

Population Genetic Analysis of Ten Geographically Isolated Tibetan Pig Populations

Peng Shang ^{1,†}, Wenting Li ^{2,†}, Zhankun Tan ¹, Jian Zhang ¹, Shixiong Dong ¹, Kejun Wang ^{2,*} and Yangzom Chamba ^{1,*}

¹ Animal Science College, Tibet Agriculture and Animal Husbandry University, Linzhi 860000, China; nemoshpmh@126.com (P.S.); tanzhankun@xza.edu.cn (Z.T.); zhj88n@126.com (J.Z.); herko1@163.com (S.D.)

² College of Animal Sciences and Veterinary Medicine, Henan Agricultural University, Zhengzhou 450001, China; liwenteng_5959@hotmail.com

* Correspondences: Wangkejun.me@163.com or wangkj@cau.edu.cn (K.W.); qbyz628@126.com (Y.C.)

† These authors contributed equally to this work.

Received: 16 March 2020; Accepted: 20 April 2020; Published: 29 July 2020

Simple Summary: Whole-genome re-sequencing data from 10 geographically isolated Tibetan pig populations were collected and analyzed in this study. Population genetic analyses, including Principal Component Analysis (PCA), phylogenetic tree, genetic differentiation, deleterious variant, contribution to meta-population genetic diversity and selective sweep were performed. Limited genetic differentiation was identified among these Tibetan pig populations. Most deleterious variants were low-frequency mutations and population specific. Contribution to the meta-population was largest in the TT population, based on gene and allelic diversity. Genes under selection were involved in hypoxia adaptation, hard palate development, facial appearance, and perception of smell.

Abstract: Several geographically isolated populations of Tibetan pigs inhabit the high-altitude environment of the Tibetan Plateau. Their genetic relationships, contribution to the pool of genetic diversity, and their origin of domestication are unclear. In this study, whole-genome re-sequencing data from 10 geographically isolated Tibetan pig populations were collected and analyzed. Population genetic analyses revealed limited genetic differentiation among the Tibetan pig populations. Evidence from deleterious variant analysis indicated that population-specific deleterious variants were the major component of all mutational loci. Contribution to the meta-population was largest in the TT (Qinghai-Tibet Plateau) population, based on gene diversity or allelic diversity. Selective sweep analysis revealed numerous genes, including RXFP1, FZD1, OR1F1, TBX19, MSTN, ESR1, MC1R, HIF3A, and EGLN2 which are involved in lung development, hard palate development, coat color, hormone metabolism, facial appearance, and perception of smell. These findings increase our understanding of the origins and domestication of the Tibetan pig, and help optimize the strategy for their conservation.

Keywords: domestication; selection; conservation; genetic diversity

1. Introduction

Tibetan pigs originated on the Tibetan Plateau, and the record of their breeding by humans can be traced back to the 7th century [1]. Tibetan pigs are adapted to hypoxia, and a low temperature, low-pressure environment due to long-term natural selection [2–5]. Considerable research has focused on their origins, physiological features, and adaptation to their high-altitude environment. mtDNA evidence has revealed the origin centers of the Tibetan pig [6], and nuclear genome analysis

has shown that the Tibetan pigs originated exclusively on the Tibetan plateau and survived in a wild state [7]. Results from SNP-chip data showed that the Tibetan pig was a domestic indigenous pig living on the Tibetan plateau and underwent genetic differentiation among three populations [4].

In this study, using whole-genome resequencing data, we analyzed the genetic relationship among the 10 geographically isolated Tibetan pig populations as well as between Tibetan pigs and wild pigs. Gene and allelic diversity of each Tibetan pig population was calculated to identify their contributions to the population pool. Finally, selective signatures in the genomes of Tibetan pigs were scanned to identify the genes involved in adaptation to the harsh environment of the Tibetan plateau.

2. Materials and Methods

2.1. Sample Collection

A total of fifty Tibetan pigs from 10 geographically isolated populations and 15 wild pigs were used for this study. As reported in a previous study, all Tibetan pigs were collected from the Qinghai-Tibet Plateau, including Chinese Tibet, Sichuan, Yunnan, and Gansu province [7,8]. As described in Mingzhou Li et al. and Huashui Ai et al. the populations are genetically unrelated, having no common ancestors within three generations [7,8]. The raw data of whole-genome resequencing was downloaded from the Sequence Read Archive (NCBI accession number available in Table S1). Detailed information on the Tibetan pig populations is shown in Table 1 and Table S1. All animals were handled, housed, and all procedures were performed following the guidelines approved by the College of Agriculture and Animal Husbandry, Tibet University.

2.2. Read Mapping and SNP Calling

Libraries with inserts of 500 bp were constructed for each pig. 100-bp paired-end reads were generated using the HiSeq2000 platform [7,8]. Raw reads were filtered (seed mismatches, palindromes simple clip—2:30:10) using Trimmomatic [9]. The BWA-MEM program, with the default parameter settings, [10] was used to map the filtered reads against the reference genome (Ensemble *Sscrofa11.1*). SAMtools (v1.9) was used to merge and sort the BAM files [11]. Duplicate reads were marked using MarkDuplicates from Picard-Tools (<http://broadinstitute.github.io/picard/>). Subsequently, SNP calling was performed for each individual pig using GATK4 HaplotypeCaller then joined using the GenomicsDBImport and GenotypeGVCFs programs from GATK4 (v4.1.2.0) [12]. To get a reliable dataset, the VariantFiltration program from GATK4 was used to filter the SNPs with the parameters “ $QD < 2.0 \parallel FS > 200.0 \parallel SOR > 10.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$ ”.

2.3. Phylogenetic Tree Construction and PCA

Using VCFtools (0.1.16), biallelic SNPs located on autosomes were identified and kept (<http://vcftools.github.io/>) [13]. Using PLINK(v1.9) [14], variants with a call rate >90% and MAF > 0.05 were identified and retained (<http://www.cog-genomics.org/plink2/>). Before constructing the phylogenetic tree, high-density SNPs were pruned using PLINK (v1.9) with the parameters “--indep-pairwise 50 5 0.2”. The phylogenetic tree was constructed using the IQ-TREE software (multicore version 1.6.12) with 1000 bootstrap iterations based on the PHYLIP format data [15]. The phylogenetic tree was visualized using the iTOL web server [16]. Principal Component Analysis (PCA) was also performed with pruned data using the SMARTPCA program of EIGENSOFT [17].

