Animal 19 (2025) 101674

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

Genomic scan of selective sweeps in an alpaca population subjected to directional selection for fibre quality traits

K.D. Arias [a], M. More [b], F. Goyache [a], A. Cruz [c,d], A. Burgos [c], G. Gutiérrez [d,e], I. Cervantes [f], J.P. Gutiérrez [f,*]

[a] Área de Genética y Reproducción Animal, SERIDA-Deva, 33394 Gijón, Asturias, Spain
[b] Facultad de Zootecnia, Universidad Nacional de Piura, Piura, Perú
[c] Pacomarca Scientific Station – Inca Tops SA. Avda. Miguel Forga 348, Arequipa, Perú
[d] Universidad Nacional Agraria La Molina. Avda. La Molina s/n, Lima, Perú
[e] Department of Animal Science, Iowa State University, Ames 50011, USA
[f] Departamento de Producción Animal, Universidad Complutense de Madrid. Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 16 November 2024
Revised 23 September 2025
Accepted 25 September 2025
Available online 4 October 2025

Keywords:
Alpaca genome
Candidate genes
Directional selection
Fibre production
Selection signatures

ABSTRACT

Improving fibre quality traits is a key objective in alpaca breeding programmes. A total of 630 alpaca individuals managed at the Pacomarca experimental farm located in Puno, Perú, and subjected to selection for fibre quality traits were genotyped using the Affymetrix Custom Alpaca genotyping array (76 508 SNPs). Two tests, the integrative Haplotype Score (**iHS**) and the number of segregating Sites by Length (**nSL**), aimed at identifying incomplete sweeps, were applied on all sample genotypes to identify selective candidate genomic regions spanning genes putatively involved in fibre production. The Cross-Population Extended Haplotype Homozygosity (**XP-EHH**) statistic was applied to two divergent subpopulations included in the total genotyped population: one less selected (reference; 49 individuals) and one highly selected (target; 127 individuals) to assess if some of the selective sweeps identified with both iHS and nSL were fixed in the target subpopulation. A total of 544 single–nucleotide polymorphisms (**SNPs**) were identified as candidate selective signals using both the iHS and nSL tests. Furthermore, 27 SNPs were identified as surrounding selective signals and further considered as candidate selective signals as well. A total of 509 candidate selective genomic regions covering 14.6 Mb of the alpaca genome were constructed. Gene-annotation enrichment analyses enabled to identify 293 candidate genes in these selective candidate genomic regions. Up to 11 candidate genes (e.g. *PTPRN*, *FRK*, *PTPN23*, and *PTPRT*) were associated with Gene Ontology (**GO**) terms related to phosphorylation and dephosphorylation and 28 candidate genes (e.g. *FOXP1*, *MITF*, *RNF10*, and *ZNF664*) with GO terms related to RNA polymerase activity. These genes have been reported to play a crucial role in the development of animal hair follicles, as well as in regulating hair growth and quality across different livestock species. Functional annotation conducted allowed to identify four significantly enriched functional term clusters including a total of 32 candidate genes. These functional clusters may be involved in the modulation of fibre growth and quality as well as skin and hair follicle development. Moreover, enrichment analyses within candidate selective genomic regions identified with XP-EHH allowed to identify 54 candidate genes, 10 of which were also identified using both the iHS and nSL tests. However, no functional clusters were identified. The results presented suggest that the genomes analysed were shaped by directional selection for fibre quality traits, involving multiple chromosomal areas of the alpaca genome. This fits well with traits expected to be mainly controlled by genes of additive effect.

Implications

High-quality fibre production in alpacas is crucial for the textile industry, enhancing the economic value of their fleece and shaping breeding decisions. The Pacomarca experimental farm has been under formal genetic evaluation and selection since 2007, with the general aim of increasing the amount of fine fibre per individual. This study suggests that directional selection for production traits may have caused selective sweep signatures on the alpaca genome, identifying new genomic regions and candidate genes potentially involved in fibre production. While these findings offer

* Corresponding author.
E-mail address: gutgar@vet.ucm.es (J.P. Gutiérrez).

https://doi.org/10.1016/j.animal.2025.101674
1751-7311/

valuable insights for breeding programmes in alpacas, further validation is needed before implementation on a broader scale.

Introduction

Alpaca (*Vicugna pacos*) is a South American camelid probably domesticated from vicuña (*Vicugna vicugna*) in the wet puna of Peru's central Andes 6 000–7 000 years ago (Wheeler et al., 2006). Present alpaca population results from a complex history: different vicuña populations may have contributed to the domesticated alpaca and there is evidence suggesting that hybridisation among llamas (*Lama glama*) and alpacas (Kadwell et al., 2001) occurred continuously until present (Varas et al., 2020; Fan et al., 2020). Whereas alpacas were domesticated mainly for fibre production, llamas were presumably domesticated from guanaco (*Lama guanicoe*) to provide draught power and meat (Wheeler et al., 2006). This continuous interbreeding between llamas and alpacas may lessen the alpacas' ability for high-quality fibre production, a concern underscored by genetic evidence suggesting that a major gene affecting fibre quality traits could be segregating in alpacas (Pérez-Cabal et al., 2010).

From a genetic point of view, the assessment of traits involved in fibre production is challenging. Although fleece weight is of major importance for producers' income (Wurzinger and Gutiérrez, 2022, McGregor, 2006), there is a need to decrease prickle factor by decreasing the percentage of medullated fibre and reducing the fibre diameter (Quispe et al., 2022; Cruz et al., 2019; Allain and Renieri, 2010; Gutiérrez et al., 2009). Fleece weight is influenced by several factors such as fibre density, staple length, and the skin surface where the fibre grows (McFadden and Neale, 1954), adding complexity to the genetic management of fibre production. Selection aimed at reducing fibre medulation has also been explored as a way to obtain valuable fibre to address market demands (Cruz et al., 2019, 2024).

The Pacomarca experimental farm, founded in 1992 by the INCA Group, is located in the department of Puno (Peruvian Altiplano) at an altitude of 4 060 m above sea. Currently, the farm raises approximately 2 400 alpacas subjected to performance recording for fleece weight and fibre quality traits. Since 2007, this alpaca population has been under formal genetic evaluation and selection with the general aim of increasing the amount of fine fibre per individual (Morante et al., 2009). To reach this goal, genetic parameters and breeding values are routinely estimated for fleece weight, diameter, fibre mean variability, fibre length, fibre medulation, shearing interval, textile value, and morphological traits (Cruz et al., 2019, 2024; Gutiérrez et al., 2009, 2011; Cervantes et al., 2010). Furthermore, the breed (either Suri or Huacaya) and the colour coat are considered when planning matings.

Directional selection for production traits may cause selective sweep signatures on the genome, characterised by reduced diversity and extensive linkage disequilibrium in different chromosomal areas (Sabeti et al., 2002). Favourable alleles increase in frequency more rapidly than expected for a neutral genetic scenario and further alter the allele frequencies of genomic variants near the beneficial mutation (Sabeti et al., 2002). Despite the scientific and economic importance of alpaca fibre production, studies aiming at the identification of selective sweeps in alpaca are still scant (see Ghildiyal et al., 2023 for a review). Marín et al. (2018) studied the genetic variation of two genes (*MC1R* and *ASIP*) involved in coat colour to gain insights into domestication and management of South American camelids. Mendoza et al. (2019), using fluorescence *in situ* hybridisation, located potential genes for alpaca fibre development and colour. Recently, More et al. (2023) analysed genotypes from alpacas belonging to three different Peruvian farms to investigate genetic factors influencing fibre diameter. Using the Cross-population Extended Haplotype Homozygosity (**XP-EHH**) statistics to compare fine (≤ 18 μm) and coarse (> 23 μm) fibre groups across the alpaca autosomes (**VPA**), they identified four candidate regions on VPA6, VPA9, and VPA29 associated with fibre diameter.

The analysis of the selected alpaca population of Pacomarca can be of interest to ascertain the genomic basis of alpaca fibre production. The goals of the current research were a) to explore the impact of a selection programme for fibre traits on the alpaca genome; and b) to contribute to the knowledge of the genomic basis of alpaca fibre production. Most analyses were carried out at the whole population level under the assumption that sweeps due to selection may still be incomplete. However, to gain insights into the effect of the selection programme on the genome, 176 individuals, out of 630, were classified into two different subpopulations (less selected and highly selected) to ascertain the existence of nearly fixed sweeps.

Material and methods

Samples

This study focuses on DNA samples obtained from 630 alpaca individuals, only including 85 males and 545 females. A complete description of the samples used is given in Supplementary Table S1. The genotyped sample included 135 parents-offspring trios. All sampled individuals belonged to the Huacaya breed and were mainly white-coated, except for two individuals that had an LFX (i.e. light beige) coat colour. Individuals were born between 2002 (one individual) and 2021 (60 individuals). Of the 630 alpacas, two subpopulations were defined: a less selected subpopulation (**LS**; 49 individuals born between 2002 and 2008) and a highly selected subpopulation (**HS**; 127 individuals born in 2020 and 2021). On average, fibre diameter was higher in the less selected population than in the highly selected population (24.16 μm ± 2.28 and 19.35 μm ± 1.62, respectively). The total pedigree available, managed in the Pacomarca experimental farm, consisted of 14 457 individuals with up to seven generations depth, born from 1992 to 2022. Using the program ENDOG v4.8 (Gutiérrez and Goyache, 2005), equivalent to discrete generations (**t**), was computed for each individual in the pedigree as the sum of overall known ancestors of the terms computed as the sum of $(1/2)^n$, where n is the number of generations separating the individual to each known ancestor (Maignel et al., 1996).

