





## Article

# Whole-Genome Resequencing Highlights the Unique Characteristics of Kecai Yaks

Yandong Kang<sup>1,2</sup>, Shaoke Guo<sup>1,2</sup>, Xingdong Wang<sup>1,2</sup> , Mengli Cao<sup>1,2</sup>, Jie Pei<sup>1,2</sup> , Ruiwu Li<sup>3</sup>, Pengjia Bao<sup>1,2</sup> , Jiefeng Wang<sup>3</sup>, Jiebu Lamao<sup>3</sup>, Dangzhi Gongbao<sup>3</sup>, Ji Lamao<sup>3</sup>, Chunnian Liang<sup>1,2</sup>, Ping Yan<sup>1,2</sup> and Xian Guo<sup>1,2,\*</sup> 

- <sup>1</sup> Key Laboratory of Yak Breeding Engineering of Gansu Province, Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences, Lanzhou 730050, China  
<sup>2</sup> Key Laboratory of Animal Genetics and Breeding on Tibetan Plateau, Ministry of Agriculture and Rural Affairs, Lanzhou 730050, China  
<sup>3</sup> Station of Animal Husbandry in Xiahe County, Xiahe 747199, China  
\* Correspondence: guoxian@caas.cn; Tel.: +86-1899-3037-854

**Simple Summary:** Kecai yaks are an important genetic resource in Gansu Province, China. These yaks, produced through the crossbreeding of Gannan yaks and wild yaks, exhibit excellent production performance and high fecundity. Despite these beneficial characteristics, the genetic characteristics, population structure, and other properties of Kecai yaks have not been effectively characterized to date. Accordingly, this study was developed to explore the genetic diversity and population structure of Kecai yaks through whole-genome resequencing. These analyses ultimately revealed the status of Kecai yaks as an independent group within the overall Chinese yak population.

**Abstract:** Kecai yaks are regarded as an important genetic resource in China owing to their high fecundity and flavorful meat. However, the genetic characteristics of Kecai yaks have not been effectively characterized to date, and the relationship between Kecai yaks and other yak breeds remains to be fully characterized. In this paper, the resequencing of the Kecai yak genome is performed leading to the identification of 11,491,383 high-quality single nucleotide polymorphisms (SNPs). Through principal component, phylogenetic, and population genetic structure analyses based on these SNPs, Kecai yaks were confirmed to represent an independent population of yaks within China. In this study, marker and functional enrichment analysis of genes related to positive selection in Kecai yak was carried out, and the results show that such selection in Kecai yaks is associated with the adaptation to alpine environments and the deposition of muscle fat. Overall, these results offer a theoretical foundation for the future utilization of Kecai yak genetic resources.

**Keywords:** Kecai yak; whole-genome resequencing; genetic diversity; molecular phylogeny



**Citation:** Kang, Y.; Guo, S.; Wang, X.; Cao, M.; Pei, J.; Li, R.; Bao, P.; Wang, J.; Lamao, J.; Gongbao, D.; et al. Whole-Genome Resequencing Highlights the Unique Characteristics of Kecai Yaks. *Animals* **2022**, *12*, 2682. <https://doi.org/10.3390/ani12192682>

Academic Editor: Marco Milanese

Received: 3 September 2022

Accepted: 3 October 2022

Published: 6 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

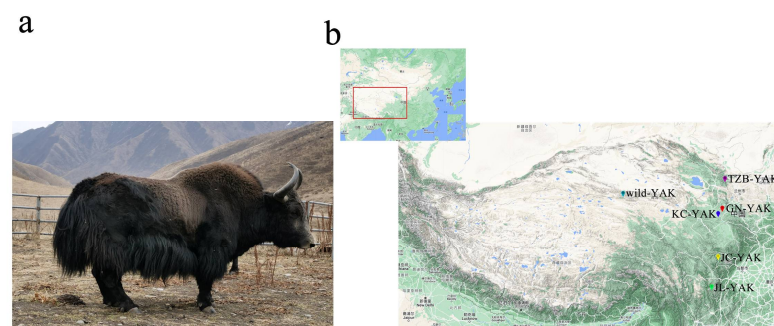
## 1. Introduction

Yaks (*Bos grunniens*) are a long-haired livestock species native to the Qinghai–Tibet Plateau and surrounding regions, living at an average altitude of over 3000 m. These yaks are uniquely adapted to high-altitude conditions such that they exhibit robust cardiopulmonary functionality, strong muscles, and good feeding performance [1]. Yaks are a critical source of animal labor, food, and other materials for local herdsmen such that they are an integral facet of life on the Qinghai–Tibet Plateau. Kecai yaks were domesticated from wild yak (*Bos mutus*) populations, and have accordingly experienced a series of population bottlenecks, artificial selection events, and continuous bidirectional gene flow [2,3].

Several studies have explored yak population structures and performed detailed analyses of the yak domestication process based on mtDNA sequences and other nucleic acid markers. By analyzing mitochondrial D-loop sequences from 250 domesticated yaks and 13 wild yaks, Guo et al. [3] were able to identify two distinct maternal branches

that diverged approximately 130,000–150,000 years ago. Through further comparisons of mitochondrial sequences from 405 domesticated yaks and 45 wild yaks, Wang et al. [4] defined a third maternal branch that only included wild yaks and further observed the effects of relaxed selection among domesticated yaks as a consequence of domestication. Wang et al. [5] further compared mitochondrial sequence data from five populations of yaks in Xinjiang and Qinghai, ultimately separating these domestic and wild yaks into three clades and proposing an update to the yak reference genome to support further research efforts. In 2012, Qiu et al. [6] sequenced and assembled the first domestic yak genome through the use of second-generation DNA sequencing technologies, gaining preliminary insight into the potential genetic basis for the ability of yaks to adapt to high-altitude living. More recently, Liu et al. [7] employed a second-generation sequencing approach to compile the first wild yak genome, thereby offering a valuable resource for further molecular biology-focused studies. Ji et al. [8] additionally leveraged PacBio long-read sequencing technologies to compile the first chromosome-level domesticated yak reference genome, enabling the in-depth analysis of genetic diversity among different yak breeds. Through the use of resequence samples from 65 domesticated yaks and 14 wild yaks, Zhang et al. [9] detected 2643 copy number variation regions (CNVRs), ultimately leading to the identification of relationships between domestication traits and variations in certain protein-coding genes, including the brain-development-related *GRIN2D* gene. Zhang et al. [10] were further able to identify genes associated with immunity, behavior, reproduction, and the nervous system through comparisons of structural variations (SVs) between domesticated and wild yaks.