2.4. Genetic Diversity and Deleterious Variants

Expected heterozygosity (H_E), observed heterozygosity (H_O), and the inbreeding coefficient for each Tibetan population were estimated using PLINK (v1.9). Nei’s minimum genetic distance and the gene frequency differentiation index (F_{ST}) between pairs of subpopulations were computed using Metapop v2 [18]. Ensembl Variant Effect Predictor tool (VEP, v98.3) was used to assess the deleterious variants across autosomes. It is well-known that the SIFT score of VEP is sensitive to the alleles that

are likely to be predicted as deleterious. Variant loci with a SIFT score close to 0 were considered deleterious [19]. Using the R UpsetR package [20] their distribution and intersecting sets among populations were analyzed and visualized.

To estimate the maximal contribution to heterozygosity of each population, total heterozygosity was partitioned into the average expected heterozygosity within subpopulations (H_s) and average Nei's minimum genetic distance between subpopulations (D_G) [21]. Similarly, total allelic diversity (A_T) was partitioned within and between the subpopulations. Average allelic diversity within subpopulations is calculated as the average number of alleles segregating in the subpopulations minus one (A_s), while average allelic diversity between subpopulations is calculated as the average number of unique alleles in a subpopulation compared to the other subpopulations averaged over all possible subpopulation pairs (D_A) [18]. Using the Metapop2 program, the above-mentioned indicators were used to estimate the expected contribution of each subpopulation to gene and allelic diversity of the meta-population by removing the subpopulation from the analysis and computing the change in diversity [18].

2.5. Selective Sweeps

Selective sweeps were scanned with sequence diversity statistic θ_π [22] and population differentiation statistics F_{ST} , using a 10 kb no-overlapping window. To identify genomic regions under selection in Tibetan pigs, genome-wide θ_π and F_{ST} were collected. Z-transformed F_{ST} and θ_π ratios were analyzed using the scale function included with the R package. Genomic signatures were defined by a high Z- F_{ST} (Top 1%) and Z- θ_π ratio ($\theta_{\pi, \text{wild}}/\theta_{\pi, \text{Tibetan}}$, Top 1%). Genes located in the selective signature were obtained using R 'biomaRt' package [23], and then subjected to GO enrichment analysis using the Panther web server (<http://www.pantherdb.org>).

3. Results and Discussion

In this study, sequencing data from 50 Tibetan pigs, from 10 geographically isolated populations of the Qinghai-Tibet Plateau, and 15 wild pigs were collected and analyzed (Tables 1 and S1). Approximately 47.4 million autosomal biallelic SNPs were ultimately identified.

Table 1. Genetic diversity in 10 Tibetan pig populations.

Population	<i>n</i>	Origin	H_o	H_e	A_R	F	f_{ii}
YNT	6	Qinghai-Tibet Plateau, Yunnan province	0.3728	0.3714	1.5902	0.0083	0.7456
SCT	6	Qinghai-Tibet Plateau, Sichuan province	0.3594	0.3760	1.5700	0.0514	0.7539
GST	4	Qinghai-Tibet Plateau, Gansu province	0.4385	0.4151	1.5256	0.0189	0.7828
AbaT	4	Qinghai-Tibet Plateau, Tibetan Qiang Autonomous Prefecture of Ngawa	0.3262	0.4178	1.5279	0.2194	0.7832
GZT	5	Qinghai-Tibet Plateau, Tibetan Autonomous Prefecture of Garzê	0.2671	0.3951	1.5150	0.3240	0.7836
TT	6	Qinghai-Tibet Plateau, Tibet	0.3324	0.3670	1.6028	0.0987	0.7409
T1	5	Qinghai-Tibet Plateau	0.2872	0.3956	1.5151	0.2742	0.7835
T2	5	Qinghai-Tibet Plateau	0.2569	0.3874	1.5574	0.3399	0.7698
T3	3	Qinghai-Tibet Plateau	0.4036	0.4682	1.4769	0.1384	0.8128
T5	6	Qinghai-Tibet Plateau	0.2379	0.3681	1.5394	0.3537	0.7712

Note: H_o , observed heterozygosity; H_e , expected heterozygosity; A_R , allelic richness; F , inbreeding coefficient. f_{ii} , Average coancestry within the population.

3.1. Population Structure Analysis

To investigate the genetic structure of 10 Tibetan pig populations and their relationship to wild pigs, we utilized principal component and phylogenetic analysis. For the first principal component (8.78%), no obvious segregation was observed among Tibetan pig populations, while significant differentiation was shown between Tibetan pig and wild pig populations (Figure 1A). Evidence from the second principal component (7.07%) demonstrated that the T1 population was significantly differentiated from others, and the GZT, SCT, and YNT populations were slightly differentiated from the others. These data suggest that genetic drift or blood mixture occurred in the T1 population. Similar results were also observed in the phylogenetic relationships, namely the GZT, SCT, and YNT populations formed a small clade, while T1 and T2 were more like wild pigs (Figure 1B). This phylogenetic pattern was similar to that reported in a previous study based on SNP chip data [4]. In that study Tibetan pigs from Yunnan and Sichuan province were grouped and found to be closer to geographic neighbors' other local breeds. The Ancient Tea Horse Road was a trade route between Tibet and Southeastern China for over 2000 years. Ancient Tea Horse Road, including the Yunnan-Tibet and Sichuan-Tibet routes, enabled commodity exchange between Tibet and Yunnan and Sichuan provinces and likely played an important role in the modern genetic makeup of Tibetan pigs [4]. Analysis of mitochondrial DNA from Tibetan pigs has provided a local origin of domestication [6], and suggests that genetic differentiation of the 10 Tibetan pig populations may be due to breed mixing from other ancient indigenous pigs or genetic drift in the current small population.