Genotyping, quality control and population structure analyses

All alpaca samples were genotyped using the Affymetrix Custom Alpaca genotyping array containing 76 508 single–nucleotide polymorphisms (**SNPs**; Calderon et al., 2021). Although genotyping was carried out in different periods, no batch effect could be ascertained (Supplementary Figure S1). Furthermore, while Pacomarca has been a closed farm under a breeding programme since 2007, there is no defined family structure among its individuals. The software Axiom™ Analysis Suite v4.0.3.3 (Thermo Fisher Scientific, Waltham, MA) was used to create both standard genotypic.ped and.map files. The physical positions of the SNPs were mapped to the reference genome *VicPac3.1* (Richardson et al., 2019). Although the assembly *VicPac3.1* covers 90% of the alpaca genome in just 103 scaffolds, with 76% of the genome mapped to the 36 pairs of the alpaca autosomes and the X chromosome (Richardson et al., 2019), preliminary annotations of the assembly predicted 22 462 coding genes and 29 337 isoforms, therefore making the *VicPac3.1* assembly useful to explore genomic features in alpacas (Richardson et al., 2019).

Autosomal SNPs with unassigned chromosomal positions were removed, and only those with a minimum genotype call rate of 0.99 were retained. Furthermore, up to 3% of individual missingness was allowed, and a sample call rate of 0.99 was considered. Using 135 parent-offspring trios genotyped, SNP array data were filtered on Mendelian errors, using the program Plink v1.9 (Chang et al., 2015), which are mainly due to systematic technical issues causing Allele-Drop-in and Allele-Drop-out events. Compared with classical editing strategies using minor allele frequency and deviation from Hardy–Weinberg proportions, this strategy has been proven to be useful in removing genotyping errors while keeping putatively informative loci and dataset size (Arias et al., 2022).

A total of 6 944 mismatches were identified and the 2 948 loci on which these Mendelian errors were ascertained, were removed from the dataset. A total of 297 monomorphic loci were removed as well. Finally, a total of 42 990 SNPs located on the 36 alpaca autosomes were analysed. Missing genotypes were phased and imputed using Beagle v5.4 (Browning et al., 2018, 2021) with default parameters. A relatedness matrix was computed using Plink v1.9 (Chang et al., 2015) and further used to perform Principal Component Analysis (**PCA**) using the '*stats*' package of R. Eigenvectors computed for each individual via PCA were used to construct dispersion plots using the library ggplot2 v3.5.1 of R (Wickham, 2016). Relatedness matrix was further used to ascertain if the molecular kinship of all pairs formed by the 135 individuals with genotyped parents included in the sample departed from the general distribution of kinship in the whole genotyped population.

Identification and assessment of positive selective sweeps

Two approaches, using three statistics, the integrative Haplotype Score (**iHS;** Voight et al., 2006), the number of segregating Sites by Length (**nSL;** Ferrer-Admetlla et al., 2014) and XP-EHH (Sabeti et al., 2007), were used to identify candidate selective signals:

The first approach used two within-population statistics, iHS and nSL, to identify candidate selective signals. The iHS statistic compares the Extended Haplotype Homozygosity (**EHH**) between derived and ancestral alleles within a population, searching for loci where the derived allele resides on a longer haplotype than the ancestral allele (Voight et al., 2006). Therefore, the iHS values are less affected by the demographic history. The nSL statistic measures haplotype lengths in terms of the number of segregating sites in the sample, making it more robust to recombination rate variations. Both iHS and nSL aim to identify incomplete sweeps, where the selected allele is not fixed in the sample. Selective signals identified within either iHS or nSL were further overlapped to identify candidate selective signals within the whole population.

The second approach aimed to assess whether some of the candidate selective sweeps identified with both iHS and nSL were fixed within a target subpopulation. Here, the less selected subpopulation, previously defined, was used as the reference, while the highly selected subpopulation was used as the target. The XP-EHH approach was performed, which compares the integrated EHH profiles at the same SNP between two genetically similar subpopulations (reference and target). This statistic assumes that the genome of the reference subpopulation has been less influenced by directional selection compared to the target population. It also incorporates information on the relationship between allele frequencies and linkage disequilibrium measurements with neighbouring alleles. Therefore, the XP-EHH test is useful to detect both ongoing and nearly fixed selection signatures, making it highly appropriate for scenarios involving complete selective sweeps. Furthermore, this statistic offers increased power even with small sample sizes (Eydivandi et al., 2021; Álvarez et al., 2020; Pickrell et al., 2009).

All analyses were carried out using the program selscan v1.0.4 (Szpiech and Hernandez, 2014), fitting default parameters: maximum Extended Haplotype Homozygosity (EHH) extension in bp ("--max-*extend*" option) 1000000, maximum gap allowed between two SNPs in bp ("--max-*gap*" option) 200000, EHH decay cutoff ("--*cutoff*" option) 0.05. The output results for each SNP were frequency-normalised over all chromosomes using the program *norm*, provided with selscan. As implemented by default in the program norm, normalised iHS and nSL values higher than |2| were used to identify SNPs under selection. Regarding the XP-EHH test, normalised values > 2 were used to identify SNPs under selection in the target population. This normalisation was carried out using default parameters as well: number of frequency bins ("–*bins*" option) 100.

Selective candidate genomic regions

Two sets of selective candidate genomic regions were constructed: one based on the candidate selective signals identified by both iHS and nSL statistics, and a second set based on the XP-EHH statistic only.

Selective candidate genomic regions were constructed as follows: The 13 797 bp (mean distance between SNPs in the analysed array) up- and downstream flanking regions of the retained SNPs within each approach used (either iHS – nSL or XP-EHH) were identified as candidate selective signals. Then, using the merge function of the BedTools software v2.31.0 (Quinlan and Hall, 2010), flanking regions of the SNPs were merged into a single selective candidate genomic region.

To improve the robustness of the identification of selective sweep, the *intersect* function of the BedTools software v2.31.0 (Quinlan and Hall, 2010) was used to overlap the set of candidate selective signals, identified by both iHS and nSL, with the potential selective signals found in their surrounding areas, identified within either iHS or nSL only. This process enabled the merging of overlapping candidate selective signals and adjacent selective signals into a single candidate genomic region, enhancing the robustness of the selection signature analyses (Arias et al., 2023; Goyache et al., 2021; Álvarez et al., 2020). The distribution of the selective candidate genomic regions identified was illustrated using the RIdeogram package version 0.2.2 of R version 4.1.2 (Hao et al., 2020).

Enrichment and functional annotation analyses

Enrichment analyses were first performed within the bounds of selective candidate genomic regions identified by each approach (iHS – nSL or XP-EHH only) to obtain two sets of candidate genes. Genes identified in both approaches were assigned as potential candidate genes. These genes were retrieved based on the annotation of the alpaca reference genome *VicPac 3.1.* (Richardson et al., 2019). Protein sequence similarity was performed using an *Ovis aries* background, as a complete *Vicugna pacos* background is not available, to confirm their classification within the same group of orthologous genes identified by Eukaryotic Genome Annotation Pipeline (NCBI). Gene ontology annotations and Kyoto Encyclopedia of Genes and Genomes pathways were obtained from the DAVID database (Huang et al., 2009).

Functional annotation clustering was then carried out using DAVID with only the potential candidate genes, with an *Ovis aries* background and fitting "highest" as the stringency criterion. An enrichment score threshold of 1.3, which is equivalent to the Fisher's exact test *P*-value of 0.05, was used to identify significantly

enriched functional clusters compared with the whole ovine reference genome background. Using enrichment scores rather than single annotation terms (clusters) enabled the identification of broader biological functions. This approach, which moves the focus of the biological function from individual genes to that of biological processes (Tipney and Hunter, 2010; Huang et al., 2009), allowed us to investigate relationships between gene ontology annotation terms.

Results

Mean pedigree depth was t = 2.3 ± 0.7, varying from t = 0 to t = 3.8. The less selected subpopulation had a pedigree depth of t = 1.3 ± 0.4 whereas that of the highly selected population was t = 3.0 ± 0.5. Fig. 1 shows the dispersion of the individuals genotyped according to the eigenvalues computed using a relatedness matrix for the first two principal components computed via PCA. Principal component 1, on the X-axis, explained 12% of the genomic variance, whereas Principal component 2, on the Y-axis, explained 7% of the variance. Fig. 1A illustrates the dispersion according to the subpopulation to which each individual was assigned. Fig. 1B illustrates the dispersion of the same first two principal components according to the trio status of the individuals. Our results suggest that PCA using a relatedness matrix did not allow to assess the existence of genetic structuring in the analysed population.

Since between-individuals relatedness may bias the results obtained, the distribution of molecular kinship of the pairs formed by the 135 individuals with their parents genotyped (e.g. those including all parents-offspring pairs in the sample) was compared with all other pairs in the population. The highest kinship value among the offspring-related pairs was 0.257, whereas this value for the other individuals was 0.273 (Fig. 2). The distribution of the molecular kinship values among those subsamples overlapped, even in the extreme values of the distribution, therefore suggesting that, if existing, possible bias of the results presented due to relatedness would be low. It is worth noting that the number of pairs with extreme values (≥ 0.25; i.e. fullsibs or parent-offspring kinship) is low (37 pairs of individuals; 0.04% out of the total pairs of the whole population).