Gannan yaks are 1 of 18 unique genetic resources in China, exhibiting advantageous characteristics including robust adaptability, nutrient-rich meat, and a palatable taste [11]. Kecai yaks were produced through several generations of crossing wild yaks with Gannan yaks. These Kecai yaks are primarily bred in Kecai Town, Xiahe County, Gannan Autonomous Prefecture, Gansu Province, and are characterized by a black body with horns and fluffy hair located on the chest and under the tail (Figure 1a). These Kecai yaks are large, exhibit high fecundity, and inherited the excellent characteristics of delicious meat of the Gannan yak. Despite their advantageous characteristics and status as an important domestic yak genetic resource, the genetic characteristics, unique traits, and population structure of Kecai yaks have not been studied. Accordingly, this study was developed with the goal of exploring Kecai yak population structure and genetic diversity through genomic resequencing and the comparison of these genomic sequences with those from other yak breeds. The overall goal of this approach is to provide a theoretical foundation for the future use of Kecai yak genetic resources based on a genome-level understanding of these animals.



**Figure 1.** The geographical distribution of yaks selected in this study. (a) Kecai yak. (b) Sampling distribution map. GN: Gannan yak; KC: Kecai yak; TZB: Tianzhu white yak; JC: Jinchuan yak; JL: Jiulong yak.

## 2. Materials and Methods

### 2.1. Ethics Statement

The China Council on Animal Care and the Ministry of Agriculture of the People's Republic of China guidelines were followed when performing all animal studies. The Animal Care and Use Committee of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences approved all yak handling protocols (Permit No: SYXK-2014-0002). Blood samples for this study were obtained from the jugular vein of live yaks following surface disinfection with topical alcohol. Iodophor was applied to the site of venipuncture after blood collection was complete to prevent infection.

### 2.2. Animals Sample Collection

For this study, 45 adult yaks of similar age, health status, and size were analyzed. A total of 9 Kecai yaks (Kecai Town, Gannan Tibetan Autonomous Prefecture, Gansu Province, China), 7 Gannan yaks (Xiahe County, Gannan Tibetan Autonomous Prefecture, Gansu Province, China), 10 Jiulong yaks (Jiulong County, Ganzi Tibetan Autonomous Prefecture, Sichuan Province, China), 10 Tianzhu white yaks (Tianzhu Tibetan Autonomous Prefecture, Gansu Province, China), and 9 Jinchuan yaks (Aba Tibetan and Qiang Autonomous Prefecture, Sichuan Province, China) were included. DNA was extracted from individual blood samples with an EasyPure Blood Genomic DNA Kit (Quanshijin Biotechnology Co., Ltd., Beijing, China). DNA purity and concentrations were measured with a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, MA, USA) based on OD<sub>260</sub>/OD<sub>280</sub> values. The integrity of samples with an OD<sub>260</sub>/OD<sub>280</sub> of 1.6–1.8 and a concentration of 50–1000 ng/μL were separated via 1% agarose gel electrophoresis (AGE) to ensure sample quality. Those samples that exhibited a single distinct AGE band were diluted to 50 ng/μL and stored at −80 °C. For sequencing analyses, these samples were fragmented into 350–500 bp segments with Covaris after which a TruSeq DNA LT Sample Prep kit was used for library construction. Briefly, DNA fragments were subjected to end repair, polyadenylation, the addition of sequencing connectors, purification, PCR amplification, and other necessary steps, with the final libraries then being used for 150 bp paired-end sequencing with the HiSeq X Ten platform (Illumina Inc., Hayward, CA, USA). Wild yak genomic data were obtained from breeding genome data generated by the European Nucleotide Archive (EMBL-EBI) (Accession number: PRJNA285834), while Kecai yak sequencing data were uploaded to the Sequence Read Archive (SRA) (Accession number: PRJNA842787). The data were compared to genomic data from other local yak populations.

### 2.3. Sequence Quality Control and Filtering

Raw reads from Kecai yaks were filtered to remove low-quality reads, base duplications, and other forms of artificial bias. Library sequencing and data analyses were partially conducted by Shanghai OE Biomedical Technology Co. Raw reads were initially obtained in the fastq format following high-throughput sequencing, after which pre-processing was performed using fastp v 0.20.0 (Shenzhen, China) [12] through four primary steps: (1) the removal of spliced sequences; (2) the removal of reads with  $\geq 5$  N (non-AGCT) bases; (3) the removal of reads with an average base mass value  $< 20$  over a sliding window of 4 bases; and (4) the removal of remaining reads  $< 75$  bp in length or with an average base mass value  $< 15$ .

### 2.4. Genome Alignment

Reference genome alignment of the filtered reads was performed using BWA v 0.7.12 [13], using the BWA mem alignment algorithm with the default parameters. The results were formatted and sorted using SAMtools v 1.9 [14], after which duplicate reads were identified using the MarkDuplicates function Picard v 2.18.17 (<https://broadinstitute.github.io/picard/>) (accessed on 15 January 2022). To enable statistical analyses, the results were then compared with Qualimap [15].

