

Article

Genome-Wide Characterization and Comparative Analyses of Simple Sequence Repeats among Four Miniature Pig Breeds

Hongyang Wang ^{1,2,†}, Yang Fu ^{3,†}, Peng Gu ⁴, Yingying Zhang ^{1,2}, Weilong Tu ^{1,2}, Zhe Chao ⁵, Huali Wu ^{1,2}, Jianguo Cao ^{1,2}, Xiang Zhou ⁶, Bang Liu ⁶, Jennifer J. Michal ⁷, Chun Fan ⁸ and Yongsong Tan ^{1,2,*}

- ¹ Institute of Animal Husbandry and Veterinary Science, Shanghai Academy of Agricultural Sciences, Shanghai 201106, China; wanghongyang@saas.sh.cn (H.W.); zhangyingying@saas.sh.cn (Y.Z.); tuweilong@saas.sh.cn (W.T.); wuhuali@saas.sh.cn (H.W.); caojianguo@saas.sh.cn (J.C.)
 - ² Shanghai Engineering Research Center of Breeding Pig, Shanghai 201302, China
 - ³ Research Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China; fuyang@saas.sh.cn
 - ⁴ Institute of Comparative Medicine & Laboratory Animal Management Center, Southern Medical University, Guangzhou 510515, China; kupper@smu.edu.cn
 - ⁵ Institute of Animal Science and Veterinary Medicine, Hainan Academy of Agricultural Sciences, Haikou 571100, China; zhanggavin@webmail.hzau.edu.cn
 - ⁶ Key Laboratory of Agricultural Animal Genetics, Breeding, and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China; zhouxian@mail.hzau.edu.cn (X.Z.); liubang@mail.hzau.edu.cn (B.L.)
 - ⁷ Department of Animal Sciences, Washington State University, Pullman, WA 99164, USA; jennifer_michal@wsu.edu
 - ⁸ Shanghai Laboratory Animal Research Center, Shanghai 201203, China; fanchun@slarc.org.cn
- * Correspondence: kyc1@saas.sh.cn; Tel.: +86-021-34505325
- † These authors contribute equally to this work.

Received: 10 August 2020; Accepted: 28 September 2020; Published: 2 October 2020



Simple Summary: Simple sequence repeats (SSRs) are present at high densities in regulatory elements, suggesting that they may affect gene function and phenotypic traits. Therefore, SSRs can be exploited in marker-assisted selection. In addition, they can be widely used as molecular markers to study genetic diversity, population structure, and evolution. While SSRs have been widely studied in many mammalian species, very little research has focused on genome-wide SSRs of miniature pigs, a small but special group of pigs that express the dwarf phenotype. Based on the SSR-enriched library building and sequencing, about 30,000 novel polymorphic SSRs for four miniature pig breeds were mapped to the Duroc pig reference genome. The four miniature pig breeds had different numbers and types of SSRs and distributions of repeat units. There were 2518 polymorphic SSRs in the intron or exon regions that were common to all four breeds and functional analyses revealed 17 genes that were associated with body size and other genes that were associated with growth and development. In conclusion, the SSRs detected in the miniature pigs in this study may provide useful genetic markers for the selection of farm animals and the polymorphic SSRs provide valuable insights into the determination of mature body size, as well as the immunity, growth and development of animals.

Abstract: Simple sequence repeats (SSRs) are commonly used as molecular markers in research on genetic diversity and discrimination among taxa or breeds because polymorphisms in these regions contribute to gene function and phenotypically important traits. In this study, we investigated genome-wide characteristics, repeat units, and polymorphisms of SSRs using sequencing data from SSR-enriched libraries created from Wuzhishan (WZS), Bama (BM), inbred Luchuan (LC) and Zangxiang (ZX) miniature pig breeds. The numbers and types of SSRs, distributions of repeat units

and polymorphic SSRs varied among the four breeds. Compared to the Duroc pig reference genome, 2518 polymorphic SSRs were unique and common to all four breeds and functional annotation revealed that they may affect the coding and regulatory regions of genes. Several examples, such as *FGF23*, *MYF6*, *IGF1R*, and *LEPROT*, are associated with growth and development in pigs. Three of the polymorphic SSRs were selected to confirm the polymorphism and the corresponding alleles through fluorescence polymerase chain reaction (PCR) and capillary electrophoresis. Together, this study provides useful insights into the discovery, characteristics and distribution of SSRs in four pig breeds. The polymorphic SSRs, especially those common and unique to all four pig breeds, might affect associated genes and play important roles in growth and development.

Keywords: SSR; repeat unit; length polymorphism; miniature pig; molecular marker

1. Introduction

Simple sequence repeats (SSRs), also known as microsatellites or short tandem repeats (STRs), consist of 2 to 6 base-pair motifs repeated several times in tandem. As a consequence of their wide distribution and high mutation rate in eukaryotic genomes [1], SSRs have been used in genetic diversity and population structure studies [2–5], for discrimination among species or breeds [6,7], in marker-assisted selection [8–10] and in evolution analysis [11]. In humans, SSRs were predicted to be bound by protein-coding transcripts, long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) and affect competing endogenous RNA crosstalk [12]. Besides being an important category of regulatory elements, polymorphic SSRs could quantitatively regulate the transcription of tissue-specific genes in the development of the frog embryo [13]. Another study showed that polymorphic SSRs play an important role in shaping splicing regulatory elements and lead to alternative splicing events in different stress environments [14]. Over the past decade, an increasing number of studies on SSR discovery and functional analysis have been conducted, providing evidence for the importance of SSRs in gene function and complex traits [15].

Considering their widely functional role, SSRs have been discovered in various taxonomies, most of which discoveries were based on reference genome scanning. However, a poorly assembled genome leads to imperfect SSRs with inaccurate repeat units or repeat number, leading to the limited use of SSRs. For that reason, a high-throughput SSR isolation method based on SSR-enriched library building and next-generation sequencing (NGS) was developed. The major probes designed to enrich the SSR sequences were validated on 13 species, resulting in the acquisition of high-quality genetic markers [16]. Until now, the method has been utilized to isolate SSRs in humans, plants, fungi, invertebrates, and birds. Although SSRs started to be used as markers for breeding projects and genetic diversity studies in pigs in the mid-1990s [17], accurate and genome-wide SSRs are lacking. To our knowledge, only one study isolated polymorphic SSRs from pooled pig breeds based on a porcine reference genome and genome resequencing data [18].

The miniature pig is considered the best model organism for the study of growth and development of animals with small body size. For instance, Wuzhishan pigs (WZS), the most famous indigenous miniature pig, are characterized by their small adult size with mature body weights of only 30 kg [19]. Some genetic mechanisms associated with poor body growth and immunity-related genes were discovered based on transcriptome analyses of liver and muscle tissues of Jeju Native and miniature pigs [20]. The miniature pig shares many anatomical and physiological features with humans and has been used as an animal model in biomedical research, resulting in great contributions to the medical advances of human beings [21]. Recently, studies on chronic renal failure [22], progressive hearing loss [23], and diabetes [24] were conducted on Bama pigs. Of the Chinese indigenous miniature pig breeds, the genome of the Wuzhishan pig was the first to be assembled at the scaffold level [19]. This breed has been widely used for research on metabolic disease [25], diphodont and craniofacial

development [26], mesenchymal stem cells [27] and corneal xenotransplantation [28]. Even so, accurate sequences of genome wide SSRs of miniature pig breeds are not currently available.

This study was aimed to discover genomic SSRs of the Wuzhishan, Bama (BM), inbred Luchuan (LC) and Zangxiang (ZX) miniature pig breeds based on an SSR-enrichment library. The distribution and functional annotation of SSRs were also compared among the four pig breeds. All the results provided molecular markers for conservation and utilization of germplasm resources of the miniature pig.

2. Materials and Methods

2.1. Ethics Statement

These experiments were carried out in accordance with local guidelines for the care of laboratory animals and were approved by the institution's ethics committee for research using laboratory animals, approval code: SN-XS-20190143.

2.2. Animals

Fifteen male pigs with distant relationships from each of four miniature pig breeds ($n = 60$) were involved in this study. Wuzhishan (WZS) pigs were obtained from Hainan Academy of Agricultural Sciences and Zangxiang (ZX) pigs were obtained from Southern Medical University, while Bama (BM) and inbred Luchuan (LC) pigs were obtained from Shanghai Academy of Agricultural Sciences. The inbred Luchuan pigs have been inbred since the 2000s. The smallest boars and gilts, with shorter body lengths than non-inbred Luchuan pigs, were selected for breeding over the past 20 years. The four pig breeds have no relationship with each other and are mainly raised in extensive or semi-extensive farming systems. Ear tissues were collected from the 60 male piglets when they were weaned at 50 days and weighed 2.5~4.0 kg. Tissues were placed in tubes containing 75% ethanol and taken back to the laboratory where they were stored at $-80\text{ }^{\circ}\text{C}$ for subsequent DNA extraction.

