

RESEARCH

Open Access



Genome-wide scan for runs of homozygosity in South American Camelids

Stefano Pallotti^{1*}, Matteo Picciolini², Marco Antonini³, Carlo Renieri⁴ and Valerio Napolioni¹

Abstract

Background Alpaca (*Vicugna pacos*), llama (*Lama glama*), vicugna (*Vicugna vicugna*) and guanaco (*Lama guanicoe*), are the camelid species distributed over the Andean high-altitude grasslands, the Altiplano, and the Patagonian arid steppes. Despite the wide interest on these animals, most of the loci under selection are still unknown. Using whole-genome sequencing (WGS) data we investigated the occurrence and the distribution of Runs Of Homozygosity (ROHs) across the South American Camelids (SACs) genome to identify the genetic relationship between the four species and the potential signatures of selection.

Results A total of 37 WGS samples covering the four species was included in the final analysis. The multi-dimensional scaling approach showed a clear separation between the four species; however, admixture analysis suggested a strong genetic introgression from vicugna and llama to alpaca. Conversely, very low genetic admixture of the guanaco with the other SACs was found. The four species did not show significant differences in the number, length of ROHs (100-500 kb) and genomic inbreeding values. Longer ROHs (> 500 kb) were found almost exclusively in alpaca. Seven overlapping ROHs were shared by alpacas, encompassing nine loci (*FGF5*, *LOC107034918*, *PRDM8*, *ANTXR2*, *LOC102534792*, *BSN*, *LOC116284892*, *DAG1* and *RIC8B*) while nine overlapping ROHs were found in llama with twenty-five loci annotated (*ERC2*, *FZD9*, *BAZ1B*, *BCL7B*, *LOC116284208*, *TBL2*, *MLXIPL*, *PHF20*, *TRNAD-AUC*, *LOC116284365*, *RBM39*, *ARFGF2*, *DCAF5*, *EXD2*, *HSPB11*, *LRRC42*, *LDLRAD1*, *TMEM59*, *LOC107033213*, *TCEANC2*, *LOC102545169*, *LOC116278408*, *SMIM15*, *NDUFAF2* and *RCOR1*). Four overlapping ROHs, with three annotated loci (*DLG1*, *KAT6B* and *PDE4D*) and three overlapping ROHs, with seven annotated genes (*ATP6V1E1*, *BCL2L13*, *LOC116276952*, *BID*, *KAT6B*, *LOC116282667* and *LOC107034552*), were detected for vicugna and guanaco, respectively.

Conclusions The signatures of selection revealed genomic areas potentially selected for production traits as well as for natural adaptation to harsh environment. Alpaca and llama hint a selection driven by environment as well as by farming purpose while vicugna and guanaco showed selection signals for adaptation to harsh environment. Interesting, signatures of selection on *KAT6B* gene were identified for both vicugna and guanaco, suggesting a positive effect on wild populations fitness. Such information may be of interest to further ecological and animal production studies.

Keywords Alpaca, Llama, Vicugna, Guanaco, South American camelids, Runs of homozygosity, Signatures of selection, Domestication

*Correspondence:

Stefano Pallotti
stefano.pallotti@unicam.it

¹Genomic And Molecular Epidemiology (GAME) Lab, School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

²SYNBIOTEC Laboratori s.r.l, Camerino, Italy

³Italian National Agency for New Technologies, Energy and Sustainable Development (ENEA), Roma, Italy

⁴School of Pharmacy and Health Products, University of Camerino, Camerino, Italy



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Four species of South American Camelids (SACs) are distributed over the Andean high-altitude grasslands, the Altiplano, and the Patagonian arid steppes [1]. These include two wild species, vicugna (*Vicugna vicugna*) and guanaco (*Lama guanicoe*), and two domestic species, alpaca (*Vicugna pacos*) and llama (*Lama glama*) [1].

Vicugna (*Vicugna vicugna*) is a wild SAC producing a very fine fibre highly demanded by international market of luxury goods [2]. The wild population underwent a strong reduction due to intense hunting since the Spanish conquest: in the 1960s, the species was at risk of extinction as the number of specimens dropped from 2 million to less than 10,000 [3]. After decades of protection policies, vicugna population recovered [3], although controlled catches are still allowed.

The other wild SAC species is the guanaco (*Lama guanicoe*), a large herbivore whose population size declined from around 30 million to 5,000 specimens during the past century [4]. Such dropping started at the end of the 19th century when European sheep farmers colonized the Patagonia region killing both indigenous groups and guanacos due to their perceived role as competitors for sheep pasture [1]. Nowadays, guanacos concentrate in few low stock or sheep-free areas [1]. As occurred with the vicugna, preventive conservation strategies were adopted, although controlled capture and shearing of the wild animals are still allowed [4].

Alpaca (*Vicugna pacos*) is a SAC domesticated around 9,000 years ago and it has been associated with humans for as long as cattle, horses, and dogs [5]. Being an important resource for Andean highlands people, alpaca is bred by Peruvian farmer for meat and fibre production [6–8]. Several hypotheses were made concerning its ancestry [1]; indeed, it has been hypothesized that alpaca descended from the vicugna [9], the guanaco [10] or as hybrid between the vicugna and the llama [11].

The other SAC domesticated by indigenous communities is llama (*Lama glama*); this species was domesticated 5,000 years ago, starting from the wild guanacos [1]. Like the alpaca, these animals are bred for meat [12] and fibre production [13], being an important economic source for the Andean population.

Despite the extensive literature on SACs, most of the loci under selection are still unknown. Indeed, both natural and artificial selection shapes genetic variation across the genomes, rising the frequency of favorable alleles and specific haplotypes over time. This process leads to the establishment of genomic region with high differentiation across breeds and species known as “selection signatures”: the analysis of these regions allows to identify loci deviating from neutrality [14].

A well-established method to detect selection signatures is the genome-wide scan for Runs Of Homozygosity

(ROHs): these are contiguous lengths of homozygous segments in the genome where the two haplotypes inherited from the parents are identical [15] and arise from a single common ancestor [16]. Genomic regions harboring selection signatures often overlap with shared ROHs regions within and across populations [17]. Moreover, the analysis of ROHs represents the state-of-the-art method for population inbreeding analyses [16].

Recently, important insights on SACs domestication using F_{ST} outliers and extended haplotype homozygosity analysis were provided by Fan and colleagues [11]. However, the identification of ROHs in SACs was never undertaken. This is mainly due to the lack of availability of genomic data. Indeed, recent efforts were only made to evaluate either the use of Bovine SNP-chip [18] or to design ad-hoc SNP-chip [19] for performing genomic studies in alpaca, thus further highlighting the need of genomic studies on SACs. However, the advent of next-generation sequencing led to an enormous amount of genomic data from several species, freely available for evolutionary and zootechnical research [20].

Starting from the whole-genome sequencing (WGS) data of 30 SACs retrieved from publicly available repositories, along with the *de novo* WGS of seven alpacas, we investigated the occurrence and the distribution of ROHs across the genome of SACs. Thus, we aim to unveil genomic selection signatures in the SACs, providing new information on the domestication and the potential convergent/divergent selection underwent by these species.

Results

Population structure

Multi-dimensional scaling (MDS) was used to visualize the relationships between the 37 SACs included in the analysis; the plot showed a clear separation between the four species (Fig. 1).

Admixture was computed by running the analysis for 2- to 4- clusters (K) (Fig. 2). The cross validation (CV) error test for each K value was performed to determine the most probable number of clusters. The CV errors were 0.648, 0.680, 0.751, and 0.846 for K=2, 3, 4, and 5, respectively.