### 2.5. Variant Detection and Annotation

SNP and InDel detection were performed using the Haplotypecaller module in GATK v.3.8.1 [16] based on comparisons between sample sequences and the reference genome (LU\_Bosgru\_v3.0). For each sample, gvcf files were generated and SNPs were detected using the GenotypeGVCFs module. Data were filtered with the following criteria: (1) The Genome Analysis Toolkit (GATK) filtering parameters for loci exclusion: QualByDepth (QD) < 2.0 || RMSMappingQuality (MQ) < 40.0 || FisherStrand (FS) > 60.0 || StrandOddsRatio (SOR) > 3.0 || MappingQualityRankSumTest (MQRankSum) < -12.5 || ReadPosRankSum < -8.0; (2) Allele type: as SNPs are generally biallelic, loci with more than 2 distinct genotypes were filtered out; (3) Loci with a minor allele frequency (MAF) < 0.05 were removed.

GATK was similarly used to analyze InDels in these data with the following screening criteria: QD < 2.0 || MQ < 40.0 || FS > 200.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.

The SnpEff program was used to annotate InDels and SNPs based on the reference genome [17]. Annotation was only performed for high-quality SNPs. SNPs were classified as being located in exonic, intronic, intergenic, upstream, downstream, or splicing regions of the genome. Those SNPs located within coding exons were additionally classified as either synonymous or non-synonymous. Variants resulting in stop-gain or stop-loss mutations were additionally annotated. InDels were classified based on whether they were conserved in-frame insertions/deletions, disruptive in-frame insertions/deletions, bidirectional gene fusions, frameshift mutations, or conservative-disruptive in-frame insertions/deletions. SNPs were specifically annotated with classifications including missense, initiator codon, and synonymous variants.

### 2.6. Population Genetic Polymorphism Analyses

To facilitate a genome-wide analysis of phylogenetic relationships, PLINK [18] was used to identify SNPs in linkage disequilibrium (LD) using the following parameters: '-indep-pairwise 50 10 0.2'. Those SNPs not exhibiting close linkage were selected, after which neighbor-joining (NJ) phylogenetic trees were generated with the NEIGHBOR program in PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>) (accessed on 15 January 2022). Feature vectors were computed with EIGENSOFT [19], which was used to extract the first four principal components for a principal component analysis (PCA). Population structure analyses were performed using ADMIXTURE [20] across a range of k values from 2 to 10, with 10 different seeds being used for 10 repeated analyses. The results were clustered using Pong [21], and optimal k values were determined based upon cross-validation error. The PopLDdecay program [22] was employed to calculate LD with the following settings: '-MaxDist 500 -MAF 0.05'.

### 2.7. Selective Sweep and Functional Enrichment Analyses

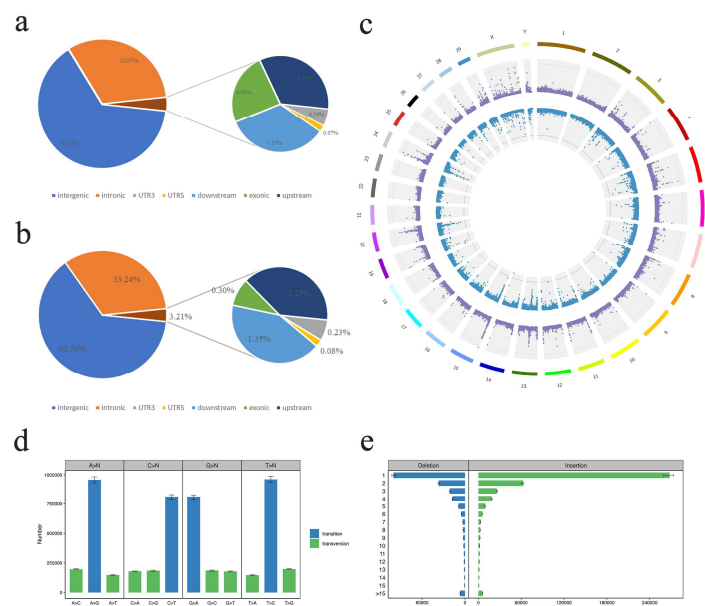
Selected regions of the genome were identified by using the  $\theta\pi$  ratio and  $F_{st}$  values for screening purposes [23], as this strategy can effectively identify sites under selective pressure, particularly when mining for functional regions related to environmental stress that often yield strong selection signals. The  $\log_2\theta\pi$  ratio and  $F_{st}$  values were calculated using VCFtools [24], with a window size of 100 kb and a step size of 10 kb. Those regions in the top and bottom 5% of  $\log_2\theta\pi$  ratio values and the top 5% of  $F_{st}$  values were identified as selected regions of the genome, and the target genes were subject to functional enrichment analyses using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional annotation tools. The numbers of genes associated with these GO terms and KEGG pathways were counted, after which a hypergeometric distribution test was used to compute the significance of gene enrichment therein.

### 3. Results

#### 3.1. Sequencing, Detection, and Annotation of Genome-Wide SNPs and InDels

In these resequencing analyses, an average of 209,142,764 raw reads were obtained per Kecaik yak sample, with 208,238,267 valid reads ( $\geq 99\%$ ) being retained following the filtering of unqualified data. Overall, 95.23% of reads with a quality value  $\geq 20$  bases (sequencing error rate  $< 0.01$ ), with a GC content of 43.14%, with good sequencing quality. After filtering, over 98.75% of these reads were successfully aligned to the reference genome (Supplementary Table S1).

The genomes of 13 wild yaks and 45 domestic yaks from the Gansu and Sichuan provinces were analyzed. After filtering, 11,491,383 valid high-quality SNPs were retained for subsequent annotation and analysis. Resequencing was achieved at an average  $6.7\times$  depth with 99.07% average genome coverage. Intergenic regions accounted for the majority (64.77%) of these data (Supplementary Table S2), while intronic regions accounted for 32.18% of the data. Exonic regions accounted for just 0.81% of the overall dataset (Figure 2a), including 46,970 and 45,006 synonymous and nonsynonymous mutations, respectively (Supplementary Table S3). The ratio of transition mutations ( $n = 31,935,918$ ) to transversion mutations ( $n = 12,776,170$ ) was 2.499, with T > C and A > G being the most common substitution mutation types (Figure 2d).

