animal

Genome-wide association study of bone mineral density trait among three pig breeds

B. Jiang[1], M. Wang[1,2], Z. Tang[1], X. Du[3], S. Feng[1], G. Ma[2], D. Ye[1], H. Cheng[1], H. Wang[3] and X. Liu[1,2†] (iD)

[1]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China; [2]Key Lab of Swine Healthy Breeding of Ministry of Agriculture and Rural Affairs, Guigang 537100, China; [3]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

*Leg weakness (**LW**) issues are a great concern for pig breeding industry. And it also has a serious impact on animal welfare. To dissect the genetic architecture of limb-and-hoof firmness in commercial pigs, a genome-wide association study was conducted on bone mineral density (**BMD**) in three sow populations, including Duroc, Landrace and Yorkshire. The BMD data were obtained by ultrasound technology from 812 pigs (including Duroc 115, Landrace 243 and Yorkshire 454). In addition, all pigs were genotyped using genome-by-sequencing and a total of 224 162 single-nucleotide polymorphisms (**SNPs**) were obtained. After quality control, 218 141 SNPs were used for subsequent genome-wide association analysis. Nine significant associations were identified on chromosomes 3, 5, 6, 7, 9, 10, 12 and 18 that passed Bonferroni correction threshold of 0.05/(total SNP numbers). The most significant locus that associated with BMD (P value = $1.92e^{-14}$) was detected at approximately 41.7 Mb on SSC6 (SSC stands for Sus scrofa chromosome). CUL7, PTK7, SRF, VEGFA, RHEB, PRKAR1A and TPO that are located near the lead SNP of significant loci were highlighted as functionally plausible candidate genes for sow limb-and-hoof firmness. Moreover, we also applied a new method to measure the BMD data of pigs by ultrasound technology. The results provide an insight into the genetic architecture of LW and can also help to improve animal welfare in pigs.*

Keywords: bone mineral density, ultrasonic, animal welfare, single-nucleotide polymorphism, candidate gene

Implications

Leg weakness in pigs is a serious problem in the pig industry. It leads to a decrease in the longevity of breeding pigs and an increase in the rate of premature culling of breeding animals in breeding programs. In addition, the production performance of pigs will be influenced dramatically. Our study will help to solve these problems through studying the genetic architecture of limb-and-hoof firmness in three commercial pig populations. It will accelerate the progress in genetic breeding, decrease the cost for the pig industry, improve animal welfare and provide useful information for the study of human osteoporosis.

Introduction

Reproduction performance and service longevity are economically important and have been intensively selected in

the global pig industry. Leg weakness (**LW**) has a great impact on fitness and service longevity of animals, which influences not only animal welfare but also production and reproduction performance. Pigs with LW are more likely to suffer from limb-and-hoof diseases and be eliminated early, which greatly limits the service longevity of excellent pigs. Some previous studies reported that better leg status would decrease involuntary culling through the genetic correlations between LW-related traits and longevity in breeding sows (de Sevilla *et al.*, 2009). Dissection of the genetic architecture of LW in pigs not only benefits the pig industry and animal welfare but also sheds insight into our understanding of human osteoporosis because pig is an ideal animal model for humans and the shape and size of pig's organ is similar to that of humans.

The factors that cause LW are mainly environment, nutrition and heritability. Heritability may have more influence on LW, because pigs are bred with the same diet and environment, according to their growth requirements in modern pig farms (Fukawa and Kusuhara, 2001). Previous studies had

† E-mail: liuxiangdong@mail.hzau.edu.cn

2443

detected a series of quantitative trait locus (**QTLs**) with LW on pig chromosomes 1, 2, 3, 10 and 11 in multiple populations (Laenoi *et al.,* 2011). Nevertheless, due to a low density of microsatellite markers, the application of these identified QTLs in pig breeding programs is hindered. With the commercialization of the Illumina PorcineSNP60 Beadchip in 2008, genome-wide association study (**GWAS**) have been widely explored to detect associations with LW (Rangkasenee *et al.,* 2013; Rothammer *et al.,* 2017). However, the study of LW compared with other traits is limited, especially in pig populations. Additionally, the methods for measuring phenotypes, including scoring, slaughtering and dual-energy X-ray absorptiometry, are not suitable (Fan *et al.,* 2009; Laenoi *et al.,* 2011), because the score system is subjective and dual-energy X-ray will affect the health of pigs. Furthermore, slaughtering pigs will finish their service directly, and it is not advisable. Therefore, applying an unharmful and objective method to assess the LW of alive pigs will benefit animal welfare and improve the accuracy of phenotype data. Bone mineral density (**BMD**) can reflect how weakness the leg is, our phenotypic data were measured objectively and harmlessly for pigs by ultrasonic bone densitometer and the genotype data were acquired by genotyping-by-sequencing (**GBS**).

