

RESEARCH ARTICLE

Genome-wide Association Study Identifies Loci for the Polled Phenotype in Yak

Chunnian Liang¹✉, Lizhong Wang²✉, Xiaoyun Wu¹, Kun Wang², Xuezhi Ding², Mingcheng Wang², Min Chu¹, Xiuyue Xie², Qiang Qiu²‡*, Ping Yan¹‡*

1 Key Laboratory of Yak Breeding Engineering Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Science, Lanzhou, China, **2** State Key Laboratory of Grassland Agro-ecosystem, College of Life Science, Lanzhou University, Lanzhou, China

✉ These authors contributed equally to this work.

‡ These authors jointly supervised this work.

* qiuqiang@lzu.edu.cn (QQ); pingyanlz@163.com (PY)



OPEN ACCESS

Citation: Liang C, Wang L, Wu X, Wang K, Ding X, Wang M, et al. (2016) Genome-wide Association Study Identifies Loci for the Polled Phenotype in Yak. PLoS ONE 11(7): e0158642. doi:10.1371/journal.pone.0158642

Editor: Ramona Natacha PENA i SUBIRÀ, University of Lleida, SPAIN

Received: March 10, 2016

Accepted: June 20, 2016

Published: July 7, 2016

Copyright: © 2016 Liang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The sequencing data for this project have been deposited in the GenBank under accession code PRJNA309369.

Funding: This study was supported by grants from the Agricultural Science and Technology Innovation Program (www.caas.cn/en/research/research_program/index.shtml); CAAS-ASTIP-2014-LIHPS-01; P.Y.), National Beef Cattle Industry Technology & System (www.beefsys.com/; CARS-38; Q.Q.), the National High Technology Research and Development Program of China (www.most.gov.cn/eng/programmes1/2013AA102505_3-2; Q.Q.), the National Natural Science Foundation of China (www.nsf.gov/).

Abstract

The absence of horns, known as the polled phenotype, is an economically important trait in modern yak husbandry, but the genomic structure and genetic basis of this phenotype have yet to be discovered. Here, we conducted a genome-wide association study with a panel of 10 horned and 10 polled yaks using whole genome sequencing. We mapped the *POLLED* locus to a 200-kb interval, which comprises three protein-coding genes. Further characterization of the candidate region showed recent artificial selection signals resulting from the breeding process. We suggest that expressional variations rather than structural variations in protein probably contribute to the polled phenotype. Our results not only represent the first and important step in establishing the genomic structure of the polled region in yak, but also add to our understanding of the polled trait in bovid species.

Introduction

The yak (*Bos grunniens*) is endemic to the Qinghai-Tibet Plateau (QTP), the largest and harshest highland in the world [1]. More than 14 million domestic yaks are currently distributed across the QTP, providing the food, shelter, fuel and transport that enable nomadic Tibetans and other pastoralists to survive in this harsh climate; indeed, the yak has become an iconic symbol of Tibet [1, 2]. They are also strongly integrated into Tibetans' socio-cultural life. Due to its important position in Tibetan daily life, yak production and its related products are the cornerstone of Tibetan animal husbandry [3, 4]. In addition, yak live on unpolluted highland pasture where they produce green, nutritional and healthy products, much valued by modern communities [3, 4].

The bovine polled phenotype, the highly desired and favorable trait in modern husbandry systems, has huge practical importance for breeders and is of special interest to geneticists [5, 6] (Fig 1A shows a horned and a polled yak). Yak horns are a major cause of injuries, particularly in feedlots and during transport [7]. Nowadays, commercial beef yaks are confined to barns and fenced-in enclosures such as pastures or corrals. More hornless yak can be

nsfc.gov.cn/; 31322052; Q.Q.), the Fok Ying Tung Education Foundation (www.hydef.edu.cn/; 151105; Q.Q.) and the Fundamental Research Funds for the Central Universities (www.moe.edu.cn/; lzujbky-2016-k04; Q.Q.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

accommodated in the same space compared to yak with horns, and the trait would reduce economic losses due to injuries to both humans and animals under these conditions [7]. Although dehorning at a young age is routinely practiced in horned breeds of yak, this method does not eradicate the problem and there are associated animal welfare concerns. Considering an autosomal dominant mode of inheritance for the polled trait, the approach also limits the ability to discriminate between heterozygous and homozygous polled animals [8, 9]. Thus, creating polled genetic markers to identify homozygous/heterozygous yak and breeding genetically polled yak based on non-invasive and high welfare methods is a promising alternative [7, 9], which would be valuable to modern yak husbandry in high altitude harsh environments [10]. In addition, identification of genes and causal variations associated with the polled phenotype will contribute to our knowledge and understanding of the molecular mechanisms that underlie horn differentiation and development in bovine species.

In cattle (*Bos taurus*), the *POLLED* locus has previously been mapped on the proximal end of bovine chromosome 1 (BTA1) [11]. More recent efforts to refine the polled locus and detect candidate causal mutations have included seeking additional microsatellite markers [12–14], BAC-based physical mapping [15], high-density SNP genotyping [16–19], targeted capture sequencing [17, 20, 21] and whole genome resequencing [16, 19]. Recently, at least two different alleles for polledness were reported in cattle, identifying an 80-kb duplication (BTA1: 1,909,352–1,989,480 bp) in Friesian original breeds (P_F allele) and a duplication of 212-bp (BTA1: 1,705,834–1,706,045 bp) in place of a 10-bp deletion (1,706,051–1,706,060 bp) in Celtic original breeds (P_C allele, a 202-bp insertion–deletion, InDel), respectively, suggesting the existence of allelic heterogeneity at the *POLLED* locus [17, 21, 22]. In addition, other sporadic mutations associated with the horn-like scurs phenotype have been described [23, 24]. Intriguingly, none of these mutations was located in known coding or regulatory regions [13, 16]. One plausible reason is that different alleles have been selected in different geographic regions or breeds, and world-wide and across-breed breeding using limited founders and artificial insemination have led to the complex inheritance pattern of the horn trait in different breeds, thus adding to the complexity of understanding the molecular basis of polledness [17]. Hence, simultaneous research in different Bovidae species needs to be undertaken to provide extra information [7, 19].

In the current breeding stage, polled yaks are developed deliberately by crossing polled cows (PP or Pp) with horned bulls (pp) by artificial insemination (S1 Fig) at the Datong Yak Breeding Farm of Qinghai Province, providing an ideal system to study the genomic structure and genetic basis of this phenotype [10]. Although our previous analysis based on an a priori candidate gene set detected associated signals, the actual location of the *POLLED* locus has so far not been confirmed as BTA1 in yak [10]. In the present study, we describe our efforts to determine the polled trait associated genome regions in yak based on whole genome sequencing; these were carried out independently from any recently published studies.