Runs of homozygosity (ROHs) and genomic inbreeding

(F_{ROH})

First, we evaluated the presence of ROHs with length 100–500 kb (Table 1). The average number of 100 kb ROHs found in the four species was 445 (± 175) for alpaca, 451 (± 287) for the guanaco, 315 (± 117) for llama and 489 (± 200) for vicugna. Concerning the length of the segments, alpacas showed the longest ROHs with an average length (in kilobases, kb) of 145,915 ($\pm 15,960$), followed by vicugna, (138,710 $\pm 2,630$), llama (136,748 $\pm 2,840$) and guanaco (134,830 $\pm 4,750$). Thus, we evaluated the

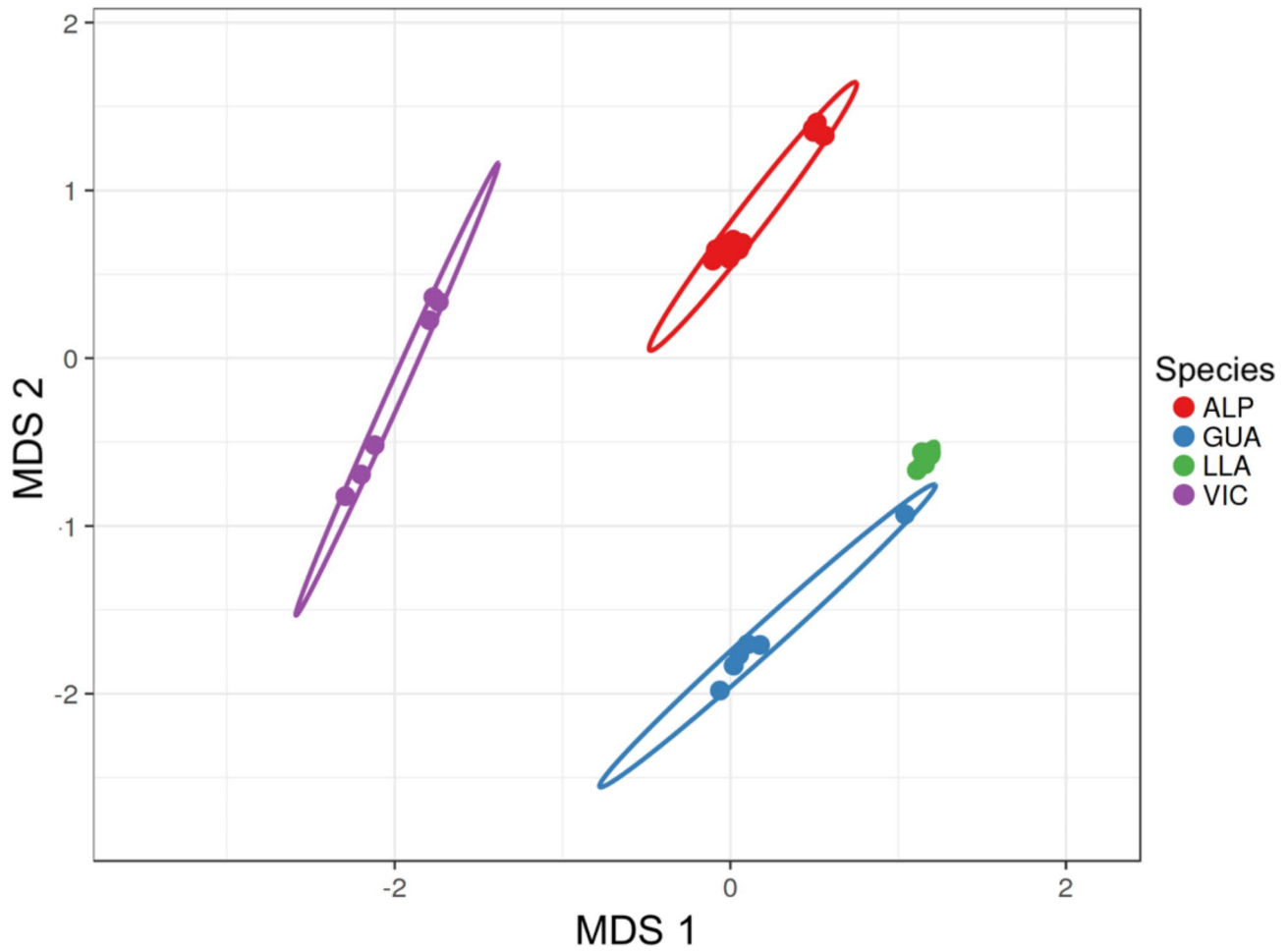


Fig. 1 MDS plot of four South American camelids plotted for the first 2 dimensions. Each dot represents an individual sample (ALP, alpaca; GUA, guanaco; LLA, llama; VIC, vicugna)

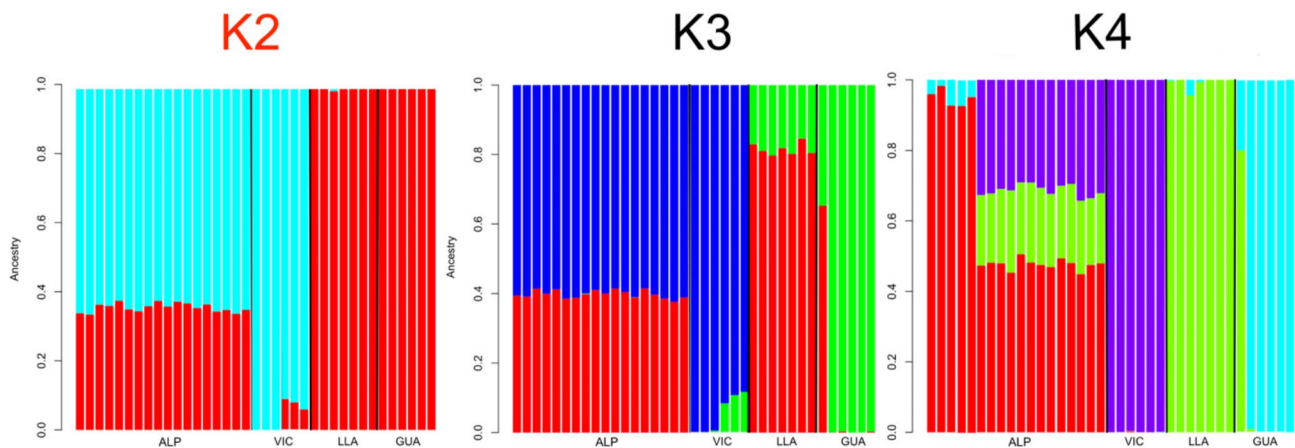


Fig. 2 Admixture results for K=2–4. The optimal number of clusters (K=2), according to the cross-validation analysis, is indicated in red (ALP, alpaca; VIC, vicugna; LLA, llama; GUA, guanaco)

presence of longer ROHs (length > 500 kb) (Table 2); these were found only in 6 out of 18 alpacas, with an average length of 580,830 ($\pm 53,62$). Guanaco and llama did not show ROHs > 500 kb; however, only one vicugna showed

a single ROH with a length of 517,702 kb. ANOVA did not show any significant differences in terms of ROHs number (P-value=0.369) and length (P-value=0.134) between the four species (Table 1).

Table 1 Distribution of ROHs and F_{ROH} (100–500 kb) in South American camelids

| Species | N | Number of ROHs | | Total kilobase distance spanned by segments | | F_{ROH} | | Number of variants in the population | Segment size (kb) | |
|---------|----|----------------|---------|---|------------------|--------------------|-------------|--------------------------------------|-----------------------|-------------------|
| | | mean | min-max | mean | min-max | mean | min-max | | mean | min-max |
| Alpaca | 18 | 445 (± 175) | 264–803 | 67,477 (± 34,829) | 31,605 – 141,769 | 0.043 (± 0.022) | 0.020–0.090 | 25,526,240 | 145,915 (± 15,957) | 133,186 – 176,550 |
| Llama | 7 | 315 (± 117) | 218–566 | 61,733 (± 41,698) | 28,622 – 78,747 | 0.028 (± 0.011) | 0.018–0.050 | 14,153,719 | 136,749 (± 2,840) | 131,292 – 139,651 |
| Vicugna | 6 | 489 (± 200) | 267–772 | 43,226 (± 16,600) | 36,585 – 107,703 | 0.043 (± 0.018) | 0.023–0.069 | 17,145,639 | 138,705 (± 2,629) | 137,024–143,715 |
| Guanaco | 6 | 451 (± 287) | 208–986 | 68,092 (± 28,883) | 26,548 – 140,133 | 0.039 (± 0.027) | 0.017–0.089 | 14,793,242 | 134,826 (± 4,752) | 127,633 – 142,122 |

N: number of animals

Table 2 Distribution of ROHs and F_{ROH} >500 kb in South American camelids

| Species | N | Number of ROHs | | Total kilobase distance spanned by segments | | Segment size (kb) | |
|---------|---|-----------------|---------|---|-------------------|-----------------------|-------------------|
| | | mean | min-max | mean | min-max | mean | min-max |
| Alpaca | 6 | 5,7 (± 4,89) | 1–15 | 3,369,405 (± 2956,820) | 508,712–8,987,700 | 580,829 (± 53,616) | 508,712 – 651,219 |
| Vicugna | 1 | 1 | - | 517,702 | - | 517,702 | - |

N: number of animals

The inbreeding value based on ROHs > 100 kb was 0.043 (± 0.022) for alpaca and vicugna, 0.039 (± 0.027) for guanaco and 0.028 (± 0.011) for llama (Table 1). ANOVA did not show significant differences in terms of F_{ROH} values between the four populations (P-value = 0.399).

