Genome-wide SNP analysis reveals the selection signatures of two indigenous buffalo breeds in Sichuan

Wei Wang1 · Donghui Fang1 · Yi Shi1 · Fang He1 · Jia Gan1 · Maozhong Fu1 · Xiaodong Deng1 · Tao Zhang2 · Hongbing Shen3 · Shengchao Zhong3 · Yueda Agu1 · Luxi Chen2 · Naza An2 · Zhiguang Zhang4 · Ji Yi1

Received: 9 November 2021 / Accepted: 27 April 2022 / Published online: 27 May 2022
© The Author(s) 2022

Abstract
Sichuan Province spawned abundant of indigenous buffalo varieties, which probably harbor valuable gene resources beneficial to the genetic improvement of buffalo. However, limited genetic information was publicly available. To better understand their selection signatures between different populations, we performed a restriction site-associated DNA sequencing (RAD-seq) to explore genome-wide SNPs among two indigenous breeds of Sichuan buffaloes. As a result, a total of 2,110,077 high-quality SNPs were finally obtained. Population genetic analysis indicated a obviously genetic differentiation between two breeds. The detection of selective genes showed that 995 and 910 protein-coding genes underwent positive selection in Yibin buffalo (GYBS) and Dechang buffalo (XCS). Further functional analysis revealed distinctly discrepant selection in two breeds. Candidate genes that positively selected from Yibin buffaloes have mainly occurred in functions closely related to meat quality, complex living environment adaption capability, and disease resistance. While they were significantly enriched in cell proliferation and cell components in Dechang buffalo, indicating the selection pressure primarily derived from the requirement of organism growth and development speed during breed formation. Our dataset constitutes a promising reservoir of genome-wide SNP markers of Sichuan buffaloes and provides potentially traits selected in different local populations. Such comprehensive genetic resources offer an unprecedented opportunity for genetic association analysis of economically important traits and precision breeding programs in buffaloes.

Keywords Buffalo · Selection signatures · SNPs · RADseq

Introduction

Domestic buffalo (*Bubalus bubalis*) is an important multi-purpose agricultural animal in tropical and subtropical regions. China has the third-largest buffalo population in the world, following India and Pakistan (Zhang et al. 2007), and holds a long history in domestication and utilization of buffalo for thousands of years (Yue et al. 2013). In the long period of utilization, a variety of buffalo breeds were produced and retained through artificial selection for desirable performance traits, such as milk or meat production, labour force and so on (Zhang et al. 2020; Liang et al. 2021). Sichuan Basin in southwestern China is positioned to the northeast of the eastern Himalayas and is a typical agricultural production region with mountains encircled. Due to the special geographical location and climatic environment, Sichuan Basin spawned two indigenous buffalo varieties, Yibin buffalo and Dechang buffalo. Yibin buffalo are better adapted to high heat environments, and Dechang buffalo grows faster. However, limited attention has been given to them, which is apparently insufficient for mining functional genes associated with important traits and the genetic breeding of buffalo.

Single nucleotide polymorphisms (SNPs) are the most abundant molecular markers and have contributed significantly to address genetic diversity and positive selection
sites of many species (Anello et al. 2016; Wang et al. 2013). Restriction site-associated DNA sequencing (RADseq) is a powerful approach for obtaining genetic variation on a genome-wide scale by a reduced representation sequencing of regions adjacent to restriction cut sites. Since the ability for providing numerous information on genetic variation and relatively low cost (Baird et al. 2008), RADseq has been widely used for markers discovery and genotyping in many species (Orita et al. 2021; Liu et al. 2019; Yang et al. 2017; Wang et al. 2018; Junge et al. 2019).

In present study, we employed RADseq approach to discover genome-wide SNPs and explored positively selected location of two representative indigenous buffalo breeds in the Sichuan Basin. These results of the present study undoubtedly facilitate a better understanding of the genetic composition of Sichuan indigenous buffaloes and provide a comprehensive set of candidate genetic regions for improving management strategies for buffalo breeding selection.

Materials and methods

Samples collection and DNA isolation

Blood samples from 20 buffaloes were collected from two representative indigenous buffalo breeds, Yibin buffalo (GYBS, n = 10, 27° 50′–29° 16′ N, 103° 36′–105° 20′ E) and Dechang buffalo (XCS, n = 10, 27° 05′–27° 36′ N, 101° 54′–102° 29′ E), in Sichuan Province. Genomics DNA for genotyping was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). The DNA qualification was performed by using three methods as follows: (1) DNA purity and integrity were evaluated with agarose gel electrophoresis, (2) DNA purity was tested using Nanodrop (Wilmington, DE, USA) (OD260/OD280) and (3) DNA concentration was measured by Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA).