Results and Discussion

Genetic variants and population structure

We sequenced 10 horned and 10 polled yaks to an average depth (raw data) of 11.2× using an Illumina HiSeq2000 instrument, resulting in a total of 6.04 billion reads comprising approximately 595Gb of sequencing data. Using BWA-MEM software [25], reads were aligned to the *B. grunniens* reference genome with an average alignment rate of 91.20%, covering 99.26% of the genome [26] (S1 Table). After SNP calling using SAMtools [27], filtering the potential PCR duplicates, removing SNPs with potential errors and correcting the misalignments around

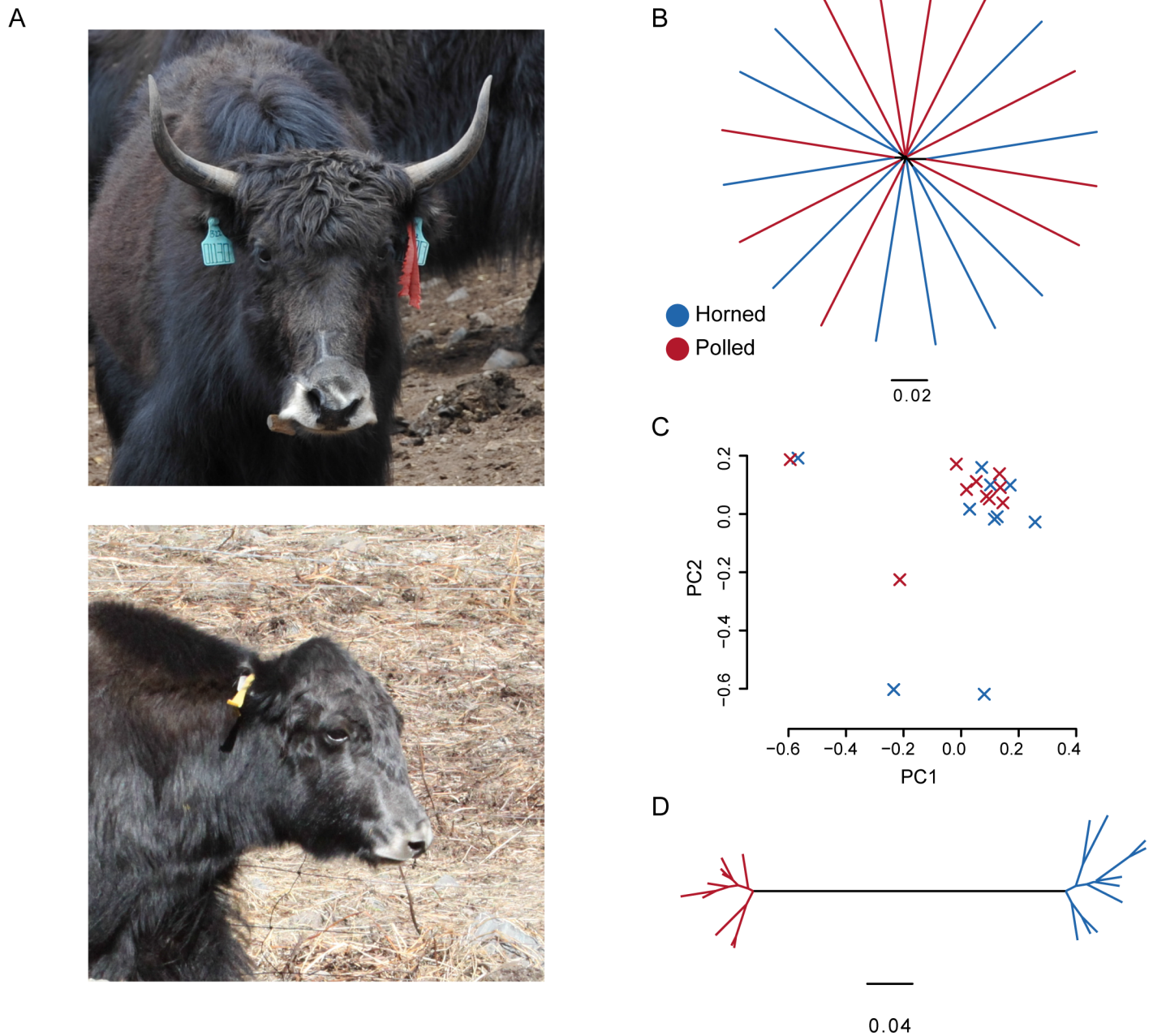


Fig 1. Phylogenetic and population structure of horned and polled yaks. (A) Photos of horned (above) and polled (below) yak herds, taken by Chunnian Liang. (B) A neighbor-joining phylogenetic tree constructed using whole-genome SNP data. The scale bar represents level of similarity. Horned (blue) and polled (red) samples are indicated. (C) Principal component (PC) analysis plots of the first two components. (D) A neighbor-joining phylogenetic tree constructed using SNP data for the GWAS region.

doi:10.1371/journal.pone.0158642.g001

InDels (details in Materials and Methods), approximately 8.4 million high quality SNPs were retained.

To examine the phylogenetic relationship between horned and polled yaks, we constructed a neighbor-joining tree based on our high-quality SNPs. The horned and polled yaks formed a mixed clade (Fig 1B), indicating that pairwise distances between horned and polled yaks were

not larger than those within each population. We also performed principle component analysis (PCA, Fig 1C and S2 Table) and population structure analysis (S2 Fig) based on the genotype data. Taken together, all of these results indicate no population genetic structure correlated with the horned/polled phenotypes, consistent with the relatively short time of breeding polled yaks. More importantly, the indistinguishable genomic background suggests that genome-wide association studies should enable high-resolution mapping of genomic regions associated with the horned/polled phenotype.

Identification of genome regions associated with the polled phenotype

Taking advantage of the yak population with no population stratification, we performed a genome-wide association study (GWAS) analysis between 10 horned and 10 polled animals using the Dominant model with PLINK [28]. This autosomal dominant Mendelian trait was mapped to a 200 kb interval between positions 1,122,103 and 1,322,666 bp on scaffold526_1 (P -values < 0.0001, which means that the probability of obtaining these frequencies by chance is very low, < 0.01%) (Fig 2A and 2B, S3 Fig). Despite the small sample size and the consequent relatively limited statistical power [29], this exclusive signal reached genome-wide significance in the GWAS analyses and appeared to align within the polled locus of BTA1 mapped in previous studies of cattle [11–19, 21]. Further, we found that the horned and polled individuals