Identification of selective signals within approach

Table 1 summarises the number of SNPs, as selective signals, identified per autosomes. Detailed information is provided in Supplementary Tables S2 and S3. After normalisation of the raw values obtained, the iHS test identified a total 1 605 SNPs as selective signals distributed across 35 of the 36 autosomes (Supplementary Table S2), with no selective signals found on VPA36. The nSL test identified 1 024 SNPs as selective signals located on 33 autosomes (Supplementary Table S3). A total of 544 selective signals overlapped using both the *iHS* and the *nSL* statistics and were further considered as candidate selective signals. No candidate selective signals were identified on VPA30, VPA31, VPA34 and VPA36. The autosome carrying the highest number of candidate selective signals was VPA8 (52 SNPs; 9.6% of the total candidate selective signals), followed by VPA3, VPA4, and VPA17, each containing between 45 and 48 candidate selective signals (from 8.3 to 8.8% of the total).

A total of 965 SNPs had normalised values higher than 2 for the XP-EHH test, meaning that these candidate selective signals were about to be fixed in the highly selected alpaca subpopulation analysed when compared with the less selected reference population (Table 1; Supplementary Table S4). Of them, a total of 106 SNPs and 33 SNPs were identified as selective signals by either the iHS or the nSL test, respectively. Notably, only 20 SNPs overlapped between the XP-EHH results and the set of candidate selective signals identified by both iHS and nSL (Table 1).

Selective candidate genomic regions

The overlap of the flanking regions of the 544 candidate selective signals, with the areas surrounding them identified using either iHS or nSL, allowed the identification of 27 SNPs adjacent to the candidate selective signals. These 571 SNPs allowed to construct a total of 509 selective candidate genomic regions (Table 2; Supplementary Table S5) covering 14.6 Mb of the alpaca genome (Table 2; Supplementary Table S5). The location of these selective candidate genomic regions across autosomes is illustrated in Fig. 3. A total of 453 selective candidate genomic regions (89%) included one SNP only, while the remaining 56 regions (11%) clustered from two to four SNPs.

These clustered genomic regions were distributed across four autosomes (VPA1, VPA3, VPA22, and VPA32), each carrying one selective candidate region, including three SNPs, covering from 27.6 to 70.7 kb of the autosomes. One candidate region gathering four SNPs (58.7 kb) was identified on VPA11. Eleven autosomes (VPA7, VPA11, VPA12, VPA18, VPA22, VPA25, VPA26, VPA28, VPA29, VPA33, and VPA35) gathered less than 10 candidate regions including a total of 60 SNPs as candidate selective signals (10.5%). Four autosomes (VPA3, VPA4, VPA8, and VPA17) gathered 180 candidate regions (35.4%) and 200 SNPs (35.0%), covering 5.2 Mb of these four chromosomes.

The 965 SNPs identified by XP-EHH allowed to construct a total of 119 selective candidate genomic regions fixed in the highly selected subpopulation (Supplementary Table S6). These regions were distributed across 23 autosomes only, covering 3.28 Mb of the alpaca genome. Notably, three autosomes (VPA2, VPA10, and VPA20) gathered 49 selective candidate genomic regions (41% of the total), covering 1.35 Mb of these three chromosomes.

Functionality of candidate genes linked to selective sweeps

Gene-annotation enrichment analysis, performed within the 509 selective candidate genomic regions identified using both iHS and nSL, allowed to identify a total of 293 candidate genes (Supplementary Table S7). These candidate genes were located in 46 different scaffolds mapped on 32 different autosomes. A full description of these 293 genes, including their identification, autosome location, and gene ontology terms to which they are associated, is given in Supplementary Table S7. Seven different autosomes (VPA3, VPA4, VPA5, VPA8, VPA16, VPA17, and VPA27) carried 15 or more candidate genes, together comprising a total of 141 genes, corresponding to 48.1% of the candidate genes.

A total of 11 candidate genes, such as the *PTPRN* (*tyrosine phosphatase receptor type N*) gene on VPA5, the *PTPN21* (*protein tyrosine phosphatase non-receptor type 21*) gene on VPA6, the *FRK* (*fyn related Src family tyrosine kinase*) and the *PTPRK* (*protein tyrosine phosphatase receptor type K*) genes on VPA8, the *PTPN23* (*protein tyrosine phosphatase non-receptor type 23*) gene on VPA17, and the *PTPRT* (*protein tyrosine phosphatase receptor type T*) gene on VPA19 were associated with gene ontology terms related with phosphorylation and dephosphorylation (Supplementary Table S7). Furthermore, a total of 28 candidate genes, such as the *FOXP1* (*forkhead box P1*) and the *MITF* (*melanocyte inducing transcription factor*) genes on VPA17, or the *RNF10* (*ring finger protein 10*) and the *ZNF664* (*zinc finger protein 664*) genes on VPA32, were associated with gene ontology terms related to RNA polymerase activity (Supplementary Table S7).

Selective candidate genomic regions identified using the XP-EHH test were annotated as well. Within the bounds, a total of

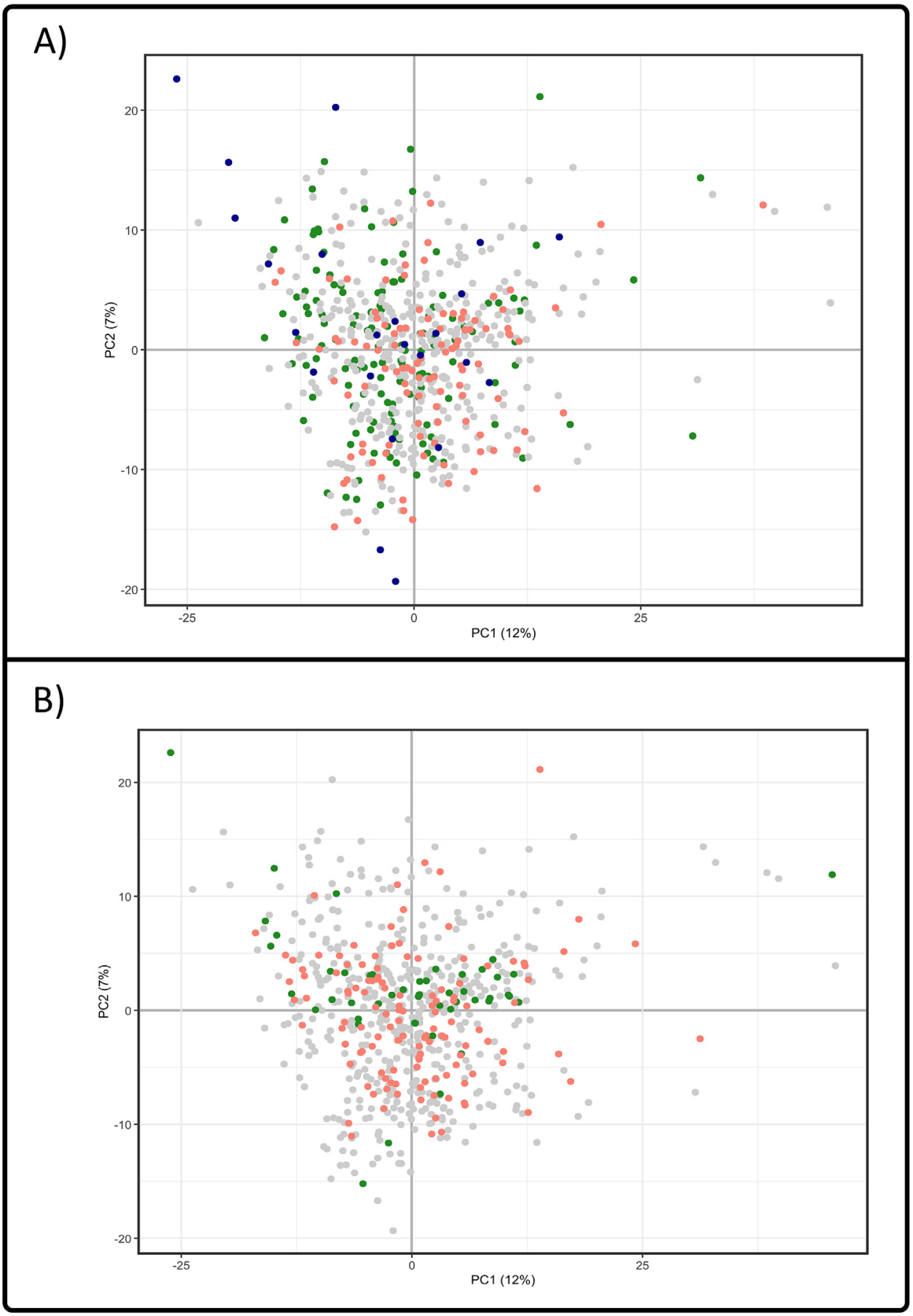

Fig. 1. Dispersion plot of individual alpaca genomes constructed according to the first two principal components identified via principal component analyses using a relatedness matrix. Principal component 1, on the X-axis, and Principal component 2, on the Y-axis, explained 12% and 7% of the total variance, respectively. Fig. 1A illustrates the dispersion of the parents-offspring trios (father, mother and offspring in blue, pink and green, respectively) and individuals with parents non-genotyped (in colour grey). Fig. 1B illustrates the individuals of the highly (target) and less (reference) selected subpopulations used for Cross-population Extended Haplotype Homozygosity (XP-EHH) analyses in pink and green, respectively, while non-selected individuals are in light grey.