Signatures of selection: overlapping ROHs

To identify potential signatures of selection, we focused on the ROHs shared by at least 70% of the population within each species (Table 3). Seven overlapping ROHs were shared by alpacas, with four located on chromosome NW_021964157.1 (chromosome 2), two on chromosome NW_021964193.1 (chromosome 17) and one on chromosome NW_021964178.1 (chromosome 12). These ROHs encompass nine loci (*FGF5*, *LOC107034918*, *PRDM8*, *ANTXR2*, *LOC102534792*, *BSN*, *LOC116284892*, *DAG1* and *RIC8B*). Nine overlapping ROHs were found in llama: three are located on chromosome NW_021964165.1 (chromosome 6), two on chromosome NW_021964196.1 (chromosome 19) and one in each of the remaining four chromosomes NW_021964160.1 (chromosome 3), NW_021964182.1 (chromosome 13), NW_021964193.1 (chromosome 17) and NW_021964195.1 (chromosome 18). A total of twenty-five loci were annotated for llama (*ERC2*, *FZD9*, *BAZ1B*, *BCL7B*, *LOC116284208*, *TBL2*, *MLXIPL*, *PHF20*, *TRNAD-AUC*, *LOC116284365*, *RBM39*, *ARFGF2*, *DCAF5*, *EXD2*, *HSPB11*, *LRRC42*, *LDLRAD1*, *TMEM59*, *LOC107033213*, *TCEANC2*, *LOC102545169*, *LOC116278408*, *SMIM15*, *NDUFAF2* and *RCOR1*). 83% of the vicugna showed four overlapping ROHs; two of them were located on chromosome NW_021964156.1 (chromosome 1), one on chromosome

NW_021964160.1 (chromosome 3) and on chromosome NW_021964175.1 (chromosome 11). Three loci were annotated (*DLG1*, *KAT6B* and *PDE4D*). Guanacos showed 3 overlapping ROHs located on chromosome NW_021964175.1 (chromosome 11), NW_021964178.1 (chromosome 12) and NW_021964223.1 (chromosome 34), respectively, with seven annotated loci (*ATP6V1E1*, *BCL2L13*, *LOC116276952*, *BID*, *KAT6B*, *LOC116282667* and *LOC107034552*).

The enrichment analysis failed to identify significantly involved molecular and functional pathways.

Discussion

Genetic relationships between the four South American Camelids species

Several studies have been performed to better understand the SACs ancestry and their genetic relationship, and diverse hypotheses on the potential crossbreeding between the four species have been made [1, 9–11]. The MDS analysis showed a clear separation between the four species, with llama and guanaco close to each other respect to the alpaca and vicugna population. One guanaco clustered with the llama group. This was also observed in the admixture analysis suggesting a hybridization between the two species. To further study the population structure of SACs, admixture analysis was carried out. Recently, starting from WGS of eight alpacas, Fan and colleagues [11] found strong genetic admixture of the alpaca with vicugna and llama. In our study the ancestry of alpaca was studied considering 18 alpacas. By running admixture analysis for K2–4, the cross-validation error test suggested that K2 was the most probable value

Table 3 Signatures of selection: overlapping ROHs ($\geq 70\%$) in each population

| SPECIES | % of population that present the ROHs | N animals presenting the ROHs | Chr | Bp1 | Bp2 | Kb | nVariants | Genes |
|---------|---------------------------------------|-------------------------------|---------------------|------------|------------|---------|-----------|---|
| Alpaca | 89 | 16 | 2 (NW_021964157.1) | 47,004,014 | 47,097,954 | 93.941 | 1,864 | <i>FGF5, LOC107034918, PRDM8</i> |
| | 83 | 15 | 2 (NW_021964157.1) | 47,223,830 | 47,288,964 | 65.135 | 1,095 | <i>ANTXR2, LOC102534792</i> |
| | 78 | 14 | 17 (NW_021964193.1) | 32,857,799 | 32,908,888 | 51.09 | 743 | <i>BSN</i> |
| | 72 | 13 | 2 (NW_021964157.1) | 47,109,781 | 47,116,834 | 7.054 | 119 | --- |
| | 72 | 13 | 2 (NW_021964157.1) | 47,301,062 | 47,372,080 | 71.019 | 1,462 | <i>ANTXR2, LOC116284892</i> |
| | 72 | 13 | 17 (NW_021964193.1) | 32,961,329 | 32,990,407 | 29.079 | 297 | <i>DAG1</i> |
| Llama | 86 | 6 | 12 (NW_021964178.1) | 18,429,922 | 18,453,728 | 23.807 | 279 | <i>RIC8B</i> |
| | 86 | 6 | 17 (NW_021964193.1) | 28,016,567 | 28,120,792 | 104.226 | 1,256 | <i>ERC2</i> |
| | 86 | 6 | 18 (NW_021964195.1) | 1,773,096 | 1,897,672 | 124.577 | 1,992 | <i>FZD9, BAZ1B, BCL7B, LOC116284208, TBL2, MLXIPL</i> |
| | 86 | 6 | 19 (NW_021964196.1) | 20,312,362 | 20,426,143 | 113.782 | 1,814 | <i>PHF20, TRNAD-AUC, LOC116284365, RBM39</i> |
| | 71 | 5 | 6 (NW_021964165.1) | 62,144,615 | 62,147,491 | 2.877 | 65 | --- |
| | 71 | 5 | 19 (NW_021964196.1) | 10,836,071 | 10,932,799 | 96.729 | 2,295 | <i>ARFGF2</i> |
| | 71 | 5 | 6 (NW_021964165.1) | 35,341,921 | 35,425,124 | 83.204 | 1,200 | <i>DCAF5, EXD2</i> |
| | 71 | 5 | 13 (NW_021964182.1) | 8,027,836 | 8,128,031 | 100.196 | 2,000 | <i>HSPB11, LRRC42, LDLRAD1, TMEM59, LOC107033213, TCEANC2, LOC102545169</i> |
| | 71 | 5 | 3 (NW_021964160.1) | 24,469,883 | 24,598,961 | 129.079 | 1,901 | <i>LOC116278408, SMIM15, NDUFAF2</i> |
| | 71 | 5 | 6 (NW_021964165.1) | 62,072,906 | 62,144,562 | 71.657 | 1,065 | <i>RCOR1</i> |
| Vicugna | 83 | 5 | 1 (NW_021964156.1) | 17,067,190 | 17,152,248 | 85.059 | 1,422 | --- |
| | 83 | 5 | 1 (NW_021964156.1) | 17,152,557 | 17,228,106 | 75.55 | 1,304 | <i>DLG1</i> |
| | 83 | 5 | 11 (NW_021964175.1) | 7,005,163 | 7,020,809 | 15.647 | 231 | <i>KAT6B</i> |
| | 83 | 5 | 3 (NW_021964160.1) | 25,397,227 | 25,469,962 | 72.736 | 1,116 | <i>PDEAD</i> |
| Guanaco | 83 | 5 | 34 (NW_021964223.1) | 21,760,321 | 21,837,816 | 77.496 | 1,406 | <i>ATP6V1E1, BCL2L13, LOC116276952, BID</i> |
| | 83 | 5 | 11 (NW_021964175.1) | 6,977,685 | 7,065,977 | 88.293 | 1,310 | <i>KAT6B</i> |
| | 83 | 5 | 12 (NW_021964178.1) | 25,197,404 | 25,272,878 | 75.475 | 1,075 | <i>LOC116282667, LOC107034552</i> |

Chr: chromosome; Bp1: starting position of the segment; Bp2: ending position of the segment; Kb: segment length; nSNP: number of variants contained in the segment

for K with alpacas clustering with vicugna and llama with guanaco. Moreover, the plot clearly showed a strong genetic introgression from vicugna and llama to alpaca. At K3 and K4 the plots revealed very low genetic admixture of the guanaco with the other species; guanacos in fact, showed a clearly separated gene pool. Only one guanaco showed genetic admixture with llama suggesting a possible hybridization between the two species as suggested by MDS plot. As already observed by Fan et al. [11], these results indicated alpaca as a possible hybrid species between llama and vicugna while not excluding the presence of a low genetic flow between guanacos and the other species.