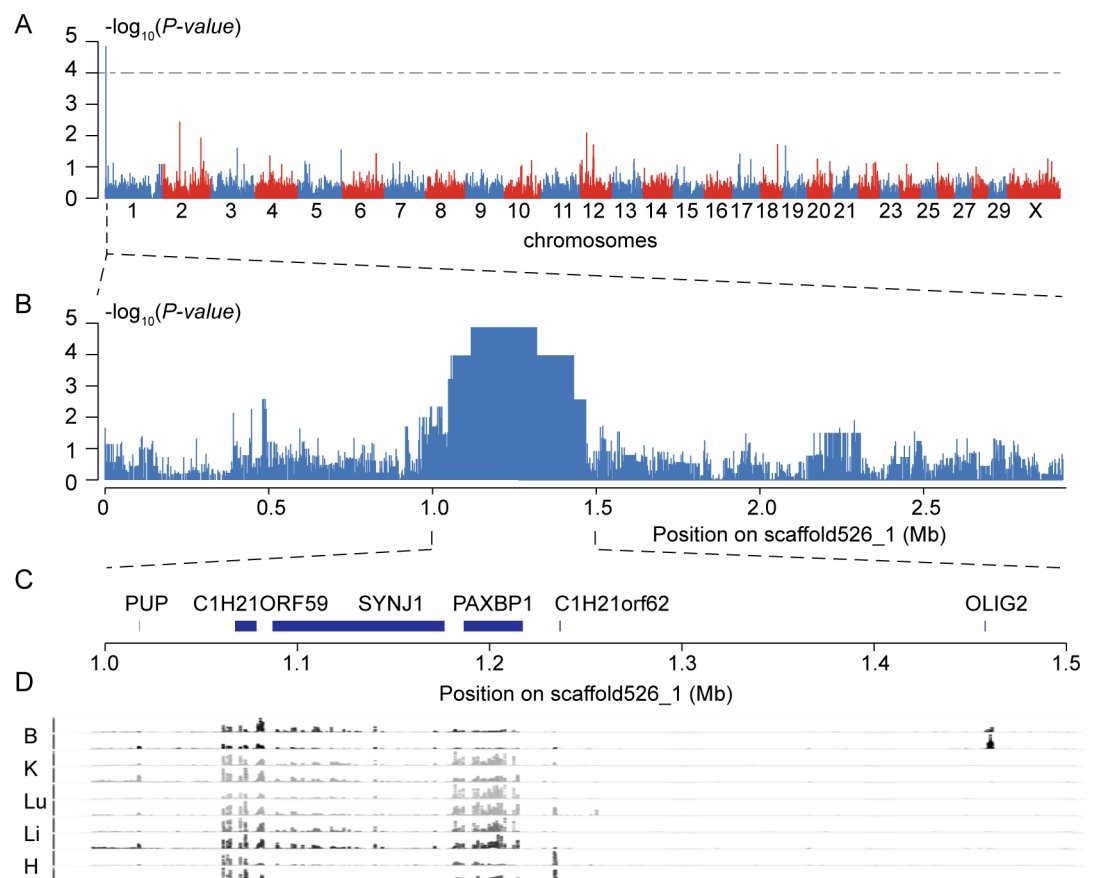


Fig 2. Associated mapping of the polled phenotype. (A) Genome-wide P values (y axis) are plotted along the genome (B) and magnification of scaffold526_1. (C) All genes around the candidate GWAS region. (D) diagram of read depths (X axes) of RNA-seq data mapping of five different tissues: brain (B), kidney (K), lung (Lu), liver (Li) and heart (H), each with two replicates (Y axes).

doi:10.1371/journal.pone.0158642.g002

clustered into two genetically distinct groups in a phylogenetic tree based on this region (Fig 1D). In contrast, there was no significant differentiation between these groups when the tree was constructed for the whole genome (Fig 1B). Moreover, most of these loci with P -values < 0.0001 (95%, 567 of 596) are entirely heterozygous (Pp) in polled animals, coinciding with the breeding practice of crossing polled cows (PP or Pp) with horned bulls (pp) by artificial insemination. Although all previous studies identified different polled mutations in different cattle breeds, all the clues lead to the same position on BTA1 [11–19, 21]; we therefore believe that the horned/polled trait in cattle and yak may share the same ontogenetic mechanism.

The region defined as the GWAS loci of the polled phenotype contained 3 protein-coding genes: *SYNJ1*, *PAXBP1* and *C1H21orf62* (Fig 2C). *SYNJ1* encodes synaptojanin 1, a key neural protein highly expressed in nerve terminals with essential roles in the regulation of synaptic vesicles in conventional synapses and hair cells [30, 31]. Recently, a histological analysis revealed that nervous tissue and hair follicle development have different features in horn buds and polled frontal skin during the development of the horn buds of bovine fetuses, implying that *SYNJ1* maybe have an important role in horn differentiation [32]. *PAXBP1* is an essential binding protein that regulates the proliferation of muscle precursor cells, which in turn, are involved in the development of normal craniofacial features and spine morphogenesis [33]. *C1H21orf62* is an uncharacterized protein. Since previous transcript profiling analyses of polled and horned tissues from cattle, using the Agilent 44 k bovine array, failed to find differential expression in any of the genes located in the *POLLED* locus [34], we sought to re-annotate the novel protein-coding genes in this region by mapping large-scale RNA-seq data for five tissues (brain, kidney, lung, liver and heart) from a previous study [35] of domestic yak. We were, however, unable to find any new gene or open reading frame (Fig 2D, details in Materials and Methods).

Previous genetic and genomic research has proposed two structural variants (P_F and P_C alleles) associated with the polled phenotype based on larger scale GWAS results in many cattle breeds [17]. We therefore examined whether these two structural variants exist in the yak genome based on our whole genome sequencing data ($> 10\times$), which should ensure quality and accuracy in the detection of structural variants [36]. Our results indicated that the yak genome does not include these two cattle candidate structural variants (S4 Fig and S3 Table), contrasting with previous studies indicating that there is extensive allelic heterogeneity of the polled trait in highly mobile bovid species. We further analyzed other structural variants from this associated region and could not identify any duplications or InDels associated with the polled trait in yak (S3 Table). Due to the small sample size, short breeding history and absence of homozygous polled (PP) individuals, our results cannot be used to refine the candidate genomic region nor to detect causal variants (causal SNPs or structural variants) in yak at present. However, we are convinced that a future study based on more samples with detailed pedigree information will narrow down the candidate region of this trait [37], and this highly confidential region should be the target of focused studies to establish the functional significance of this key trait in domestic yak.