GBS library construction and high-throughput sequencing

After a preliminary in silico analysis of the buffalo genome sequences, the genomic DNA from individual samples was first normalized to a concentration of 50 ng/μl and digested with restriction enzyme EcoRI. RAD libraries were prepared as the protocol described in Baird et al. (Baird et al. 2008). Briefly, after ligating with a uniquely barcoded sequencing adaptor (Solexa P1 Adapter), samples were pooled and randomly sheared. Fragments of 300 bp to 700 bp were collected by gel purification and ligated to a second adapter (P2) with divergent ends. Products were further amplified by polymerase chain reaction (PCR), and amplicons were sequenced on the Illumina HiSeq Platform with 150 bp paired-end reads (Novogene Co. Ltd., Beijing).

Quality filtering, SNP calling, and annotation

According to the official pipeline (https://www.illumina.com/), initials image data obtained by high throughput sequencer were converted into sequence files in FASTQ format. To ensure the reliability of sequence data, raw data were trimmed using the following criteria: removing reads containing adapter sequences, removing reads containing unidentified nucleotides (N) representing ≥ 10% of total length; removing reads with > 50% of bases with quality scores ≤ 5. If any sequence of the paired reads was identified as low quality, both pairs were discarded. At the same time, Q20, Q30, GC content, and duplication level of the sequencing data were also calculated. After these steps, clean reads with high quality were obtained and used for the following analyses.

High-quality reads remaining were mapped to the reference genome (Low et al. 2019), https://www.ncbi.nlm.nih.gov/assembly/GCF_003121395.1) using BWA software (Li and Durbin 2009) with the command ‘mem -t 4 -k 32 -M’. SNP calling was performed on a population scale by using Bayesian methods implemented in the SAM tools package (Li et al. 2009). Filtering was performed to improve the quality of SNP obtained as follows: SNP quality value was more than twenty, the number of reads supported by SNP was in the range of 1/3 to 5 × of average sequencing depth, and the distance of two adjacent SNPs ≥ 5. Annotating and prioritizing of all genetic variants in genome sequencing data were carried out by ANNOVAR software (Wang et al. 2010).

PCA analysis

Based on the high-quality SNPs dataset, genetic diversity and relationship among populations were calculated. The phylogenetic tree was constructed by using the neighbor-joining method in software TreeBest v1.9.2 with 1000 bootstrap replicates. The principal component analysis (PCA) was performed based on SNPs among individual genomes using GCTA (Yang et al. 2011).

Screening for selection signatures and candidate genes analysis

Genetic diversity and population differentiation patterns of two populations in genome-wide were characterized by using nucleotide diversity (π) and population differentiation statistic (Fst) calculated by a sliding window approach (a 40 kb window sliding with a step size of 20 kb). Selection signatures of population-specific genomic regions between XCS and GYBS were measured as the genomic regions...
with a significantly high $F_{st}$ (top 5%) and $\theta$ ratio ($\log_2 (\pi_1) - \log_2 (\pi_2)$, top 5%) values. Genes located in these regions were annotated for functional ontology terms and biological pathways by using Gene Ontology (GO) (Gene et al. 2000) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto 2000).

**Results**

**Sequencing and variation calling**

To investigate the genetic structure and discover the potential regions under artificial selection in the formation process of indigenous breeds, 20 buffalo individuals from two representative indigenous breeds, GYBS and XCS, in Sichuan Basin were selected for RAD-tags sequencing. After raw data filtering and quality control, a total of 1.06 billion paired-end clean reads were generated, ranging from 44.03 million to 64.49 million, with a mean of 53.10 million reads per individual (Additional file 1). Calling of SNP genotypes was carried out by aligning each individual sequence reads to the buffalo reference genome published previously (Low et al. 2019) with the BWA software (Li and Durbin 2009). After removal of redundancy by using SAMtools (Li et al. 2009), over 97.9% of the generated sequence reads mapped to reference genome ranging from 97.01 to 98.94% for each sample. The average sequence depth was $\sim 11 \times$ within a range of 10.25 to 12.56 per individual and coverage was approximately 30% (27.67–31.29%) of the total length of reference genome with 1 \times coverage depth (Additional file 1).