The aim of our study was to identify single-nucleotide polymorphism (**SNP**) associated with BMD across three populations, including Duroc, Landrace and Yorkshire pigs, by GWAS method.

Materials and methods

Animals and phenotypic measurements

Pigs were not raised or treated distinguishingly in any way for the purpose of this study, and for this reason, the ethical statement was not needed. A total of 812 purebred sows (115 Duroc, 243 Landrace and 454 Yorkshire) were included in this study. They are all from a large-scale pig farm in Guangxi, China. To avoid the impact of nutrition and environment, these pigs with different parties were fed with the same diet in the same fenced environment. The floor type of this pig farm was leaky, and the pig house was cooled by water curtain and ventilated by negative pressure. At the beginning of the measurement, we recorded the individual identification (**ID**), breed, age and parity in detail. Then we collected the ear tissue of each pig from June to November 2017 for subsequent genotyping. The BMD was measured using an ultrasonic bone densitometer from the company Sunlight, Petah Tikva City, Israel. The measurement time of BMD is 1 day after sow parturition, and the target position was the metatarsal bone of the left back leg. The measurement method had been described in detail in Shanks (2016). The result measured by the ultrasonic bone densitometer is the speed of sound (**SOS**) in bone, rather than the actual BMD value. However, we can use the SOS as the BMD value because of the positive correlation between SOS

and BMD based on previous studies (Peter *et al.,* 1993; Moris *et al.,* 1995). According to the classified recommendation of World Health Organization in 1994, we classified these sows with their BMD data. The criteria for judging the weakness of legs is shown in Supplementary Table S1. Bone strength: $SOS > A + SD$; normal bone: $A - SD < SOS < A + SD$; less bone: $A - 2.5SD < SOS < A - SD$ and osteoporosis: $SOS < A - 2.5SD$.

Excel 2016 software was used to record the phenotype data and statistics was run with the number of individuals and the mean, variance, CV, maximum and minimum of BMD. Breed, parity and age were taken as factors affecting BMD. Multivariate variance analysis was applied to analyze which factor impacts the metatarsal BMD of sows more, and multiple comparisons were performed among sows of different breeds, parities and ages using R Studio software (version 3.5.3) (Racine, 2012).

Genotyping-by-sequencing

The ear tissue samples that acquired from 812 purebred sows were stored in 75% alcohol. DNA was extracted using TIANamp Genomic DNA kits (Tiangen Biotech, Beijing, China), and the DNA concentration was measured with spectrophotometer (Nanodrop 2000; Thermo Fisher Scientific, Wilmington, DE, USA). Meanwhile, the quality of DNA was tested by agarose gel electrophoresis. After that the DNA was diluted to 100 ng/μl for genotyping with HiSeq X Ten (Illumina, Inc., San Diego, CA, USA).

Single-nucleotide polymorphism calling and genotype estimation

The quality control of low-quality data for the initial sequence mainly includes two steps: remove the joint reads and the low-quality base reads. Cutadapt software (version 1.5) was applied to cut the joint reads and adjust the error rate to 0.1. Low-quality reads were removed using Sickle software (Joshi and Fass, 2011). The conditions that satisfied the exclusion criteria were as follows: the mass value of each base was not less than 10, and the length of reads was not less than 20. If the first base of reads was N, it would be removed.

All clean reads of individuals were aligned against the NCBI Sus scrofa 11.1 genome (https://www.ncbi.nlm.nih.gov/) using Burrows-Wheeler Aligner (**BWA**) (Li and Durbin, 2009). Then the results of the alignment were filtered, and only the reads in which both ends were aligned to the same location were retained. Then Picard (http://broadinstitute.github.io/picard/) was used to sort the data that passed the quality control and remove the repetitive reads generated by PCR amplification.

The Genome Anlysis Toolkit (**GATK**) software (version 3.8.0) (Mckenna *et al.,* 2010) was used to perform partial multiple alignment after eliminating the repetitive reads. The aim was to reconstruct the regions including sequence with insertion or deletion (Indel). Detecting variation depends heavily on the quality of the bases; therefore, the

base quality score needs to be recalibrated (**BQSR**) after partial multiple alignment. The BQSR is an important (or even the only) indicator to evaluate whether the bases are correct. The HaplotypeCaller module of GATK software was used to detect and filter the variation in samples.