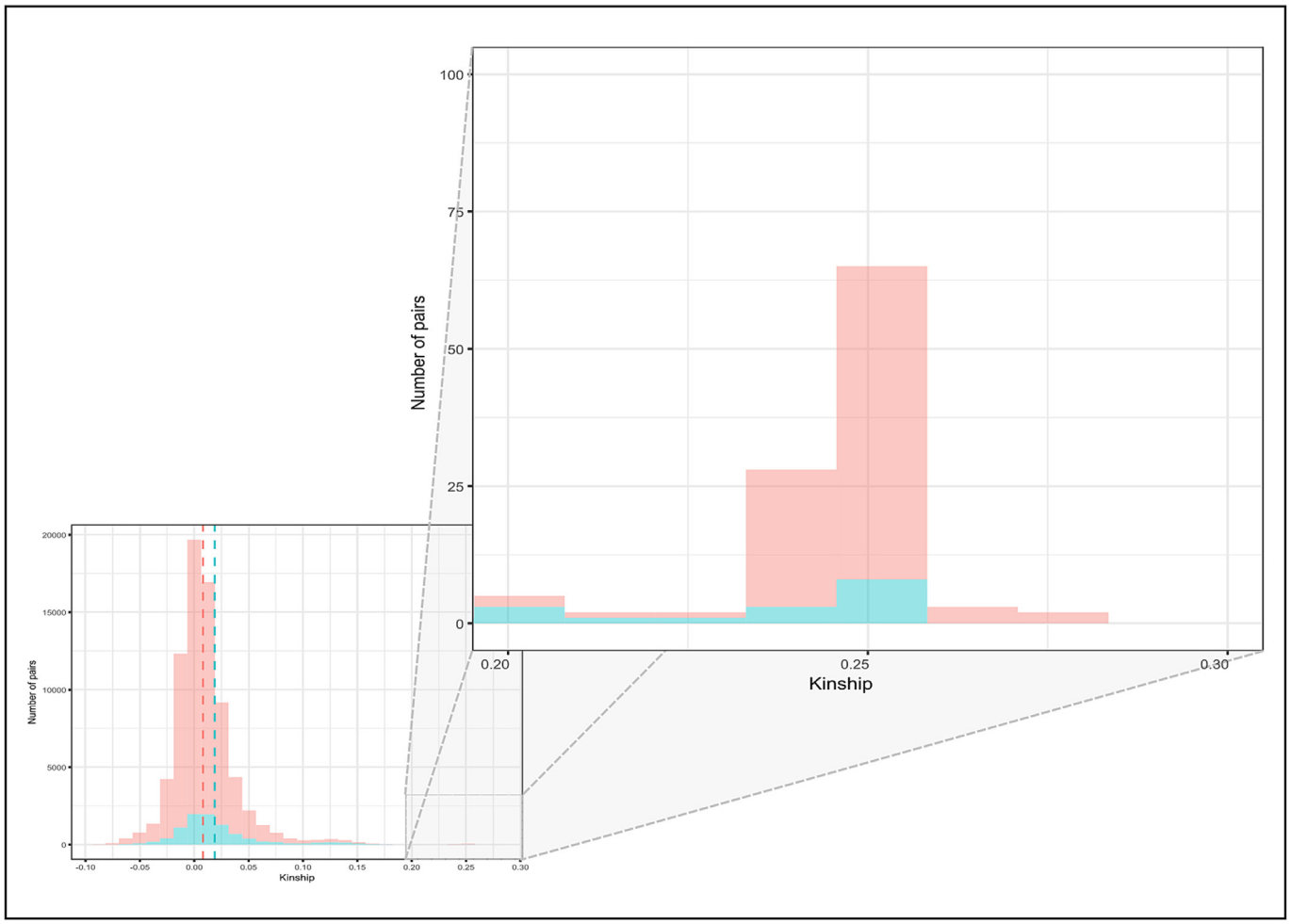

Fig. 2. Distribution of the molecular kinship values among the individuals included in the alpaca sample genotyped. The kinship values of the pairs formed by the 135 individuals with their parents genotyped are in blue, whereas the other are in red. The distribution of the extreme values computed is zoomed up to facilitate the assessment of the overlap.

54 different potential candidate genes were identified (Supplementary Table S6). Of these, ten candidate genes overlapped with those candidate genes identified when using both iHS and nSL statistics (Supplementary Table S7). These ten candidate genes, such as the *WDHD1* (*WD repeat and HMG-box DNA binding protein 1*) gene on VPA6, the *EYA4* (*EYA transcriptional coactivator and phosphatase 4*) and the *SYNE1* (*spectrin repeat containing nuclear envelope protein 1*) genes on VPA8, the *FOXP1* (*forkhead box P1*) gene on VPA17, or *NOTCH3* (*notch receptor 3*) gene on VPA22, were mainly involved in DNA replication and DNA and RNA transcription.

Functional annotation conducted within each approach, allowed to identify 22 different functional term clusters (Supplementary Table S8) when iHS – nSL approach was performed while no functional clusters were identified when XP-EHH statistic was used. Of them, only five functional term clusters were significantly enriched (enrichment score higher than 1.3; Table 3). Annotation Cluster 1 (enrichment score = 2.50) included six genes located on six different autosomes (*PTPRT*, *PTPRN*, *PTPN21*, *PTPRK*, *PTPN23*, and *PTPN13*) coding tyrosine-protein phosphatases. Annotation Cluster 2 (enrichment score = 2.19) included nine different genes (*SH3GL3*, *CACNB3*, *DOCK3*, *RUSC2*, *CASKIN2*, *SH3PXD2B*, *SRGAP3*, *MPP7*, *FRK*, and *SH3D19*), located on eight different chromosomes, belonging to the *SH3* (*Src homology 3*) protein domain (IPR036028:SH3 and IPR001452:SH3). Annotation Clusters 3 and 4 (enrichment scores of 1.50 and 1.48, respectively) included four (*EGFLAM*, *LAMA2*, *SLIT2*, and *CELSR3*) and nine (*WDHD1*, *WDR70*, *ERCC8*, *WDR31*, *SCAP*, *WDFY2*, *AAMP*, *CDC40*, and *DCAF7*) genes, mainly located on VPA3, VPA8, and VPA17, encoding laminins. Finally, Annotation Cluster 5 (enrichment score = 1.33) included four genes (*PAPPA*, *SVEP1*, *CSMD2*, and *CSMD1*), located on VPA4, VPA13, and VPA26, encoding proteins involved in the complement activation. Two of the genes validated using XP-EHH statistics, the *WDHD1* (*WD repeat and HMG-box DNA binding protein 1*) gene and the *CSMD1* (*CUB and Sushi multiple domains 1*) gene, were included in Annotation Clusters 4 and 5, respectively.

Discussion

The current analyses have been carried out using a selected population formed by related individuals. Therefore, genomic data are far from being independent and could be biased due to relatedness and family structuring. Although our dataset included a non-negligible number of related individuals, PCA based on a relatedness matrix computed would support that family structure, if existent, is not evident (Fig. 1). Furthermore, the distribution of the molecular kinship values of the offspring-related pairs clearly overlaps with that of the individuals with no parents in sample. We consider that, in this scenario, the results obtained are not biased due to relatedness. In any case, if bias existed, it would con-

Table 1
Number of SNPs (with percentage in brackets), per alpaca autosome, identified as in putative selection, based on normalised values higher than |2| using the integrative Haplotype Score (iHS) and the number of segregating Sites by Length (nSL) statistics. The number of SNPs identified by both tests and, therefore, considered candidate SNPs is given as well. The table also reports the number of SNPs (with percentage in brackets) identified as in putative selection (normalised values higher than 2) in the highly selected alpaca subpopulation using the Cross-population Extended Haplotype Homozygosity (XP-EHH). The number of SNPs identified by XP-EHH coinciding with those identified via iHS, nSL (or both tests) is also given.