Single nucleotide polymorphisms (SNPs) detection for each sample was performed by using SAMtools (Li et al. 2009) based on this high-quality sequencing data. After strict quality control and filtration, 2,110,077 SNPs were identified from 20 individuals. Based on the well-assembled reference genome, clean SNPs were classified to chromosomes ($n = 2,092,682$) and un-anchored contigs (17,395). The detailed SNP distribution information on the chromosome of the reference genome was showed in Fig. 1a. The number of SNPs was the most on chromosome 1 (NC_037545.1) and the least on chromosome 24 (NC_037568.1), accounting for 8.21% and 1.67% of the SNPs, respectively. SNP density varied among chromosomes, and chromosome 13 (NC_037557.1) had the highest density with 91.9 SNPs per 100 Kb, and chromosome X (NC_037569.1) had the lowest density with 31.9 SNPs per 100 Kb. The transitions and transversions SNPs were 1,206,600 and 903,477, respectively, resulting in an overall transition/transversion ratio of 1.34 (Fig. 1b). The SNP was further annotated by ANNOVAR software (Wang et al. 2010) (Table 1) to get more detailed information on these variants. Among these variants detected, most SNPs were identified as located in intergenic regions (1,270,286) and intronic regions (786,162). Further, 10,812 SNPs were located upstream and 11,889 downstream within 1 Kb regions of genes. Since their positions are in the transcription start or end, these SNPs probably play a role in transcriptional regulation (Guo et al. 2016). Coding regions exhibited relatively lower diversity levels with
11,737 SNPS identified in exonic regions, including 5002 non-synonymous mutations, 6650 synonymous and 85 exonic-stop gain or loss, resulting in a non-synonymous to synonymous substitution ratio of 0.75.

### Population genetic structure analysis

In order to explore the phylogenetic relationships and genetic differentiation within and between populations, the high-quality SNPs information was used to build a phylogenetic tree by the neighbor-joining methods (bootstrap values = 1000) based on the distance matrix calculated by TreeBest version 1.9.2 software (Fig. 2a). The results showed that two distinct taxa were clustered with the obvious separation of accessions from GYBS and XCS. This result indicated that two breeds have probably undergone different artificial selection during domestication.

PCA is a statistical method that could convert a large number of measurements into a few independent factors and explain the differences of different groups by the divergence of these factors (Reich et al. 2008). To access the genetic separation of two buffalo breeds and discovery of population structure, PCA analysis was also calculated by using the genome wide SNPs data in this study by using GCTA software (Yang et al. 2011) (Fig. 2b). PCA analysis revealed a clear separation of individuals collected from GYBS and XCS by the PC1 and was consistent with the result from phylogenetic tree analysis. In the PCA plot, individuals from XCS exhibit a wider range of distribution than those from GYBS in PC2, suggesting more genetic diversity among these samples.

### Screening for selective regions

To investigate the potential artificial selection between two indigenous buffalo breeds during the domestication,
we focused on variation in genes or genomic regions that were possibly under selection in buffalo breeds. A population genetic methodology of fixation (Fst) and genetic diversity (θπ) was selected and calculated for each population by using a 50 kb sliding window with a 10 kb step size. As described previously (Wang et al. 2021) that the genomic regions undergone positive selection invariably exhibit specific patterns of variation as high FST and lower levels of θπ, the selected regions for the GYBS and XCS populations were detected with a significantly high Fst (top 5%) and θπ ratio (log2 (π1) − log2 (π2), top 5%) values. Results suggested a low level of differentiation in most of the genomic areas. Further analysis indicated 995 and 910 protein-coding genes were located in the selected regions in GYBS and XCS, respectively (Fig. 3a and b).

To thoroughly understand domestication processes and selective site, GO and KEGG enrichment analysis were conducted after functional annotation (Fig. 3c–f, Additional file 2). GO enrichment analysis showed 229 and 160 GO terms were significantly enriched for GYBS and XCS, respectively (Additional file 2). As to the GYBS (Additional file 2a), genes involved in protein processing were significantly represented in the biological process, as the 'protein processing' and 'protein maturation' were the most significantly enriched terms in the biological process, which probably suggesting the protein processing were significantly selected in this breed. It is noteworthy that many biotic and abiotic stimulus-related GO terms were also exhibited strong positive selection signal, including 'response to external stimulus', 'detection of temperature stimulus' and so on. This result probably indicates that GYBS were confronted with a complex living environment and underwent relatively strong environmental pressure, including biotic and abiotic originated during the domestication. Moreover, the neuron development and mitochondria-related functions were also highly presented with several GO terms significantly enriched. In the cellular component, membrane-related functional terms were mostly enriched, including 'membrane', 'membrane part' and 'integral component of membrane'. For the molecular function classification, 'antioxidant activity' and 'adenosine deaminase activity' were the most enriched GO terms, which probably also indicated the capacity for immunity was selected since both the 'antioxidant activity' and 'adenosine deaminase activity' were closely related to organism immune (Li et al. 2015; Hitoglou et al. 2001). The top thirty significant GO terms were listed in Fig. 3c. KEGG is an important resource for understanding higher-order functional meanings (Kanehisa et al. 2004). To gain the major pathway under selection, KEGG enrichment analysis were also executed. The result showed 'linoleic acid metabolism' pathway, which was known for its function in fat content regulation, meat quality improvement (Li et al. 2012), was significantly selected (Fig. 3d).