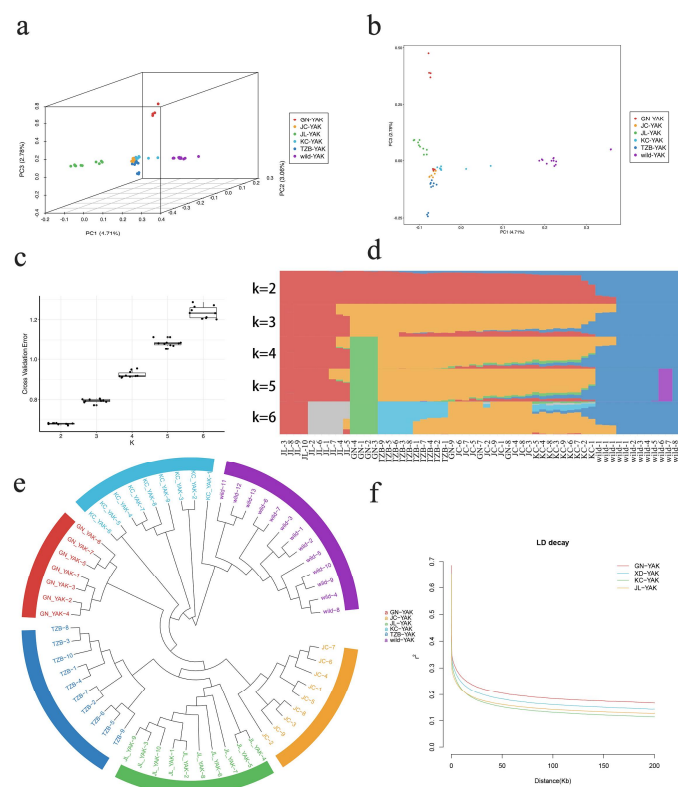


**Figure 2.** Kecaik yak SNP and InDel statistics. (a) SNP annotation results. (b) InDel annotation results. (c) SNP and InDel distributions on individual chromosomes, where the first circle represents the yak chromosome, the purple dots in the second circle represent the SNP density, and the blue dots in the third circle represent the indel density. (d) SNP transition and transversion distributions. (e) InDel insertion and deletion statistics.

Following the filtering of InDel data for all yak populations, 1,404,453 InDel loci were obtained and associated genes were annotated. Of these InDels, 63.35% of the indels are located in the intergenic region, 33.13% in the intron region, and only 0.30% in the exon region (Figure 2b). Among the indels located in the exon region, 1217 induced frameshift deletions, 2108 induced frameshift insertions, and 432 and 342, respectively, induced non-frameshift deletions and non-frameshift insertions (Supplementary Table S4). Deletions and insertions of less than 4 bp accounted for 82.87% and 91.28%, respectively (Figure 2e). SNP and InDel distributions across Kecaik yak chromosomes are represented with Circos plots shown in Figure 2c.

### 3.2. Population-Level Analyses of Genetic Polymorphisms and LD Decay

To explore the genetic population structure of Kecaik yaks and their relationship with other yak populations, genome-wide polymorphism analyses were performed for both domesticated and wild yaks. Phylogenetic trees developed based upon these genome-wide resequencing data indicated that each group of yaks clustered separately, with Kecaik yaks being located immediately between wild and Gannan yaks, with some Kecaik yaks being genetically similar to wild yaks but wholly separate from Gannan yaks, consistent with their identity as a separate group (Figure 3e). In a principal component analysis, samples from Kecaik yaks, Tianzhu white yaks, Jiulong yaks, and wild yaks were grouped together (Figure 3a). Specifically, Kecaik yak samples clustered in close proximity to those from Jinchuan and Gannan yaks, suggesting that they share a similar genetic background (Figure 3a). In further analyses of population structure, at  $k = 2$ , all of yaks were separated into two groups: wild yaks and other yaks. When  $k = 3$ , Jiulong yaks were classified separately from other yaks. When  $k \geq 4$ , Gannan yaks were similarly separated from other yaks, and Gannan and Kecaik yaks were separated into distinct groups (Figure 3d). LD decay patterns in these different yak populations were analyzed based on the correlations between allele frequencies and genomic distances ( $r^2$ ) for SNP pairs. In this analysis, the slowest decay was observed for Gannan yaks, followed by Jiulong and Xueduo yaks, whereas the most rapid decay was observed for Kecaik yaks. These results suggest that, relative to these other yak populations, Kecaik yaks have been subject to a lesser degree of domestication and to less intense selection (Figure 3f).