2.3. Dna Extraction and Sequencing Based on Simple Sequence Repeat (SSR)-Enriched Library

Genomic DNA was isolated from all samples using a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The quantity and quality of the extracted DNA were assessed using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and agarose gel electrophoresis (1%), respectively. For each breed, equal amounts of DNA from the 15 pigs were pooled and used for the SSR-enriched library preparation. The protocol of the SSR-enriched library building was similar to that of a previous study [29]. In short, the pooled genomic DNA was digested to small fragments and a standard genomic library was built with a 400-bp insert size. Next, eight biotin-labeled oligonucleotides were used to hybridize SSR repeat sequences in the genomic library and the resulting four libraries from the four pig breeds were sequenced on an Illumina MiSeq platform at Shanghai Personal Biotechnology Co., Ltd. (Shanghai, China). The eight probes which have been described in a previous study were designed to enrich sequences with the following motifs: $(AG)_{10}$, $(AC)_{10}$, $(AAC)_8$, $(ACG)_8$, $(AAG)_8$, $(AGG)_8$, $(ACAT)_6$ and $(ATCT)_6$ [16]. The raw sequence data in fastq format is deposited in the Sequence Read Archive (SRA) and the reviewer link is <https://dataview.ncbi.nlm.nih.gov/object/PRJNA628105?reviewer=5mfadppoucgn41cpgj2idujl86>.

2.4. Data Treatment and SSRs Scanning

For each library, paired-end data (2×250 bp) were generated from the sequencing platform in fastq format. AdapterRemoval software (v2.1.7) [30] was used to remove adapters and low-quality reads. First, the Q value of the base pair (bp) was scanned with 5 bp sliding window and 1 bp sliding step for all reads. In a window, if the average Q value was less than 20 or the Q value of the last base pair was less than 2, the base pair next to the last and the previous base pair was kept. Second, paired reads were removed if the length of one of the pairs was less than 50 bp. After quality filtering, FLASH software (V1.2.11) [31] was utilized to combine read1 and read2 from each of the paired reads

and used to generate longer sequences with the following criteria: (1) Min overlap: 100; (2) Max mismatch density: 0.1; (3) Allow “outie” pairs: false; (4) Cap mismatch quals: false. SSRs were scanned and counted for the pig reference genome (*Sscrofa11.1*, GenBank: GCA_000003025.6) and the combined sequences from each of the four datasets using the MIncroSAtellite (MISA) script [32,33]. In this step, the most important parameter is minimum repeat, which was defined as 6, 5, 5, 5 and 5 for di-, tri-, tetra-, penta-, and hexanucleotides, respectively [34]. Distributions of repeat units for the pig reference genome and four pig breeds were drawn using R software (3.6.1).

2.5. Analysis of Polymorphic SSRs and Functional Annotation

Polymorphic SSRs depended on SSR length polymorphism (SSLP) and were discovered from SSR-containing sequences that were obtained from the combined sequences. First of all, an in-house Perl script was utilized to identify and mask SSRs with “R” in the SSR-containing sequences. In this step, if the length of the flanking sequence of SSR was less than 20 bp, the sequence would be removed for the reason that it could not be accurately used for similarity analysis. After that, clustering was performed based on similarity of the flanking sequence using CD-HIT software [35]. Similarity and coverage were 90% and 70%, respectively. Other parameters were defined as 1 for gap and 0 for gep-ext. For the clustering results, another in-house Perl script was used to identify SSLP. If only one type of length existed in a cluster, the corresponding SSLP would be defined as 1. If the length of the SSR had two types, the SSLP of the SSR would be defined as 2, and so on. Finally, we obtained the polymorphic SSRs and SSLP for each type of SSR.

SSRs with an SSLP more than 1 were selected and alignment was performed based on the flanking sequences of SSRs in the corresponding cluster. The flanking sequences longer than 20 bp were extracted and mapped to the reference genome (*Sscrofa11.1*) using Burrows–Wheeler Alignment software [36]. According to chromosome coordinates of the mapped SSRs, overlapping was analyzed to find common and specific SSRs among the four pig breeds using the UpSetR package [37]. SSRs annotation and associated functional genes were discovered using annotated files from the Ensembl database (*Sus_scrofa.Sscrofa11.1.97*). Functional enrichment analysis was performed using the clusterProfiler [38] package and corresponding database (*org.Ss.eg.db*, V3.10.0).

2.6. Designing Primers and Experimental Validation

Based on the flanking sequences of SSRs with SSLP more than 1, primer pairs were designed using Primer3 (v2.3.6) [39]. Three primer pairs were chosen to detect alleles of the SSR in all 60 pigs at high resolution using fluorescence polymerase chain reaction (PCR) and capillary electrophoresis. First of all, we checked specific amplification and length of PCR products using normal PCR followed by agarose gel electrophoresis. After that, forward primers were fluorescence-labeled with HEX at the 5′ end as described in a previous study [40]. Fluorescence PCR were performed on ABI-2720 thermal cycle (Applied Biosystems, Foster City, CA, USA) and each 25 μL reaction contained 1 μL of each primer (10 μM), 1 μL of template DNA, 2 μL 10 × buffer, 0.5 μL dNTP, 0.5 μL Taq enzyme and 14 μL ddH₂O. Cycling conditions were 95 °C for 4 min, followed by 10 cycles with 60 °C for 30 s, 72 °C for 30 s and 95 °C for 30 s, followed by 25 cycles with 52 °C for 30 s, 72 °C for 30 s and 72 °C for 7 min. The final amplicons were subjected to capillary electrophoresis (ABI-3730XL, Applied Biosystems, Foster City, CA, USA) and the output data was analyzed by GeneMapper software (V2.2.0).

3. Results

3.1. Overview of SSRs and Repeat Units in the Pig Reference Genome

Using MISA software and the parameters described above, we scanned the pig reference genome (*Sscrofa11.1*) to discover the profiles of the SSRs and repeat units in the pig. We discovered a total of 471,287 SSRs, including 290,373 dinucleotide repeats (Di-SSRs), 82,517 trinucleotide repeats (Tri-SSRs), 83,936 tetranucleotide repeats (Tetra-SSRs), 11,545 pentanucleotide repeats (Penta-SSRs)

and 2916 hexanucleotide repeats (Hexa-SSRs). All the SSRs occupied 0.4% of the reference genome (Table S1). The size and proportion of the repeat units were also discovered in different types of SSR. Among Di-SSRs, AC/GT (52.2%) was the most common repeat unit and CG/CG (0.5%) was the least abundant repeat unit. For Tri-SSRs, the most abundant type was AAC/GTT (50.2%), followed by 22.7% and 9.1% for AAT/ATT and AAG/CTT, respectively. In Tetra-SSRs, AAAT/ATTT (34.6%), AAAG/CTTT (21.1%) and AAAC/GTTT (15.0%) were the three major types and occupied 70.7% of all the tetranucleotide repeat units. A/T-rich motifs were the main types in Penta-SSRs, which was similar to the tetranucleotide repeats. However, the most abundant repeat unit for Hexa-SSRs was AACCT/AGGGT (42.2%), followed by ACAGCC/CTGTGG (21.4%) (Figure 1).