					XP-EHH coincidences with		
Autosome	iHS	nSL	Candidate	XP-EHH	iHS	nSL	Both tests
1	94 (5.9)	39 (3.8)	20 (3.7)				
2	157 (9.8)	58 (5.7)	31 (5.7)	81 (8.4)	9	2	1
3	89 (5.5)	70 (6.8)	48 (8.8)	24 (2.5)	3	2	2
4	85 (5.3)	66 (6.4)	48 (8.8)	20 (2.1)	1		
5	103 (6.4)	48 (4.7)	23 (4.2)	64 (6.6)	6	1	1
6	68 (4.2)	37 (3.6)	20 (3.7)	89 (9.2)	9	2	2
7	33 (2.1)	11 (1.1)	4 (0.7)	12 (1.2)	2		
8	110 (6.9)	101 (9.9)	52 (9.6)	42 (4.4)	4	2	1
9	47 (2.9)	25 (2.4)	11 (2.0)	9 (0.9)	1		
10	57 (3.6)	32 (3.1)	19 (3.5)	71 (7.4)	11	3	2
11	62 (3.9)	22 (2.1)	10 (1.8)	32 (3.3)	1		
12	95 (5.9)	14 (1.4)	6 (1.1)	20 (2.1)	9		
13	40 (2.5)	27 (2.6)	12 (2.2)	41 (4.2)			
14	33 (2.1)	25 (2.4)	16 (2.9)	67 (6.9)	5	1	1
15	41 (2.6)	39 (3.8)	25 (4.6)	2 (0.2)			
16	34 (2.1)	35 (3.4)	15 (2.8)	17 (1.8)			
17	71 (4.4)	90 (8.8)	45 (8.3)	40 (4.1)	3	5	2
18	16 (1.0)	9 (0.9)	6 (1.1)	8 (0.8)			
19	29 (1.8)	16 (1.6)	11 (2.0)	5 (0.5)			
20	54 (3.4)	16 (1.6)	10 (1.8)	65 (6.7)	27	3	3
21	30 (1.9)	24 (2.3)	14 (2.6)	55 (5.7)	4	5	1
22	32 (2.0)	17 (1.7)	4 (0.7)	8 (0.8)	1	2	
23	18 (1.1)	29 (2.8)	12 (2.2)	3 (0.3)			
24	13 (0.8)	28 (2.7)	11 (2.0)	13 (1.3)			
25	30 (1.9)	20 (2.0)	8 (1.5)	8 (08)			
26	20 (1.2)	13 (1.3)	7 (1.3)	18 (19)	1	1	1
27	37 (2.3)	39 (3.8)	21 (3.9)	30 (31)	2	1	1
28	10 (0.6)	11 (1.1)	3 (0.6)	6 (0.6)	1	1	1
29	11 (0.7)	5 (0.5)	2 (0.4)	7 (0.7)		1	
30	14 (0.9)			10 (1.0)	1		
31	2 (0.1)	3 (0.3)		18 (1.9)			
32	41 (2.6)	42 (4.1)	25 (4.6)	31 (3.2)	4	1	1
33	13 (0.8)	6 (0.6)	2 (0.4)	18 (1.9)			
34	9 (0.6)			21 (2.2)	1		
35	7 (0.4)	7 (0.7)	3 (0.6)	10 (1.0)			
Total	1 605	1 024	544	965	106	33	20

SNP = single–nucleotide polymorphisms.

tribute to accomplish to the general aim of the paper: to test if the selection programme of the Pacomarca population shaped the alpaca genome. Therefore, the population analysed can be considered informative on how directional selection can have shaped the genome of the alpacas managed in the Pacomarca's Experimental Station.

In the Pacomarca scenario analysed, favourable selection for an ancestral allele cannot be discarded. Therefore, our strategy for the identification of candidate selective signals and further construction of selective candidate genomic regions, considering both large positive and negative iHS and nSL scores, was potentially informative. Although negative iHS scores would mean that a derived allele would have swept up in frequency and, therefore, could be considered of more interest, large positive values can indicate that ancestral alleles hitchhike with the selected site (Álvarez et al., 2020; Schlamp et al., 2016). A similar rationale can be applied to the recombination-free counterpart of iHS, the nSL test. However, for the assessment of selective sweeps identified with both iHS and nSL statistics, large positive XP-EHH scores were considered because our interest was to ascertain if some of the selective sweeps identified with both iHS or nSL could have led to the fixation in the more recent (and 'highly' selected) alpaca subpopulation analysed. Furthermore, although the number of candidate selective signals identified by XP-EHH was low, these results suggest that directional selection is acting and, in some cases, the alpaca genome is shaped due to its effects.

The definition of the subpopulations (less selected and highly selected) used to contrast the effect of the selection programme was carried out using birth years, assuming that pedigree tends to accumulate with time in real closed populations. However, the separation of both subpopulations in generation terms is low ($t < 2$) and, as expected, the number of potential selective signals identified by XP-EHH as fixed in the target subpopulation was low as well. These results also mirror the average differences in performance between the subpopulations fitted: the diameter of fibre for the less selected population was higher than the highly selected population (24.16 μm ± 2.28 and 19.35 μm ± 1.62, respectively).

Biological importance of the potentially candidate genes and functional clusters identified

The candidate genes identified are enriched for biological functions, namely phosphorylation and RNA transcription and replication, that are known to be related to fibre production in other species (Kalds et al., 2022; Zhao et al., 2017, 2020; Yue et al., 2015; Geng et al., 2013).

Table 2
Summary, per alpaca autosome, of the selective candidate genomic regions considered under selection using both the integrative Haplotype Score (iHS) and the number of segregating Sites by Length (nSL) tests. The number of candidate areas, the number of SNPs and the total length (in kb) of the sweeps identified are given.

Autosome	Candidate areas	SNPs	Length
1	16	20	480.0
2	29	32	840.3
3	45	49	1 316.6
4	42	48	1 235.6
5	23	25	634.7
6	20	22	551.9
7	4	4	110.4
8	52	54	1 434.9
9	10	12	302.1
10	19	19	524.3
11	8	12	251.9
12	6	6	165.6
13	12	13	331.1
14	14	17	427.2
15	22	26	646.0
16	15	15	413.9
17	41	49	1 207.5
18	6	6	165.6
19	11	11	303.5
20	10	10	275.9
21	13	15	362.0
22	4	6	110.4
23	12	13	331.1
24	10	13	300.1
25	7	8	219.1
26	7	7	193.2
27	20	22	569.0
28	3	3	82.8
29	2	2	55.2
32	21	26	653.5
33	2	2	55.2
35	3	4	82.8
Total	509	571	14 633.0

SNP = single–nucleotide polymorphisms.

The phosphorylation of proteins is a reversible form of post-translational modification that regulates such biological processes as cell signalling, gene expression, molecular recognition, and metabolism (Zhai et al., 2008). Phosphorylation affects the development of animal hair follicles, hair growth, and hair quality (Chen et al., 2021; Foitzik et al., 2000) due to its major role in the assembly of keratin filaments (Herbert et al., 1997; Ku et al., 1996). The family of genes encoding keratin-associated proteins (*KAP*) is widely involved in most wool production traits in sheep (Kalds et al., 2022).

Differences in gene expression, linked to RNA activity, have been reported to influence major traits in sheep such as wool growth (Liu et al., 2014), fleece resistance to rot (Smith et al., 2010), wool diameter (Zhao et al., 2017, 2020; Yue et al., 2015), or wool pigmentation (Peñagaricano et al., 2012) in sheep, fibre production in cashmere goat (Dong et al., 2013; Geng et al., 2013), and wool fibre diameter in Angora rabbits (Huang et al., 2024).

This general interpretation of the biological importance of the set of candidate genes identified for fibre-related traits was confirmed by cluster analyses.

The tyrosine-protein phosphatase-related Annotation Clusters 1 and 2 are likely to be acting on fibre production as well. Some genes belonging to the tyrosine phosphatase-encoding family, particularly *PTPN3*, have been associated with wool production traits in Iranian sheep and Chinese Merino (Mohamadipoor Saadatabadi et al., 2021; Wang et al., 2014) because of their role in the maintenance of tight junction integrity between the cell membrane and the cytoskeleton.

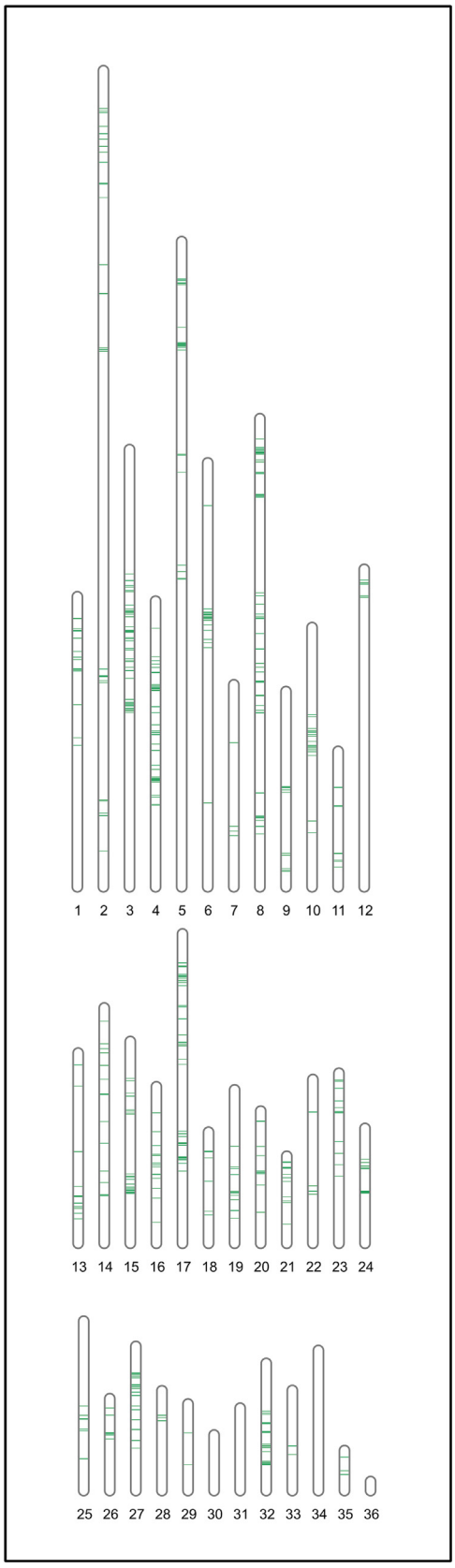

Fig. 3. Ideogram illustrating, per alpaca autosome, the distribution of the 509 selective candidate genomic regions identified using the integrative Haplotype Score (iHS) and the number of segregating Sites by Length (nSL) statistics in the Pacomarca alpaca population.