ROHs and genomic inbreeding

Applying different length criteria for ROHs detection can reveal information about population demography across a range of time frames, although the minimum segments length defined to identify a ROHs is yet set arbitrarily [21]. The similar amount of ROHs detected in the four

species suggests a loss of genetic diversity of the wild species from a historical founder effect or genetic bottleneck [21] potentially due to the extensive hunting during the last century [1]; on the other, this data hint a low selective pressure for the two domestic species. Moreover, alpaca showed longer ROHs (100–500 kb), with an average length of 145,92 kb ($\pm 15,96$) compared to the other species where the average length ranged from 138,71 ($\pm 2,63$) to 134,83 ($\pm 4,75$). Although ANOVA did not show significant difference in the average length, probably attributable to the small sample size, the result suggests a more recent inbreeding for alpaca compared to other species [21]. This was further confirmed by the scanning for long ROHs (>500 kb) which were detected exclusively in alpaca samples except for one vicugna which showed a single ROHs 517,702 kb long. Long ROHs are expected in populations that experienced recent inbreeding [22]. It should be stressed that even though the F_{ROH} values based on ROHs 100-500 kb were slightly higher for alpaca respect to the other species, ANOVA did not

show any significant difference in terms of F_{ROH} values between the four population. The results suggest a low selective pressure in alpaca and llama whose inbreeding values are comparable to those of the wild populations. However further studies in a larger population are needed to validate the results.

Overlapping ROHs: signature of selection

Genes under selection in SACs include loci selected for production traits (such as body conformation, fertility, and maternal traits) as well as for natural adaptation to the environment.

In alpaca, the overlapping ROHs showed loci potentially selected for adaptation to environment such as *BSN* and *RIC8B*, both involved in sensory processing of sound and olfactory signaling pathway [23–25] and stress related genes such as *ANTXR2*, *PRDM8* and *FGF5*. Signatures of selection on the latter three genes were already observed by Fan and colleagues [11], which suggested that *FGF5* and *ANTXR2* may play a key role in regulating hypoxia stress, while *PRDM8* was proposed as a novel gene associated with hypoxic adaptation. In addition, *ANTXR2* was also found under climate-mediated selection in human, sheep [26] and cattle [27, 28]. Not surprisingly, the signatures of selection in alpaca also include loci influencing reproduction efficiency and production traits such as fiber and meat quality. In this regard, *ANTXR2* was involved in reproductive regulatory processes in Bactrian camel [29], also influencing the development of primordial germ cells and reproductive organs, thus impacting on the age at first calving [30]. Similarly, *DAG1* affects fertility in dairy cattle being associated with sire conception rate [31] and regulation of spermatogenesis [32]. It must also be noted that signatures of selection included *FGF5*, a gene involved in the modulation of alpaca hair coat length [8]. Finally, alpaca genome was characterized by the presence of overlapping ROHs that harbor loci influencing meat attributes such as *RIC8B* [33, 34] and *DAG1* [28, 35].

In llama, most of the selection signatures encompasses loci involved in the regulation of reproduction and maternal traits: some are involved in the lipid metabolism during lactation, as *ARFGEF2* and *FZD9* which are known for their role in fatty acid transport [36] adipocytes differentiation [37] and quantity of milk produced [38]. Similarly, the *PHF20* gene is involved in milk production [39] and lactating efficiency [40], while *RCORI* is a candidate gene implicated in lactose synthesis and milk yield [41]. Moreover, signatures of selections in llama suggested selection for efficient reproductive performances in harsh environments. *EXD2* for example, it has been involved in reproductive performance of cattle under heat stress [42]. In addition, overlapping ROHs comprise genes affecting pregnancy and fertility (*FZD9*) [43], heifer early

calving (*PHF20* and *RBM39*) [44], resistance to uterine disease traits in first parity (*ERC2*) [45] and ovulation rate (*BCL7B*) [46]. Finally, signals of selection were found in *ARFGEF2*, a candidate gene associated with low fertility in single kid cashmere goat [47]. Like alpaca, signatures of selection for carcass quality traits were detected also in llama. In fact, *ARFGEF2* was found to be under selection in dromedary camel and proposed as determinant of camel body weight [48] while other selected loci such as *BCL7B*, *MLXIPL* and *NDUF2* are implied in fat deposition and depth [46, 49–52]. Two loci associated with animal domestication and behavior traits, were found under selection in llama. *ERC2* is involved in the neurotransmitter release, and it was shown to be related to behavioural changes such as reduction in fear [53]. Similarly, *BAZ1B* was recently associated with the domestication syndrome by influencing the development of the neural crest in Zebrafish [54]; moreover, the gene was also involved in the human self-domestication hypothesis as a master regulator of the modern human face [55]. Signal of selection for harsh climate condition in llama was found in genes involved in cellular response to heat stress (*HSPB11*) [56] and adaptation to UV exposure through pigmentation mechanism (*ERC2*) [57]. Finally, *LDLRAD1* is known to be potentially involved in resistance against bacterial infection [58], while *RBM39* has been proposed as target of putative selective sweep in swine being involved in RNA splicing and RNA processing [59].

As expected, the two wild species showed a lower number of loci under selective pressure (Table 3) and overlapping ROHs were identified only over three and four loci for vicugna and guanaco, respectively. Three genes under selection in wild SACs are known to be involved in reproduction. In the vicugna, *DLG1*, already proposed as putative signature of selection for reproductive traits in Iranian dromedary camels [60], was also associated with fertility in other species being implied in oocyte polarization during maturation in cattle [61] and ovary development and litter size in sheep [62]. Similarly, *ATP6V1E1* and *BCL2L13* were found to be under selection in guanaco; these genes are known to play an important role in mouse and cattle fertility [63, 64]. Four loci indicate adaptation driven by environmental conditions in wild SACs. In vicugna, overlapping ROHs in *DLG1* suggest selection for feeding efficiency in poor pasture quality [65] while the selection signals found in *PDE4D* hint the adaptation of the species to UV radiation exposure through the modulation of pigmentation and/or eye-protective phenotypes. Indeed, this gene was recently described as essential in signaling pathways of melanin encompassing signatures of selection for variation in pigmentation in Groningen White Headed cattle [66] and goat breeds from Southern China [67]. Potential environmental adaptation was also observed in guanaco as the species

showed overlapping ROHs in *BCL2L13*, a gene selected for heat stress tolerance [68]. Environmental selection pressure for heat stress tolerance of the guanaco was also suggested by the presence of selection signals in *BID* gene that was found differentially expressed during heat stress in cattle [69–71] and buffalo [72]. Finally, it should be noted that both wild species present signatures of selection in *KAT6B*, a gene previously known to be associated with carcass traits and leg morphology in cattle [73]. A recent genetic association study suggested that a rare *KAT6B* haplotype is responsible for lower weight and height of the Russian Yaroslavl cattle, a breed adapted to the harsh conditions of the Yaroslavl region of eastern Russia. Authors suggested that the selected haplotype may be the result of an historical positive selection under the harsh environmental conditions and low feeding base and that the lack of such mutations in most other cattle breeds hints its negative effect in other environmental conditions and/or negative selection by humans [74]. Similarly, it can be assumed that environmental selective pressure has driven the shaping of guanaco and vicugna genome leading to the signature of selection detected in *KAT6B*. This assumption is also corroborated by the selective signal for high feeding efficiency [64] found in *DLG1* for vicugna.