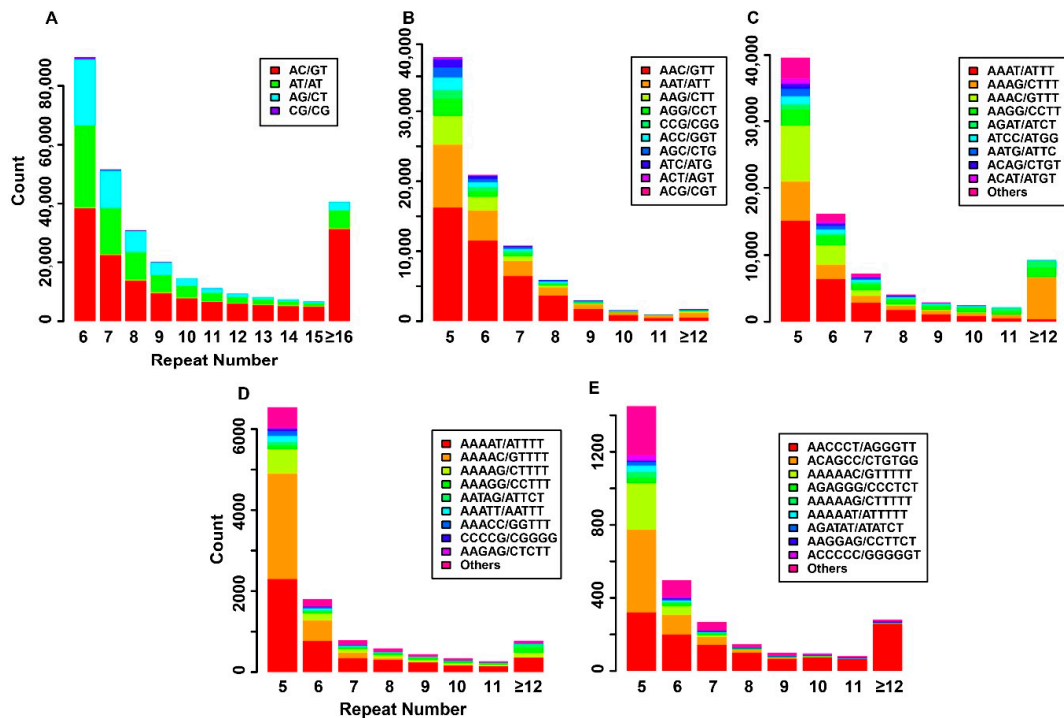


Figure 1. Distributions of repeat units for different simple sequence repeat (SSR) types in the pig reference genome. (A–E) The distributions of repeat units in dinucleotide repeat (Di-SSR), trinucleotide repeat (Tri-SSR), tetranucleotide repeat (Tetra-SSR), pentanucleotide repeat (Penta-SSR) and hexanucleotide repeat (Hexa-SSR). The X axis represents repeat number, Y axis represents the count of the repeat unit corresponding to different colors.

3.2. SSR Discovery from Four Miniature Pig Breeds

A total of 60.6 million raw reads were obtained from four datasets. 54.7 million (90.3%) reads with an average length of 232 bp were left after quality filtering using AdapterRemoval software (v2.1.7). According to the overlapped and mismatched reads, we combined a total of 47.9 million (87.6%) reads utilizing FLASH software (V1.2.11), and obtained 6.6, 4.3, 3.8 and 9.0 million combined sequences for Wuzhishan (WZS), Bama (BM), inbred Luchuan (LC) and Zangxiang (ZX) pigs, respectively (Table 1). In all four datasets, we found the length of the sequences ranged from 100 bp to 500 bp and sequences with 300 bp in length were the most abundant (Figure S1). The raw SSRs data generated from MISA software is displayed in File S1, which showed that the number of SSRs was greatest in WZS, followed by BM and LC, while ZX had the least number of SSRs. In the four pig breeds, Di-SSRs were far more frequent (75.7%, 70.5%, 75.5% and 51.5% for WZS, BM, LC and ZX, respectively) than other SSR types, followed by Tri- and Tetra-SSRs, which is similar to the proportion of different SSR types in the reference genome as described above (Table 1).

Table 1. Statistics of four datasets and SSRs discovered from four pig breeds.

| Items | WZS ¹ | BM ¹ | LC ¹ | ZX ¹ |
|---|------------------|-----------------|-----------------|-----------------|
| Raw reads | 17,083,436 | 10,335,212 | 10,466,408 | 22,759,784 |
| High quality reads | 15,473,282 | 9,660,700 | 9,255,072 | 20,279,502 |
| Total length of high quality reads (base pair, bp) | 3,549,570,313 | 2,367,772,949 | 2,010,410,274 | 4,756,156,243 |
| Combined sequences | 6,658,339 | 4,369,729 | 3,885,823 | 9,043,156 |
| Total length of combined reads (bp) | 2,103,692,216 | 1,435,035,633 | 1,177,952,198 | 2,948,109,147 |
| Dinucleotide repeats (Di-SSRs) | 4,127,838 | 1,843,882 | 2,216,458 | 951,588 |
| Trinucleotide repeat (Tri-SSRs) | 586,364 | 264,912 | 316,662 | 101,743 |
| Tetranucleotide repeat (Tetra-SSRs) | 234,811 | 125,631 | 134,298 | 122,506 |
| Pentanucleotide repeat (Penta-SSRs) | 33,676 | 23,977 | 18,932 | 46,849 |
| Hexanucleotide repeat (Hexa-SSRs) | 17,693 | 7088 | 9198 | 2382 |

¹ WZS, BM, LC and ZX represent Wuzhishan, Bama, inbred Luchuan and Zangxiang pigs.

3.3. Frequency of Repeat Units

For each type of SSR, the frequency of the repeat units at the position of combined sequences was checked and most of the repeat units were located in proximal sequences in all datasets (Figure S2). Furthermore, the distributions of the number of repeat units were calculated in each type of SSR (Table S2) and a comparison was performed between the four miniature pig breeds and the reference genome. In the four pig breeds, AC/GT and AAC/GTT repeats, for Di- and Tri-SSRs, respectively, were more common than others in the corresponding SSR type, which was similar to trends observed in the reference genome. However, in the four pig breeds, AGAT/ATCT, AATAG/ATTCT and AAGGAG/CCTTCT were the most abundant repeat units for Tetra-, Penta- and Hexa-SSRs, respectively, which were different than the distributions of repeat units in the reference genome.

We selected all the repeat units for Di- and Tri-SSRs and the top 10 repeat units for Tetra-, Penta- and Hexa-SSRs and compared the proportion based distribution models among the four miniature pig breeds (Figure 2). There were no differences in Di- and Penta-SSRs among the four pig breeds, which showed a similar distribution. Special distributions were found in the ZX pig, which showed a high abundance of ACT/AGT in Tri-SSRs and ACAG/CTGT and AAGG/CCTT in Tetra-SSRs. In comparison, the other three pig breeds had similar distribution models. There were extremely diverse distributions of repeat units of Hexa-SSR among the four pig breeds.

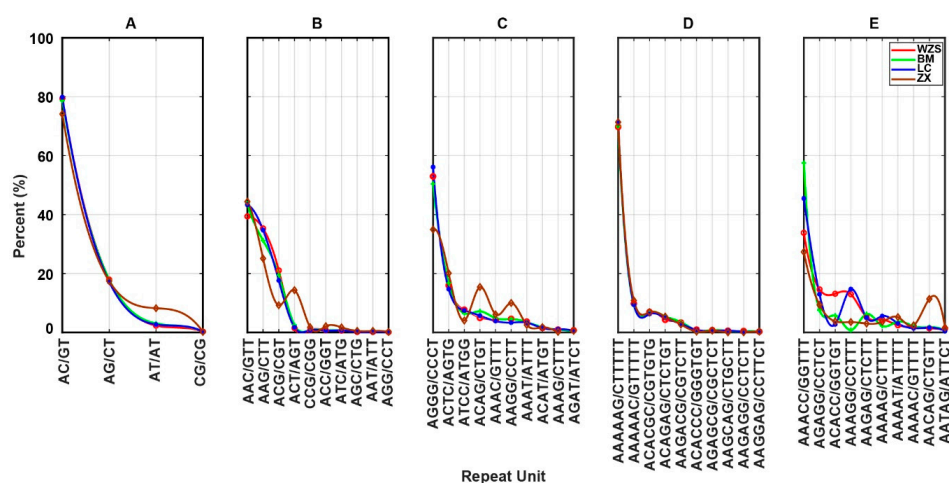


Figure 2. Distribution models of different repeat units among four miniature pig breeds. (A–E) Reading from left to right are the distributions of repeat units in Di-SSR, Tri-SSR, Tetra-SSR, Penta-SSR and Hexa-SSR, respectively. The X axis represents types of repeat unit, Y axis represents percentage of repeat units in corresponding total SSRs. The red, green, blue and brown dots connected by lines represent WZS, BM, LC and ZX, respectively.

3.4. Polymorphic and Functional SSRs in Four Miniature Pig Breeds

We discovered SSR length polymorphisms (SSLPs) in all SSRs examined. A summary of the total clusters and corresponding SSLP is displayed in Table 2 and the details are shown in File S2. We focused on 60,020, 70,886, 63,968 and 42,400 clusters containing SSRs with SSLP more than 1 for the WZS, BM, LC and ZX pig breeds, respectively (Table 2). Among them, 19,957, 14,099, 20,671 and 14,120 clusters containing 26,393, 17,722, 28,387 and 16,886 SSRs, respectively, were mapped to the reference genome. The details of total SSRs with SSLP and mapped clusters are shown in File S3. According to the results of the overlapping analysis among the four pig breeds, 5173, 2802, 5969 and 4463 clusters were specific for the WZS, BM, LC and ZX pig breeds, respectively, and 2518 clusters were common among all four pig breeds (Figure 3).