The same analyses were conducted for XCS simultaneously. Enriched GO terms of genes positively selected in XCS (Fig. 3e and Additional file 2b) showed relatively obvious differences with those in GYBS. In biological process, 'assembly of actomyosin apparatus involved in cytokinesis' and 'actomyosin contractile ring assembly' were the two most significant categories. Several actomyosin and cellular processes GO terms, including 'cytokinetic process', 'regulation of cytokinetic process', and et al., were also enriched in addition and accounted for the vast majority proportion of total enriched functional terms in biological process. This result is obviously different from that in GYBS and probably suggests the growth and development progress received significantly positive selection. It’s worth noting that a ‘NFkB’ related GO term, which has been confirmed to play a vital role in cell proliferation and is related to many human diseases (Chandrakesan et al. 2013), exhibited a strong positive selection signal. Additionally, nervous and perception system, membrane process and energy related biological process were also significantly presented. For cellular component, 'Roundabout signaling pathway', followed by 'ciliary basal body' and 'membrane part' were the top 3 significant functional terms. Besides membrane-related function terms, ciliary-related components were also enriched, which perhaps suggests cell morphogenesis was selected during breed formation. As to molecular function, 'ErbB-2 class receptor binding' was the most significant term, followed by 'phospholipid scramblase activity' and 'guanyl-nucleotide exchange factor activity'. Since ERBB receptors are thought to mediate signals for cell proliferation, differentiation, migration, or apoptosis (Tvorogov and Carpenter 2004), this result further confirmed the growth and development progress was undergone strong positive selection. KEGG analysis showed ‘Phenylalanine, tyrosine and tryptophan biosynthesis’ pathway was significantly enriched (Fig. 3f). As reported previously, this pathway appeared closely related to brain development (Fernstrom 2013; Parker and Brochtie 2011), presumably, the nervous system development was probably an underlying important domestication trait.

Discussion

Sichuan Basin is located in southwestern China and is surrounded by Tibet Plateau, Yunnan-Kweichow Plateau, Daba mountain, and Huaying mountain. Since the wide range of elevation ladder and complex terrain, Sichuan harbors substantial diversity in geography and climate. Meanwhile, it is also a historic agricultural area. During the long-term practice of agriculture, Sichuan Basin spawns diverse local
The Most Enriched GO Terms (XCS−vs−GYBS.0.05.GYBS)

- Linoleic acid metabolism
- Phenylalanine, tyrosine and tryptophan biosynthesis

Statistics of Pathway Enrichment XCS−vs−GYBS.0.05.GYBS

The Most Enriched GO Terms (GYBS−vs−XCS.0.05.XCS)

- Assembly of actin-myosin contractile ring
- Cyclic nucleotide−gated ion channel activity
- Nervous system development

Statistics of Pathway Enrichment GYBS−vs−XCS.0.05.XCS
breeds of animals to match complex geographical environments and applications. Water buffalo is a representative one of the numerous domestic draught animals in this area and always employed as an important labor force in the rice planting process, as well as a provider of milk, meat, horns, and skin (Yue et al. 2013; Nanda and Nakao 2003). It is recognized that close inbreeding frequently occurred during breed formation, which would result in the fixation of some key characteristics (Gorssen et al. 2020). During the long process of local breeding process, buffaloes have undergone intense selection pressures for this development of interesting traits and gradually formed self-specific characteristics distinct among breeds. Thus, these distinctive buffalo breeds undoubtedly fixed diverse precious gene resources for desirable performance traits and favors, a great potential for animal breeding.