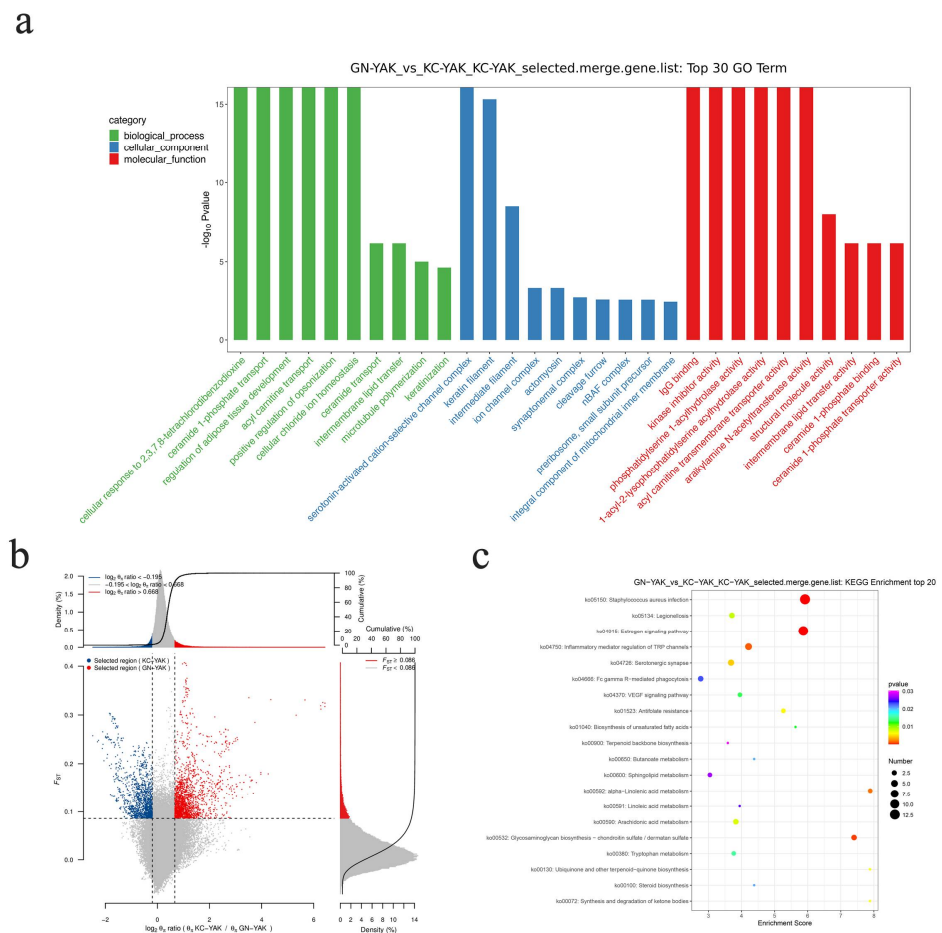
**Figure 3.** Yak population stratification analyses. (a,b) Yak principal component analysis results represented in three dimensions (a) and two dimensions (b). Different colors correspond to different yak varieties. (c) Cross-validation error results at different  $k$  values. (d) Yak population structure analysis. (e) Yak phylogenetic analysis. (f) Yak LD decay analyses.

### 3.3. Analyses of Genetic Diversity

The genetic differentiation index ( $F_{st}$ ) is frequently used in population genetics studies as a means of gauging the degree of genetic differentiation between two populations. When comparing the Xueduo and Kecaik yak populations, the average  $F_{st}$  value was 0.00523,

whereas it rose to 0.01998 when comparing Gannan and Kecai yaks. Xueduo yaks are primarily distributed throughout the Henan County located in Huangnan Prefecture, Qinghai Province, in the eastern region of the Qinghai–Tibet Plateau. Ganna yaks are primarily distributed in the Gannan Tibetan Autonomous Prefecture regions of Gansu Province. Greater differentiation was evident between Gannan and Kecai yaks relative to that observed between Xueduo and Kecai yaks. Higher levels of nucleotide diversity correspond to a higher degree of genetic variation within a given population. In these resequencing analyses, nucleotide diversity was 0.00121 in Gannan yaks, 0.00117 in Jiulong yaks, 0.00132 in Kecai yaks, and 0.00127 in Xueduo yak, suggesting that Kecai yaks exhibit the highest levels of nucleotide diversity among analyzed yak populations.

The  $\theta\pi$  ratio and  $F_{st}$  values were next used to screen the areas of the genome under selection using defined thresholds ( $\log_2\theta\pi$  ratio > 0.668,  $F_{st} \geq 0.086$ ); values above these thresholds indicate positive selection and can be divided into groups with unique characteristics. When comparing Gannan and Kecai yak samples, the selection area under positive selection was found to be greater than that under negative selection (Figure 4b).



**Figure 4.** Analyses of the enrichment and functional annotation of the Kecai yak genome under genetic selection. (a) GO enrichment results for selected genes detected in the Kecai yak genome. (b) A location map comparing selected areas of the genome between Kecai and Gannan yaks. The selected area was determined based on the top 5% area of  $F_{st}$  values and the  $\theta\pi$  ratio, with a window of 200 kb and a step size of 20 kb. Blue data points correspond to the selected area identified in Kecai yak samples. Red data points correspond to the selected area identified in Gannan yak samples. (c) KEGG pathway enrichment results for the genes in the selected regions of the Kecai yak genome.

### 3.4. Enrichment Analysis

Genome-wide variations in the Gannan and Kecai yak populations were next analyzed, with the results being used for GO enrichment analyses of the 1882 identified genes

(Supplementary Table S5). These genes were enriched in biological processes including cellular processes, metabolic processes, and responses to stimuli, cellular components including organelles, macromolecular complexes, synaptic fractions, and molecular functions, including catalytic activity, enzyme regulator activity, and molecular transducer activity (Figure 4a). In KEGG enrichment analyses, these genes were found to be closely associated with several biologically important pathways (Supplementary Table S6), including the *Staphylococcus aureus* infection, terpene skeleton biosynthesis, sphingolipid metabolism, linoleic acid metabolism, and FcγR-mediated phagocytosis pathways (Figure 4c). Additionally, in the pathways involved in muscle fat development, a candidate gene *ACOX1* related to fat deposition was found.

#### 4. Discussion

Yaks are an invaluable genetic resource in the Qinghai–Tibet Plateau and surrounding high-altitude regions. Several reports estimate that yaks were initially domesticated from wild yak populations approximately 7300 years ago [25,26]. Gannan yaks represent a local genetic resource unique to the Qinghai–Tibet Plateau. Following prolonged natural selection and artificial breeding, these yaks exhibit a stable population structure, robust stress resistance, and excellent reproductive performance. Gannan-yak-derived products have a geographical indication and are highly desirable owing to their production of high-quality milk and meat [11]. Kecai yaks are a recently described genetic resource generated through the breeding of Gannan yaks with wild yaks. These Kecai yaks are larger than other yak breeds, and retained the excellent reproductive and production performance of Gannan yaks. To better understand the evolution and genetic characteristics of these Kecai yaks in the present study, genomic resequencing was performed and the resultant data were compared with genomic sequences from other yaks to establish the overall yak population structure. In addition, the genetic diversity of Kecai yaks was analyzed, and the unique characteristics of Kecai yaks and other yak breeds were evaluated.