Table 2. SSR length polymorphisms (SSLPs) and corresponding number of SSRs in four pig breeds.

| Items ¹ | WZS | BM | LC | ZX |
|--------------------------|-----------|-----------|-----------|-----------|
| SSR-containing sequences | 3,804,507 | 1,849,907 | 2,061,668 | 1,448,491 |
| Total clusters | 377,558 | 398,960 | 433,945 | 453,677 |
| SSLP = 1 | 317,538 | 328,074 | 369,977 | 411,277 |
| SSLP = 2 | 37,841 | 44,188 | 41,098 | 30,024 |
| SSLP = 3 | 13,861 | 16,865 | 14,879 | 8542 |
| SSLP = 4 | 5420 | 6341 | 5210 | 2650 |
| SSLP = 5 | 1840 | 2221 | 1779 | 747 |
| SSLP = 6 | 567 | 704 | 525 | 202 |
| SSLP = 7 | 206 | 257 | 182 | 75 |
| SSLP = 8 | 88 | 103 | 101 | 39 |
| SSLP = 9 | 60 | 61 | 50 | 29 |
| SSLP ≥ 10 | 137 | 146 | 144 | 92 |

¹ SSLP means SSR length polymorphism.

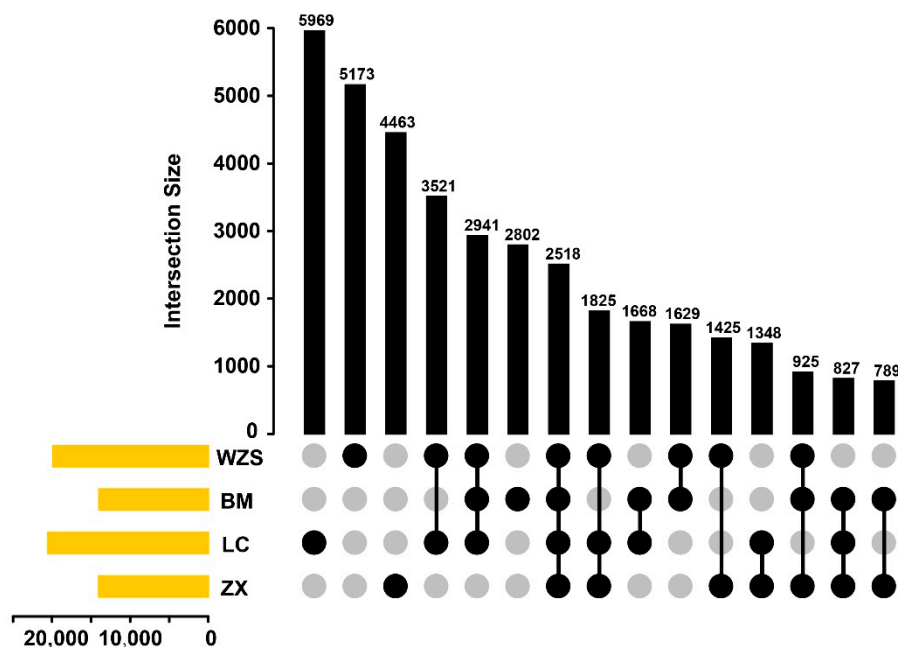


Figure 3. Overlapping analysis of SSRs among four pig breeds. The horizontal frames (in yellow) represent the total number of mapped SSRs with an SSLP more than 1 corresponding to the four pig breeds. The vertical frames (in black) show the SSR number corresponding to the bottom intersection groups.

We merged the 2518 common clusters to annotate and ascertain the universal functions of the SSRs in the four pig breeds. Results showed that most were located in intergenic regions (63.0~65.4%)

and 80, 357 and 436 clusters overlapped with 5' untranslated region (5' UTR), 3' untranslated region (3' UTR) and the coding sequence (CDS), respectively. The results illustrate that polymorphic SSRs were commonly found in noncoding regions and the rest of the SSRs were located in exons, which might affect the function of associated genes. For the SSRs located in the exons, functional enrichment analysis of associated genes was conducted and we found most of the genes were involved in cell–cell signaling, peptide hormone secretion and other biological processes with *p*-value less than 0.01 (Figure 4). Finally, we identified the functional genes corresponding to the polymorphic SSRs with repeat units. Most of these genes were associated with bone remodeling, muscle development and immunity and are described in Table 3 and Table S3.

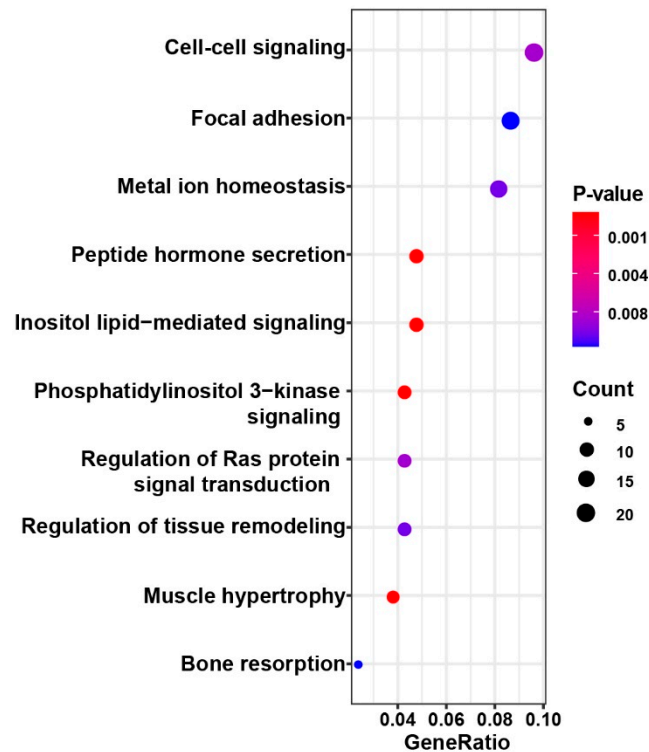


Figure 4. Enrichment analysis of functional genes affected by polymorphic SSRs.

Table 3. Summary information of functional SSRs (SSLP ≥ 2) and associated genes.

| Genomic Coordinate of SSR Flanking Sequence ¹ | Repeat Sequence | Gene | Functional Region ² | Gene Function |
|--|---|---------------|--------------------------------|---|
| 1:137690845_137690967 | (GCG) ₆ , (GCG) ₈ , | <i>IGF1R</i> | 5' UTR | Muscle development [41] |
| 3:23673114_23673194 | (GT) ₁₃ , (CA) ₁₅ | <i>IGSF6</i> | 5' UTR | Inflammatory disease [42] |
| 3:73527603_73527759 | (AC) ₉ , (GT) ₁₃ , (GT) ₁₅ , (GT) ₁₇ , (GT) ₁₈ , (GT) ₂₀ | <i>BMP10</i> | 3' UTR | Skeletal development [43] |
| 4:93179209_93179533 | (TG) ₁₅ , (CA) ₁₆ , (CA) ₁₇ , (TG) ₁₈ , (CA) ₂₂ | <i>PEAR1</i> | 3' UTR | Platelet activation [44], myoblast differentiation [45] |
| 4:102696359_102696552 | (CA) ₂₂ , (GT) ₂₃ | <i>SPAG17</i> | CDS | Sperm motility [46] |

Table 3. Cont.

| Genomic Coordinate of SSR Flanking Sequence ¹ | Repeat Sequence | Gene | Functional Region ² | Gene Function |
|--|--|---------|--------------------------------|----------------------------|
| 5:62835807_62836251 | (AC) ₁₂ , (TG) ₁₃ , (AC) ₁₄ , (AC) ₁₆ , (GT) ₁₇ | GDF3 | CDS | BMP inhibiting [47] |
| 5:66028468_66028506 | (GT) ₁₅ , (GT) ₁₆ , (TG) ₁₇ , (GT) ₁₈ | FGF23 | 5' UTR | Bone remodeling [48] |
| 5:100760758_100761162 | (GT) ₁₄ , (AC) ₁₈ | MYF6 | 3' UTR | Muscle development [49] |
| 6:146979743_146980171 | (CA) ₁₈ , (TG) ₁₉ , (CA) ₂₀ , (CA) ₂₂ , (CA) ₂₃ | LEPROT | CDS | Fat content [50] |
| 9:136235771_136236041 | (AC) ₁₇ , (AC) ₁₉ | SPATA48 | CDS | Spermatogenesis [51] |
| 13:111022378_111022649 | (GT) ₁₅ , (GT) ₁₆ | TNFSF10 | 3' UTR | Cell apoptosis [52] |
| 14:88520086_88520470 | (CA) ₁₅ , (GT) ₁₈ , (CA) ₂₁ , (GT) ₂₂ , (TG) ₂₃ , (TG) ₂₄ , (TG) ₂₅ , (TG) ₂₆ , (TG) ₂₈ | GDF10 | 3' UTR | TGFβ superfamily [53] |

¹ genomic coordinate corresponding to chromosome: start_end. ² 5' UTR, 3'UTR and CDS mean 5' untranslated region, 3' untranslated region and coding sequence, respectively.