The SNPs that satisfy the following conditions were retained:

$$QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 4.0 || ReadPOSRanSum < -8.0;$$

The parameter of Indel variation is:

$$QUAL < 30.0 || QD < 2.0 || FS > 200.0 | SOR > 10.0 || InbreedingCoeff < -0.8 || ReadPOSRankSum < 20.0 || MQ < 40.0 || MQRankSum < -12.5.$$

Quality control and imputation
A total of 224 162 SNPs were obtained from GBS, and quality control was carried out by Plink (version 1.9) (Purcell *et al.,* 2007). Briefly, SNPs with minor allele frequencies of less than 0.01 or a low call rate of less than 90% were removed. Samples with low call rates of less than 80% were also removed. Missing genotype data were imputed by Beagle (version 4.0) program (Browning and Browning, 2007). Then quality control was carried out again. Finally, a set of 218 141 SNPs was selected for subsequent genome-wide association analysis.

Genome-wide association analysis
According to the results of multivariate analysis of variance and previous studies, we used the FarmCPU model (Liu *et al.,* 2016) of the R package MVP developed by Xiaolei Liu (https://github.com/XiaoleiLiuBio/rMVP) to perform a genome-wide association analysis. The FarmCPU model used a fixed effects model and a random effects model as the loop. The fixed effects model was used to detect genetic markers *m* one by one. Potential association sites were included in the model as covariates to control false positive. Fixed effects model can be written as follows:

$$y_i = M_{i1}b_1 + M_{i2}b_2 + \ldots + M_{it}b_t + S_{ij}d_j + e_i$$

where y_i is the observation value of the ith individual; $M_{i1}, M_{i2} \ldots, M_{it}$ as the genotype of potential association sites were added to the model, which is empty in the first iteration; $b_1, b_2, \ldots, b_t$ as the corresponding effect value of potential association sites were added to the model; S_{ij} is the genotype of the jth genetic marker of the ith individual; d_j is the corresponding effect value of S_{ij} and e_i is a residual vector which obeys a normal distribution with zero mean and variance of σ_e^2.

The random-effects model uses the SUPER algorithm to optimize the potential association sites of different combinations using the *P* value and location information of the genetic markers. Random effects model can be written mathematically as follows:

$$y_i = u_i + e_i$$

where y_i is the observation value of the ith individual; u_i is the total genetic effect of the ith individual and e_i is a residual vector which obeys a normal distribution with zero mean 0 and variance of σ_e^2.

The expected value of the total genetic effect is 0, and the variance and covariance matrix of the individual's total genetic effects is $G = 2K\sigma_a^2$, where σ_a^2 is an unknown genetic variance and K is a kinship matrix calculated from the significant association sites of different combinations.

In our study, we took breed as fixed effect because it is an intrinsic characteristic, parity and age that can be changeable were added as covariates in the fixed effect model for association test. In addition, pseudo quantitative trait nucleotides (QTNs) were added as covariates. The fixed effects model in our study can be written as:

$$y = Ab_a + M_tb_t + S_jd_j + e$$

where *y* is phenotypic observation vector (BMD); A is a matrix of fixed effects, including breed, parity and age; M_t is the genotype matrix of *t* pseudo QTNs that was used as fixed effects; b_a and b_t are the relevant matrices designed for *A* and M_t, respectively; S_j is the ith marker to be tested and d_j is the corresponding effect; *e* is the residual effect vector and obeys a normal distribution with zero mean and variance of σ_e^2.

The random effect model in our study is written as:

$$y = u + e$$

where *y* and *e* stay the same as in fixed effects model and *u* is the genetic effect.

Gene ontology
The associated gene stable IDs located in the 1 Mb region on either side of significant SNPs were extracted from the Ensembl genome browser (http://asia.ensembl.org/index.html). Gene ontology enrichment analysis was performed using the database for annotation, visualization and integrated discovery (**DAVID**) (https://david.ncifcrf.gov/).