A phylogenetic tree was constructed by SNP, and it was found that wild yak and domestic yak were mainly divided into two different branches. Because Kecai yak is the offspring of a branch of Gannan yak crossed with wild yak, some Kecai yak are genetically close to wild yak. The genetic similarities between wild and Kecai yak populations are likely attributable to the fact that Kecai yaks were genetically distinguishable from wild and Gannan yak populations. When analyzing yak population structures at  $k = 2$ , these yaks were clustered into wild and domesticated yak samples in line with prior research [25,27]. When  $k \geq 4$ , Gannan yak samples formed an independent subgroup distinct from other domesticated yak samples. Compared with Gannan yak, Jiulong yak and Xueduo yak, the measured  $F_{st}$  values of harvestable yak were all lower than 0.05, indicating that there was gene exchange among yak populations. Notably, Kecai yak samples exhibited a higher degree of nucleic acid diversity relative to other domestic yak subpopulations. These characteristics offer insights into the unique properties of Kecai yaks.

Genome-wide analyses revealed that over 50% of the selected regions of the Kecai and Gannan yak genomes were under positive selection, while further supporting the classification of Kecai yaks as a distinct yak subgroup. GO analyses of genes annotated in these selected regions revealed their enrichment in biological processes and molecular functions related to transfer between glycolipid membranes, including glycolipid transfer protein (GLTP) involved in ceramide transport, membrane interlipid transfer, intermembrane lipid transfer activity, ceramide 1-phosphate binding, and ceramide 1-phosphate transporter activities. GLTP is a soluble 23.8 kDa protein that can accelerate selective intermembrane glycolipid transfer [28–30]. Glycosphingolipids (GSLs) function to stimulate DNA synthesis, cell growth, differentiation, and the sorting and transport of proteins, in addition to influencing eukaryotic processes, such as adhesion, cell–cell recognition, and development [31–33]. GLTP can bind and transfer a wide range of GSLs, functioning as a de facto sensor for levels of GlcCer within cells and regulating normal cell sphingolipid homeostasis. Selected genes identified in this study were significantly enriched in the



regulation of adipose tissue development [34,35]. The metabolism of both fat and muscle tissue is critical for the maintenance of an appropriate energy balance in yaks. Given their prolonged evolution under high-altitude alpine conditions, yaks exhibit a unique body structure and corresponding genetic changes that govern adipose tissue development [36]. The genetic differences identified in this study suggest that Kecaï yaks exhibit greater advantages with respect to their ability to adapt to the environment of alpine regions relative to other domesticated yaks.

KEGG enrichment analyses indicated that genes in selected regions of the genome were most strongly enriched in the *Staphylococcus aureus* infection, estrogen signaling, and glycosaminoglycan biosynthesis–chondroitin sulfate/pidan sulfate pathways. Amino acids are essential taste-active compounds that play a central role in establishing the flavor of meat. The unsaturated fatty acid biosynthesis, linolenic acid metabolism, arachidonic acid metabolism, and linoleic acid metabolism pathways detected in this study are involved in muscle fat development, and ACOX1 is an acyl-CoA oxidase associated with fatty acid oxidative decomposition. ACOX1 is a highly conserved and unique enzyme that is highly expressed in the liver, with lower levels of expression being evident in the kidney, brain, and adipose tissue consistent with its importance in the context of fat metabolism [37,38]. Many studies have demonstrated a close relationship between the ACOX1 genes and loci related to quantitative traits that influence daily gain, birth weight, backfat thickness, and fatty acid composition [39–41]. As such, ACOX1 represents a candidate gene related to fat deposition that plays important roles in fatty acid metabolism and may be associated with the excellent muscle fat deposition and meat quality observed in Kecaï yaks.

## 5. Conclusions

In summary, a whole-genome resequencing strategy was used in this paper to explore the genetic diversity and population structure of Kecaï yaks. The SNP markers used for these analyses revealed that Kecaï yaks are genetically distinguishable from wild and Gannan yak populations. These Kecaï yaks also exhibited evidence of the functional enrichment of positively selected genes relative to Gannan yaks, suggesting that these genes may play a role in the meat quality of these animals and their adaptation to alpine environments. The comprehensive analysis showed that the Kecaï yak has unique genetic characteristics. This study provides a theoretical basis for the protection and utilization of Kecaï yak resources.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ani12192682/s1>, Table S1: The results of quality control Table S2: The results of mapping summary; Table S3: The results of SNP comment type statistics; Table S4: The results of InDel comment type statistics; Table S5: The results of GO enrichment analysis; Table S6: The results of KEGG enrichment analysis.

**Author Contributions:** Data curation, Y.K. and S.G.; Investigation, R.L., J.W., J.L. (Jiebu Lamao), D.G. and J.L. (Ji Lamao); Writing—Original draft, Y.K., X.W. and S.G.; Writing—Review and editing, P.B., C.L., M.C., Y.K. and J.P.; Funding acquisition, P.Y. and X.G. All authors contributed to the interpretation of the results and writing of the article. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was financially supported by the National Key Research and Development Program of China (The funding number: 2021YFD1600205; The funder: P.Y. and X.G.), China Agriculture Research System of MOF and MARA (The funding number: CARS-37; The funder: P.Y. and X.G.) and Innovation Project of Chinese Academy of Agricultural Sciences (The funding number: 25-LZIHPS-01; The funder: P.Y. and X.G.).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Lanzhou Institute of Husbandry and Pharmaceutical Sciences Chinese Academy of Agricultural Sciences (protocol code: SYXK-2014-0002, date: 4 January 2014).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The genome data of wild yak were selected from the breeding genome data published by the European Nucleotide Archive (EMBL-EBI) (Accession number: PRJNA285834), and the sequencing data of Kecal yak were uploaded to the Sequence Read Archive (SRA) (Accession number: PRJNA842787).