Characterization of the polled interval

Considering the breeding practice developed by crossing polled cows (PP or Pp) with horned bulls (pp) (Schematic shown in S1 Fig), the candidate region identified could be expected to exhibit specific signatures of recent artificial selection in the polled population, including a high proportion of heterozygous, significantly differentiated nucleotide diversity levels and long-range haplotype homozygosity [38]. Based on these principles, we examined five different

parameters to evaluate detailed genetic polymorphism and differentiation between horned and polled yaks: nucleotide diversity (π), the proportion of shared and private SNPs, F_{ST} (population-differentiation statistic), d_{xy} (mean pairwise comparisons of the nucleotide difference between groups) and the linkage disequilibrium (LD). Population-specific estimates of π showed that the level of nucleotide diversity is higher in polled yaks than in horned yaks in the 200 kb GWAS region (Fig 3A), although a similar level of nucleotide diversity was observed in the rest of genome (mean pairwise nucleotide diversity of π_{horned} : 0.00139 ± 0.00081 ; π_{polled} : 0.00136 ± 0.00077 , S5 Fig). In addition, this region showed an elevated proportion of private

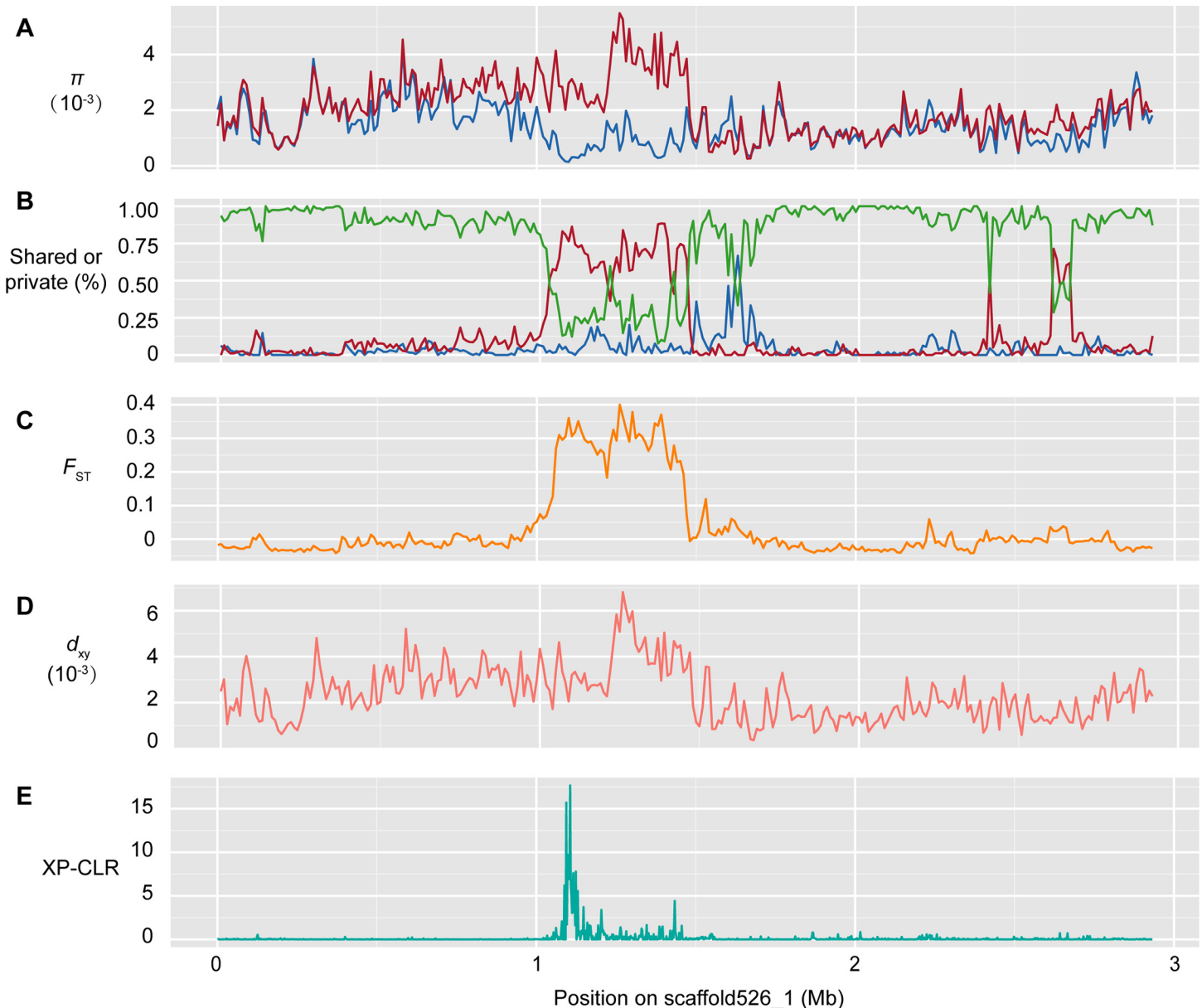


Fig 3. Distribution of population genomic parameters along scaffold526_1. The plots show: (A) the nucleotide diversity (π , blue for horned and red for polled yaks) for each population; (B) the proportion of shared polymorphisms among sites that are polymorphic in at least one population (green), the proportion of private polymorphisms among sites that are polymorphic within populations (blue for horned and red for polled yaks), private and shared polymorphisms shown in the same panel; (C) F_{ST} ; (D) d_{xy} ; and (E) XP-CLR of scaffold526_1.