Table 3
Significantly enriched functional term clusters (enrichment score higher than 1.3) for genes identified within selective candidate genomic regions in the alpaca population from the Pacomarca experimental farm, following DAVID analysis. A full description of the genes included in functional Annotation Clusters can be found in Supplementary Table S7.

Functional Cluster (Enrichment score)	Term	Description	*P*-Value	Fold change	Gene name (alpaca autosome no.)
Cluster 1 (2.50)	INTERPRO	IPR000242:PTP_cat	2.66×10^{-4}	10.4	*PTPRT* (19), *PTPRN* (5), *PTPN21* (6), *PTPRK* (8), *PTPN23* (17), *PTPN13* (2)
	SMART	SM00194:PTPc	6.21×10^{-4}	8.6	
	INTERPRO	IPR003595:Tyr_Pase_cat	1.37×10^{-3}	7.3	
	SMART	SM00404:PTPc_motif	4.56×10^{-3}	5.5	
	UP_SEQ_FEATURE	DOMAIN:Tyrosine-protein phosphatase	7.14×10^{-3}	5.0	
	INTERPRO	IPR000387:Tyr_Pase_dom	7.73×10^{-3}	4.9	
	UP_SEQ_FEATURE	DOMAIN:Tyrosine–specific protein phosphatases	1.01×10^{-2}	4.6	
	INTERPRO	IPR029021:Prot-tyrosine_phosphatase-like	1.61×10^{-2}	4.1	
Cluster 2 (2.19)	INTERPRO	IPR036028:SH3-like_dom_sf	5.54×10^{-3}	3.4	*CACNB3* (12), *DOCK3* (17), *RUSC2* (4), *CASKIN2* (16), *SH3PXD2B* (22), *SRGAP3* (17), *MPP7* (35), *FRK* (8), *SH3D19* (2)
	INTERPRO	IPR001452:SH3_domain	6.30×10^{-3}	3.3	
	UP_SEQ_FEATURE	DOMAIN:SH3	7.45×10^{-3}	3.2	
Cluster 3 (1.50)	UP_SEQ_FEATURE	DOMAIN:Laminin G	2.47×10^{-2}	6.4	*EGFLAM* (3), *LAMA2* (8), *SLIT2* (2), *CELSR3* (17)
	INTERPRO	IPR001791:Laminin_G	2.57×10^{-2}	6.3	
	SMART	SM00282:LamG	4.96×10^{-2}	4.8	
Cluster 4 (1.48)	UP_SEQ_FEATURE	DOMAIN:Laminin G	2.47×10^{-2}	6.4	*WDHD1* (6), *WDR70* (3), *ERCC8* (3), *WDR31* (4), *SCAP* (17), *WDFY2*, *AAMP* (5), *CDC40* (8), *DCAF7* (16)
	INTERPRO	IPR001791:Laminin_G	2.57×10^{-2}	6.3	
	SMART	SM00282:LamG	4.96×10^{-2}	4.8	
Cluster5 (1.33)	UP_KW_DOMAIN	KW-0768~Sushi	3.21×10^{-2}	5.7	*PAPPA* (4), *SVEP1* (4), *CSMD2* (13), *CSMD1* (26)
	INTERPRO	IPR035976:Sushi/SCR/CCP_sf	4.20×10^{-2}	5.2	
	INTERPRO	IPR000436:Sushi_SCR_CCP_dom	4.20×10^{-2}	5.2	
	UP_SEQ_FEATURE	DOMAIN:Sushi	4.93×10^{-2}	4.8	
	SMART	SM00032:CCP	7.64×10^{-2}	4.0	

Furthermore, the SH3 domains (Annotation Cluster 2) are eukaryotic protein domains that participate in a plethora of cellular processes including signal transduction, proliferation, and cellular movement (Tatárová et al., 2012). Several studies indicate that tyrosine phosphorylation could play a significant role in the regulation of SH3 domains and, therefore, genes belonging to Annotation Cluster 1 may interact with the series of candidate genes involved in phosphorylation in the modulation of fibre growth and quality (Chen et al., 2021; Foitzik et al., 2000).

The potentially candidate genes included Annotation Clusters 3 and 4 encode laminines. Laminines are well known to be involved in skin and hair follicle development, due to their roles in basal keratinocytes proliferation and blood vessel function that can affect oxygen supply and, therefore, hair morphogenesis (Wegner et al., 2016; Li et al., 2002). Laminine-coding genes have been reported as candidates for wool-related traits in Chinese Merino sheep (Wang et al., 2014).

Although genes coding proteins belonging to the Sushi domain are known to have functions in the central nervous system (Kraus et al., 2006), the functionality of Annotation Cluster 5 can be explained in consistency with the selection for fibre production of the alpaca population analysed as well: the *PAPPA2* gene is involved in the proliferation of dermal papilla cells (Wu et al., 2021) and the regulation of wool fibre diameter in sheep (Yue et al., 2023); the *SVEP1* gene is involved in blood flow regulation by modulating the vascular tone (Morris et al., 2022); the highly similar CUB and sushi multiple domains 1 (*CSMD1*) and 2 (*CSMD2*) have been proved to be overexpressed in epithelial tissues and can be involved in their development via regulating the complement activation and inflammation (Kraus et al., 2006).

Overall, although the current analysis focuses on the ascertainment of how the alpaca genome was shaped due to the application of a directional selection programme for fibre traits, it cannot be discarded that some of the selective candidate genomic regions identified harbour genes that are beneficial to other traits different than fibre traits. For instance, one of the candidate genes identified, the melanocyte–inducing transcription factor (*MITF*) gene (Supplementary Table S7), involved in coat colour, could suggest that a more or less hidden selection for the more commercial white fibre (Cruz et al., 2021) may be acting to some extent in the Pacomarca alpaca population analysed. A decreased *MITF* gene expression suppresses skin pigmentation (Yokoyama et al., 2008). Other genetically correlated traits, such as type or size traits (Cervantes et al., 2010), may also be favoured by the application of Pacomarca's selection programme. It is worth noting that none of the genes suggested as possible candidates for fibre traits in alpacas in previous reports (see Ghildiyal et al., 2023, for a review) were identified in our population. The current results can constitute a template for further studies aiming at the characterisation of the genetic basis of alpaca fibre production.

In summary, the current research adds to the understanding of the genomics of fibre production in a species that has not been studied enough yet. Our results suggest that the genome of the alpaca population analysed was shaped by directional selection, involving multiple genomic areas across most autosomes, which fits well with traits that are expected to be mainly controlled by genes of additive effect. The genomic areas under selection identified suggest that physiological patterns involving phosphorylation and gene expression (RNA-polymerases) can be of major importance in fibre production in this species. However, the results obtained are far from being conclusive: the SNP array available is of medium density only and the current reference alpaca genome is just a draft where the assembly *VicPac3.1* covers 90% of the alpaca genome in just 103 scaffolds (Richardson et al., 2019). A further developed version of that reference genome will allow a more accurate identification of chromosomal areas under selection than that provided by scaffolds. Further development and refinement of the broad genomic map of the alpaca will make it possible to confirm the results of the current study. This research provides a new set of genomic areas and potential candidate genes allowing researchers to focus on them to obtain new insights on the genetic basis of alpaca fibre production.

Supplementary material

Supplementary Material for this article (https://doi.org/10.1016/j.animal.2025.101674) can be found at the foot of the online page, in the Appendix section.

Ethics approval

Alpaca blood samples for DNA isolation were obtained following the Peruvian National Law No. 30407, "Animal Protection and Welfare Law", in effect in Peru since 7 January 2016. The Dean of the College of Animal Science approved the protocol *in lieu* of the UNALM "Ethics Committee for Scientific Research" No. 0345-2018-CU-UNALM—22 October 2018.

Data and model availability statement

None of the data were deposited in an official repository. Genotypes are property of the INCA Group and are subjected to embargo. However, they may be available upon reasonable request to the corresponding author.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) did not use any AI and AI-assisted technologies.

Author ORCIDs

K. D. Arias: https://orcid.org/0000-0001-6710-4208.
M. More: https://orcid.org/0000-0001-8677-993X.
F. Goyache: https://orcid.org/0000-0002-6872-1045.
A. Cruz: https://orcid.org/0000-0002-9462-4986.
G. Gutiérrez: https://orcid.org/0000-0002-1896-0048.
I. Cervantes: https://orcid.org/0000-0002-0750-6480.
J. P. Gutiérrez: https://orcid.org/0000-0003-3142-2356.

CRediT authorship contribution statement

K.D. Arias: Writing – original draft, Formal analysis, Conceptualisation. **M. More:** Writing – review & editing, Investigation. **F. Goyache:** Writing – original draft, Supervision, Investigation, Conceptualisation. **A. Cruz:** Writing – review & editing, Resources, Data curation. **A. Burgos:** Resources, Data curation. **G. Gutiérrez:** Writing – original draft, Conceptualisation. **I. Cervantes:** Writing – review & editing. **J.P. Gutiérrez:** Writing – original draft, Conceptualisation.

Declaration of interest

The authors declare that this research was conducted with no financial and non-financial competing interests.

Acknowledgments

None.

Financial support statement

This research received no specific grant from any funding agency, commercial or not-for-profit section.