**Acknowledgments:** The authors would like to thank their brothers and sisters for their help in the experiments and all the reviewers who participated in the review.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Schaller, G.B.; Liu, W. Distribution, Status, and Conservation of Wild Yak *Bos Grunniens*. *Biol. Conserv.* **1996**, *76*, 1–8. [[CrossRef](#)]
2. Wang, Z.; Shen, X.; Liu, B.; Su, J.; Yonezawa, T.; Yu, Y.; Guo, S.; Ho, S.Y.W.; Vilà, C.; Hasegawa, M.; et al. Phylogeographical Analyses of Domestic and Wild Yaks Based on Mitochondrial DNA: New Data and Reappraisal: Phylogeographical Patterns of Yaks. *J. Biogeogr.* **2010**, *37*, 2332–2344. [[CrossRef](#)]
3. Guo, S.; Savolainen, P.; Su, J.; Zhang, Q.; Qi, D.; Zhou, J.; Zhong, Y.; Zhao, X.; Liu, J. Origin of Mitochondrial DNA Diversity of Domestic Yaks. *BMC Evol. Biol.* **2006**, *6*, 73. [[CrossRef](#)] [[PubMed](#)]
4. Wang, Z.; Yonezawa, T.; Liu, B.; Ma, T.; Shen, X.; Su, J.; Guo, S.; Hasegawa, M.; Liu, J. Domestication Relaxed Selective Constraints on the Yak Mitochondrial Genome. *Mol. Biol. Evol.* **2011**, *28*, 1553–1556. [[CrossRef](#)] [[PubMed](#)]
5. Wang, X.; Pei, J.; Bao, P.; Cao, M.; Guo, S.; Song, R.; Song, W.; Liang, C.; Yan, P.; Guo, X. Mitogenomic Diversity and Phylogeny Analysis of Yak (*Bos Grunniens*). *BMC Genom.* **2021**, *22*, 325. [[CrossRef](#)]
6. Qiu, Q.; Zhang, G.; Ma, T.; Qian, W.; Ye, Z.; Cao, C.; Hu, Q.; Kim, J.; Larkin, D.M.; Auvil, L.; et al. The Yak Genome and Adaptation to Life at High Altitude. *Nat. Genet.* **2012**, *44*, 946–949. [[CrossRef](#)]
7. Liu, Y.; Luo, J.; Dou, J.; Yan, B.; Ren, Q.; Tang, B.; Wang, K.; Qiu, Q. The Sequence and de Novo Assembly of the Wild Yak Genome. *Sci. Data* **2020**, *7*, 66. [[CrossRef](#)]
8. Ji, Q.; Xin, J.; Chai, Z.; Zhang, C.; Dawa, Y.; Luo, S.; Zhang, Q.; Pingcuo, Z.; Peng, M.-S.; Zhu, Y.; et al. A Chromosome-scale Reference Genome and Genome-wide Genetic Variations Elucidate Adaptation in Yak. *Mol. Ecol. Resour.* **2021**, *21*, 201–211. [[CrossRef](#)]
9. Zhang, X.; Wang, K.; Wang, L.; Yang, Y.; Ni, Z.; Xie, X.; Shao, X.; Han, J.; Wan, D.; Qiu, Q. Genome-Wide Patterns of Copy Number Variation in the Chinese Yak Genome. *BMC Genom.* **2016**, *17*, 379. [[CrossRef](#)]
10. Zhang, S.; Liu, W.; Liu, X.; Du, X.; Zhang, K.; Zhang, Y.; Song, Y.; Zi, Y.; Qiu, Q.; Lenstra, J.A.; et al. Structural Variants Selected during Yak Domestication Inferred from Long-Read Whole-Genome Sequencing. *Mol. Biol. Evol.* **2021**, *38*, 3676–3680. [[CrossRef](#)]
11. Wang, X.; Zhou, H.; Hickford, J.G.H.; Li, S.; Wang, J.; Liu, X.; Hu, J.; Luo, Y. Variation in the Yak Lipin-1 Gene and Its Association with Milk Traits. *J. Dairy Res.* **2020**, *87*, 166–169. [[CrossRef](#)] [[PubMed](#)]
12. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)] [[PubMed](#)]
13. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
14. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
15. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* **2016**, *32*, 292–294. [[CrossRef](#)]
16. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
17. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila Melanogaster* Strain W1118; Iso-2; Iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)]
18. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
19. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nat. Genet.* **2006**, *38*, 904–909. [[CrossRef](#)]
20. Alexander, D.H.; Novembre, J.; Lange, K. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)]
21. Behr, A.A.; Liu, K.Z.; Liu-Fang, G.; Nakka, P.; Ramachandran, S. Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data. *Bioinformatics* **2016**, *32*, 2817–2823. [[CrossRef](#)]
22. Zhang, C.; Dong, S.-S.; Xu, J.-Y.; He, W.-M.; Yang, T.-L. PopLDdecay: A Fast and Effective Tool for Linkage Disequilibrium Decay Analysis Based on Variant Call Format Files. *Bioinformatics* **2019**, *35*, 1786–1788. [[CrossRef](#)] [[PubMed](#)]