Limitations

The present study presents some limitations which must be discussed. First, the small sample size did not allow a wider identification of overlapping ROHs. We should acknowledge a paucity of publicly available genomic data on SACs, particularly for vicugna and guanacos, which are protected species, with their sampling being extremely difficult [3, 4]. Second, we considered exclusively the identification of a ROH island present in more than 70% of the species studied as signature of selection. However, further tests (e.g., Extended Haplotype Homozygosity, haplotype based FLK) are applicable to confirm the presence of real genetic selection. Further studies, on a larger sample size, will be necessary to confirm this preliminary evidence. Moreover, individuals from livestock population often showed high level of genetic relatedness which could led to population structure and bias in the estimation of the signature of selection. To reduce this bias, we filtered out from our sample individuals with identity-by-descent (IBD) $PI-HAT \geq 0.5$. A recurrent limitation is the lack of consensus in establishing the criteria to define the ROHs [75] and the minimum segments length defined to identify a ROHs is often set arbitrarily [76]. In SACs these parameters were never defined. To overcome this concern, the parameters to detect the ROHs were set by averaging the results from previous work on livestock and pet species reviewed by Meyermans [16].

Finally, we should acknowledge that using sequence-based gene variants, coupled with a small sample size, may lead somewhat to false-positive genotypes since we could not apply a Hardy-Weinberg Equilibrium test that is a common procedure applied to filter out gene variants badly genotyped at genome-wide level. Nevertheless, our stringent quality control pipeline regarding the overall genotyping rate (>99%) and the removal of related animals (IBD>0.5), may somehow balance the pitfalls related to the sequence-based variant calling.

Conclusions

Our results on SACs ancestry support the findings provided by Fan and colleagues [11] which demonstrated that alpaca genome showed signals of genetic introgression from vicugna and llama.

The identification of ROHs along the SACs genome showed a comparable level of genetic variability between the four species; however, population bottleneck can be assumed for vicugna and guanaco while low selection pressure, in terms of genetic inbreeding, was found for alpaca and llama.

The main aim of our work was to detect signatures of selection along the four SAC species. Not surprisingly, the identification of the overlapping ROHs showed divergent selection between the populations. Alpaca and llama hint a selection driven by environment as well as by farming purpose while llama and vicugna showed selection signal for adaptation to harsh environment. Interestingly, signatures of selection on *KAT6B* gene were identified for both vicugna and guanaco, suggesting a positive effect on wild populations fitness. Such information may be of interest to further ecological and animal production studies.

Methods

Sample collection

A total of fifty-five WGS were used for the study. Forty-eight samples were generated by previous projects (PRJNA233565, PRJNA340289, PRJNA512907, PRJNA612032 and PRJNA685331) and were retrieved from the NCBI Sequence Read Archive (SRA) (Supplementary Table 1). Seven in-house alpaca samples were sequenced *de novo*: skin biopsies were performed as described in Pallotti et al. [8]. Genomic DNA was isolated using the Genomic DNA Isolation Kit (Norgen Biotek Corp.), according to the manufacturer's instructions. The library preparation was carried out at Genomix4Life (Salerno, Italy) using the Illumina DNA Prep Kit (Illumina) followed by a 150 bp sequencing at paired-end mode, using the Illumina NovaSeq 6000 System.

The final sample consisted of seven newly sequenced Peruvian alpacas and 48 publicly available WGS samples (28 alpacas, 7 llamas, 7 guanacos and 6 vicugnas). The

public SRA files were downloaded to our server and converted to FASTQ file.

WGS quality control

The quality of the FASTQ files was checked using FastQC [77] and the adapter trimming was performed with Trimmomatic [78]. Read pairs were mapped to the alpaca reference genome *VicPac3.1* [5] using Burrows-Wheeler Alignment MEM (BWA-MEM) [79]. The X chromosome and the unplaced-scaffold sequences were removed from the reference genome FASTA file before performing the alignment. Seventeen alpaca samples and one guanaco sample were removed because of a genotyping rate lower than 99%. The remaining 38 samples (19 alpacas, 7 llamas, 6 vicugna and 6 guanacos) were included in the study and used for further analysis. The sequence length ranged from 100 to 150 bp with a sequencing depth coverage rate ranging from 15 to 63X (Supplementary Table 1). BAM files were further processed using the Genome Analysis Toolkit (GATK, v3.4) [80] and the HaplotypeCaller method was used for variant calling. The resulting VCF containing the genomic variant calling of the 38 samples was converted to PLINK file using VCFtools [81].

To have independent samples and reduce the underlying population structure potentially biasing the estimation of ancestry and ROHs, subjects' pairs with Identity-By-Descent (IBD) $PI-HAT \geq 0.5$ were identified using PLINK 1.9 [82]; only one alpaca sample was removed due to $PI-HAT$ value of 0.5. The remaining 37 samples (18 alpacas, 7 llamas, 6 vicugna and 6 guanacos) showed $PI-HAT$ values ≤ 0.125 and were used for the detection of ROHs.

Population structure analysis

To perform population structure analysis, we removed from the joint called variant file all the variants with a minor allele frequency (MAF) $< 5\%$ and we pruned the remaining variants by Linkage Disequilibrium (LD) using the PLINK command “*--indep-pairwise 1,500 150 0.1*”. Starting from 40,360,774 variants, 20,557,058 variants were removed due to low MAF ($< 5\%$) and further 19,769,677 variants were pruned due to high LD ($r^2 > 0.9$). Finally, 34,039 variants and 37 samples were used to run the Multi-Dimensional Scaling (MDS) and the admixture analyses. MDS was performed using PLINK 1.9 [81]. The results were plotted using ClustVis [83]. Population structure analysis was performed on ADMIXTURE (Version 1.23) with $K=2, 3$, and 4. The correct value for K was determined according to ADMIXTURE's cross-validation procedure [84] and the results were plotted using R [85].

Runs of homozygosity (ROHs) and genomic inbreeding

(F_{ROH})

To detect ROHs, the dataset was not pruned for low MAF ($< 0.5\%$) or high LD ($r^2 > 0.9$), as suggested by Meyermans [16]. Since the criteria to detect the ROHs in SACs was never defined, we set the parameters by averaging the results from previous works on livestock and pet species reviewed by Meyermans [16]. Based on that, the following PLINK parameters were used: “*--homozyg --homozyg-kb 100 (or 500) --homozyg-snp 15 --homozyg-gap 500 --homozyg-window-missing 1 --homozyg-window-het 3*”. The “*--homozyg-window-het*” and the minimum length of 100 kb (“*--homozyg-kb 100*”) were chosen according to the information provided by Quinodoz [86] (2021) and Harder [87]. For the identification of overlapping ROHs we kept the ones overlapping in at least 70% of the sample, according to each species. All the loci in the selected ROHs were manually annotated using NCBI Genome Data Viewer (<https://www.ncbi.nlm.nih.gov/genome/gdv/>) starting from the genomic coordinates contained in the Plink files.

The inbreeding coefficient based on ROHs (F_{ROH}) was computed as suggested by McQuillan [88] (2008) as:

$$F_{ROH} = \frac{L_{ROH}}{L_{aut}}$$

where L_{ROH} is the total length of all ROHs in the individual's genome, and L_{aut} is the length of the autosomal genome.

ANalysis Of Variance (ANOVA) was used in IBM SPSS Statistics 21 software to test for significant differences in the number and length of ROHs as well as for the F_{ROH} values between the four species, setting a statistical significance threshold of $P < 0.05$.

Gene-based enrichment analysis

For each species, gene enrichment analyses were performed according to the genes encompassing the overlapping ROHs with the web-based tool Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 [89], which allows for the investigation of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [90] and Gene Ontology (GO) for biological processes [22].

Abbreviations

| | |
|------|-------------------------|
| WGS | Whole-genome sequencing |
| ROHs | Runs Of Homozygosity |
| SACs | South American Camelids |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09547-3>.

Supplementary Material 1

Acknowledgements

Not applicable.

Authors' contributions

S.P.: study conception, analysis, drafting of the manuscript; M.P.: analysis, critical review of the manuscript; M.A.: critical review of the manuscript; C.R.: critical review of the manuscript; V.N.: study conception, analysis, drafting of the manuscript, critical review of the manuscript.

Funding

Not applicable.

Data Availability

The VCF file from the *de novo* WGS of seven alpacas generated and analysed during the current study are available in the EVA (European Variation Archive) repository under the project PRJEB61878 [<https://www.ebi.ac.uk/eva/?eva-study=PRJEB61878>].

The forty-eight samples generated by previous projects and used in this study are available in the NCBI Sequence Read Archive (SRA) [PRJNA233565, PRJNA340289, PRJNA512907, PRJNA612032 and PRJNA685331].