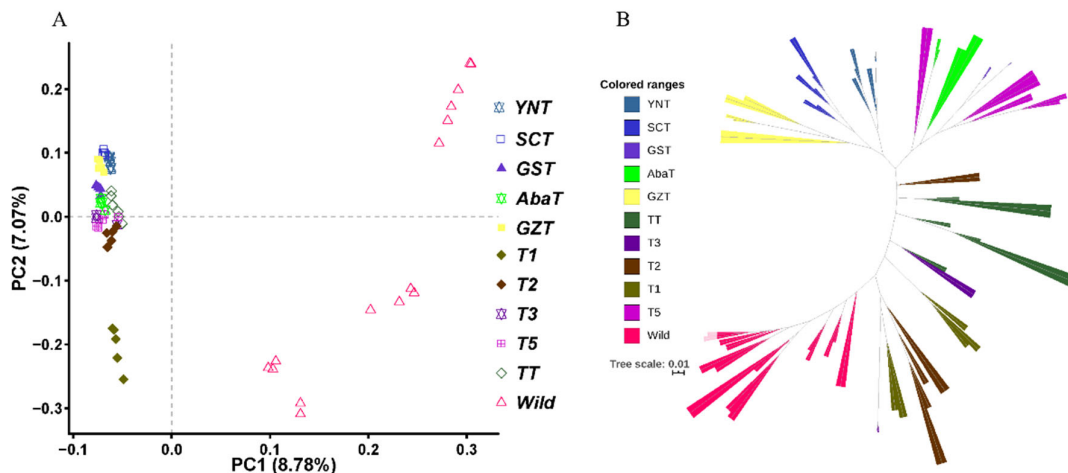


Figure 1. Principal component analysis (A) and the phylogenetic tree (B) of 10 geographically isolated Tibetan pig populations.

3.2. Genetic Diversity and Differentiation

Pair-wise genetic differentiation index (F_{ST}) and pair-wise Nei's minimum genetic distance (D_{nei}) were examined among the Tibetan pig populations (Figure 2). Lower differentiation values were found among GZT, SCT, and YNT populations based on F_{ST} (0.081, 0.1024 and 0.1066) and D_{nei} (0.1499, 0.1885 and 0.1952). The greatest differentiation was observed between the GST and T3 populations based on F_{ST} (0.1904) and D_{nei} (0.3261), while the least differentiation was between T2 and TT, based on F_{ST} (0.0775) and between TT and YNT based on D_{nei} (0.1473). The smallest averages of differentiation and genetic distance were between TT and the other populations (average F_{ST} , 0.0999; average D_{nei} , 0.1880). This result is consistent with the population structure, as the TT population is located in a relatively central position of the PCA and phylogenetic tree (Figure 1). The largest averages of differentiation and genetic distance were between T3 and the other populations (average F_{ST} , 0.1545; average D_{nei} , 0.2869). The second-largest averages of F_{ST} and D_{nei} were between T1 and the

other populations (average F_{ST} , 0.1421; average D_{nei} , 0.2522). Collectively, these results are consistent with the population structure.

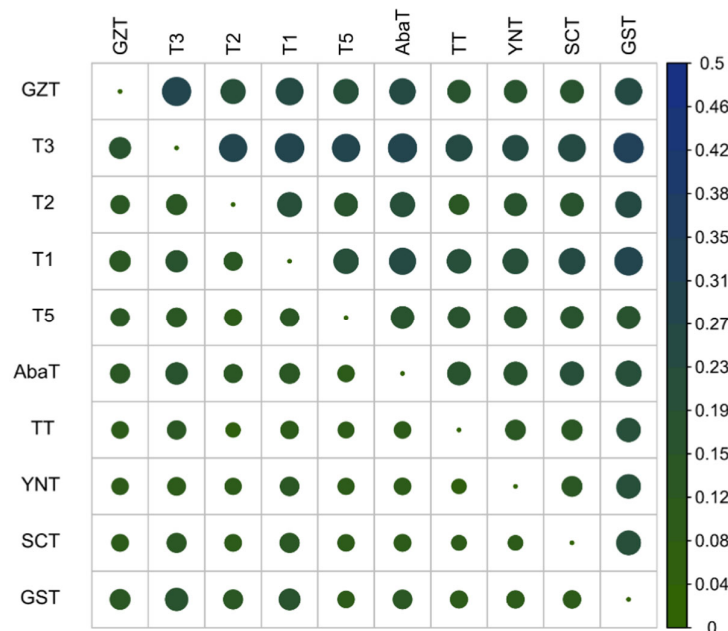


Figure 2. Heatmap plot of Nei's minimum genetic distance D_{nei} (upper) and gene frequency differentiation index F_{ST} (lower).

Genetic diversity (H_o , H_e , A_R and F) was calculated for each population based on whole-genome biallelic SNPs (Table 1). The range of values by each measure was 0.2379–0.4385 (H_o), 0.3678–0.4178 (H_e) and 1.4769–1.6028 (A_R). The maximum H_o was in the GST population (0.4385), followed by T3 (0.4036), and the minimum was in the T5 population (0.2379). The maximum H_e was in the T3 population (0.4682), followed by AbaT (0.4178), and the minimum was in the TT population (0.3670). Some of the differences observed between H_e and H_o are due to different levels of inbreeding in each population. To assess the level of inbreeding within a population, average inbreeding coefficients for each were calculated. The range of inbreeding coefficients was 0.0083 (YNT) to 0.3537 (T5) (Table 1). The greatest inbreeding coefficient and lowest heterozygosity were observed in the T5 population. Although estimation of these indicators can be biased due to the small population size, whole-genome SNP data from unrelated samples would reduce this bias to some extent.

3.3. Deleterious Variants and Their Contribution to Gene and Allele Diversity

Conservation management aims to maintain genetic diversity to maximize survival potential and fitness [19]. Often deleterious variants are used to estimate the fitness of a population. Here we used the Ensembl Variant Effect Predictor to identify putative damaging mutations across the genome; variants with a SIFT score close to 0 were designated as such, and a total of 8490 putative deleterious variants were identified. The frequency distribution of deleterious variants is presented in Figure 3A and Table S2, most are low-frequency mutations, ranging from 0 to 0.59. Allele frequency of over half the deleterious variants was 0 to 0.05 and the majority of mutations were purine to pyrimidine bases (Figure 3A). Using whole-genome resequencing data, Bosse et al. identified 3129 and 3468 low-frequency putative deleterious variants in *Sus cebifrons* and *Pietrain* pig populations respectively [19]. Similarly, we found that the number of deleterious variants varied among the Tibetan pig populations, ranging from 1606 (T3) to 3499 (TT); only 229 of these were present in all populations (Figure 3B and Table S2). Most deleterious variants were population-specific, which ranged from 652 (TT) to 187 (T3) (Figure 3B). Furthermore, the pair-wise variant number of intersection sets were all less than 100 except for the T2-TT pair. Subsequently, the functional

annotation of each deleterious variant was collected, after removing redundancy, over 3000 genes were subject to gene ontology (GO) enrichment analysis (Table S3). The results indicated that many of these genes functioned in sensory perception of smell, detection of stimulus, sensory perception of chemical stimulus, nervous system processes, carboxylic acid metabolic processes, and double-strand break repair (Table S3).