3.5. Experiment Validation Using Fluorescence Polymerase Chain Reaction (PCR) and Capillary Electrophoresis

Three primers were selected to detect polymorphic SSRs in the 60 pigs to confirm the sequencing results (Table 4). For the first locus, our predicted result showed that five variations located in the region ranged from 272,578,714 bp to 272,578,954 bp of chromosome 1, and the corresponding SSRs consisted of an (AC) repeat unit ranging from 12 to 17 repeats. Capillary electrophoresis analysis of PCR amplicons confirmed that five alleles (except for one rare allele with 226 bp in length) 224, 228, 230, 232 and 234 bp in length existed in all four pig breeds (Figure 5). We verified that six alleles occurred in each of two other loci, which was confirmed with the polymorphic SSRs (Figures S3 and S4). The raw data from capillary electrophoresis and alleles are displayed in Table S4.

Table 4. Information and validation of 3 SSRs in 60 pigs.

| Loci ¹ | Primer Sequences (5'–3') ² | Repeat Sequence ² | Allele Size/bp ² | NO. of Alleles ³ |
|----------------------------------|---|---|-----------------------------|-----------------------------|
| chr1: 272,578,714–272,578,954 | F: TACACACCACGCGTGACCT R: ACAACGGTTGCTGCTTTTTC | (AC) ₁₂ , (AC) ₁₄ , (AC) ₁₅ , (AC) ₁₆ , (AC) ₁₇ | 207–275 | 5 |
| chr11:70,376,652–70,376,765 | F: CCAGCTTCCAGTTTCGTGT R: AGATTGTGAGTGCCGAGATG | (GT) ₁₄ , (GT) ₁₅ , (GT) ₁₆ , (GT) ₁₇ , (GT) ₁₈ , (GT) ₁₉ | 77–149 | 6 |
| chr18:1,858,964–1,859,153 | F: CTCCAAGGTCAAAGCCAAA R: CGACTTGGGGTTTCCTAACA | (AC) ₁₃ , (AC) ₁₄ , (AC) ₁₅ , (AC) ₁₆ , (AC) ₁₇ , (AC) ₁₈ | 154–226 | 6 |

¹ Based on reference genome. ² Based on SSR-containing sequences. ³ Based on the fragment analysis of 60 pigs on an ABI-3730XL.

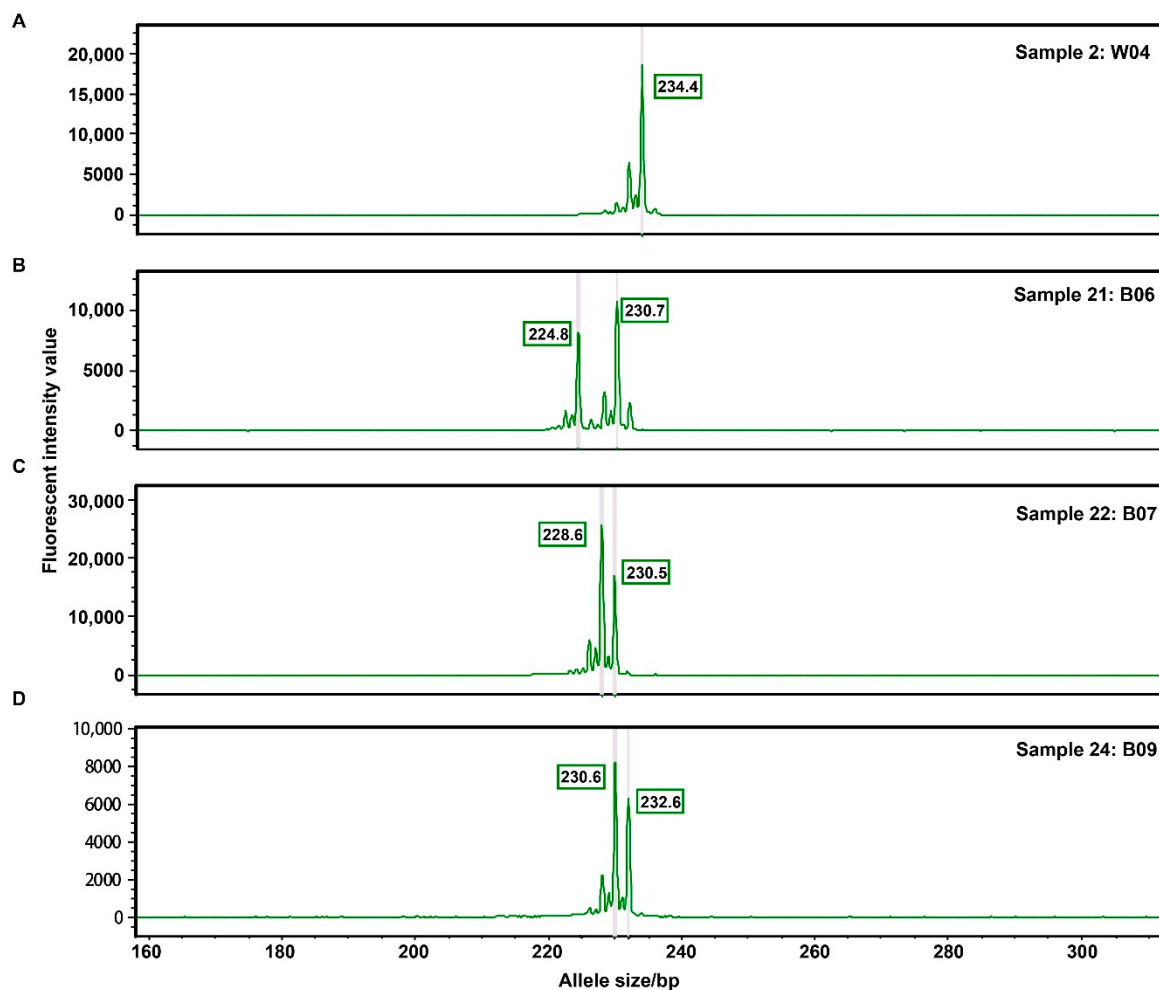


Figure 5. Different alleles of SSR located in chr1:272,578,714-272,578,954. The output data from capillary electrophoresis was analyzed by GeneMapper software (V2.2.0) and generated the allele report for 60 pigs. (A–D) Five alleles 224, 228, 230, 232 and 234 bp in length existed in chr1:272,578,714-272,578,954 bp are exemplified in corresponding pig individuals. The X axis represents allele size, Y axis represents fluorescent intensity for different allele size, one or two peaks on the green line represents homozygosity and heterozygosity, respectively. The sample number W04 means the individual in the Wuzhishan pig breed. The sample number B06, B07 and B09 mean the individuals in Bama pig breed.

4. Discussion

Because of the rapid development of NGS, SSRs have been discovered through scanning the reference genome and genotyping based on a large set of genome resequencing data in pigs. Here, for the first time, an SSR-enriched library was built, sequenced and analyzed to describe characteristics of SSRs in four miniature pig breeds, including different types, distribution of repeat units, polymorphism and function, providing accurate genetic markers for pig breeding and polymorphic SSRs for gene function analysis.

Based on the SSR-enriched library, we obtained an average of 1,225,072 SSRs in the four pig breeds, which is less than the number of SSRs in the MicroSatellite DataBase (MSDB) [54]. In addition, Hexa- and Tetra-SSR were the most abundant types in the MSDB (56%) and another previous study (31.3%) [18], respectively. However, Di-SSRs were far more frequent than other SSR types and similar trends were found for the reference genome in our study. The difference between our study and previous results is probably because of the minimum repeat size used for the SSR scanning. The most commonly used methods for SSR scanning contain MISA, Tandem Repeats Finder [55] and other

custom scripts [56] based on Python or Perl, which are based on similar principles in terms of minimum repeat size. The two studies defined minimum repeats as 6, 4, 3, 3, 3 (or 2) for di-, tri-, tetra-, penta-, and hexanucleotides, respectively. In the present, the minimum repeat size was set to 5 for repeat units longer than 2, the same as other studies [34,57–59], which led to a smaller number of Hexa- and Tetra-SSR. Moreover, previous studies report that mononucleotide repeats are most frequent in eukaryotes, followed by dinucleotide repeats, while trinucleotide repeats are more abundant in prokaryotic genomes [60–62].