References

Allain, D., Renieri, C., 2010. Genetics of fibre production and fleece characteristics in small ruminants, Angora rabbit and South American camelids. Animal 4, 1472–1481. https://doi.org/10.1017/S1751731110000029.
Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N.A., Goyache, F., 2020. Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. Scientific Reports 10, 2824. https://doi.org/10.1038/s41598-020-59839-x.
Arias, K.D., Álvarez, I., Gutiérrez, J.P., Fernández, I., Menéndez, J., Menéndez-Arias, N. A., Goyache, F., 2022. Understanding Mendelian errors in SNP arrays data using a Gochu Asturcelta pig pedigree: genomic alterations, family size and calling errors. Scientific Reports 12, 19686. https://doi.org/10.1038/s41598-022-24340-0.
Arias, K.D., Lee, H., Bozzi, R., Álvarez, I., Gutiérrez, J.P., Fernandez, I., Menéndez, J., Beja-Pereira, A., Goyache, F., 2023. Ascertaining the genetic background of the Celtic-Iberian pig strain: a signatures of selection approach. Journal of Animal Breeding and Genetics 141, 96–112. https://doi.org/10.1111/jbg.12829.
Browning, B.L., Zhou, Y., Browning, S.R., 2018. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics 103, 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015.
Browning, B.L., Tian, X., Zhou, Y., Browning, S.R., 2021. Fast two-stage phasing of large-scale sequence data. The American Journal of Human Genetics 108, 1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.
Calderon, M., More, M.J., Gutierrez, G.A., Ponce de León, F.A., 2021. Development of a 76k alpaca (Vicugna pacos) single nucleotide polymorphisms (SNPs) microarray. Genes 12, 291. https://doi.org/10.3390/genes12020291.
Cervantes, I., Pérez-Cabal, M.A., Morante, R., Burgos, A., Salgado, C., Nieto, B., Goyache, F., Gutiérrez, J.P., 2010. Genetic parameters and relationships between fibre and type traits in two breeds of Peruvian alpacas. Small Ruminant Research 88, 6–11. https://doi.org/10.1016/j.smallrumres.2009.10.016.
Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7. https://doi.org/10.1186/s13742-015-0047-8.
Chen, Y., He, D., Li, Y., Luo, F., Zhang, M., Wang, J., Chen, L., Tao, J., 2021. A study of the phosphorylation proteomic skin characteristics of Tan sheep during the newborn and er-mao stages. Tropical Animal Health and Production 54, 30. https://doi.org/10.1007/s11250-021-02899-6.
Cruz, A., Morante, R., Gutiérrez, J.P., Torres, R., Burgos, A., Cervantes, I., 2019. Genetic parameters for medullated fiber and its relationship with other productive traits in alpacas. Animal 13, 1358–1364. https://doi.org/10.1017/S1751731118003282.
Cruz, A., Yucra, A., Gutiérrez, G.A., Burgos, A., Morante, R., Gutiérrez, J.P., Cervantes, I., Wurzinger, M., 2021. Colorimetry analysis of coat color and its relationship with fiber traits in alpacas. Animal 15, 100219. https://doi.org/10.1016/j.animal.2021.100219.
Cruz, A., Murillo, Y., Burgos, A., Yucra, A., Morante, R., Quispe, M., Quispe, C., Quispe, E., Gutiérrez, J.P., 2024. Genetic parameters for different types of medullated fibre in alpacas. Journal of Animal Breeding and Genetics 141, 521–530. https://doi.org/10.1111/jbg.12861.
Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., Wang, W., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature Biotechnology 31, 135–141. https://doi.org/10.1038/nbt.2478.
Eydivandi, S., Roudbar, M.A., Karimi, M.O., Sahana, G., 2021. Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia. Scientific Reports 11, 2834. https://doi.org/10.1038/s41598-021-82625-2.
Fan, R., Gu, Z., Guang, X., Marín, J.C., Varas, V., González, B.A., Wheeler, J.C., 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, M.W., Dong, C., 2020. Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. Genome Biology 21, 159. https://doi.org/10.1186/s13059-020-02080-6.
Ferrer-Admetlla, A., Liang, M., Korneliussen, T., Nielsen, R., 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. Molecular Biology and Evolution 31, 1275–1291. https://doi.org/10.1093/molbev/msu077.
Foitzik, K., Lindner, G., Mueller-Roever, S., Maurer, M., Botchkareva, N., Botchkarev, V., Handjiski, B., Metz, M., Hibino, T., Soma, T., Paolo Dotto, G., Paus, R., 2000. Control of murine hair follicle regression (catagen) by TGF-β1 in vivo. The FASEB Journal 14, 752–760. https://doi.org/10.1096/fasebj.14.5.752.
Geng, R., Yuan, C., Chen, Y., 2013. Exploring differentially expressed genes by RNA-Seq in cashmere goat (Capra hircus) skin during hair follicle development and cycling. PLoS One 8, e62704. https://doi.org/10.1371/journal.pone.0062704.
Ghildiyal, K., Panigrahi, M., Kumar, H., Rajawat, D., Nayak, S.S., Lei, C., Bhushan, B., Dutt, T., 2023. Selection signatures for fiber production in commercial species: a review. Animal Genetics 54, 3–23. https://doi.org/10.1111/age.13272.
Goyache, F., Pérez-Pardal, L., Fernández, I., Traoré, A., Menéndez-Arias, N.A., Álvarez, I., 2021. Ancient autozygous segments subject to positive selection suggest adaptive immune responses in West African cattle. Gene 803, 145899. https://doi.org/10.1016/j.gene.2021.145899.

Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. Journal of Animal Breeding and Genetics 122, 172–176. https://doi.org/10.1111/j.1439-0388.2005.00512.x.
Gutiérrez, J.P., Goyache, F., Burgos, A., Cervantes, I., 2009. Genetic analysis of six production traits in Peruvian alpacas. Livestock Science 123, 193–197. https://doi.org/10.1016/j.livsci.2008.11.006.
Gutiérrez, J.P., Varona, L., Pun, A., Morante, R., Burgos, A., Cervantes, I., Pérez-Cabal, M.A., 2011. Genetic parameters for growth of fiber diameter in alpacas. Journal of Animal Science 89, 2310–2315. https://doi.org/10.2527/jas.2010-3746.
Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., Chen, J., 2020. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Computer Science 6, e251.
Herbert, B.R., Molloy, M.P., Yan, J.X., Gooley, A.A., Bryson, W.G., Williams, K.L., 1997. Characterisation of wool intermediate filament proteins separated by micropreparative two-dimensional electrophoresis. Electrophoresis 18, 568–572. https://doi.org/10.1002/elps.1150180339.
Huang, D., Ding, H., Wang, Y., Wang, X., Zhao, H., 2024. Integration analysis of hair follicle transcriptome and proteome reveals the mechanisms regulating wool fiber diameter in Angora rabbits. International Journal of Molecular Sciences 25, 3260. https://doi.org/10.3390/ijms25063260.
Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4, 44–57. https://doi.org/10.1038/nprot.2008.211.
Kadwell, M., Fernandez, M., Stanley, H.F., Baldi, R., Wheeler, J.C., Rosadio, R., Bruford, M.W., 2001. Genetic analysis reveals the wild ancestors of the llama and the alpaca. Proceedings Biological Sciences 268, 2575–2584. https://doi.org/10.1098/rspb.2001.1774.
Kalds, P., Zhou, S., Gao, Y., Cai, B., Huang, S., Chen, Y., Wang, X., 2022. Genetics of the phenotypic evolution in sheep: a molecular look at diversity-driving genes. Genetics Selection Evolution 54, 61. https://doi.org/10.1186/s12711-022-00753-3.
Kraus, D.M., Elliott, G.S., Chute, H., Horan, T., Pfenninger, K.H., Sanford, S.D., Foster, S., Scully, S., Welcher, A.A., Holers, V.M., 2006. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. Journal of Immunology 176, 4419–4430. https://doi.org/10.4049/jimmunol.176.7.4419.
Ku, N.O., Liao, J., Chou, C.F., Omary, M.B., 1996. Implications of intermediate filament protein phosphorylation. Cancer Metastasis Reviews 15, 429–444. https://doi.org/10.1007/BF00054011.
Li, S., Harrison, D., Carbonetto, S., Fässler, R., Smyth, N., Edgar, D., Yurchenco, P.D., 2002. Matrix assembly, regulation, and survival functions of laminin and its receptors in embryonic stem cell differentiation. The Journal of Cell Biology 157, 1279–1290. https://doi.org/10.1083/jcb.200203073.
Liu, N., Li, H., Liu, K., Yu, J., Bu, R., Cheng, M., De, W., Liu, J., He, G., Zhao, J., 2014. Identification of skin-expressed genes possibly associated with wool growth regulation of Aohan fine wool sheep. BMC Genetics 15, 144. https://doi.org/10.1186/s12863-014-0144-1.
Maignel, L., Boichard, D., Verrier, E., 1996. Genetic variability of French dairy breeds estimated from pedigree information. Interbull Bulletin 14, 49–56.
Marín, J.C., Rivera, R., Varas, V., Cortés, J., Agapito, A., Chero, A., Chávez, A., Johnson, W.E., Orozco-terWengel, P., 2018. Genetic variation in coat colour genes MC1R and ASIP provides insights into domestication and management of South American camelids. Frontiers in Genetics 9, 487. https://doi.org/10.3389/fgene.2018.00487.
McFadden, Wm.D., Neale, P.E., 1954. A meter for studying quantitative wool differences among sheep. Journal of Animal Science 13, 229–233. https://doi.org/10.2527/jas1954.131229x.
McGregor, B.A., 2006. Production, attributes and relative value of alpaca fleeces in southern Australia and implications for industry development. Small Ruminant Research, Special Issue: South American Camelids 61, 93–111. https://doi.org/10.1016/j.smallrumres.2005.07.001.
Mendoza, M.N., Raudsepp, T., Alshanbari, F., Gutiérrez, G., Ponce de León, F.A., 2019. Chromosomal localization of candidate genes for fiber growth and color in Alpaca (Vicugna pacos). Frontiers in Genetics 10, 583. https://doi.org/10.3389/fgene.2019.00583.
Mohamadipoor Saadatabadi, L., Mohammadabadi, M., Amiri Ghanatsaman, Z., Babenko, O., Stavetska, R., Kalashnik, O., Kucher, D., Kochuk-Yashchenko, O., Asadollahpour Nanaei, H., 2021. Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. BMC Veterinary Research 17, 369. https://doi.org/10.1186/s12917-021-03077-4.
Morante, R., Goyache, F., Burgos, A., Cervantes, I., Pérez-Cabal, M.A., Gutiérrez, J.P., 2009. Genetic improvement for alpaca fibre production in the Peruvian Altiplano: the Pacomarca experience. Animal Genetic Resources/resources Génétiques Animales/recursos Genéticos Animales 45, 37–43. https://doi.org/10.1017/S1014233909990307.
More, M., Veli, E., Cruz, A., Gutiérrez, J.P., Gutiérrez, G., Ponce de León, F.A., 2023. Genome-wide association study of fiber diameter in alpacas. Animals 13, 3316. https://doi.org/10.3390/ani13213316.
Morris, G.E., Denniff, M.J., Karamanavi, E., Andrews, S.A., Kostogrys, R.B., Bountziouka, V., Ghaderi-Najafabadi, M., Shamkhi, N., McConnell, G., Kaiser, M.A., Carleton, L., Schofield, C., Kessler, T., Rainbow, R.D., Samani, N.J., Webb, T. R., 2022. The integrin ligand SVEP1 regulates GPCR-mediated vasoconstriction via integrins α9β1 and α4β1. British Journal of Pharmacology 179, 4958–4973. https://doi.org/10.1111/bph.15921.
Peñagaricano, F., Zorrilla, P., Naya, H., Robello, C., Urioste, J.I., 2012. Gene expression analysis identifies new candidate genes associated with the development of black skin spots in Corriedale sheep. Journal of Applied Genetics 53, 99–106. https://doi.org/10.1007/s13353-011-0066-9.
Pérez-Cabal, M.A., Cervantes, I., Morante, R., Burgos, A., Goyache, F., Gutiérrez, J.P., 2010. Analysis of the existence of major genes affecting alpaca fiber traits. Journal of Animal Science 88, 3783–3788. https://doi.org/10.2527/jas.2010-2865.
Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B. S., Barsh, G.S., Myers, R.M., Feldman, M.W., Pritchard, J.K., 2009. Signals of recent positive selection in a worldwide sample of human populations. Genome Research 19, 826–837. https://doi.org/10.1101/gr.087577.108.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
Quispe, E., Quispe, M., Quispe, C., Poma, A., Paucar-Chanca, R., Cruz, A., McGregor, B. A., 2022. Relationships between the incidence and degree of medullation with the diameter of alpaca fibers evaluated using a novel device based on artificial intelligence. The Journal of the Textile Institute 114, 1016–1031. https://doi.org/10.1080/00405000.2022.2105110.
Richardson, M.F., Munyard, K., Croft, L.J., Allnutt, T.R., Jackling, F., Alshanbari, F., Jevit, M., Wright, G.A., Cransberg, R., Tibary, A., Perelman, P., Appleton, B., Raudsepp, T., 2019. Chromosome-level alpaca reference genome VicPac3.1 improves genomic insight into the biology of new world camelids. Frontiers in Genetics 10, 586. https://doi.org/10.3389/fgene.2019.00586.
Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837. https://doi.org/10.1038/nature01140.
Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. https://doi.org/10.1038/nature06250.
Schlamp, F., van der Made, J., Stambler, R., Chesebrough, L., Boyko, A.R., Messer, P. W., 2016. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. Molecular Ecology 25, 342–356. https://doi.org/10.1111/mec.13485.
Smith, W.J., Li, Y., Ingham, A., Collis, E., McWilliam, S.M., Dixon, T.J., Norris, B.J., Mortimer, S.I., Moore, R.J., Reverter, A., 2010. A genomics-informed, SNP association study reveals FBLN1 and FABP4 as contributing to resistance to fleece rot in Australian Merino sheep. BMC Veterinary Research 6, 27. https://doi.org/10.1186/1746-6148-6-27.
Szpiech, Z.A., Hernandez, R.D., 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Molecular Biology and Evolution 31, 2824–2827. https://doi.org/10.1093/molbev/msu211.
Tatárová, Z., Brábek, J., Rösel, D., Novotný, M., 2012. SH3 domain tyrosine phosphorylation – sites, role and evolution. PLoS One 7, e36310. https://doi.org/10.1371/journal.pone.0036310.
Tipney, H., Hunter, L., 2010. An introduction to effective use of enrichment analysis software. Human Genomics 4, 202. https://doi.org/10.1186/1479-7364-4-3-202.
Varas, V., Vásquez, J.P., Rivera, R., Longo, A., Valdecantos, P.A., Wheeler, J.C., Johnson, W.E., Marín, J.C., 2020. Interbreeding among South American camelids threatens species integrity. Journal of Arid Environments 181, 104249. https://doi.org/10.1016/j.jaridenv.2020.104249.
Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biology 4, e72.
Wang, Z., Zhang, H., Yang, H., Wang, S., Rong, E., Pei, W., Li, H., Wang, N., 2014. Genome-wide association study for wool production traits in a Chinese merino sheep population. PLoS One1 9, e107101. https://doi.org/10.1371/journal.pone.0107101.
Wegner, J., Loser, K., Apsite, G., Nischt, R., Eckes, B., Krieg, T., Werner, S., Sorokin, L., 2016. Laminin α5 in the keratinocyte basement membrane is required for epidermal–dermal intercommunication. Matrix Biology 56, 24–41. https://doi.org/10.1016/j.matbio.2016.05.001.
Wheeler, J.C., Chikhi, L., Bruford, M.W., 2006. Genetic analysis of the origins of domestic South American camelids. In: Zeder, M.A., Bradley, D., Emshwiller, E., Smith, B.D. (Eds.), Documenting Domestication: NEw Genetic and Archaeological Paradigms. University of California Press, Berkeley, CA, USA, pp. 329–341. https://doi.org/10.1525/9780520932425-026.
Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY, USA.
Wu, T., Wang, S., Jin, Q., Lv, X., Sun, W., 2021. PAPPA2 promote the proliferation of dermal papilla cells in Hu sheep (Ovis aries) by regulating IGFBP5. Genes 12, 1490. https://doi.org/10.3390/genes12101490.
Wurzinger, M., Gutiérrez, G., 2022. Alpaca breeding in Peru: from individual initiatives towards a national breeding programme? Small Ruminant Research 217, 106844. https://doi.org/10.1016/j.smallrumres.2022.106844.
Yokoyama, S., Feige, E., Poling, L.L., Levy, C., Widlund, H.R., Khaled, M., Kung, A.L., Fisher, D.E., 2008. Pharmacologic suppression of MITF expression via HDAC inhibitors in the melanocyte lineage. Pigment Cell & Melanoma Research 21, 457–463. https://doi.org/10.1111/j.1755-148X.2008.00480.x.
Yue, Y., Guo, T., Liu, J., Guo, J., Yuan, C., Feng, R., Niu, C., Sun, X., Yang, B., 2015. Exploring differentially expressed genes and natural antisense transcripts in sheep (Ovis aries) skin with different wool fiber diameters by digital gene expression profiling. PLoS One 10, e0129249. https://doi.org/10.1371/journal.pone.0129249.

Yue, L., Lu, Z., Guo, T., Liu, J., Yang, B., Yuan, C., 2023. Proteome analysis of alpine merino sheep skin reveals new insights into the mechanisms involved in regulating wool fiber diameter. International Journal of Molecular Sciences 24, 15227. https://doi.org/10.3390/ijms242015227.

Zhai, B., Villén, J., Beausoleil, S.A., Mintseris, J., Gygi, S.P., 2008. Phosphoproteome analysis of Drosophila melanogaster embryos. Journal of Proteome Research 7, 1675–1682. https://doi.org/10.1021/pr700696a.

Zhao, J., Liu, N., Liu, K., He, J., Yu, J., Bu, R., Cheng, M., De, W., Liu, J., Li, H., 2017. Identification of genes and proteins associated with anagen wool growth. Animal Genetics 48, 67–79. https://doi.org/10.1111/age.12480.

Zhao, J., Qin, H., Xin, J., Liu, N., Han, R., Perez-Campo, F.M., Li, H., 2020. Discovery of genes and proteins possibly regulating mean wool fibre diameter using cDNA microarray and proteomic approaches. Scientific Reports 10, 7726. https://doi.org/10.1038/s41598-020-64903-7.