doi:10.1371/journal.pone.0158642.g003

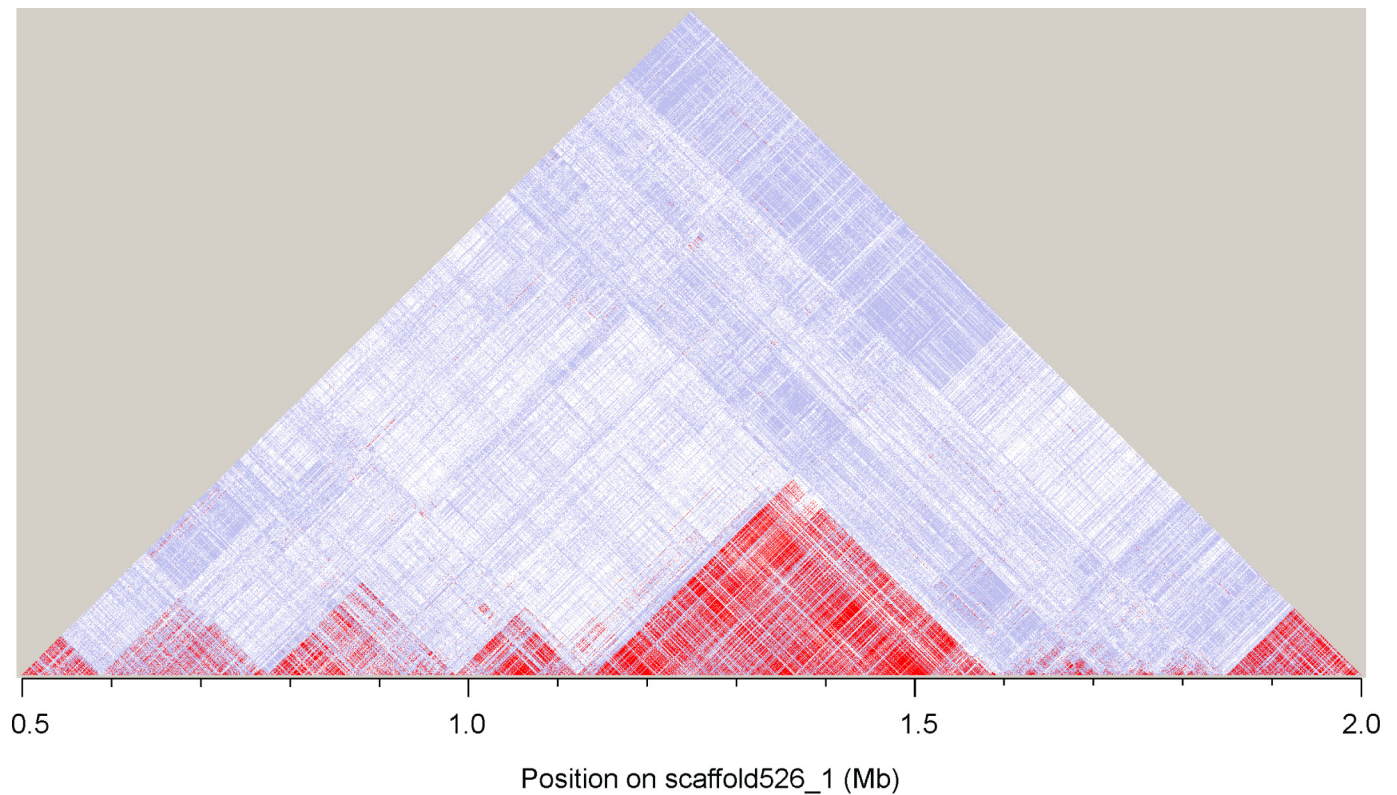


Fig 4. Haplotype block at linkage disequilibrium along the scaffold526_1.

doi:10.1371/journal.pone.0158642.g004

SNPs in polled yaks and a reduced level of shared SNPs between horned and polled yaks (Fig 3B), which also accord with the breeding approach. Furthermore, we found that the GWAS region implicated in the polled phenotype showed striking genetic differentiation between the horned and polled individuals in the F_{ST} analysis (Fig 3C, with a mean F_{ST} of only 0.0006, S6 Fig). The mean pairwise nucleotide difference between-group comparisons (d_{xy}) showed a divergence peak (more than 0.006) compared with the flanking regions (Fig 3D, with whole genome level of 0.0013 ± 0.0006). Linkage analysis for this scaffold also revealed a higher linkage disequilibrium (LD) value with one haplotype block of 450 kb (1.14~1.59Mb, Fig 4), which was probably defined by the causal allele and its linked neighbor variants. Taken together, these results indicate that this candidate genomic region tends to be highly diverged and exhibits clear signals of selection. As a complementary approach, we used a likelihood method (the cross-population composite likelihood ratio, XP-CLR [39]) to scan for extreme allele frequency differentiation over extended linked regions and found this region to have elevated XP-CLR values (Fig 3E), which means that the polled trait was affected by selection during breeding activities.

Potential genetic basis and future directions

Despite the fact that the *POLLED* locus has been easily mapped, fine characterization of this locus and a definitive description of the molecular basis of horn ontogenesis has proved more difficult than expected [17, 22]. One important reason is that complex genomic structural variations and a possible regulatory effect rather than nonsynonymous variations in protein sequence are the major contributors to the differentiation of the horned/polled trait [16, 17, 19,

[21]. A recent study reported the success of production of hornless dairy cattle using genome editing technology involving introgressing the Celtic original candidate *POLLED* allele (202-bp insertion–deletion), but the potential functional effect of this sequence variant remains unknown [22]. To date, none of the causative variants have been located in known coding or regulatory regions and no genes with a high probable impact have been identified in cattle [13, 16]. To test this result in yak, we scanned the SNPs within the coding regions of the *POLLED* locus and were unable to identify any nonsynonymous substitutions, consistent with the situation previously observed in cattle. In addition, differential expression studies among horned and polled cattle failed to reveal differences in gene expression located in the *POLLED* locus, but several genes outside this region showed a high level of expression divergence [19].

By combining our results with those of earlier studies, we suggest that unknown regulatory sequences and cis-regulation elements may reside in the *POLLED* locus, thus influencing horn development. Furthermore, the horned/polled trait is developmentally-related, and such traits are likely to be involved in the complex interaction of many genes [16, 19, 35]. This indicates that a high level of genetic heterogeneity is expected and that different species or breeds may have developed this phenotype as a result of different genomic strategies [17]. For example, sequence changes in the *POLLED* locus may affect the transcription factor and noncoding RNA binding, coordination of histone modifications and other chromatin remodeling activities, which can lead to transcription changes to horn-related genes in the *POLLED* locus and other regions of the genome [40]. Therefore, we must emphasize the important role of gene regulation in horn development. Future studies should focus on identification of novel regulated elements in the *POLLED* locus and involve an examination of detailed expression patterns across different horn developmental stages, which should reveal the precise mechanism responsible for the polled trait.

Conclusions

The importance of breeding polled yak has grown considerably due to animal welfare issues and the needs of modern yak husbandry. Herein, we report the first genome-wide association study of the *POLLED* phenotype in which we identified a 200-kb genomic region responsible for this economically important trait in yak. However, we need to point out that the candidate region was inevitably large because of the small sample sizes, and our current data were insufficient to detect causal variants. Further research based on larger sample sizes will be necessary to obtain more reliable estimates and refine the genomic loci that contribute to this trait. We found that this region was under artificial selection and the characterizations of the *POLLED* locus were concordant with the breeding process. Combined with previous results in cattle, we further suggest that expressional variations other than structural variations in protein are the major causes of the polled phenotype. The results of our study represent a critical advance towards the delimitation of a genomic region for further functional study and provide new insights into the genetic basis of the polled trait in yak and other bovine species.

Materials and Methods

Sample collection and sequencing

We randomly selected 10 horned and 10 polled individuals (S1 Table) of Datong yaks from a large herd ($n > 2000$, from the Datong Yak Breeding Farm of Qinghai Province (37°15'35.6"N, 101°22'24.0"E). For each yak, genomic DNA was extracted from blood samples using a standard Phenol/Chloroform method. The quality and integrity of the extracted DNA was controlled by the A260/A280 ratio and agarose gel electrophoresis. Paired-end sequencing libraries with an insert size of 500 bp were constructed according to the Illumina manufacturer's

instructions, for sequencing on the Hiseq 2000 platform, and paired-end reads were generated. Sequencing and base calling were performed according to the standard Illumina protocols. All individuals were sequenced to an average raw read sequencing depth of 11.2× assuming a genome size of 2.66 Gb. All experimental protocols were approved by ethical committees of the Datong Yak Breeding Farm of Qinghai Province.

Sequence quality checking and mapping

We performed a per-base sequence quality checks [41] and low quality reads of the following types were filtered out: (i) Reads with $\geq 10\%$ unidentified nucleotides (N); (ii) Reads for which more than 65% of the read length had a phred quality score ≤ 7 ; (iii) Reads with more than 10 bp aligned to the adapter, allowing 2 bp mismatches; and (iv) duplicate reads. Reads were also trimmed if they had three consecutive base pairs with a phred quality score of 13 or below, and discarded if they were shorter than 45 bp.