Consistent with previous results [18], AC/GT and AAC/GTT were the most abundant repeat units in the pig for Di- and Tri-SSR, respectively. In contrast, GC-containing SSR, such as CG/GC and ACG/CGT, accounted for a small percentage in the reference genome and the four pig breeds, which is similar to reports in other species [63–66]. The bias against GC sequences in the process of library building and sequencing might explain why the GC-SSRs were relatively rare, however, eight probes which contained (ACG)₈ and (AGG)₈ were used to hybridize the GC-containing SSR in this study and should have ensured comprehensive genome-wide SSR enrichment. Therefore, GC-containing SSRs are infrequent and have fewer polymorphisms, explaining why the GC enrichment sequence is always associated with functionality [67]. Furthermore, AGAT/ATCT, AATAG/ATTCT and AAGGAG/CCTTCT were the most abundant repeat units in the four pig breeds for Tetra-, Penta- and Hexa-SSRs, respectively, which was different from the reference genome and due to the fact that the repeat units used for enrichment were over-represented in these SSRs, in particular AAG, AGG and ATCT. Nevertheless, all different types of repeat unit were discovered and displayed different distributions among the four pig breeds.

SSR-based genotyping has been used to study genetic diversity and breed identification within pigs, and most of the studies depending on SSR markers were developed in the domestic pig [17]. However, SSRs and primers developed from different species or breeds always lead to most SSRs with no polymorphism of interest. At the genome-wide level, 1,620,469 SSRs were discovered in the pig reference genome (Duroc) and only 16,527 SSRs displayed high polymorphism in a total of 102 pigs, including 8 Chinese domestic pig breeds and 6 commercial pig breeds [18]. In the current study, 60,020, 70,886, 63,968 and 42,400 SSRs with SLP more than 1 were discovered for the WZS, BM, LC and ZX pig breeds, respectively, providing genetic markers for further analysis. In addition, frequency analysis of repeat units showed different distributions for Hexa-SSRs among the four pig breeds, and specific distributions of Tri- and Tetra-SSRs in the ZX pigs. We speculated that distribution analysis of repeat units combined with validation of SSR polymorphism on a population scale might accurately discriminate among pig breeds.

Body size is one of the most important traits for the research of growth and development and improving production in farm animals. Based on the wide functions of SSRs in genes and traits described above, we examined polymorphic SSRs in genes associated with body size in four miniature pig breeds. Interestingly, about 17 genes involved in body size are affected by polymorphic SSRs. In humans, mutations in *IGF1*, *SHOX*, *GHRHR*, *ZBTB38* and *PIT1* genes can explain part of height variation. We found two variations of 6 and 8 repeats of the (GCG) repeat unit that affect the 5' UTR of *IGF1* gene. Polymorphic SSRs were also discovered in the introns of *ZBTB7C*, *ZBTB16* and *ZBTB20*, which belong to the Krueppel C2H2-type zinc-finger protein family. The coding sequence of the *LEPROT* gene is affected by polymorphic SSRs with (CA) repeat units ranging from 18 to 23 in our findings. A previous study confirmed that the *LEPROT* gene was related to the fat content of Duroc pigs [50] and had a role in the leptin receptor which was related to the reduction of body size in domestic fowl [68]. Compared to the 11 genes related to small body size in the Chinese Debaos Pony [69], we found six genes belonging to the FACIT (fibril-associated collagens with interrupted triple helices) collagen family, including *COL6A6*, *COL8A1*, *COL25A1*, *COL12A1*, *COL11A1* and *COL14A1*, and other genes such as *FGF14*, *FGF23*, *GDF3*, *BMP10*, *LEMD1* and *PCSK6* that are associated with bone and muscle development also had polymorphic SSRs. For such complex quantitative traits such as body size, the *IGF1* gene contributes to only 16% of height variation in humans and the majority of body size

in dogs. The genes and their corresponding contribution to body size still need to be discovered in pigs. However, the polymorphic SSRs and associated genes discovered in this study might provide some useful information that may contribute to future understanding of mature body size in pigs.

5. Conclusions

In summary, we built and sequenced an SSR-enrichment library and analyzed SSRs at the genome-wide level. We described unique SSR characteristics among four miniature pig breeds, including frequency of SSR type and distribution models of the repeat units. Polymorphic SSRs that were common to the four pig breeds were discovered and annotated, revealing that functional polymorphic SSRs might be related to the growth and development of the miniature pig through their effects on associated genes. The SSRs discovered from this study supplement the genetic variation information of the pig genome and molecular markers of the miniature pig. The established method might provide a reference for SSR analysis, identification of different species and breeds, and a genome-wide association study based on SSRs in the future.

Supplementary Materials: The raw sequence data generated from Illumina MiSeq platform are deposited in the Sequence Read Archive (SRA) and the reviewer link is <https://dataview.ncbi.nlm.nih.gov/object/PRJNA628105?reviewer=5mfadppoucgn41cpgj2idujl86>. The supplementary files were uploaded to Zenodo (10.5281/zenodo.3963831, <https://zenodo.org/deposit/3963831>) during the manuscript submission process. The following are available online at <http://www.mdpi.com/2076-2615/10/10/1792/s1>. **File S1:** SSRs discovered from MISA software for four pig breeds, **File S2:** Sequences of total clusters and corresponding predicted polymorphism of SSRs for four pig breeds, **File S3:** Information of the mapped SSRs and corresponding primers designed for four pig breeds. Supplementary figures and tables were uploaded during the manuscript submission process, **Figure S1:** Length distribution of combined sequences in four datasets, **Figure S2:** Start position of repeat units at combined sequences for data of four breeds, **Figure S3:** Different alleles of SSR located in chr11:70,376,652-70,376,765, **Figure S4:** Different alleles of SSR located in chr18:1,858,964-1,859,153, **Table S1:** Summary description of SSRs discovered in pig reference genome (*SScrofa11.1*), **Table S2:** Details of repeat units in each types of SSRs, **Table S3:** Annotation results of common SSRs existed among four pig breeds and functional genes affected by SSRs, **Table S4:** Alleles of three polymorphic SSRs detected in 60 pigs. All data were obtained from at least three independent experiments.

Author Contributions: Conceptualization, H.W. (Hongyang Wang) and Y.T.; data curation, H.W. (Hongyang Wang) and Y.F.; formal analysis, H.W. (Hongyang Wang) and Y.F.; funding acquisition, H.W. (Hongyang Wang) and Y.T.; methodology, H.W. (Hongyang Wang), Y.F., Y.T., X.Z., and B.L.; project administration, H.W. (Hongyang Wang), Y.F. and Y.T.; resources, P.G., W.T., Z.C., H.W. (Huali Wu) and J.C.; supervision, Y.T.; validation, Y.Z. and H.W. (Huali Wu); visualization, H.W. (Hongyang Wang) and Y.F.; writing—original draft, H.W. (Hongyang Wang) and Y.F.; writing—review and editing, X.Z., B.L., J.J.M., C.F. and Y.T. All authors read and approved the final manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (31900416) and Shanghai Committee of Science and Technology (18140900501) and Postdoctoral Science Foundation of China (2019M651548) and Hainan Provincial Natural Science Foundation of China (2019RC358).

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

References

- Toth, G.; Gáspári, Z.; Jurka, J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **2000**, *10*, 967–981. [[CrossRef](#)]
- Seyoum, M.; Du, X.M.; He, S.P.; Jia, Y.H.; Pan, Z.; Sun, J.L. Analysis of genetic diversity and population structure in upland cotton (*Gossypium hirsutum* L.) germplasm using simple sequence repeats. *J. Genet.* **2018**, *97*, 513–522. [[CrossRef](#)]
- Park, D.H.; Sa, K.J.; Lim, S.E.; Ma, S.J.; Lee, J.K. Genetic diversity and population structure of *Perilla frutescens* collected from Korea and China based on simple sequence repeats (SSRs). *Genes Genom.* **2019**, *41*, 1329–1340. [[CrossRef](#)]
- Chombe, D.; Bekele, E. Genetic diversity analysis of cultivated Korarima [*Aframomum corrorima* (Braun) P.C.M. Jansen] populations from southwestern Ethiopia using inter simple sequence repeats (ISSR) marker. *J. Biol. Res. (Thessalon)* **2018**, *25*, 1. [[CrossRef](#)]