Result

Statistical analysis of metatarsal bone mineral density
In order to determine the main and interactive effects of breed, parity and age on metatarsal BMD in sows, variance analysis was performed among the three varieties of sow. The result (Table 1) showed that breed, parity and the

Table 1 *Variance analysis of the metatarsal BMD in the Duroc, Landrace and Yorkshire pig breeds*

Category	*P* value	Significance
Breed	$1.80e^{-14}$	***
Parity	$1.10e^{-08}$	***
Age	0.26846	–
Breed × parity	0.00025	***
Breed × age	0.59825	–
Parity × age	0.81771	–
Breed × Parity × Age	0.55147	–

'***' = indicates a highly significant difference; '–' = indicates no significant difference; BMD = bone mineral density.

interaction effects between breed and parity on metatarsal BMD of sows were very significant ($P < 0.01$).

Then the mean, SD, CV, maximum and minimum values of metatarsal BMD of different breeds were summarized. From the result (Supplementary Table S2), we know that the metatarsal BMD of Duroc was the lowest, and the average ultrasound value was 4130.33 m/s. The maximum and minimum BMD values of Landrace were the highest among the three breeds, indicating that Landrace have the best limb-and-hoof firmness among the three breeds. Then we make multiple comparisons using Tukey honest significance test (**HSD**) method. The result (Supplementary Table S3) showed a very significant difference in metatarsal BMD between Duroc and Landrace as well as Yorkshire ($P < 0.01$), but no significant difference was observed between Landrace and Yorkshire ($P > 0.05$).

Age was divided into different age-groups at an interval of 100 days for statistical analysis of the mean, SD, CV, maximum and minimum values of metatarsal BMD in different age-groups. The result (Supplementary Table S4) showed that the highest metatarsal BMD was age-group 8 (SOS = 4695 m/s), and the lowest BMD was age-group 5 (SOS = 3388 m/s). Then we make multiple comparisons using Tukey HSD method to find out which age-groups are significantly different. The results of multiple comparisons among different age-groups (Supplementary Table S5) showed that there is very significant difference in metatarsal BMD between age-group 1 and three age groups (age-group 6, 8 and 9) ($P < 0.01$). And there is very significant difference in metatarsal BMD between age-group 2 and age-group 6 ($P < 0.01$). Also, age-group 3 has significant difference in metatarsal BMD with three age groups (age-group 5, 6 and 8) ($P < 0.01$).

According to the different parities, the mean, SD, CV, and maximum and minimum values of metatarsal BMD were calculated. The result (Supplementary Table S6) showed that the mean of metatarsal BMD was the largest in the eighth parity (SOS = 4365 m/s) and the smallest in the second parity (SOS = 4203 m/s). Only three sows had reproduced six times, so there was no further analysis with this parity. The metatarsal BMD of the sows with first and second parities was significantly lower than that of the sows with third, fourth, fifth, seventh and eighth parities ($P < 0.01$). Then we make
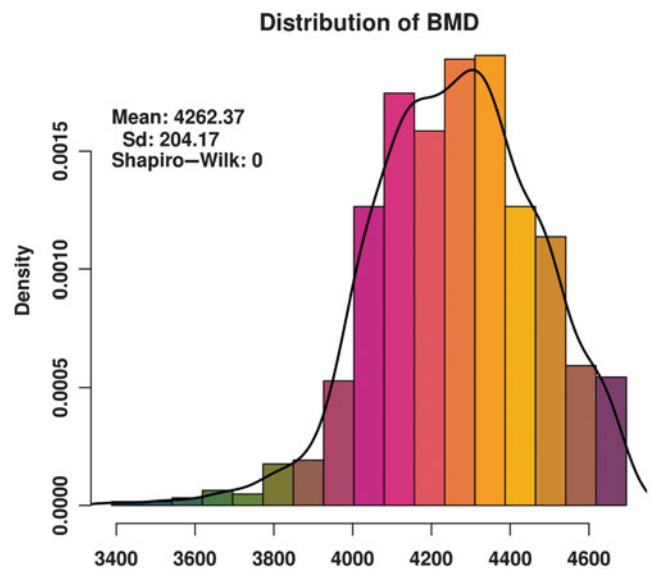

Figure 1 Histogram of the bone mineral density (BMD) distribution among three varieties (Duroc, Landrace and Yorkshire) of sow. Standard deviation is abbreviated as SD. The x-axis indicates the speed of sound in the metatarsal, that is, BMD. The y-axis indicates the density of experimental animals. The Shapiro-Wilk test shows that the phenotype data satisfy a normal distribution.

multiple comparisons using Tukey HSD method to find out which parities are significantly different. The results (Supplementary Table S7) showed that the metatarsal BMD of the first parity has very significant difference with two parities (the fourth and the fifth parity) ($P < 0.01$). And there is aslo very significant difference in metatarsal BMD between the second parity and four parities (the third, the fourth, the fifth and the seventh parity) ($P < 0.01$).