The pair-end sequence reads were mapped to the *B. grunniens* reference genome using BWA-MEM [25] (0.7.10-r789) with default parameters. The *picard* software (<http://broadinstitute.github.io/picard/>, version 1.92) was subsequently used to assign read group information containing library, lane, and sample identity. Duplicated reads were filtered and index files were built for reference as well as bam files using SAMtools (0.1.19). The Genome Analysis Toolkit (GATK, version 2.6-4-g3e5ff60) [42] was used to perform local realignment of reads to enhance the alignments in the vicinity of InDel polymorphisms. Realignment was performed with GATK in two steps. The first step used the RealignerTargetCreator to identify regions where realignment was needed, and the second step used IndelRealigner to realign the regions found in the first step, generating a realigned mapping file for each individual. The overall mapping rate of reads to the reference genome was 91.20%, with average read depths of 10.2× (10.06× to 10.37×). On average, across all samples, the reads covered 99.26% of the genome (S1 Table).

SNP calling

After alignment, we performed SNP calling using a Bayesian approach as implemented in the package SAMtools. Realigned regions were piped to SAMtools and reformatted into pileup files for SNP identification. Sequence variants from pileups were then condensed into a variant call format (VCF) file using BCFtools [27] (0.1.19). The genotype with maximum posterior probability was picked as the genotype for that locus.

The threshold of SNP calling was set to 20 for both base quality and mapping quality. SNPs were discarded based on the following conditions: (i) quality less than 20; (ii) those with too low (total depth $< 2 \times 20$) or too high (total depth $> 40 \times 20$) a depth (possibly bad assembly or repetitive regions); (iii) 5 bp around InDels; (iv) those occurring in a cluster (more than three SNPs with 10 bp); (v) failure in the exact test for Hardy-Weinberg equilibrium at $P < 0.001$; or (vi) those with $> 50\%$ missing genotype data with the population.

Phasing and linkage disequilibrium

The program Beagle [43] (version: r1196) was used to infer the haplotype phase and impute missing alleles with default parameters. Linkage disequilibrium (pairwise r^2 statistic) was calculated using Haploview [44] (v4.2) software with the parameters '-dprime -maxDistance 1000 -minMAF 0.05 -hwcutoff 0.001 -missingCutoff 0.5 -minGeno 0.6'.

Phylogenetic relationship and population structure analysis

To assess recent relationships between samples, we calculated pairwise estimates of Identity-By-State (IBS) scores [28]. We found no possible duplicate (IBS>0.9) that showed high pairwise genetic similarity with another sampled individual, indicating that these 20 individuals, as unrelated samples, can be used in the downstream analyses.

Neighbor-joining trees were constructed with PHYLIP (v3.696, <http://evolution.genetics.washington.edu/phylip.html>) using the matrix of pairwise genetic distances ('—cluster—distance-matrix' of PLINK v1.07). The SmartPCA program from the EIGENSOFT [45] (v5.0.1) package was used to perform principle component analysis on the individuals that we sequenced with default parameters. A Tracy–Widom test was used to determine the significance level of the eigenvectors and no significant eigenvectors were found (S2 Table). In addition, ADMIXTURE [46] (v1.23, with default parameters) was used to infer the population substructure among the samples with number for population grouping parameter K set from 1 to 3.

Association analysis using PLINK

To map the poll trait loci, we performed a case-control analysis between horned and polled yaks using the Dominant model with PLINK [28] (v1.07, with parameters '—model—model-dom—fisher'). This Dominant model assumes that an effect on phenotype is only seen if you have at least one copy of the minor allele. It categorizes individuals into two groups based on whether they have at least one minor allele A (either Aa or AA) or no copies of the minor allele (aa). Fisher's exact test was used to analyze genotypic differences between the 10 horned and 10 polled samples.

Population genetic statistics

The nucleotide diversity (π) and population-differentiation statistic (F_{ST}) were calculated using VCFtools [47] (v0.1.12a) with a sliding window approach (50 kb window sliding in 10 kb steps).

d_{xy} was calculated as follows:

$$d_{xy} = \sum_{ij} x_i y_i d_{ij}$$

where, in two populations, x and y, d_{ij} measures the number of nucleotide differences between the i^{th} haplotype from x and the j^{th} haplotype from y.

XP-CLR values were calculated with default parameters using XP-CLR (v1.0).

Re-annotating the associated region using RNA-seq data

To find out whether there were new genes or open reading frames within our GWAS loci, we mapped previous RNA-seq data of domestic yak [32] to scaffold526_1 of the yak genome (as described in the 'Sequence quality checking and mapping' section). Read depths of RNA from five tissues (brain, kidney, lung, liver and heart, each with two duplicates) were calculated using SAMtools (parameters of 'samtools depth') and visualized as shown in Fig 2D.

Checking P_F , P_C alleles and other structural variants

To check the existence of a P_F allele in yak, we mapped our sequencing reads to the BTA1 sequence (UMD3.1 genome build, downloaded from Ensembl release77, <http://www.ensembl.org/>). Sequencing depths for each sample around the genome region near the P_F allele were

calculated using SAMtools. The relative depths of each sample were calculated and no P_F allele (i.e. 80 kb duplication) was found in polled yaks (S4 Fig).

Structural variants (insertions, deletions, tandem duplications and inversions) were discovered using Pindel [48] (v0.2.5a3), which uses a pattern growth approach to identify the break-points of these variants from paired-end short reads (with default parameters and '-c 1 -T 30 -l -k'). No P_C allele was found in polled yaks and no other structural variants were found associated with polled phenotypes.

Supporting Information

S1 Fig. Schematic of breeding practice for Datong yaks. Sexuality is indicated by a circle (cow) or a square (bull), genotypes are indicated by different colors (PP, orange; Pp, green; pp, blue).

(TIF)

S2 Fig. Population structure plots with $K = 1-3$. The y axis quantifies the proportion of the individual's genome from inferred ancestral populations, and the x axis shows the different populations. The CV error of each run is given in parentheses.

(TIF)

S3 Fig. QQ-plot of GWAS P values.

(TIF)

S4 Fig. Relative depth of horned (blue) and polled (red) yaks around the P_F allele on BTA1. The green frame indicates the region of the P_F allele.

(TIF)

S5 Fig. Genome-wide distribution of π_{horned} (a) and π_{polled} (b).

(TIF)

S6 Fig. Genome-wide distribution of F_{ST} .

(TIF)

S1 Table. Overview of sample information and sequencing statistics.

(XLS)

S2 Table. Tracy-Widom (TW) statistics and P -values for the first five eigenvalues in the PCA. No significant P values.

(XLS)

S3 Table. Results of structural variants discovery.

(XLS)

Acknowledgments

We thank Dr. Tao Ma, Dr. Yongzhi Yang and Dr. Bingbing Liu for their helpful comments and suggestions about this project. Special thanks to the farmers and researchers who bred the yaks from the Datong Yak Breeding Farm of Qinghai Province.

Author Contributions

Conceived and designed the experiments: QQ PY. Performed the experiments: CL LW XW KW XD MW MC XX. Analyzed the data: CL LW XW KW XD MW MC XX. Contributed

reagents/materials/analysis tools: CL LW XW KW XD MW MC XX. Wrote the paper: CL LW QQ PY.