In present study, to clarify the genetic population structure and selective signals of buffalo in the Sichuan Basin, we collected 20 individuals representing two different indigenous buffalo breeds with ten samples of each. We conducted a RAD-Seq to analyze population structure and positive selection region during breeding process. A total of 2,110,077 SNP loci were detected in all samples. As to our knowledge, this is the first time to provide such a huge SNP data set in indigenous buffalo breeds in Sichuan and further illustrate the abundant diversity of indigenous buffalo breeds. Additionally, the population structure was determined with the phylogenetic tree and principal component analysis. The results showed that the GYBS and XCS were clustered in two different branches on the phylogenetic tree map. The PCA map further supported the result with GYBS and XCS separate obviously by PC1, suggesting a significant genetic differentiation between two populations. It is noteworthy that the XCS exhibits a much wider distribution on the PCA plot, which probably indicates a higher genetic diversity of XCS. Although the limited number of samples and populations were analyzed in the present study, all the results implied a distinct genetic differentiation among the different populations, and these indigenous breeds are likely to harbor huge genetic varieties resources, which deserve our further exploration and discussion.

To detect positive selection signatures in different buffalo breeds, we searched the genome of the buffalo for regions with a significantly high Fst (top 5%) and $\theta_\pi$ ratio ($\log_2 (\theta_\pi)$, top 5%) values. As a result, 995 and 910 protein-coding genes in GYBS and XCS were identified as located in the selected regions. Functional annotation and enrichment results showed these two breeds underwent distinctly different evolutionary scenarios under different selections. As to GYBS, functional enrichment analysis revealed that many candidate genes positively selected are related to protein processing, meat quality improvement, and stress response (Additional file 2a), which likely reflect the main objective of selection for this breed is to improve the meat quality and increase the adaption capability for complex living environment and disease resistance. While in the XCS (Additional file 2b), these genes were significantly enriched in cell proliferation and cell morphogenesis, suggesting the primary selection pressure of XCS was presumably derived from the speed requirement of growth and development. Additionally, energy metabolism and neuron development appeared to be selected in both breeds.

**Conclusions**

This study collected two indigenous buffalo breeds in Sichuan Basin and employed RADseq to explore selection signatures across genome. Based on the 2,110,077 high-quality SNPs identified from two populations, the genetic structure was analyzed. Result showed significant genetic differentiation between two breeds. Analysis of selective regions identified 995 and 910 positive selection genes in GYBS and XCS, respectively. Functional annotation and enrichment analysis showed these two breeds seemingly undergone different selective pressure with that the GYBS were mainly selected in functions closely related to meat quality, complex living environment adaption, and disease resistance, while selection pressure in XCS possibly derived from speed requirement of organism growth and development during breed formation. Overall, the majority of genome-wide SNP markers among two breeds of Sichuan indigenous buffaloes, as well as the considerable number of genes for traits of interest reported here, will serve as a valuable resource to accelerate genetic research and precision breeding programs in buffaloes.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s12686-022-01275-2.

**Acknowledgements** We thank Liangshan Autonomous Prefecture, Yibin Animal Husbandry Technology Promotion Center, and Xuzhou District Animal Husbandry Technology Extension Station for performing many hours of fieldwork and providing samples.

**Author contributions** WW wrote the manuscript and performed analyses together with DF, YS, and FH. WW conceived the study, DF, JG, XD and AY obtained funds for the project. MJ and RE performed laboratory work.
Funding This study was funded by the Sichuan Province Science and Technology Planning Project (2021YFYZ0001) and the Sichuan Beef Cattle Innovation Team of the Modern agricultural technology system (SCCXTD-2021-13).

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request and can be shared after agreeing on their use with Sichuan Animal Science Academy.

Code availability All software used in this study is outlined throughout the Material and Methods section. Custom scripts are provided in the supplementary material.

Declarations

Conflict of interest The authors declare they do not have any conflict of interest.