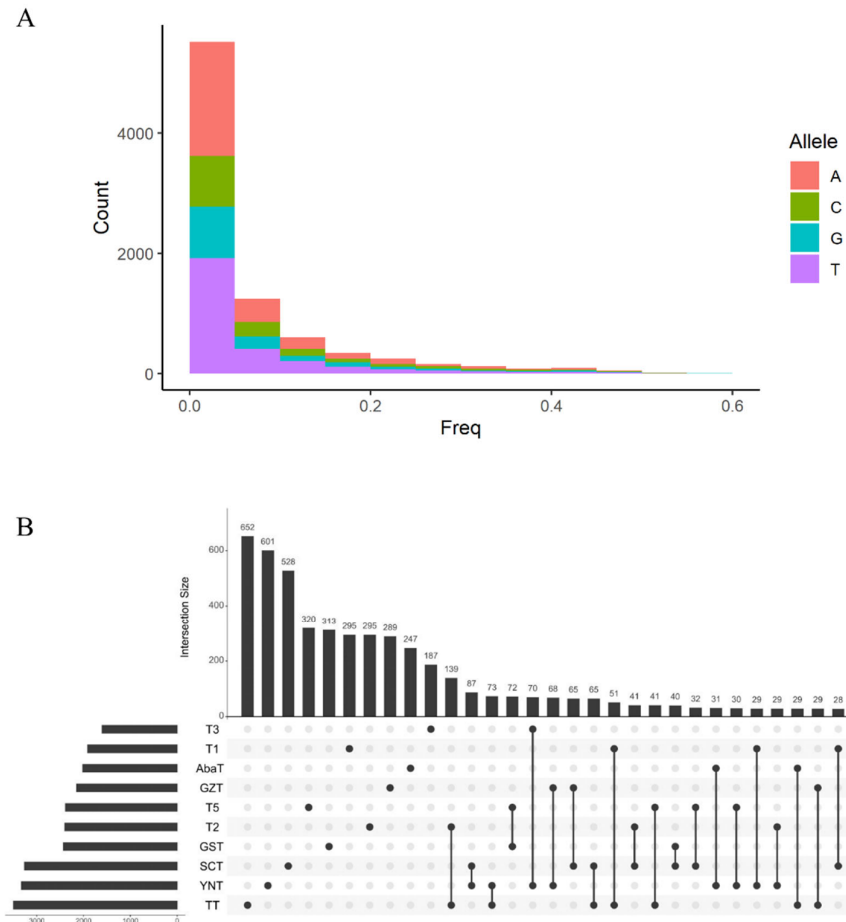


Figure 3. Distribution of deleterious variants. The frequency spectrum of deleterious variants (A). Intersect plot of deleterious variant among 10 Tibetan pig populations (B).

To realize the adaptive potential of populations to new environmental challenges, the maintenance of genetic diversity is a priority for conservation management [24]. Genetic diversity is usually estimated by expected heterozygosity under Hardy–Weinberg equilibrium and by allele diversity [21]. These methods are complementary since heterozygosity is sensitive to selection effects [25] and allelic diversity is sensitive to bottleneck effects [26]. In this study, methods of maximizing globally expected heterozygosity and allelic diversity were employed to compute the contribution of each population. The contribution of each population was estimated by removing a subpopulation from the meta-population and calculating the change in gene diversity and allelic diversity [27,28]. A positive contribution value represents the loss of diversity after removing a subpopulation, while a negative value means a gain of diversity, as described by Metapop2 [18]. As shown in Figure 4, the rank of contribution to total diversity among populations was roughly the same for total heterozygosity and total allelic diversity. Removal of the TT population revealed the largest loss to total heterozygosity, H_T , (0.7528%), followed by YNT (0.7056%) (Figure 4A). Similarly, removal of the TT population revealed the largest loss to total allelic diversity, A_T , (1.0644%), followed by YNT (0.8696%). A gain of H_T was the largest in the T5 population (−0.2737%), while the gain of A_T was the largest in the T3 population (−0.6132%). T3 had the greatest negative contribution to both H_S

(−0.9713%) and A_s (−0.9923%), implying this population had low within-population diversity (Figure 4A, B). This could be explained by its low allelic richness (1.4769) and the highest average co-ancestry (0.8128) (Table 1). The TT population presented the largest positive contribution to both H_s (1.3501%) and A_s (0.9260%), suggesting this population had the highest within-population diversity. This also could be explained by its high allelic richness (1.6028) and low average co-ancestry (0.7409) (Table 1). In conservation programs, generally maximizing expected heterozygosity is equivalent to minimizing average co-ancestry within populations [29]. However, the TT population had the lowest negative contribution to D_G (−0.5973%) as its average Nei's distance against others (0.1881) was the smallest among the populations (Figure 2). The TT population contributed positively to D_A , which is derived from its high A_R (1.6028) (Table 1). This explains why its removal brings a loss in between-population gene diversity but a gain in between-population allelic diversity. Optimum contributions of each subpopulation to the meta-population pool were demonstrated based on maximal expected heterozygosity and total number of alleles (Figure 4C). Expectedly, the TT population made the largest contribution to gene diversity (H) and allele diversity (K). For gene diversity, the TT, YNT, T1, SCT, and GST populations contributed more than 10%, to allele diversity, TT, AbaT and SCT contributed more than 10% while the others accounted for ~9% of the total. In aggregate, these data suggest that a conservation plan for the Tibetan pig would at least initially start by focusing on protecting the TT population.

