
L.g.guanicoe
L.g.guanicoe
L.g.guanicoe
L.glama
L.glama
L.glama
V.pacos
V.v.mensalis
V.v.vicugna
V.pacos
V.v.mensalis
V.v.vicugna
V.pacos
V.v.mensalis
V.v.vicugna
B
More closely related to
H1
H2
C75
C72
C69
C21
-0.50
-0.25
0.00
0.25
0.50
D-score
H1
H2
L.g.cacsilensis
L.g.cacsilensis
L.g.guanicoe
L.g.guanicoe
L.glama
L.glama
C
C50
C39
C38
C20
-0.50
-0.25
0.00
0.25
0.50
D-score
H1
H2
V.v.mensalis
V.v.mensalis
V.v.vicugna
V.pacos
V.v.vicugna
V.pacos