References

1. Wiener G, Jianlin H, Ruijun L. The Yak. 2nd ed: FAO Regional Office for Asia and the Pacific Food and Agriculture Organization of the United Nations, Bangkok, Thailand; 2003.
2. Schaller GB, Liu W. Distribution, status, and conservation of wild yak *Bos grunniens*. *Biological Conservation*. 1996; 76(1):1–8.
3. Guo X, Long R, Kreuzer M, Ding L, Shang Z, Zhang Y, et al. Importance of functional ingredients in yak milk-derived food on health of Tibetan nomads living under high-altitude stress: a review. *Critical reviews in food science and nutrition*. 2014; 54(3):292–302. doi: [10.1080/10408398.2011.584134](https://doi.org/10.1080/10408398.2011.584134) PMID: [24188303](https://pubmed.ncbi.nlm.nih.gov/24188303/)
4. Han Jianlin R, C., Hanotte, O., McVeigh, C. and Rege, J.E.O., editor Yak production in central Asian highlands. Proceedings of the third international congress on yak held in Lhasa, PR China, 4–9 September 2000; Lhasa, P.R. China: ILRI (International Livestock Research Institute), Nairobi, Kenya.
5. Götz KU, Luntz B, Robeis J, Edel C, Emmerling R, Buitkamp J, et al. Polled Fleckvieh (Simmental) cattle—Current state of the breeding program. *Livestock Science*. 2015; 179:80–5.
6. Windig JJ, Hoving-Bolink RA, Veerkamp RF. Breeding for polledness in Holstein cattle. *Livestock Science*. 2015; 179:96–101.
7. Prayaga KC. Genetic options to replace dehorning in beef cattle—a review. *Australian Journal of Agricultural Research*. 2007; 58(1):1–8.
8. Brenneman RA, Davis SK, Sanders JO, Burns BM, Wheeler TC, Turner JW, et al. The polled locus maps to BTA1 in a *Bos indicus* x *Bos taurus* cross. *The Journal of Heredity*. 1996; 87(2):156–61. PMID: [8830095](https://pubmed.ncbi.nlm.nih.gov/8830095/)
9. Graf B, Senn M. Behavioural and physiological responses of calves to dehorning by heat cauterization with or without local anaesthesia. *Applied Animal Behaviour Science*. 1999; 62(2–3):153–71.
10. Liu WB, Liu J, Liang CN, Guo X, Bao PJ, Chu M, et al. Associations of single nucleotide polymorphisms in candidate genes with the polled trait in Datong domestic yaks. *Animal genetics*. 2014; 45(1):138–41. doi: [10.1111/age.12081](https://doi.org/10.1111/age.12081) PMID: [24033474](https://pubmed.ncbi.nlm.nih.gov/24033474/)
11. Georges M, Drinkwater R, King T, Mishra A, Moore SS, Nielsen D, et al. Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nature genetics*. 1993; 4(2):206–10. PMID: [8348158](https://pubmed.ncbi.nlm.nih.gov/8348158/)
12. Harlizius B, Tammen I, Eichler K, Eggen A, Hetzel DJ. New markers on bovine chromosome 1 are closely linked to the polled gene in Simmental and Pinzgauer cattle. *Mammalian genome: official journal of the International Mammalian Genome Society*. 1997; 8(4):255–7.
13. Mariasegaram M, Harrison BE, Bolton JA, Tier B, Henshall JM, Barendse W, et al. Fine-mapping the POLL locus in Brahman cattle yields the diagnostic marker CSAFG29. *Animal genetics*. 2012; 43(6):683–8. doi: [10.1111/j.1365-2052.2012.02336.x](https://doi.org/10.1111/j.1365-2052.2012.02336.x) PMID: [22497221](https://pubmed.ncbi.nlm.nih.gov/22497221/)
14. Schmutz SM, Marquess FL, Berryere TG, Moker JS. DNA marker-assisted selection of the polled condition in Charolais cattle. *Mammalian genome: official journal of the International Mammalian Genome Society*. 1995; 6(10):710–3.
15. Wunderlich KR, Abbey CA, Clayton DR, Song Y, Schein JE, Georges M, et al. A 2.5-Mb contig constructed from Angus, Longhorn and horned Hereford DNA spanning the polled interval on bovine chromosome 1. *Animal genetics*. 2006; 37(6):592–4. PMID: [17121607](https://pubmed.ncbi.nlm.nih.gov/17121607/)
16. Allais-Bonnet A, Grohs C, Medugorac I, Krebs S, Djari A, Graf A, et al. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PloS one*. 2013; 8(5):e63512. doi: [10.1371/journal.pone.0063512](https://doi.org/10.1371/journal.pone.0063512) PMID: [23717440](https://pubmed.ncbi.nlm.nih.gov/23717440/)
17. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Gopel KH, et al. Bovine polledness—an autosomal dominant trait with allelic heterogeneity. *PloS one*. 2012; 7(6):e39477. doi: [10.1371/journal.pone.0039477](https://doi.org/10.1371/journal.pone.0039477) PMID: [22737241](https://pubmed.ncbi.nlm.nih.gov/22737241/)
18. Seichter D, Russ I, Rothammer S, Eder J, Forster M, Medugorac I. SNP-based association mapping of the polled gene in divergent cattle breeds. *Animal genetics*. 2012; 43(5):595–8. doi: [10.1111/j.1365-2052.2011.02302.x](https://doi.org/10.1111/j.1365-2052.2011.02302.x) PMID: [22497248](https://pubmed.ncbi.nlm.nih.gov/22497248/)
19. Wiedemar N, Tetens J, Jagannathan V, Menoud A, Neuenschwander S, Bruggmann R, et al. Independent polled mutations leading to complex gene expression differences in cattle. *PloS one*. 2014; 9(3):e93435. doi: [10.1371/journal.pone.0093435](https://doi.org/10.1371/journal.pone.0093435) PMID: [24671182](https://pubmed.ncbi.nlm.nih.gov/24671182/)
20. Glatzer S, Merten NJ, Dierks C, Wohlke A, Philipp U, Distl O. A single nucleotide polymorphism within the gene perfectly coincides with polledness in Holstein cattle. *PloS one*. 2013; 8(6):e67992. doi: [10.1371/journal.pone.0067992](https://doi.org/10.1371/journal.pone.0067992) PMID: [23805331](https://pubmed.ncbi.nlm.nih.gov/23805331/)