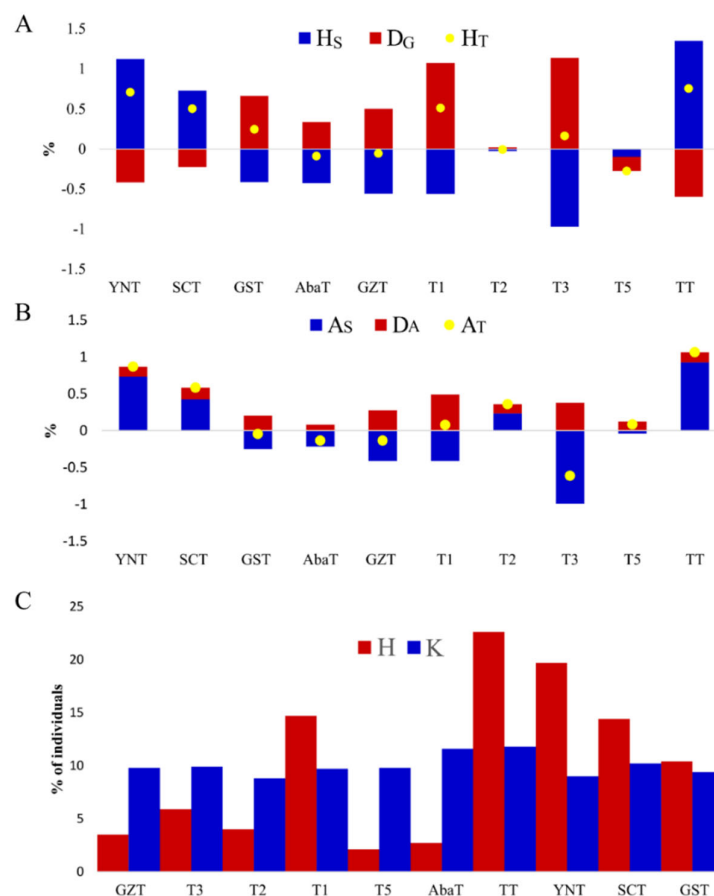


Figure 4. Contribution to genetic diversity of 10 Tibetan pig populations. Contribution to gene diversity (A) and allele diversity (B) H_s and A_s indicate within-population gene and allele diversity, respectively. D_G and D_A indicate between-population gene and allele diversity, respectively. H_T and A_T indicate total gene and allele diversity, respectively. Contribution of individuals from each population to a pool with maximal gene diversity (H) and allele diversity (K) (C).

3.4. Selective Sweep

Tibetan pigs originated on the Tibetan Plateau, and have undergone long term natural selection in that harsh environment [5]. Here, in order to investigate the effects of domestication and selection on the Tibetan pig genome, we performed a selective sweep analysis of the whole genome. Since the extent of LD (Linkage disequilibrium) is less than 10 kb in Chinese domestic pigs, we set the selective windows size at 10 kb [1,30]. Over 220,000 non-overlapping windows were scanned for selective signatures. Selective signatures were identified in 863 windows by combining $Z\text{-}F_{ST}$ (Top 1%) and, $Z\text{-}\theta_{\pi}$ ratio ($\theta_{\pi, \text{wild}}/\theta_{\pi, \text{Tibetan}}$, Top 1%) (Figure 5A). After functional analysis of selective signatures, 384 genes were identified (Table S4). Results from GO analysis identified genes involved in lung connective tissue development and lung development, hard palate development and craniofacial suture morphogenesis, sensory perception of smell, negative regulation of cell growth involved in cardiac muscle cell development, response to oxygen-containing compound, growth hormone secretion and animal organ morphogenesis. (Figure 5B and Table S5).