Declarations

Ethics approval and consent to participate

Alpaca's skin biopsies were obtained in 2008 according to the guidelines of the Animal Ethics Committee of the University of Camerino.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Received: 27 April 2023 / Accepted: 31 July 2023

Published online: 21 August 2023

References

- Vilá B, Arzamendia Y. South american Camelids: their values and contributions to people. *Sustain Sci*. 2022;17(3):707–24. <https://doi.org/10.1007/s11625-020-00874-y>.
- Yacobaccio H. The historical relationship between people and the Vicuña. In: Gordon IJ, editor. *The Vicuña*. Boston: Springer; 2009.
- Lichtenstein G, Vilá B. Vicuna use by andean communities: an overview. *MRD*. 2003;23(2):198–201.
- Carmanchahi PD, Ovejero R, Marull C, Lopez GC, Schroeder N, Jahn GA, ..., Somoza GM. Physiological response of wild guanacos to capture for live shearing. *Wildl Res*. 2011;38(1):61–8.
- Richardson MF, Munyard K, Croft LJ, Allnutt TR, Jackling F, Alshanbari F, Jevit M, Wright GA, Cransberg R, Tibary A, Perelman P, Appleton B, Raudsepp T. Chromosome-level Alpaca Reference Genome VicPac3.1 improves genomic insight into the Biology of New World Camelids. *Front Genet*. 2019;10:586. <https://doi.org/10.3389/fgene.2019.00586>.
- Pallotti S, Chandramohan B, Pediconi D, Nocelli C, La Terza A, Renieri C. Interaction between the melanocortin 1 receptor (MC1R) and agouti signalling protein genes (ASIP), and their association with black and brown coat colour phenotypes in peruvian alpaca. *Ital J Anim Sci*. 2020;19(1):1518–22.
- Pallotti S, Pacheco C, Valbonesi A, Antonini M. A comparison of quality of the fleece and follicular activity between sheared and non-sheared yearling alpacas (Vicugna pacos). *Small Rumin Res*. 2020;192:106243.
- Pallotti S, Pediconi D, Subramanian D, Molina MG, Antonini M, Morelli MB, Renieri C, La Terza A. Evidence of post-transcriptional readthrough regulation in FGF5 gene of alpaca. *Gene*. 2018;647:121–8. <https://doi.org/10.1016/j.gene.2018.01.006>.
- Marín JC, Romero K, Rivera R, Johnson WE, González BA. Y-chromosome and mtDNA variation confirms independent domestications and directional hybridization in South American camelids. *Anim Genet*. 2017;48(5):591–5. <https://doi.org/10.1111/age.12570>.
- Kadwell M, Fernandez M, Stanley HF, Baldi R, Wheeler JC, Rosadio R, Bruford MW. Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc Biol Sci*. 2001;268(1485):2575–84. <https://doi.org/10.1098/rspb.2001.1774>.
- Fan R, Gu Z, Guang X, Marín JC, Varas V, González BA, Wheeler JC, Hu Y, Li E, Sun X, Yang X, Zhang C, Gao W, He J, Munch K, Corbett-Detig R, Barbato M, Pan S, Zhan X, Bruford MW, Dong C. Genomic analysis of the domestication and post-spanish conquest evolution of the llama and alpaca. *Genome Biol*. 2020;21(1):159. <https://doi.org/10.1186/s13059-020-02080-6>.
- Pérez P, Maino M, Guzmán R, Vaquero A, Köbrich C, Pokniak J. Carcass characteristics of llamas (*Lama glama*) reared in Central Chile. *Small Rumin Res*. 2000;37(1–2):93–7. [https://doi.org/10.1016/s0921-4488\(99\)00127-3](https://doi.org/10.1016/s0921-4488(99)00127-3).
- Wurzinger M, Willam A, Delgado J, Nürnberg M, Zárata AV, Stemmer A, Ugarte G, Sölkner J. Design of a village breeding programme for a llama population in the high Andes of Bolivia. *J Anim Breed Genet*. 2008;125(5):311–9. <https://doi.org/10.1111/j.1439-0388.2007.00713.x>.
- Onzima RB, Upadhyay MR, Doekes HP, Brito LF, Bosse M, Kanis E, Groenen MAM, Crooijmans RPMA. Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan Goat Breeds. *Front Genet*. 2018;9:318. <https://doi.org/10.3389/fgene.2018.00318>.
- Mastrangelo S, Tolone M, Sardina MT, Sottile G, Suter AM, Di Gerlando R, Portolano B. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genet Sel Evol*. 2017;49(1):84. <https://doi.org/10.1186/s12711-017-0360-z>.
- Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*. 2020;21(1):94. <https://doi.org/10.1186/s12864-020-6463-x>.
- Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS ONE*. 2017;12(5):e0176780. <https://doi.org/10.1371/journal.pone.0176780>.
- More M, Gutiérrez G, Rothschild M, Bertolini F, Ponce de León FA. Evaluation of SNP genotyping in Alpacas using the bovine HD genotyping beadchip. *Front Genet*. 2019;10:361. <https://doi.org/10.3389/fgene.2019.00361>.
- Calderon M, More MJ, Gutierrez GA, Ponce de León FA. Development of a 76k Alpaca (*Vicugna pacos*) single nucleotide polymorphisms (SNPs) microarray. *Genes (Basel)*. 2021;12(2):291. <https://doi.org/10.3390/genes12020291>.
- Pallotti S, Piras IS, Marchegiani A, Cerquetella M, Napolioni V. Dog-human translational genomics: state of the art and genomic resources. *J Appl Genet*. 2022;63(4):703–16. <https://doi.org/10.1007/s13353-022-00721-z>.
- Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest Sci*. 2014;166:26–34.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium*. *Nat Genet*. 2000;25(1):25–9. <https://doi.org/10.1038/75556>.
- Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. The GeneCards suite: from Gene Data Mining to Disease Genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54(1301–). <https://doi.org/10.1002/cpbi.5>.
- Zhuang H, Matsunami H. Synergism of accessory factors in functional expression of mammalian odorant receptors. *J Biol Chem*. 2007;282(20):15284–93. <https://doi.org/10.1074/jbc.M700386200>.
- Verbeugt C, Wilkin F, Tarabichi M, Gregoire F, Dumont JE, Chatelain P. Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS ONE*. 2014;9(5):e96333. <https://doi.org/10.1371/journal.pone.0096333>.
- Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li MH, Ajmone Marsan P. Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol*. 2014;31(12):3324–43. <https://doi.org/10.1093/molbev/msu264>.
- Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, Casabianca F, Casu S, Ciampolini R, Coeur D'Acier A, Coquelle C, Delgado JV, El-Beltagi A, Hadjipavlou G, Joussetin E, Landi V, Lauvie A, Lecomte P, Ligda C, Marinthe C, Martinez A, Mastrangelo S, Menni D, Moulin CH, Osman MA, Pineau O, Portolano B, Rodellar C, Saïdi-Mehtar N, Sechi T, Sempéré G, Thévenon S, Tsiokos D, Laloë D, Gautier M. A genomic map of climate adaptation