Genome-wide association analysis
The BMD of these animals basically conforms to a normal distribution (Figure 1) and satisfies the GWAS phenotype data requirement. The distribution of SNPs on chromosomes (Supplementary Figure S1) shows that the coverage and quality of the SNPs is enough for subsequent genome-wide association analysis.

The association analyses were conducted using the FarmCPU model of the R package MVP. From the quantile-quantile (**Q-Q**) plot (Figure 2a), the y-axis indicates observed value ($-\log_{10}(P\text{-value})$) and the x-axis indicates expected value ($-\log_{10}$(probability value of uniform distribution)). We can see a very high-fitting precision between observed value and expected value when the *P*-value is more than $10e^{-4}$. And a separation appears seriously when the *P*-value is less than $10e^{-4}$. It indicates a significant natural selection between phenotype and genotype. And the lambda value of the Q-Q plot is 0.973, indicating a slight population stratification. The Manhattan plot (Figure 2b) shows that nine SNPs significantly associated with BMD were found on SSC3, SSC5, SSC6, SSC7, SSC9, SSC10, SSC12 and SSC18 (**SSC** stands for *Sus scrofa* chromosome). The details are shown in Table 2, and these loci that related to BMD are reported for the first
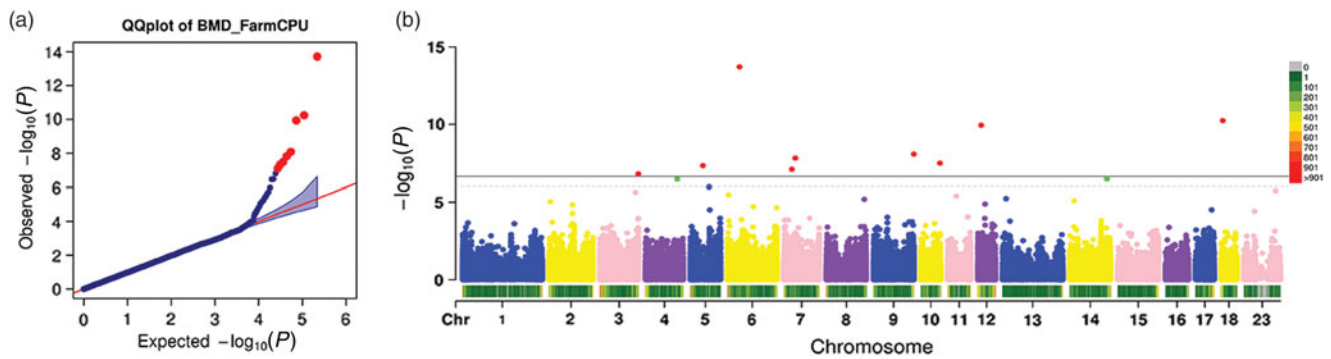
Figure 2 (a) Quantitle-quantitle (Q-Q) plots of the observed *P* values for bone mineral density (BMD) of three populations (Duroc, Landrace and Yorkshire). The Q-Q plot shows the observed $-\log_{10} P$ values (y-axis) and the expected $-\log_{10} P$ values (x-axis). And the lambda value on Q-Q plot is 0.973. (b) Manhattan plot of the observed *P* values for BMD of three populations (Duroc, Landrace and Yorkshire). The horizontal black line in the Manhattan plot indicates the genome-wide (2.29×10^{-7}) threshold. FarmCPU: a model for genome-wide association study and the abbreviation of fixed and random model circulating probability unification.

Table 2 *SNPs significantly associated with BMD obtained by GWAS among three varieties (Duroc, Landrace and Yorkshire) of sow*

SNP ID	SSC	Position	*P* value
S3_132353359	3	132 353 359	$1.58e^{-07}$
S5_43562333	5	43 562 333	$4.65e^{-08}$
S6_41723048	6	41 723 048	$1.92e^{-14}$
S7_27195548	7	27 195 548	$7.97e^{-08}$
S7_38899457	7	38 899 457	$1.52e^{-08}$
S9_139161514	9	139 161 514	$8.35e^{-09}$
S10_6616592	10	66 165 902	$3.23e^{-08}$
S12_11767894	12	11 767 894	$1.14e^{-10}$
S18_6712259	18	6 712 259	$5.71e^{-11}$

BMD = bone mineral density; GWAS = genome-wide association study; SNP ID = single-nucleotide polymorphism identification, where 'S + chromosome + position' is the name of the marker; SSC = Sus scrofa chromosome.

time as far as we know. But some genes that located adjacent to significant SNPs are reported in previous studies, and we will show them in the discussion section.