5. Silva, A.V.; Nascimento, A.L.; Vitoria, M.F.; Rabbani, A.R.; Soares, A.N.; Ledo, A.S. Diversity and genetic stability in banana genotypes in a breeding program using inter simple sequence repeats (ISSR) markers. *Genet. Mol. Res.* **2017**, *16*. [[CrossRef](#)] [[PubMed](#)]
6. Rebala, K.; Rabtsava, A.A.; Kotova, S.A.; Kipen, V.N.; Zhurina, N.V.; Gandzha, A.I.; Tsybovsky, I.S. STR Profiling for Discrimination between Wild and Domestic Swine Specimens and between Main Breeds of Domestic Pigs Reared in Belarus. *PLoS ONE* **2016**, *11*, e0166563. [[CrossRef](#)] [[PubMed](#)]
7. Lorenzini, R.; Fanelli, R.; Tancredi, F.; Siclari, A.; Garofalo, L. Matching STR and SNP genotyping to discriminate between wild boar, domestic pigs and their recent hybrids for forensic purposes. *Sci. Rep.* **2020**, *10*, 3188. [[CrossRef](#)] [[PubMed](#)]
8. Chandran, S.; Pukalenty, B.; Adhimoolam, K.; Manickam, D.; Sampathrajan, V.; Chocklingam, V.; Eswaran, K.; Arunachalam, K.; Joikumar Meetei, L.; Rajasekaran, R.; et al. Marker-Assisted Selection to Pyramid the Opaque-2 (O2) and beta-Carotene (crtRB1) Genes in Maize. *Front. Genet.* **2019**, *10*, 859. [[CrossRef](#)]
9. Ma, P.; Xu, H.; Xu, Y.; Song, L.; Liang, S.; Sheng, Y.; Han, G.; Zhang, X.; An, D. Characterization of a Powdery Mildew Resistance Gene in Wheat Breeding Line 10V-2 and Its Application in Marker-Assisted Selection. *Plant. Dis.* **2018**, *102*, 925–931. [[CrossRef](#)] [[PubMed](#)]
10. Leite, D.C.; Pinheiro, J.B.; Campos, J.B.; Di Mauro, A.O.; Uneda-Trevisoli, S.H. QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. *Genet. Mol. Res.* **2016**, *15*. [[CrossRef](#)] [[PubMed](#)]
11. Sawicki, J.; Baczkiewicz, A.; Buczkowska, K.; Gorski, P.; Krawczyk, K.; Mizia, P.; Myszczyński, K.; Slipiko, M.; Szczecińska, M. The Increase of Simple Sequence Repeats during Diversification of Marchantiidae, An Early Land Plant Lineage, Leads to the First Known Expansion of Inverted Repeats in the Evolutionarily-Stable Structure of Liverwort Plastomes. *Genes (Basel)* **2020**, *11*. [[CrossRef](#)] [[PubMed](#)]
12. Witkos, T.M.; Krzyzosiak, W.J.; Fiszer, A.; Koscińska, E. A potential role of extended simple sequence repeats in competing endogenous RNA crosstalk. *RNA Biol.* **2018**, *15*, 1399–1409. [[CrossRef](#)] [[PubMed](#)]
13. Flickinger, R. Polymorphism of simple sequence repeats may quantitatively regulate gene transcription. *Exp. Cell Res.* **2020**. [[CrossRef](#)] [[PubMed](#)]
14. Joy, N.; Beevi, Y.P.M.; Soniya, E.V. A deeper view into the significance of simple sequence repeats in pre-miRNAs provides clues for its possible roles in determining the function of microRNAs. *BMC Genet.* **2018**, *19*, 29. [[CrossRef](#)] [[PubMed](#)]
15. Gymrek, M.; Willems, T.; Guilmatre, A.; Zeng, H.; Markus, B.; Georgiev, S.; Daly, M.J.; Price, A.L.; Pritchard, J.K.; Sharp, A.J.; et al. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat. Genet.* **2016**, *48*, 22–29. [[CrossRef](#)]
16. Malausa, T.; Gilles, A.; Meglecz, E.; Blanquart, H.; Duthoy, S.; Costedoat, C.; Dubut, V.; Pech, N.; Castagnone-Sereno, P.; Delye, C.; et al. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol. Ecol. Resour.* **2011**, *11*, 638–644. [[CrossRef](#)]
17. Conyers, C.M.; Allnutt, T.R.; Hird, H.J.; Kaye, J.; Chisholm, J. Development of a microsatellite-based method for the differentiation of European wild boar (*Sus scrofa scrofa*) from domestic pig breeds (*Sus scrofa domestica*) in food. *J. Agric. Food Chem.* **2012**, *60*, 3341–3347. [[CrossRef](#)]
18. Liu, C.; Liu, Y.; Zhang, X.; Xu, X.; Zhao, S. Characterization of porcine simple sequence repeat variation on a population scale with genome resequencing data. *Sci. Rep.* **2017**, *7*, 2376. [[CrossRef](#)]
19. Fang, X.; Mou, Y.; Huang, Z.; Li, Y.; Han, L.; Zhang, Y.; Feng, Y.; Chen, Y.; Jiang, X.; Zhao, W.; et al. The sequence and analysis of a Chinese pig genome. *Gigascience* **2012**, *1*, 16. [[CrossRef](#)]
20. Ghosh, M.; Sharma, N.; Gera, M.; Kim, N.; Sodhi, S.S.; Pulicherla, K.; Huynh, D.; Kim, D.C.; Zhang, J.; Kwon, T.; et al. The first comprehensive description of the expression profile of genes involved in differential body growth and the immune system of the Jeju Native Pig and miniature pig. *Amino Acids* **2019**, *51*, 495–511. [[CrossRef](#)]
21. Schulze-Tanzil, G.; Silawal, S.; Hoyer, M. Anatomical feature of knee joint in Aachen minipig as a novel miniature pig line for experimental research in orthopaedics. *Anat. Anz.* **2020**, *227*, 151411. [[CrossRef](#)] [[PubMed](#)]
22. Liu, H.F.; Li, H.; Bai, G.; Zhang, Q.Z.; Gao, X.; Liu, T.; Wang, H.B. Establishment of Renal Failure Models by Laparoscopy in Bama Pigs Which Underwent Partial Nephrectomy and Radical Contralateral Nephrectomy. *J. Vet. Res.* **2019**, *63*, 447–455. [[CrossRef](#)] [[PubMed](#)]

23. Yao, J.; Zeng, H.; Zhang, M.; Wei, Q.; Wang, Y.; Yang, H.; Lu, Y.; Li, R.; Xiong, Q.; Zhang, L.; et al. OSBPL2-disrupted pigs recapitulate dual features of human hearing loss and hypercholesterolaemia. *J. Genet. Genom.* **2019**, *46*, 379–387. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, L.; Huang, Y.; Wang, M.; Guo, Y.; Liang, J.; Yang, X.; Qi, W.; Wu, Y.; Si, J.; Zhu, S.; et al. Development and Genome Sequencing of a Laboratory-Inbred Miniature Pig Facilitates Study of Human Diabetic Disease. *iScience* **2019**, *19*, 162–176. [[CrossRef](#)]
25. Cai, Z.; Yu, C.; Fu, D.; Pan, Y.; Huang, J.; Rong, Y.; Deng, L.; Chen, J.; Chen, M. Differential metabolic and hepatic transcriptome responses of two miniature pig breeds to high dietary cholesterol. *Life Sci.* **2020**, *250*, 117514. [[CrossRef](#)] [[PubMed](#)]
26. Wang, F.; Li, G.; Wu, Z.; Fan, Z.; Yang, M.; Wu, T.; Wang, J.; Zhang, C.; Wang, S. Tracking diphodont development in miniature pigs in vitro and in vivo. *Biol. Open* **2019**, *8*. [[CrossRef](#)] [[PubMed](#)]
27. Huang, D.; Cao, H.; Wang, S.; Zheng, L.; Chen, Z.; Wen, X.; Zhang, S.; Xiang, Y.; Gao, Y. [Isolation and culture of adipose-derived mesenchymal stem cells from inbred line miniature pig of Wuzhishan and their biological characteristics]. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **2019**, *44*, 297–306. [[PubMed](#)]
28. Zhiqiang, P.; Cun, S.; Ying, J.; Ningli, W.; Li, W. WZS-pig is a potential donor alternative in corneal xenotransplantation. *Xenotransplantation* **2007**, *14*, 603–611. [[CrossRef](#)]
29. Lepais, O.; Bacles, C.F.E. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimization in *Acacia harpophylla* F. Muell. Ex Benth. *Mol. Ecol. Resour.* **2011**, *11*, 711–724. [[CrossRef](#)]
30. Schubert, M.; Lindgreen, S.; Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **2016**, *9*, 88. [[CrossRef](#)]
31. Magoc, T.; Salzberg, S.L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [[CrossRef](#)] [[PubMed](#)]
32. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
33. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
34. Zhao, H.; Yang, L.; Peng, Z.; Sun, H.; Yue, X.; Lou, Y.; Dong, L.; Wang, L.; Gao, Z. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rep.* **2015**, *5*, 8018. [[CrossRef](#)] [[PubMed](#)]
35. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]
36. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A J. Integr. Boil.* **2012**, *16*, 284–287. [[CrossRef](#)]
39. Untergasser, A.; Cutcutache, L.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)]
40. Joachim, A.; Ruttkowski, B.; Palmieri, N. Microsatellite Analysis of Geographically Close Isolates of *Cystoisospora suis*. *Front. Vet. Sci.* **2019**, *6*, 96. [[CrossRef](#)]
41. Wang, L.; Zhang, G.; Lin, F.; Jiang, B.; Dong, F.; Liu, H. Expression of the insulin-like growth factor system in skeletal muscle during embryonic and postnatal development in the first filial generation pigs from Erhualian and Yorkshire reciprocal crosses. *Gen. Comp. Endocrinol.* **2011**, *173*, 56–62. [[CrossRef](#)] [[PubMed](#)]
42. King, K.; Moody, A.; Fisher, S.A.; Mirza, M.M.; Cuthbert, A.P.; Hampe, J.; Sutherland-Craggs, A.; Sanderson, J.; MacPherson, A.J.; Forbes, A.; et al. Genetic variation in the IGSF6 gene and lack of association with inflammatory bowel disease. *Eur. J. Immunogenet.* **2003**, *30*, 187–190. [[CrossRef](#)] [[PubMed](#)]
43. Wu, X.; Shi, W.; Cao, X. Multiplicity of BMP signaling in skeletal development. *Ann. N. Y. Acad. Sci.* **2007**, *1116*, 29–49. [[CrossRef](#)]