- in Mediterranean cattle breeds. *Mol Ecol*. 2019;28(5):1009–29. <https://doi.org/10.1111/mec.15004>.
28. Zhang T, Wang T, Niu Q, Xu L, Chen Y, Gao X, Gao H, Zhang L, Liu GE, Li J, Xu L. Transcriptional atlas analysis from multiple tissues reveals the expression specificity patterns in beef cattle. *BMC Biol*. 2022;20(1):79. <https://doi.org/10.1186/s12915-022-01269-4>.
 29. Wang Q, Zhang Q, Gan Z, Li H, Yang Y, Zhang Y, Zhao X. Screening for reproductive biomarkers in bactrian camel via iTRAQ analysis of proteomes. *Reprod Domest Anim*. 2020;55(2):189–99. <https://doi.org/10.1111/rda.13607>.
 30. Utsunomiya YT, Machado MA, VERNEQUE RDS & SILVA M. Genome Wide Scan for Age at First Calving in Gyr Dairy Cattle. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 10., 2014, Vancouver. Proceedings... Champaign: American Society of Animal Science, 2014.
 31. Rezende FM, Dietsch GO, Peñagaricano F. Genetic dissection of bull fertility in US Jersey dairy cattle. *Anim Genet*. 2018;49(5):393–402. <https://doi.org/10.1111/age.12710>.
 32. Wu S, Mipam T, Xu C, Zhao W, Shah MA, Yi C, Luo H, Cai X, Zhong J. Testis transcriptome profiling identified genes involved in spermatogenic arrest of cattleyak. *PLoS ONE*. 2020;15(2):e0229503. <https://doi.org/10.1371/journal.pone.0229503>.
 33. Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S, Oh SJ, Lee HK, Kim H. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mamm Genome*. 2017;28(11–12):528–41. <https://doi.org/10.1007/s00335-017-9715-6>.
 34. Peripolli E, Metzger J, de Lemos MVA, Stafuzza NB, Kluska S, Olivieri BF, Feitosa FLB, Berton MP, Lopes FB, Munari DP, Lôbo RB, Magnabosco CU, Di Croce F, Osterstock J, Denise S, Pereira ASC, Baldi F. Autozygosity islands and ROH patterns in Nelore lineages: evidence of selection for functionally important traits. *BMC Genomics*. 2018;19(1):680. <https://doi.org/10.1186/s12864-018-5060-8>.
 35. Liu R, Liu X, Bai X, Xiao C, Dong Y. Different expression of lipid metabolism-related genes in Shandong black cattle and Luxi cattle based on transcriptome analysis. *Sci Rep*. 2020;10(1):21915. <https://doi.org/10.1038/s41598-020-79086-4>.
 36. Valour D, Degrelle SA, Ponter AA, Giraud-Delville C, Campion E, Guyader-Joly C, Richard C, Constant F, Humblot P, Ponsart C, Hue I, Grimard B. Energy and lipid metabolism gene expression of D18 embryos in dairy cows is related to dam physiological status. *Physiol Genomics*. 2014;46(2):39–56. <https://doi.org/10.1152/physiolgenomics.00091.2013>.
 37. Pan C, Wang S, Yang C, Hu C, Sheng H, Xue X, Hu H, Lei Z, Yang M, Ma Y. Genome-wide identification and expression profiling analysis of wnt family genes affecting adipocyte differentiation in cattle. *Sci Rep*. 2022;12(1):489. <https://doi.org/10.1038/s41598-021-04468-1>.
 38. Ahmad SM, Bhat B, Bhat SA, Yaseen M, Mir S, Raza M, Iquebal MA, Shah RA, Ganai NA. SNPs in mammary gland epithelial cells unraveling potential difference in milk production between Jersey and Kashmiri cattle using RNA sequencing. *Front Genet*. 2021;12:666015. <https://doi.org/10.3389/fgene.2021.666015>.
 39. Massender E, Oliveira HR, Brito LF, Maignel N, Jafarika M, Baes CF, Sullivan B, Schenkel FS. Genome-wide association study for milk production and conformation traits in Canadian Alpine and Saanen dairy goats. *J Dairy Sci*. 2023;106(2):1168–89. <https://doi.org/10.3168/jds.2022-22223>.
 40. Ben Meir YA, Daddam JR, Kra G, Kamer H, Portnick Y, Levin Y, Zachut M. Proteomic analysis of adipose tissue revealing differentially abundant proteins in highly efficient mid-lactating dairy cows. *Sci Rep*. 2022;12(1):1–15.
 41. Li X, Yuan L, Wang W, Zhang D, Zhao Y, Chen J, Xu D, Zhao L, Li F, Zhang X. Whole genome re-sequencing reveals artificial and natural selection for milk traits in East Friesian sheep. *Front Vet Sci*. 2022;9:1034211. <https://doi.org/10.3389/fvets.2022.1034211>.
 42. Sigdel A, Liu L, Abdollahi-Arpanahi R, Aguilar I, Peñagaricano F. Genetic dissection of reproductive performance of dairy cows under heat stress. *Anim Genet*. 2020;51(4):511–20. <https://doi.org/10.1111/age.12943>.
 43. Neupane M, Geary TW, Kiser JN, Burns GW, Hansen PJ, Spencer TE, Neibergs HL. Loci and pathways associated with uterine capacity for pregnancy and fertility in beef cattle. *PLoS ONE*. 2017;12(12):e0188997. <https://doi.org/10.1371/journal.pone.0188997>.
 44. Mota LFM, Carvajal AB, Bernardes PA, Buzanskas ME, Baldi F, Lôbo RB, Munari DP. Integrating genome-wide association study and pathway analysis reveals physiological aspects affecting heifer early calving defined at different ages in Nelore cattle. *Genomics*. 2022;114(4):110395. <https://doi.org/10.1016/j.ygeno.2022.110395>.
 45. May K, Sames L, Scheper C, König S. Genomic loci and genetic parameters for uterine diseases in first-parity Holstein cows and associations with milk production and fertility. *J Dairy Sci*. 2022;105(1):509–24. <https://doi.org/10.3168/jds.2021-20685>.
 46. Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. A whole-genome association study for pig reproductive traits. *Anim Genet*. 2012;43(1):18–26. <https://doi.org/10.1111/j.1365-2052.2011.02213.x>.
 47. Wang K, Liu X, Qi T, Hui Y, Yan H, Qu L, Lan X, Pan C. Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (*Capra hircus*). *Genomics*. 2021;113(1 Pt 1):142–50. <https://doi.org/10.1016/j.ygeno.2020.11.024>.
 48. Bitaraf Sani M, Zare Harofte J, Banabazi MH, Esmailkhanian S, Shafei Naderi A, Salim N, Teimoori A, Bitaraf A, Zadehrahmani M, Burger PA, Landi V, Silawi M, Taghipour Sheshdeh A, Faghihi MA. Genomic prediction for growth using a low-density SNP panel in dromedary camels. *Sci Rep*. 2021;11(1):7675. <https://doi.org/10.1038/s41598-021-87296-7>.
 49. Graugnard DE, Berger LL, Faulkner DB, Looor JJ. High-starch diets induce precocious adipogenic gene network up-regulation in longissimus lumborum of early-weaned Angus cattle. *Br J Nutr*. 2010;103(7):953–63. <https://doi.org/10.1017/S0007114509992789>.
 50. Moisé SJ, Shike DW, Faulkner DB, Meteere WT, Keisler D, Looor JJ. Central Role of the PPAR γ Gene Network in coordinating beef cattle intramuscular adipogenesis in response to Weaning Age and Nutrition. *Gene Regul Syst Bio*. 2014;8:17–32. <https://doi.org/10.4137/GRSB.511782>.
 51. English AM, Waters SM, Cormican P, Byrne CJ, Fair S, Kenny DA. Effect of early calf-hood nutrition on the transcriptomic profile of subcutaneous adipose tissue in Holstein-Friesian bulls. *BMC Genomics*. 2018;19(1):281. <https://doi.org/10.1186/s12864-018-4681-2>.
 52. Garza Hernandez D, Mucha S, Banos G, Kaseja K, Moore K, Lambe N, Yates J, Bunger L. Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep. *Animal*. 2018;12(5):915–22. <https://doi.org/10.1017/S1751731117002488>.
 53. Mastrangelo S, Ben Jemaa S, Sottile G, Casu S, Portolano B, Ciani E, Pilla F. Combined approaches to identify genomic regions involved in phenotypic differentiation between low divergent breeds: application in sardinian sheep populations. *J Anim Breed Genet*. 2019;136(6):526–34. <https://doi.org/10.1111/jbg.12422>.
 54. Torres-Pérez JV, Anagnini S, Mech AM, Havelange W, García-González J, Fraser SE, Vallortigara G, Brennan CH. *baz1b* loss-of-function in zebrafish produces phenotypic alterations consistent with the domestication syndrome. *iScience*. 2022;26(1):105704. <https://doi.org/10.1016/j.isci.2022.105704>.
 55. Zanella M, Vitriolo A, Andirko A, Martins PT, Sturm S, O'Rourke T, Laugsch M, Malerba N, Skaros A, Trattaro S, Germain PL, Mihailovic M, Merla G, Rada-Iglesias A, Boeckx C, Testa G. Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. *Sci Adv*. 2019;5(12):eaaw7908. <https://doi.org/10.1126/sciadv.aaw7908>.
 56. Stamperna K, Dovolou E, Giannoulis T, Kalemkeridou M, Nanas I, Dadouli K, Moutou K, Mamuris Z, Amiridis GS. Developmental competence of heat stressed oocytes from Holstein and Limousine cows matured in vitro. *Reprod Domest Anim*. 2021;56(10):1302–14. <https://doi.org/10.1111/rda.13993>.
 57. Rafiepour M, Ebrahimie E, Vahidi MF, Salekdeh GH, Niazi A, Dadpasand M, Liang D, Si J, Ding X, Han J, Zhang Y, Qanbari S. Whole-genome resequencing reveals Adaptation Prior to the divergence of Buffalo Subspecies. *Genome Biol Evol*. 2021;13(1):evaa231. <https://doi.org/10.1093/gbe/evaa231>.
 58. Du Z, Sun Y, Wang J, Liu H, Yang Y, Zhao N. Comprehensive RNA-Seq profiling of the lung transcriptome of bashbay sheep in response to experimental *Mycoplasma ovipneumoniae* infection. *PLoS ONE*. 2020;15(7):e0214497. <https://doi.org/10.1371/journal.pone.0214497>.
 59. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LA, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012;491(7424):393–8. <https://doi.org/10.1038/nature11622>.
 60. Khalkhali-Evrigh R, Hedayat N, Ming L, Jirimutu. Identification of selection signatures in iranian dromedary and bactrian camels using whole genome sequencing data. *Sci Rep*. 2022;12(1):9653. <https://doi.org/10.1038/s41598-022-14376-7>.