Gene ontology

Genes within 1 Mb distance on either side of lead SNPs were downloaded from the Ensembl genome browser. The GO enrichment analysis was performed using DAVID, and the secondary classification bar plot and the bubble plot of GO results are shown in Supplementary Figure S2 and Figure 3, respectively. These genes are mainly enriched in these GO terms such as ATP binding, lipid transport and ATPase activity. To identify candidate genes related to BMD, we manually queried the NCBI database and PubMed to find genes related to bone development, bone cell differentiation or bone metabolism. Eventually, gene cullin protein family member 7 (*CUL7*), tyrosine protein kinase 7 (*PTK7*), serum response factor (*SRF*), vascular endothelial growth factor A (*VEGFA*), Ras homologs (*RHEB*), protein kinase A regulatory subunit 1α (*PRKAR1A*) and thrombopoietin (*TPO*) were selected as candidate genes related to BMD. The details are shown in Table 3.

Discussion

The BMD as a parameter to evaluate LW is conducive for genetic improvement. At present, the methods that evaluate LW are mostly visual evaluation, which is subjective. In order to fully identify markers related to LW, if the data generated by subjective evaluation methods were integrated, the measurement error will be further amplified and the result will be more inaccurate. The evaluation of leg appearance is usually classified into different grades according to the degree of injury or firmness, and the classification may also lead to a large CV. The CV of the hind limb structure score was as high as 40% in Laenoi's experimental animals (Laenoi *et al.*, 2011), and the CV of the limb-and-hoof structure score was 6.09% to 15.38% in Guo's experimental population (Guo *et al.*, 2009). In addition, the swelling of the joint is considered one of the criteria in the leg structure score, but it is related to the environmental factors such as hard floor and humid air. The methods of the gait score and joint injury score emphasize that the evaluator should strictly obey the evaluation criteria which indicate a fluctuation in the appearance score to some extent. Quantitative traits are greatly affected by phenotype data. The accuracy of phenotype data directly affects the accuracy of the results. A correct and objective evaluation method is a basis of genetic improvement in limb-and-hoof firmness.