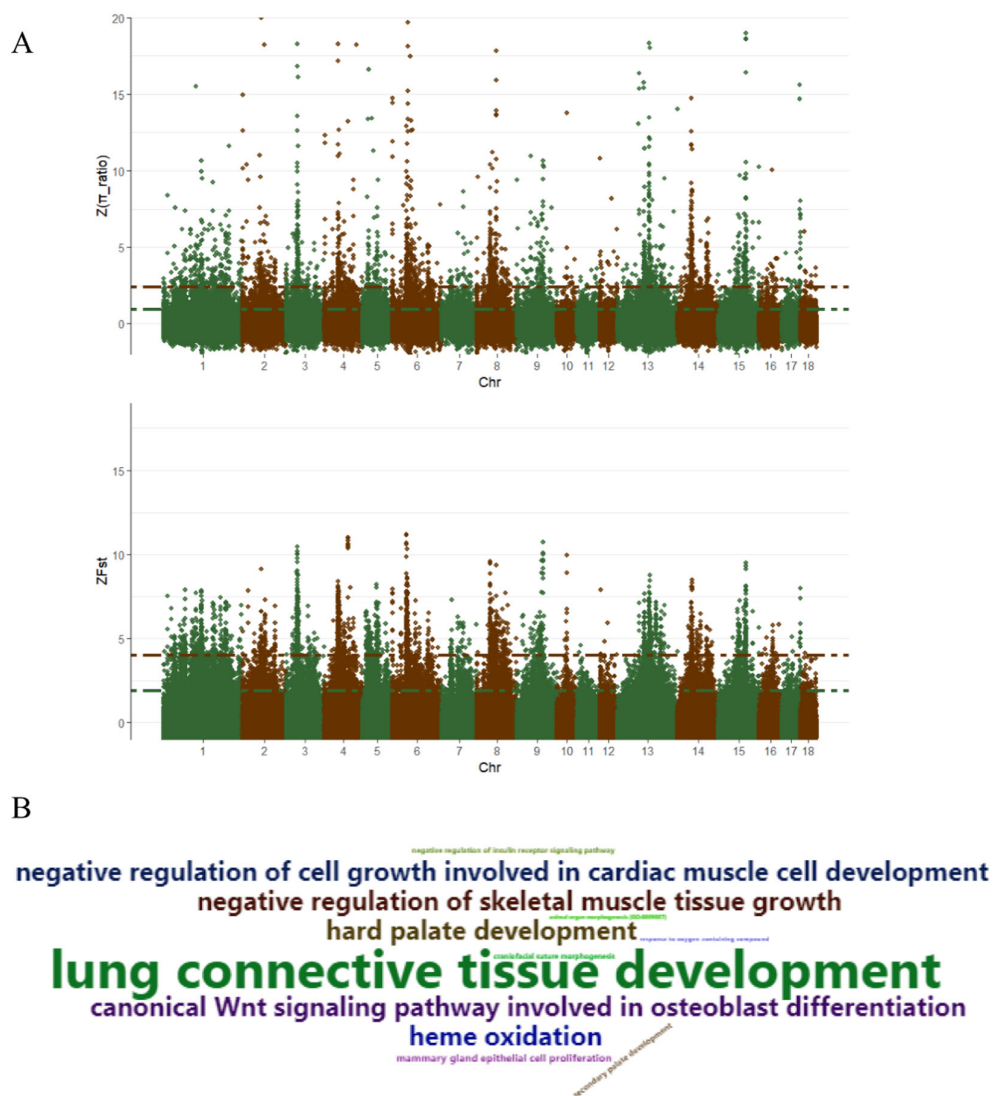


Figure 5. Genome regions of selective signature and GO enrichment analysis. Distribution of $Z\text{-}F_{ST}$ and $Z\text{-}\theta_{\pi}$ ratio ($\theta_{\pi, \text{wild}}/\theta_{\pi, \text{Tibetan}}$) (A). Each dot represents a 10 kb non-overlapping window, red and green lines represent the 95th and 99th quantile, respectively. Wordcloud plot showed the interesting GO enrichment terms (B).

In this study, we focused on some of the difference between wild and Tibetan pigs. These included facial appearance, coat color, hormone metabolism, growth, and hypoxia adaptation-related processes. Three genes, *FZD1*, *DLG1* and *MMP25*, were involved in hard palate development and craniofacial suture morphogenesis. *FZD1* and its receptor are expressed ventral to the telencephalon and in periocular mesenchyme during the development of the face [31]. Loss of function variants in *DLG1* are associated with non-syndromic discontinuous cleft lip and palate [32]. *MMP25* expression appears to have a crucial role throughout all stages of palatal development [33]. These genes may have played an important role in face or palatal changes during domestication of Tibetan pig. A peak of selective signal ($Z\text{-}F_{ST}$ 7.69 and $Z\text{-}\theta_{\pi}$ ratio 318.74) was identified in the region Chr 6:180001:190000. The *MC1R* gene is located in this selective region, which is reported to contribute to the black color of Chinese domestic pig [34]. *TBX19* is involved in the selection effect of domestication on the pig genome [35]. *TBX19* plays a crucial role in activating pituitary cell differentiation and induces proopiomelanocortin expression, which leads to generating the adrenocorticotrophic hormone working in the hypothalamic–pituitary–adrenal axis [35,36]. Five cytochrome P450 genes, *CYP2F1*, *CYP2A19*, *CYP2B22*, *CYP2B22*, and *CYP2S1*, and *ENSSSCG00000028627*, and a lncRNA *ENSSSCG00000041224* are located in a 120 kb selective region. *ENSSSCG00000028627* is located in the peak of selective signal ($Z\text{-}F_{ST}$ 8.15 and $Z\text{-}\theta_{\pi}$ ratio 4.23). These genes function in the synthesis of steroids, cholesterol and other lipids [37]. It suggests that hormone metabolic processes were likely under positive selection in Tibetan pigs. Unexpectedly, *MSTN* and *ESR1* were selected in the Tibetan pig in relative to wild pig. *MSTN*, a interesting gene, negatively regulates skeletal muscle cell proliferation and differentiation. Mutation in *MSTN* is associated with muscle hypertrophy [38]. *MSTN* was selected in lowland domestic pigs relative to wild pig [35]. *ESR1* was also selected in Chinese domestic pigs relative to wild pigs [39], participating in the reproduction process [40]. Long-standing trade likely introduced the lowland domestic pig to Tibetan plateau where they interbred with the Tibetan pig [4]. The *MSTN* and *ESR1* genes might be the result of gene flow from lowland pigs to Tibetan pigs. There were also three genes participating in sensory perception of smell, *OR1F1*, *ENSSSCG00000037040*, and *ENSSSCG00000034150*. *OR1F1*, an olfactory receptor, was confirmed to function in odor perception activation [41]. Previous publications also demonstrated Tibetan pigs lost some olfactory receptor genes across the genome [7,42].

Some interesting genes involved in lung connective tissue development and lung development included *RXFP1*, *CIC*, *MAPK8IP3*, *CCDC39*, and *SPRY1* (Table S4). *RXFP1* is directly associated with pulmonary function; low expression of this gene is found in diseased lungs [43]. *RXFP1*, a relaxin family peptide receptor, protects astrocytes from hypoxia [44]. *HIF3A* and *EGLN2* were selected in Tibetan pigs. These genes are involved in the hypoxia pathway in Tibetan cattle [45]. Unexpectedly, relatively fewer hypoxia adaptation related genes were identified in selective sweep analysis. After literature mining, we found that *EPAS1* is considered a crucial gene in hypoxia adaptation of Tibetan pig [1,4,7]. However, no significant selective signal was identified near the *EPAS1* gene in this study. These previous studies aimed to find the candidate genes involved in hypoxia adaptation of Tibetan pig and used Chinese lowland native pigs as the control [1,4,7]. This study, however, aimed to identify the domestication region of the Tibetan pig genome, and so used wild pigs as the control. What we found were many phenotype-related genes, including coat color, facial appearance, sensory perception of smell, growth, and hormone metabolism.

Additionally, the integrated analysis identified three potential deleterious genes selected in Tibetan pigs relative to wild pigs. *DUOX2* is a transmembrane protein located at the surface of thyroid follicular cells. Mutation in this gene is associated with congenital hypothyroidism [46]. *SULT1B1* is a protein gene involved in cytochrome P450 metabolism pathway [37]. *SCAPER* mutation causes autosomal recessive retinitis pigmentosa with intellectual disability [47]. These data suggest that domestication had negative, as well as positive, effects on the genome of Tibetan pig.