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

References

Anello M, Daverio MS, Romero SR, Rigalt F, Silbestro MB, Vidal-Rioja L et al (2016) Genetic diversity and conservation status of managed vicuña (Vicugna vicugna) populations in Argentina. Genetica 144:85–97
Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7
Chandrakasan P, Jakkula LUMR, Ahmed I, Roy B, Anant S, Umar S (2013) Differential effects of β-catenin and NF-kB interplay in the regulation of cell proliferation, inflammation and tumorigenesis in response to bacterial infection. PLoS ONE 8:e79432–e79432
Fernstrom JD (2013) Large neutral amino acids: dietary effects on brain neurotransmitter and function. Amino Acids 45:419–430
Gene T & Consortium O (2000) Gene ontology: tool for the. Gene Expr. https://doi.org/10.1038/75556
Guo X, Fang Q, Ma C, Zhou B, Wan Y, Jiang R (2016) Whole-genome resequencing of Xishuangbanna fighting chicken to identify signatures of selection. Genet Sel Evol 48:1–6
Gorssen W, Meyermans R, Buys N, Janssens S (2020) SNP genotypes reveal breed structure, selection signatures and highly inbred regions in Piétrain pigs. Anim Genet 51:32–42
Hitagnostou S, Zatzistilianou M, Gougoustamou D, Athanassiadou F, Kotisis A, Catrini D (2001) Adenosine deaminase activity and its isoenzyme pattern in patients with juvenile rheumatoid arthritis and systemic lupus erythematosus. Clin Rheumatol 20:411–416
Junge C, Donnellan SC, Huveneers C, Bradshaw CJA, Simon A, Drew M et al (2019) Comparative population genomics confirms little population structure in two commercially targeted carcharhinid sharks. Mar Biol 166:1–15
Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30
Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:277–280
Liang D, Zhao P, Si J, Fang L, Pairo-Castineira E, Hu X et al (2021) Genomic analysis revealed a convergent evolution of LINE-1 in coat color: a case study in water buffaloes (Bubalus bubalis). Mol Biol Evol 38:1122–1136
Liu C, Chen H, Ren Z, Zhang C, Yang X (2019) Population genetic analysis of the domestic Bactrian camel in China by RAD-seq. Ecol Evol 9:11232–11242
Low WY, Tearle R, Bickhart DM, Rosen BD, Kingan SB, Swale T et al (2019) Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. Nat Commun 10:1–11
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
Li S, Tan HY, Wang N, Zhang JI, Lao L, Wong CW et al (2015) The role of oxidative stress and antioxidants in liver diseases. Int J Mol Sci 16:26087–26124
Li RG, Wang XP, Wang CY, Ma MW, Li FC (2012) Performance, meat quality and fatty acid metabolism response of growing meat rabbits to dietary linoleic acid. Asian-Austral J Anim Sci 25:1169–1177
Nanda AS, Nakao T (2003) Role of buffalo in the socioeconomic development of rural Asia: current status and future prospects. Anim Sci 74:443–455
Orita R, Nagano Y, Kawamura Y, Kimura K, Kobayashi G (2021) Genetic diversity and population structure of razor clam Sinonovacula constricta in Ariake Bay, Japan, revealed using RAD-Seq SNP markers. Sci Rep. https://doi.org/10.1038/s41598-021-87395-5
Parker G, Brodtie H (2011) Mood effects of the amino acids tryptophan and tyrosine: “Food for Thought” III Parker and Brodtie Mood effects of amino acids. Acta Psychiatr Scand 124:417–426
Reich D, Price AL, Patterson N (2008) Principal component analysis of genetic data. Nat Genet 40:491–492
Tvrogov D, Carpenter G (2004) Epidermal growth factor receptor family. Encycl Biol Chem 2:51–55
Wang GD, Zhai W, Yang HC, Fan RX, Cao X, Zhong L et al (2013) The genomics of selection in dogs and the parallel evolution between dogs and humans. Nat Commun. https://doi.org/10.1038/ncomms2814
Wang W, Gan J, Fang D, Tang H, Wang H, Yi J et al (2018) Genome-wide SNP discovery and evaluation of genetic diversity among six Chinese indigenous cattle breeds in Sichuan. PLoS ONE 13:1–10
Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:1–7
Wang X, Zhang Z, Mammola S, Ho ALFC, Zhang Y, Qin G et al (2021) Exploring ecological specialization in pipefish using genomic, morphometric and ecological evidence. Divers Distrib 27:1393–1406
Yue XP, Li R, Xie WM, Xu P, Chang TC, Liu L et al (2013) Phylogeography and domestication of Chinese Swamp Buffalo. PLoS ONE 8:1–13
Yang A, Chen D, Wang P, Fu Y, Zhang Z (2017) Characterization of novel microsatellite markers of the Emei Shan Lioicichla using restriction site-associated DNA sequencing. Avian Res 8:1–8
Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82
Zhang Y, Colli L, Barker JSF (2020) Asian water buffalo: domestication, history and genetics. Anim Genet 51:177–191
Zhang Y, Sun D, Yu Y, Zhang Y (2007) Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers. Anim Genet 38:569–575

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