Low BMD is the main cause of bone fracture and osteomalacia. The pain of limb-and-hoof fracture is more severe than hoof injury and joint swelling (Jensen *et al.*, 2012). Pigs with fractures are difficult to cure and are eliminated frequently, and even if they are not eliminated, their reproductive performance will decrease significantly. Adult pigs with larger body weight need higher requirement for limb-and-hoof firmness. Improving BMD can help reduce the probability of elimination due to limb-and-hoof fracture. Fracture is influenced by genetic factor to a great extent. Blott showed that the *ZNF804A* gene is a candidate gene related to fracture (Blott *et al.*, 2014), and Li found three QTLs related to BMD affect fracture severity (Li *et al.*, 2002). In addition, 11 QTLs affecting lower extremity
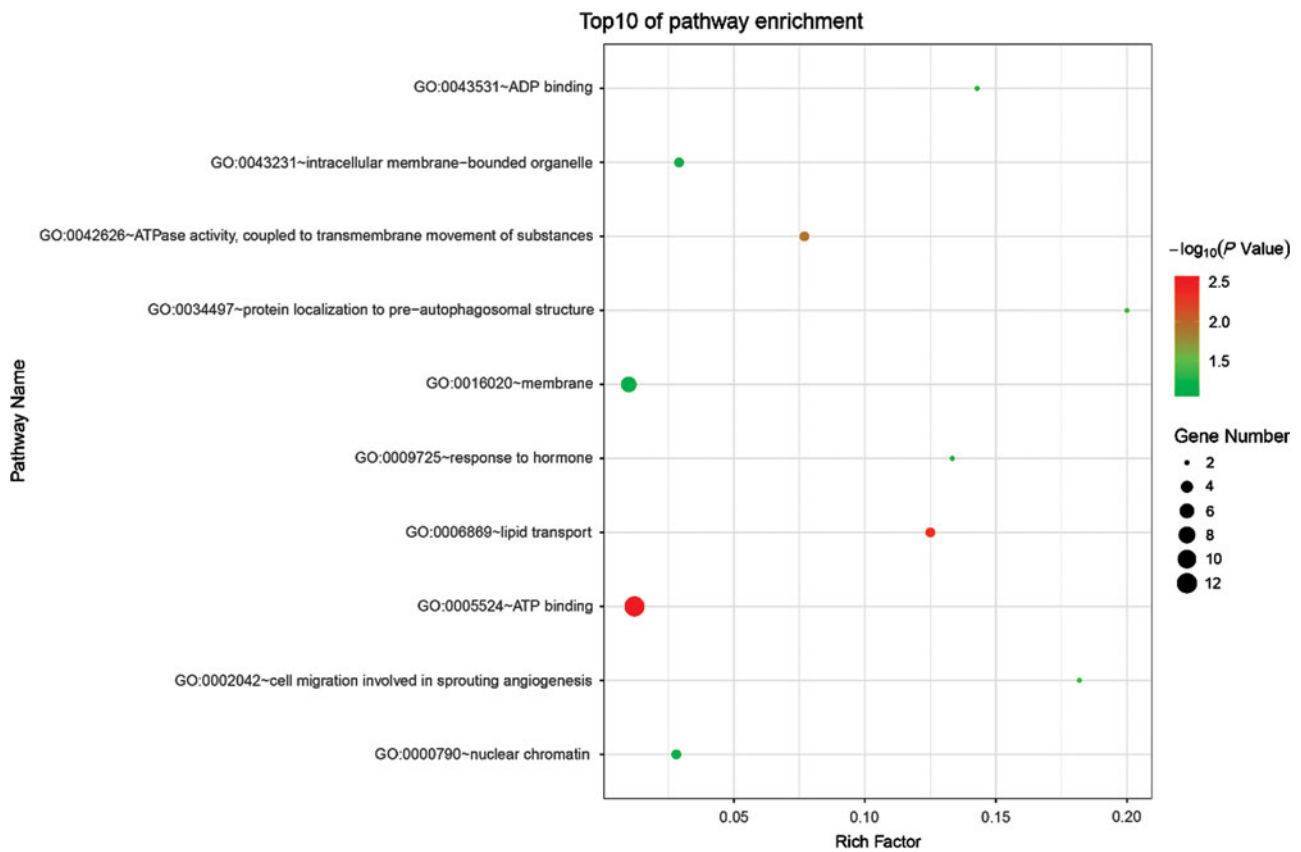
2447

Figure 3 Top 10 enriched gene ontology (GO) terms for bone mineral density (BMD) in sow. Rich factor indicates the percentage of genes associated with the pathways with respect to the total number of genes in the pathway; gene number indicates the quantity of genes associated with these pathways.

Table 3 *Important candidate genes related to BMD among three varieties* (Duroc, Landrace and Yorkshire) *of sow*

Gene name	Gene ID	SSC	Position	Description
CUL7	ENSSSCG00000001657	7	38 097 171 to 38 116 855	Cullin 7
PTK7	ENSSSCG00000001660	7	38 135 877 to 38 208 138	Protein tyrosine kinase 7
SRF	ENSSSCG00000001661	7	38 213 063 to 38 223 384	Serum response factor
VEGFA	ENSSSCG00000001695	7	38 746 055 to 38 762 342	Vascular endothelial growth factor A
RHEB	ENSSSCG00000016434	18	5 779 212 to 5 822 257	Ras homolog, mTORC1 binding
PRKAR1A	ENSSSCG00000017259	12	11 661 872 to 11 681 435	Protein kinase cAMP-dependent type I regulatory subunit alpha
TPO	ENSSSCG00000033193	3	132 384 820 to 132 408 943	Thyroid peroxidase

BMD = bone mineral density; SSC = Sus scrofa chromosome; mTORC1 = mammalian target of rapamycin complex 1; cAMP = cyclic adenosine monophosphate.

weakness were found by limb-and-hoof score and measuring BMD (Laenoi *et al.,* 2011). Moreover, the CV of BMD in our phenotype data was approximately 6.00% similar to that of Laenoi's experimental animals. Therefore, BMD as an index for evaluating limb-and-hoof firmness is of great significance for genetic improvement and breeding.