4. Conclusions

Based on population genetic analysis, the Tibetan pigs living in various regions of the Tibetan Plateau have undergone slight genetic differentiation. Analysis of the distribution of deleterious

variants indicated that population-specific deleterious variants accounted for the major component of loci with negative effects. Results from contribution to genetic diversity analysis demonstrated that the TT population made the highest contribution to gene diversity and allele diversity to the meta-population. Selective sweep analysis revealed that numerous genes, including *RXFP1*, *FZD1*, *OR1F1*, *TBX19*, *MSTN*, *ESR1*, *MC1R*, *HIF3A*, and *EGLN2* are involved in hypoxia adaptation, coat color, hormone metabolism, facial appearance, and perception of smell. These findings help us understand the origin and domestication of the Tibetan pigs, as well as optimizing conservation programs.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Sample information of 10 Tibetan pig populations, Table S2: Annotation information and frequency distribution of deleterious variants of Tibetan pigs, Table S3: GO enrichment analysis of deleterious variants, Table S4: Gene list of selective regions, Table S5: GO enrichment analysis of genes under selection.

Author Contributions: Conception and design of study: K.W. and Y.C. Performance of experiments and analysis of data: K.W., P.S., W.L., Z.T., J.Z. and S.D. Writing and revision of manuscript: K.W., P.S. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Central Guiding Local Projects of China (YDZX20195400004426), the Central Government Special Funds Support for the Development of Local Universities (ZZXT2019-02), and the Foundation of He' nan Educational Committee (19A230006).

Conflicts of Interest: We declare that we have no conflict of interest.

References

1. Ma, Y.F.; Han, X.M.; Huang, C.P.; Zhong, L.; Adeola, A.C.; Irwin, D.M.; Xie, H.B.; Zhang, Y.P. Population Genomics Analysis Revealed Origin and High-altitude Adaptation of Tibetan Pigs. *Sci. Rep.* **2019**, *9*, 11463, doi:10.1038/s41598-019-47711-6.
2. Zhang, B.; Qiangba, Y.; Shang, P.; Wang, Z.; Ma, J.; Wang, L.; Zhang, H. A Comprehensive MicroRNA Expression Profile Related to Hypoxia Adaptation in the Tibetan Pig. *PLoS ONE* **2015**, *10*, e0143260, doi:10.1371/journal.pone.0143260.
3. Zhang, B.; Chamba, Y.; Shang, P.; Wang, Z.; Ma, J.; Wang, L.; Zhang, H. Comparative transcriptomic and proteomic analyses provide insights into the key genes involved in high-altitude adaptation in the Tibetan pig. *Sci. Rep.* **2017**, *7*, 3654, doi:10.1038/s41598-017-03976-3.
4. Ai, H.; Yang, B.; Li, J.; Xie, X.; Chen, H.; Ren, J. Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genom.* **2014**, *15*, 834, doi:10.1186/1471-2164-15-834.
5. Shang, P.; Li, W.; Liu, G.; Zhang, J.; Li, M.; Wu, L.; Wang, K.; Chamba, Y. Identification of lncRNAs and Genes Responsible for Fatness and Fatty Acid Composition Traits between the Tibetan and Yorkshire Pigs. *Int. J. Genom.* **2019**, *2019*, 5070975, doi:10.1155/2019/5070975.
6. Yang, S.; Zhang, H.; Mao, H.; Yan, D.; Lu, S.; Lian, L.; Zhao, G.; Yan, Y.; Deng, W.; Shi, X.; et al. The local origin of the Tibetan pig and additional insights into the origin of Asian pigs. *PLoS ONE* **2011**, *6*, e28215, doi:10.1371/journal.pone.0028215.
7. Li, M.; Tian, S.; Jin, L.; Zhou, G.; Li, Y.; Zhang, Y.; Wang, T.; Yeung, C.K.; Chen, L.; Ma, J.; et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **2013**, *45*, 1431–1438, doi:10.1038/ng.2811.
8. Ai, H.; Fang, X.; Yang, B.; Huang, Z.; Chen, H.; Mao, L.; Zhang, F.; Zhang, L.; Cui, L.; He, W.; et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **2015**, *47*, 217–225, doi:10.1038/ng.3199.
9. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
10. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.
11. 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, doi:10.1093/bioinformatics/btp352.
12. 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, doi:10.1101/gr.107524.110.

13. 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, doi:10.1093/bioinformatics/btr330.
14. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **2015**, *4*, 7, doi:10.1186/s13742-015-0047-8.
15. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589, doi:10.1038/nmeth.4285.
16. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259, doi:10.1093/nar/gkz239.
17. 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, doi:10.1038/ng1847.
18. Lopez-Cortegano, E.; Perez-Figueroa, A.; Caballero, A. metapop2: Re-implementation of software for the analysis and management of subdivided populations using gene and allelic diversity. *Mol. Ecol. Resour.* **2019**, *19*, 1095–1100, doi:10.1111/1755-0998.13015.
19. Bosse, M.; Megens, H.J.; Madsen, O.; Crooijmans, R.P.; Ryder, O.A.; Austerlitz, F.; Groenen, M.A.; de Cara, M.A. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Res.* **2015**, *25*, 970–981, doi:10.1101/gr.187039.114.
20. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**, *33*, 2938–2940, doi:10.1093/bioinformatics/btx364.
21. Lopez-Cortegano, E.; Pouso, R.; Labrador, A.; Perez-Figueroa, A.; Fernandez, J.; Caballero, A. Optimal Management of Genetic Diversity in Subdivided Populations. *Front. Genet.* **2019**, *10*, 843, doi:10.3389/fgene.2019.00843.
22. Nei, M.; Li, W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 5269–5273, doi:10.1073/pnas.76.10.5269.
23. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191, doi:10.1038/nprot.2009.97.
24. Steffen, W.; Richardson, K.; Rockstrom, J.; Cornell, S.E.; Fetzer, I.; Bennett, E.M.; Biggs, R.; Carpenter, S.R.; de Vries, W.; de Wit, C.A.; et al. Planetary boundaries: Guiding human development on a changing planet. *Science* **2015**, *347*, 1259855, doi:10.1126/science.1259855.
25. Frankham, R. Introduction to quantitative genetics (4th edn). *Am. J. Hum. Genet.* **1996**, *12*, 280.
26. Leberg, P.L. Estimating allelic richness: Effects of sample size and bottlenecks. *Mol. Ecol.* **2002**, *11*, 2445–2449, doi:10.1046/j.1365-294x.2002.01612.x.
27. Petit, R.J.; Mousadik, A.E.; Pons, O. Identifying Populations for Conservation on the Basis of Genetic Markers. *Conserv. Biol.* **1998**, *12*, 844–855.
28. Caballero, A.; Toro, M.A. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* **2002**, *3*, 289–299, doi:10.1023/A:1019956205473.
29. Toro, M.; Pérez-Enciso, M. Optimization of selection response under restricted inbreeding. *Genet. Sel. Evol.* **1990**, *22*, 93–90.
30. Amaral, A.J.; Megens, H.J.; Crooijmans, R.P.; Heuven, H.C.; Groenen, M.A. Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* **2008**, *179*, 569–579, doi:10.1534/genetics.107.084277.
31. Geetha-Loganathan, P.; Nimmagadda, S.; Antoni, L.; Fu, K.; Whiting, C.J.; Francis-West, P.; Richman, J.M. Expression of WNT signalling pathway genes during chicken craniofacial development. *Dev. Dyn.* **2009**, *238*, 1150–1165, doi:10.1002/dvdy.21934.
32. Demeer, B.; Revencu, N.; Helaers, R.; Gbaguidi, C.; Dakpe, S.; Francois, G.; Devauchelle, B.; Bayet, B.; Vikkula, M. Likely Pathogenic Variants in One Third of Non-Syndromic Discontinuous Cleft Lip and Palate Patients. *Genes* **2019**, *10*, 833, doi:10.3390/genes10100833.
33. Paiva, K.B.S.; Maas, C.S.; Dos Santos, P.M.; Granjeiro, J.M.; Letra, A. Extracellular Matrix Composition and Remodeling: Current Perspectives on Secondary Palate Formation, Cleft Lip/Palate, and Palatal Reconstruction. *Front. Cell Dev. Biol.* **2019**, *7*, 340, doi:10.3389/fcell.2019.00340.
34. Fang, M.; Larson, G.; Ribeiro, H.S.; Li, N.; Andersson, L. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* **2009**, *5*, e1000341, doi:10.1371/journal.pgen.1000341.