44. Nanda, N.; Bao, M.; Lin, H.; Clauser, K.; Komuves, L.; Quertermous, T.; Conley, P.B.; Phillips, D.R.; Hart, M.J. Platelet endothelial aggregation receptor 1 (PEAR1), a novel epidermal growth factor repeat-containing transmembrane receptor, participates in platelet contact-induced activation. *J. Biol. Chem.* **2005**, *280*, 24680–24689. [[CrossRef](#)] [[PubMed](#)]
45. Cui, Y.F.; Yan, Y.Q.; Liu, D.; Pang, Y.S.; Wu, J.; Li, S.F.; Tong, H.L. Platelet endothelial aggregation receptor-1 (PEAR1) is involved in C2C12 myoblast differentiation. *Exp. Cell Res.* **2018**, *366*, 199–204. [[CrossRef](#)]
46. Kazarian, E.; Son, H.; Sapao, P.; Li, W.; Zhang, Z.; Strauss, J.F.; Teves, M.E. SPAG17 Is Required for Male Germ Cell Differentiation and Fertility. *Int. J. Mol. Sci.* **2018**, *19*, 1252. [[CrossRef](#)]
47. Levine, A.J.; Brivanlou, A.H. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development* **2005**, *133*, 209–216. [[CrossRef](#)]
48. Just, F.; Reyer, H.; Murani, E.; Ponsuksili, S.; Oster, M.; Wimmers, K. Genetic variants of major genes contributing to phosphate and calcium homeostasis and their association with serum parameters in pigs. *J. Appl. Genet.* **2018**, *59*, 325–333. [[CrossRef](#)]
49. Fan, H.; Cinar, M.U.; Phatsara, C.; Tesfaye, D.; Tholen, E.; Looft, C.; Schellander, K. Molecular mechanism underlying the differential MYF6 expression in postnatal skeletal muscle of Duroc and Pietrain breeds. *Gene* **2011**, *486*, 8–14. [[CrossRef](#)]
50. Henriquez-Rodriguez, E.; Bosch, L.; Tor, M.; Pena, R.N.; Estany, J. The effect of SCD and LEPR genetic polymorphisms on fat content and composition is maintained throughout fattening in Duroc pigs. *Meat Sci.* **2016**, *121*, 33–39. [[CrossRef](#)]
51. Zhang, J.; Yan, R.; Wu, C.; Wang, H.; Yang, G.; Zhong, Y.; Liu, Y.; Wan, L.; Tang, A. Spermatogenesis-associated 48 is essential for spermatogenesis in mice. *Andrologia* **2018**, *50*, e13027. [[CrossRef](#)] [[PubMed](#)]
52. Hung, C.M.; Liu, L.C.; Ho, C.T.; Lin, Y.C.; Way, T.D. Pterostilbene Enhances TRAIL-Induced Apoptosis through the Induction of Death Receptors and Downregulation of Cell Survival Proteins in TRAIL-Resistance Triple Negative Breast Cancer Cells. *J. Agric. Food Chem.* **2017**, *65*, 11179–11191. [[CrossRef](#)] [[PubMed](#)]
53. Cunningham, N.S.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Reddi, A.H.; Lee, S.J. Growth/differentiation factor-10: A new member of the transforming growth factor-beta superfamily related to bone morphogenetic protein-3. *Growth Factors* **1995**, *12*, 99–109. [[CrossRef](#)] [[PubMed](#)]
54. Avvaru, A.K.; Saxena, S.; Sowpati, D.T.; Mishra, R.K. MSDB: A Comprehensive Database of Simple Sequence Repeats. *Genome Biol. Evol.* **2017**, *9*, 1797–1802. [[CrossRef](#)]
55. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)]
56. Avvaru, A.K.; Sowpati, D.T.; Mishra, R.K. PERF: An exhaustive algorithm for ultra-fast and efficient identification of microsatellites from large DNA sequences. *Bioinformatics* **2018**, *34*, 943–948. [[CrossRef](#)]
57. Wang, X.T.; Zhang, Y.J.; Qiao, L.; Chen, B. Comparative analyses of simple sequence repeats (SSRs) in 23 mosquito species genomes: Identification, characterization and distribution (Diptera: Culicidae). *Insect Sci.* **2018**, *26*, 607–619. [[CrossRef](#)]
58. Liu, L.; Qin, M.; Yang, L.; Song, Z.; Luo, L.; Bao, H.; Ma, Z.; Zhou, Z.; Xu, J. A genome-wide analysis of simple sequence repeats in *Apis cerana* and its development as polymorphism markers. *Gene* **2017**, *599*, 53–59. [[CrossRef](#)]
59. Mladineo, I.; Trumbic, Z.; Radonic, I.; Vrbatovic, A.; Hrabar, J.; Buselic, I. Anisakis simplex complex: Ecological significance of recombinant genotypes in an allopatric area of the Adriatic Sea inferred by genome-derived simple sequence repeats. *Int. J. Parasitol.* **2017**, *47*, 215–223. [[CrossRef](#)]
60. Sharma, P.C.; Grover, A.; Kahl, G. Mining microsatellites in eukaryotic genomes. *Trends Biotechnol.* **2007**, *25*, 490–498. [[CrossRef](#)]
61. Murat, C.; Riccioni, C.; Belfiori, B.; Cichocki, N.; Labbe, J.; Morin, E.; Tisserant, E.; Paolocci, F.; Rubini, A.; Martin, F. Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers. *Fungal Genet. Biol.* **2011**, *48*, 592–601. [[CrossRef](#)] [[PubMed](#)]
62. Qian, J.; Xu, H.; Song, J.; Xu, J.; Zhu, Y.; Chen, S. Genome-wide analysis of simple sequence repeats in the model medicinal mushroom *Ganoderma lucidum*. *Gene* **2013**, *512*, 331–336. [[CrossRef](#)] [[PubMed](#)]
63. Xu, J.; Liu, L.; Xu, Y.; Chen, C.; Rong, T.; Ali, F.; Zhou, S.; Wu, F.; Liu, Y.; Wang, J.; et al. Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Res.* **2013**, *20*, 497–509. [[CrossRef](#)] [[PubMed](#)]

64. Xiao, J.; Zhao, J.; Liu, M.; Liu, P.; Dai, L.; Zhao, Z. Genome-Wide Characterization of Simple Sequence Repeat (SSR) Loci in Chinese Jujube and Jujube SSR Primer Transferability. *PLoS ONE* **2015**, *10*, e0127812. [[CrossRef](#)]
65. Zhu, H.; Song, P.; Koo, D.H.; Guo, L.; Li, Y.; Sun, S.; Weng, Y.; Yang, L. Genome wide characterization of simple sequence repeats in watermelon genome and their application in comparative mapping and genetic diversity analysis. *BMC Genom.* **2016**, *17*, 557. [[CrossRef](#)]
66. Wang, X.; Yang, S.; Chen, Y.; Zhang, S.; Zhao, Q.; Li, M.; Gao, Y.; Yang, L.; Bennetzen, J.L. Comparative genome-wide characterization leading to simple sequence repeat marker development for *Nicotiana*. *BMC Genom.* **2018**, *19*, 500. [[CrossRef](#)] [[PubMed](#)]
67. Benjamini, Y.; Speed, T.P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **2012**, *40*, e72. [[CrossRef](#)]
68. Cole, R.K. An autosomal dwarfism in the domestic fowl. *Poult. Sci.* **2000**, *79*, 1507–1516. [[CrossRef](#)]
69. Kader, A.; Li, Y.; Dong, K.; Irwin, D.M.; Zhao, Q.; He, X.; Liu, J.; Pu, Y.; Gorkhali, N.A.; Liu, X.; et al. Population Variation Reveals Independent Selection toward Small Body Size in Chinese Debao Pony. *Genome Biol. Evol.* **2015**, *8*, 42–50. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).