23. Chen, H.; Patterson, N.; Reich, D. Population Differentiation as a Test for Selective Sweeps. *Genome Res.* **2010**, *20*, 393–402. [[CrossRef](#)] [[PubMed](#)]
24. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
25. Qiu, Q.; Wang, L.; Wang, K.; Yang, Y.; Ma, T.; Wang, Z.; Zhang, X.; Ni, Z.; Hou, F.; Long, R.; et al. Yak Whole-Genome Resequencing Reveals Domestication Signatures and Prehistoric Population Expansions. *Nat. Commun.* **2015**, *6*, 10283. [[CrossRef](#)] [[PubMed](#)]
26. Lan, D.; Ji, W.; Xiong, X.; Liang, Q.; Yao, W.; Mipam, T.-D.; Zhong, J.; Li, J. Population Genome of the Newly Discovered Jinchuan Yak to Understand Its Adaptive Evolution in Extreme Environments and Generation Mechanism of the Multirib Trait. *Integr. Zool.* **2021**, *16*, 685–695. [[CrossRef](#)]
27. Lan, D.; Xiong, X.; Mipam, T.-D.; Fu, C.; Li, Q.; Ai, Y.; Hou, D.; Chai, Z.; Zhong, J.; Li, J. Genetic Diversity, Molecular Phylogeny, and Selection Evidence of Jinchuan Yak Revealed by Whole-Genome Resequencing. *G3 (Bethesda)* **2018**, *8*, 945–952. [[CrossRef](#)]
28. Abe, A.; Sasaki, T. Purification and Some Properties of the Glycolipid Transfer Protein from Pig Brain. *J. Biol. Chem.* **1985**, *260*, 11231–11239. [[CrossRef](#)]
29. Brown, R.E.; Jarvis, K.L.; Hyland, K.J. Purification and Characterization of Glycolipid Transfer Protein from Bovine Brain. *Biochim. Biophys. Acta* **1990**, *1044*, 77–83. [[CrossRef](#)]
30. Metz, R.J.; Radin, N.S. Purification and Properties of a Cerebroside Transfer Protein. *J. Biol. Chem.* **1982**, *257*, 12901–12907. [[CrossRef](#)]
31. Airene, T.T.; Kidron, H.; Nymalm, Y.; Nylund, M.; West, G.; Mattjus, P.; Salminen, T.A. Structural Evidence for Adaptive Ligand Binding of Glycolipid Transfer Protein. *J. Mol. Biol.* **2006**, *355*, 224–236. [[CrossRef](#)] [[PubMed](#)]
32. van Meer, G.; Wolthoorn, J.; Degroote, S. The Fate and Function of Glycosphingolipid Glucosylceramide. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2003**, *358*, 869–873. [[CrossRef](#)] [[PubMed](#)]
33. Degroote, S.; Wolthoorn, J.; van Meer, G. The Cell Biology of Glycosphingolipids. *Semin. Cell Dev. Biol.* **2004**, *15*, 375–387. [[CrossRef](#)] [[PubMed](#)]
34. Kamlekar, R.K.; Simanshu, D.K.; Gao, Y.; Kenoth, R.; Pike, H.M.; Prendergast, F.G.; Malinina, L.; Molotkovsky, J.G.; Venyaminov, S.Y.; Patel, D.J.; et al. The Glycolipid Transfer Protein (GLTP) Domain of Phosphoinositol 4-Phosphate Adaptor Protein-2 (FAPP2): Structure Drives Preference for Simple Neutral Glycosphingolipids. *Biochim. Biophys. Acta* **2013**, *1831*, 417–427. [[CrossRef](#)] [[PubMed](#)]
35. Malinina, L.; Simanshu, D.K.; Zhai, X.; Samygin, V.R.; Kamlekar, R.; Kenoth, R.; Ochoa-Lizarralde, B.; Malakhova, M.L.; Molotkovsky, J.G.; Patel, D.J.; et al. Sphingolipid Transfer Proteins Defined by the GLTP-Fold. *Q. Rev. Biophys.* **2015**, *48*, 281–322. [[CrossRef](#)]
36. Ji, H.; Wang, H.; Ji, Q.; Ji, W.; Luo, X.; Wang, J.; Chai, Z.; Xin, J.; Cai, X.; Wu, Z.; et al. Differential Expression Profile of MicroRNA in Yak Skeletal Muscle and Adipose Tissue during Development. *Genes Genom.* **2020**, *42*, 1347–1359. [[CrossRef](#)]
37. Nöhammer, C.; El-Shabrawi, Y.; Schauer, S.; Hiden, M.; Berger, J.; Forss-Petter, S.; Winter, E.; Eferl, R.; Zechner, R.; Hoefler, G. cDNA Cloning and Analysis of Tissue-Specific Expression of Mouse Peroxisomal Straight-Chain Acyl-CoA Oxidase. *Eur. J. Biochem.* **2000**, *267*, 1254–1260. [[CrossRef](#)]
38. Morais, S.; Knoll-Gellida, A.; André, M.; Barthe, C.; Babin, P.J. Conserved Expression of Alternative Splicing Variants of Peroxisomal Acyl-CoA Oxidase 1 in Vertebrates and Developmental and Nutritional Regulation in Fish. *Physiol. Genom.* **2007**, *28*, 239–252. [[CrossRef](#)]
39. Clop, A.; Ovilo, C.; Perez-Enciso, M.; Cercos, A.; Tomas, A.; Fernandez, A.; Coll, A.; Folch, J.M.; Barragan, C.; Diaz, I.; et al. Detection of QTL Affecting Fatty Acid Composition in the Pig. *Mamm. Genome* **2003**, *14*, 650–656. [[CrossRef](#)]
40. Casas-Carrillo, E.; Prill-Adams, A.; Price, S.G.; Clutter, A.C.; Kirkpatrick, B.W. Mapping Genomic Regions Associated with Growth Rate in Pigs. *J. Anim. Sci.* **1997**, *75*, 2047–2053. [[CrossRef](#)]
41. Yue, G.; Schröffel, J., Jr.; Moser, G.; Bartenschlager, H.; Reiner, G.; Geldermann, H. Linkage and QTL Mapping for Sus Scrofa Chromosome 12. *J. Anim. Breed. Genet.* **2003**, *120*, 95–102. [[CrossRef](#)]