61. Fonseca PAS, Suárez-Vega A, Cánovas A. Weighted gene correlation Network Meta-Analysis reveals functional candidate genes Associated with High- and Sub-Fertile Reproductive performance in beef cattle. *Genes (Basel)*. 2020;11(5):543. <https://doi.org/10.3390/genes11050543>.
62. Hernández-Montiel W, Martínez-Núñez MA, Ramón-Ugalde JP, Román-Ponce SI, Calderón-Chagoya R, Zamora-Bustillos R. Genome-wide Association Study reveals candidate genes for litter size traits in Pelibuey Sheep. *Anim (Basel)*. 2020;10(3):434. <https://doi.org/10.3390/ani10030434>.
63. Wu PH, Fu Y, Cecchini K, Özata DM, Arif A, Yu T, Colpan C, Gainetdinov I, Weng Z, Zamore PD. The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nat Genet*. 2020;52(7):728–39. <https://doi.org/10.1038/s41588-020-0657-7>.
64. Matsuyama S, Nakamura S, Minabe S, Sakatani M, Takenouchi N, Sasaki T, Inoue Y, Iwata H, Kimura K. Deterioration of mitochondrial biogenesis and degradation in the endometrium is a cause of subfertility in cows. *Mol Reprod Dev*. 2023;90(3):141–52. <https://doi.org/10.1002/mrd.23670>.
65. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freely HC, Hansen SL, Yampara-Iquise H, Johnson KA, Kerley MS, Kim J, Loy DD, Marques E, Neiberghs HL, Schnabel RD, Shike DW, Spangler ML, Weaver RL, Garrick DJ, Taylor JF. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*. 2017;18(1):386. <https://doi.org/10.1186/s12864-017-3754-y>.
66. Gonzalez-Prendes R, Ginja C, Kantanen J, Ghanem N, Kugonza DR, Makgahlala ML, Groenen MAM, Crooijmans RPMA. Integrative QTL mapping and selection signatures in Groningen White Headed cattle inferred from whole-genome sequences. *PLoS ONE*. 2022;17(10):e0276309. <https://doi.org/10.1371/journal.pone.0276309>.
67. Sun X, Jiang J, Wang G, Zhou P, Li J, Chen C, Liu L, Li N, Xia Y, Ren H. Genome-wide association analysis of nine reproduction and morphological traits in three goat breeds from Southern China. *Anim Biosci*. 2023;36(2):191–9. <https://doi.org/10.5713/ab.21.0577>.
68. Fleming DS, Weigend S, Simianer H, Weigend A, Rothschild M, Schmidt C, Ashwell C, Persia M, Reecy J, Lamont SJ. Genomic Comparison of Indigenous African and Northern European Chickens Reveals Putative Mechanisms of Stress Tolerance Related to Environmental Selection Pressure. *G3 (Bethesda)*. 2017;7(5):1525–37. <https://doi.org/10.1534/g3.117.041228>.
69. Kolliv V, Upadhyay RC, Singh D. Peripheral blood leukocytes transcriptomic signature highlights the altered metabolic pathways by heat stress in zebu cattle. *Res Vet Sci*. 2014;96(1):102–10. <https://doi.org/10.1016/j.rvsc.2013.11.019>.
70. Choudhary R, Goud TS, Kumar A, Sharma AK, Singh SV, Upadhyay RC, Mohanty AK, Kumar S. Heat stress induced adaptation in melanocytes is dependent on the level of melanin and reduction of apoptosis. *J Dermatol Sci*. 2017;85(3):250–2. <https://doi.org/10.1016/j.jdermsci.2016.12.012>.
71. Sammad A, Luo H, Hu L, Zhu H, Wang Y. Transcriptome reveals Granulosa cells coping through Redox, inflammatory and metabolic mechanisms under Acute Heat stress. *Cells*. 2022;11(9):1443. <https://doi.org/10.3390/cells11091443>.
72. Yadav A, Singh KP, Singh MK, Saini N, Palta P, Manik RS, Singla SK, Upadhyay RC, Chauhan MS. Effect of physiologically relevant heat shock on development, apoptosis and expression of some genes in buffalo (*Bubalus bubalis*) embryos produced in vitro. *Reprod Domest Anim*. 2013;48(5):858–65. <https://doi.org/10.1111/rda.12175>.
73. Van den Berg I, Fritz S, Rodriguez S, Rocha D, Boussaha M, Lund MS, Boichard D. Concordance analysis for QTL detection in dairy cattle: a case study of leg morphology. *Genet Sel Evol*. 2014;46(1):31. <https://doi.org/10.1186/1297-9686-46-31>.
74. Ruvinskiy D, Igoshin A, Yurchenko A, Ilina AV, Larkin DM. Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. *Anim Genet*. 2022;53(5):680–4. <https://doi.org/10.1111/age.13230>.
75. Ku CS, Naidoo N, Teo SM, Pawitan Y. Regions of homozygosity and their impact on complex diseases and traits. *Hum Genet*. 2011;129(1):1–15. <https://doi.org/10.1007/s00439-010-0920-6>.
76. Martin CA, Sheppard EC, Illera JC, Suh A, Nadachowska-Brzyska K, Spurgin LG, Richardson DS. Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an island-colonizing bird. *Mol Ecol*. 2023;32(8):1972–89. <https://doi.org/10.1111/mec.16865>.
77. Andrews S. FastQC: a quality control tool for high throughput sequence data. (2010).
78. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
79. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
80. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
82. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795>.
83. Metsalu T, Vilo J. *Nucleic Acids Res*. 2015;43(W1):W566–70. <https://doi.org/10.1093/nar/gkv468>. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap.
84. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64. <https://doi.org/10.1101/gr.094052.109>.
85. Team R, Development Core. "A language and environment for statistical computing." <https://www.R-project.org>, 2009.
86. Quinodoz M, Peter VG, Bedoni N, Royer Bertrand B, Cisarova K, Salmaninejad A, Sepahi N, Rodrigues R, Piran M, Mojarad M, Pasdar A, Ghanbari Asad A, Sousa AB, Coutinho Santos L, Superti-Furga A, Rivolta C. AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. *Nat Commun*. 2021;12(1):518. <https://doi.org/10.1038/s41467-020-20584-4>.
87. Harder AM, Kirksey KB, Mathur S, Willoughby JR. Detectability of runs of homozygosity is influenced by analysis parameters as well as population-specific demographic history. *bioRxiv*, 2022–09.
88. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF. Runs of homozygosity in European populations. *Am J Hum Genet*. 2008;83(3):359–72. <https://doi.org/10.1016/j.ajhg.2008.08.007>. Epub 2008 Aug 28. Erratum in: *Am J Hum Genet*. 2008;83(5):658.
89. Jiao X, Sherman BT, Huang da W, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*. 2012;28(13):1805–6. <https://doi.org/10.1093/bioinformatics/bts251>.
90. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. 2012;40(Database issue):D109–14. <https://doi.org/10.1093/nar/gkr988>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.