21. Rothhammer S, Capitan A, Mullaart E, Seichter D, Russ I, Medugorac I. The 80-kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. *Genetics, selection, evolution: GSE*. 2014; 46:44. doi: [10.1186/1297-9686-46-44](https://doi.org/10.1186/1297-9686-46-44) PMID: [24993890](https://pubmed.ncbi.nlm.nih.gov/24993890/)
22. Carlson DF, Lancto CA, Zang B, Kim E-S, Walton M, Oldeschulte D, et al. Production of hornless dairy cattle from genome-edited cell lines. *Nature Biotechnology*. 2016; 34(5):479–81. doi: [10.1038/nbt.3560](https://doi.org/10.1038/nbt.3560) PMID: [27153274](https://pubmed.ncbi.nlm.nih.gov/27153274/)
23. Capitan A, Allais-Bonnet A, Pinton A, Marquant-Le Guienne B, Le Bourhis D, Grohs C, et al. A 3.7 Mb deletion encompassing ZEB2 causes a novel polled and multisystemic syndrome in the progeny of a somatic mosaic bull. *PloS one*. 2012; 7(11):e49084 doi: [10.1371/journal.pone.0049084](https://doi.org/10.1371/journal.pone.0049084) PMID: [23152852](https://pubmed.ncbi.nlm.nih.gov/23152852/)
24. Capitan A, Grohs C, Weiss B, Rossignol MN, Reverse P, Eggen A. A newly described bovine type 2 scurs syndrome segregates with a frame-shift mutation in TWIST1. *PloS one*. 2011; 6(7):e22242. doi: [10.1371/journal.pone.0022242](https://doi.org/10.1371/journal.pone.0022242) PMID: [21814570](https://pubmed.ncbi.nlm.nih.gov/21814570/)
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
26. Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, et al. The yak genome and adaptation to life at high altitude. *Nature genetics*. 2012; 44(8):946–9. doi: [10.1038/ng.2343](https://doi.org/10.1038/ng.2343) PMID: [22751099](https://pubmed.ncbi.nlm.nih.gov/22751099/)
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007; 81(3):559–75. PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
29. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics*. 2012; 10(2):117–22.
30. Drouet V, Lesage S. Synaptojanin 1 mutation in Parkinson's disease brings further insight into the neuropathological mechanisms. *BioMed research international*. 2014; 2014:289728. doi: [10.1155/2014/289728](https://doi.org/10.1155/2014/289728) PMID: [25302295](https://pubmed.ncbi.nlm.nih.gov/25302295/)
31. Trapani JG, Obholzer N, Mo W, Brockerhoff SE, Nicolson T. Synaptojanin1 is required for temporal fidelity of synaptic transmission in hair cells. *PLoS genetics*. 2009; 5(5):e1000480. doi: [10.1371/journal.pgen.1000480](https://doi.org/10.1371/journal.pgen.1000480) PMID: [19424431](https://pubmed.ncbi.nlm.nih.gov/19424431/)
32. Wiener DJ, Wiedemar N, Welle MM, Drogemuller C. Novel features of the prenatal horn bud development in cattle (*Bos taurus*). *PloS one*. 2015; 10(5):e0127691. doi: [10.1371/journal.pone.0127691](https://doi.org/10.1371/journal.pone.0127691) PMID: [25993643](https://pubmed.ncbi.nlm.nih.gov/25993643/)
33. Ohadi M, Valipour E, Ghadimi-Haddadan S, Namdar-Aligoodarzi P, Bagheri A, Kowsari A, et al. Core promoter short tandem repeats as evolutionary switch codes for primate speciation. *American journal of primatology*. 2015; 77(1):34–43. doi: [10.1002/ajp.22308](https://doi.org/10.1002/ajp.22308) PMID: [25099915](https://pubmed.ncbi.nlm.nih.gov/25099915/)
34. Mariasegaram M, Reverter A, Barris W, Lehnert SA, Dalrymple B, Prayaga K. Transcription profiling provides insights into gene pathways involved in horn and scurs development in cattle. *BMC genomics*. 2010; 11:370. doi: [10.1186/1471-2164-11-370](https://doi.org/10.1186/1471-2164-11-370) PMID: [20537189](https://pubmed.ncbi.nlm.nih.gov/20537189/)
35. Wang K, Yang Y, Wang L, Ma T, Shang H, Ding L, et al. Different gene expressions between cattle and yak provide insights into high-altitude adaptation. *Animal genetics*. 2016; 47(1):28–35. doi: [10.1111/age.12377](https://doi.org/10.1111/age.12377) PMID: [26538003](https://pubmed.ncbi.nlm.nih.gov/26538003/)
36. Sims D, Sudbery I, Illott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nature reviews Genetics*. 2014; 15(2):121–32. doi: [10.1038/nrg3642](https://doi.org/10.1038/nrg3642) PMID: [24434847](https://pubmed.ncbi.nlm.nih.gov/24434847/)
37. Xu X, Dong GX, Hu XS, Miao L, Zhang XL, Zhang DL, et al. The genetic basis of white tigers. *Current biology*. 2013; 23(11):1031–5. doi: [10.1016/j.cub.2013.04.054](https://doi.org/10.1016/j.cub.2013.04.054) PMID: [23707431](https://pubmed.ncbi.nlm.nih.gov/23707431/)
38. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varrilly P, Shamovsky O, et al. Positive natural selection in the human lineage. *Science*. 2006; 312(5780):1614–20. PMID: [16778047](https://pubmed.ncbi.nlm.nih.gov/16778047/)
39. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome research*. 2010; 20(3):393–402. doi: [10.1101/gr.100545.109](https://doi.org/10.1101/gr.100545.109) PMID: [20086244](https://pubmed.ncbi.nlm.nih.gov/20086244/)
40. Lan X, Witt H, Katsumura K, Ye Z, Wang Q, Bresnick EH, et al. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic acids research*. 2012; 40(16):7690–704. PMID: [22675074](https://pubmed.ncbi.nlm.nih.gov/22675074/)
41. Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nature Communications*. 2015; 6:10283. doi: [10.1038/ncomms10283](https://doi.org/10.1038/ncomms10283) PMID: [26691338](https://pubmed.ncbi.nlm.nih.gov/26691338/)

42. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011; 43(5):491–8. doi: [10.1038/ng.806](https://doi.org/10.1038/ng.806) PMID: [21478889](https://pubmed.ncbi.nlm.nih.gov/21478889/)
43. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics*. 2007; 81(5):1084–97. PMID: [17924348](https://pubmed.ncbi.nlm.nih.gov/17924348/)
44. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harbor protocols*. 2009; 2009(10):pdb ip71.
45. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS genetics*. 2006; 2(12): e190. PMID: [17194218](https://pubmed.ncbi.nlm.nih.gov/17194218/)
46. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*. 2009; 19(9):1655–64. doi: [10.1101/gr.094052.109](https://doi.org/10.1101/gr.094052.109) PMID: [19648217](https://pubmed.ncbi.nlm.nih.gov/19648217/)
47. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27(15):2156–8. doi: [10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330) PMID: [21653522](https://pubmed.ncbi.nlm.nih.gov/21653522/)
48. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25(21):2865–71. doi: [10.1093/bioinformatics/btp394](https://doi.org/10.1093/bioinformatics/btp394) PMID: [19561018](https://pubmed.ncbi.nlm.nih.gov/19561018/)