35. Zhu, Y.; Li, W.; Yang, B.; Zhang, Z.; Ai, H.; Ren, J.; Huang, L. Signatures of Selection and Interspecies Introgression in the Genome of Chinese Domestic Pigs. *Genome Biol. Evol.* **2017**, *9*, 2592–2603, doi:10.1093/gbe/evx186.
36. Liu, J.; Lin, C.; Gleiberman, A.; Ohgi, K.A.; Herman, T.; Huang, H.P.; Tsai, M.J.; Rosenfeld, M.G. Tbx19, a tissue-selective regulator of POMC gene expression. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8674–8679, doi:10.1073/pnas.141234898.
37. Baldwin, W.S.; Marko, P.B.; Nelson, D.R. The cytochrome P450 (CYP) gene superfamily in *Daphnia pulex*. *BMC Genom.* **2009**, *10*, 169, doi:10.1186/1471-2164-10-169.
38. Saunders, M.A.; Good, J.M.; Lawrence, E.C.; Ferrell, R.E.; Li, W.H.; Nachman, M.W. Human adaptive evolution at Myostatin (GDF8), a regulator of muscle growth. *Am. J. Hum. Genet.* **2006**, *79*, 1089–1097, doi:10.1086/509707.
39. Yang, B.; Cui, L.; Perez-Enciso, M.; Traspov, A.; Crooijmans, R.; Zinovieva, N.; Schook, L.B.; Archibald, A.; Gatphayak, K.; Knorr, C.; et al. Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet. Sel. Evol.* **2017**, *49*, 71, doi:10.1186/s12711-017-0345-y.
40. Li, W.T.; Zhang, M.M.; Li, Q.G.; Tang, H.; Zhang, L.F.; Wang, K.J.; Zhu, M.Z.; Lu, Y.F.; Bao, H.G.; Zhang, Y.M.; et al. Whole-genome resequencing reveals candidate mutations for pig prolificacy. *Proc. Biol. Sci.* **2017**, *284*, doi:10.1098/rspb.2017.2437.
41. Malnic, B.; Godfrey, P.A.; Buck, L.B. The human olfactory receptor gene family. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2584–2589, doi:10.1073/pnas.0307882100.
42. Wang, M.S.; Yang, H.C.; Otecko, N.O.; Wu, D.D.; Zhang, Y.P. Olfactory genes in Tibetan wild boar. *Nat. Genet.* **2016**, *48*, 972–973, doi:10.1038/ng.3631.
43. Bahudhanapati, H.; Tan, J.; Dutta, J.A.; Strock, S.B.; Sembrat, J.; Alvarez, D.; Rojas, M.; Jager, B.; Prasse, A.; Zhang, Y.; et al. MicroRNA-144-3p targets relaxin/insulin-like family peptide receptor 1 (RXFP1) expression in lung fibroblasts from patients with idiopathic pulmonary fibrosis. *J. Biol. Chem.* **2019**, *294*, 5008–5022, doi:10.1074/jbc.RA118.004910.
44. Willcox, J.M.; Summerlee, A.J. Relaxin protects astrocytes from hypoxia in vitro. *PLoS ONE* **2014**, *9*, e90864, doi:10.1371/journal.pone.0090864.
45. Wu, D.D.; Ding, X.D.; Wang, S.; Wojcik, J.M.; Zhang, Y.; Tokarska, M.; Li, Y.; Wang, M.S.; Faruque, O.; Nielsen, R.; et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2018**, *2*, 1139–1145, doi:10.1038/s41559-018-0562-y.
46. Abe, K.; Narumi, S.; Suwanai, A.S.; Adachi, M.; Muroya, K.; Asakura, Y.; Nagasaki, K.; Abe, T.; Hasegawa, T. Association between monoallelic TSHR mutations and congenital hypothyroidism: A statistical approach. *Eur. J. Endocrinol.* **2018**, *178*, 137–144.
47. Tatour, Y.; Sanchez-Navarro, I.; Chervinsky, E.; Hakonarson, H.; Gawi, H.; Tahsin-Swafiri, S.; Leibur, R.; Lopez-Molina, M.I.; Fernandez-Sanz, G.; Ayuso, C.; et al. Mutations in SCAPER cause autosomal recessive retinitis pigmentosa with intellectual disability. *J. Med. Genet.* **2017**, *54*, 698–704.