The GBS is a simplified genome sequencing technology. The genomic DNA was digested and two ends of the digested fragment were sequenced with high-throughput sequencing. The genotype obtained by GBS is accurate, reaching 92.6% and consistent with that of the chip of more than 99% (Pértille *et al.,* 2016; Wang *et al.,* 2017). Moreover, the

SNPs obtained by GBS are distributed evenly in the functional area of chromosomes and can represent the genomic information of individuals well (Gurgul *et al.,* 2018). In addition, GBS is economical and efficient and has high throughput. Currently, commercial SNP chips are usually used for sequencing, but the density of the SNP chips is low and the number of markers is limited while the price of high-density SNP chips or resequencing is more expensive which makes it difficult to sequence for large-scale populations. Although advanced imputation algorithm provides a more powerful tool for a question of low coverage in GWAS, the accuracy of imputation is a bit different for various

imputation algorithm. And the cost of sequencing is very high for some species with large genome. As we all know, whole-genome sequencing has a wide coverage, which can detect all genetic information in the individual genome with high accuracy. And it will have a wider application in the future with the development of commercial chips and the progress of technology compared with GBS. But GBS has a bit of an advantage to some extent now, it provides high-density markers and has a low cost for GWAS. Therefore, GBS can be used as an important sequencing technology in pig genetic research. In general, the raw genetic difference that was screened are obtained by mutagenizing or sampling from natural population. Any phenotypic difference that was identified is connected back to the underlying causative loci by various mapping approaches including QTL mapping. The GWAS is a powerful tool for connecting the genotype and phenotype in this perspective. Ibeagha-Awemu identified several significant SNPs and candidate genes associated with milk production traits and breast function in dairy cows by GBS (Ibeagha-Awemu *et al.*, 2016). The SNPs related to teat number, litter size, litter weight and other complex reproductive traits were identified in different pig populations based on the GBS technology in some researchers' studies (Tan *et al.*, 2017).

Nine SNPs related to BMD were significantly identified on eight chromosomes by GWAS. The reason why nine association sites were distributed on eight chromosomes might be that the GWAS method is not appropriate in certain circumstances or the genetic determinism is relatively complex. Previous study showed that GWAS involve the detection and interpretation of epistasis. However, many epistasis detection algorithms cannot be directly applied into GWAS as many combinations of genetic components are present in only a small amount of samples or even none at all (Zhou *et al.*, 2016). Linear mixed models (**LMMs**) are widely used in GWAS to account for population structure and relatedness, for both continuous and binary traits. But LMMs are generally inappropriate for analyzing binary traits when population stratification leads to violation of the LMM's constant-residual variance assumption (Chen *et al.*, 2016). There is no population stratification for our experimental animal, according to the front analysis result. In addition, our samples are relatively large to endure the accuracy of our result. After enrichment analysis and annotation for all candidate genes, the genes *CUL7*, *PTK7*, *SRF*, *VEGFA*, *RHEB*, *PRKAR1A* and *TPO* related to bone development, bone differentiation and bone metabolism were considered candidate genes related to BMD. The genes *CUL7*, *PTK7*, *SRF* and *VEGFA* are related to bone development and bone homeostasis. Patients with human autosomal recessive disease 3-M syndrome grow slowly and suffer from facial deformities and skeletal dysplasia (Shaikh *et al.*, 2019). Mutation or deletion of *CUL7* is one of the main causes of 3-M syndrome (Dan *et al.*, 2011; Li *et al.*, 2014). Individuals with homozygous deletion of exon 18 of the *CUL7* gene have short and thin skeletons at birth, and they will suffer from hip joint loss and scapula and joint deformation in adulthood (Meazza *et al.*, 2013). Idiopathic scoliosis is

caused by vertebral deformities. Mutation in *PTK7* leads to abnormal Wnt signal pathway in zebra fish, leading to scoliosis (Hayes *et al.*, 2014). The *SRF* regulates and adheres to genes that relate to migration by gathering cardiac hormone transcription factors that are regulated by G-actin. And the absence of function will lead to defects in the fetal cytoskeleton (Watts *et al.*, 2008). *VEGFA* is an important regulator of osteogenic differentiation, bone development and bone homeostasis (Duan *et al.*, 2015). It is necessary for osteoblast precursor differentiation to regulate the gathering of perichondral vascular and osteoblast differentiation by increasing beta-catenin expression and inhibiting the Notch pathway. Analysis of histomorphology and protein levels in mice with type 2 diabetes mellitus indicated that *VEGFA* played an important role in angiogenesis, osteogenesis and bone remodeling, and *VEGFA* was an important regulatory target for osteogenesis in individuals with type 2 diabetes mellitus (Wallner *et al.*, 2015). In addition, the absence of *VEGFA* promotes macrophage aggregation and angiogenesis and promotes the maturation and mineralization of osteoblasts to heal and repair the damaged bone (Hu *et al.*, 2016). The cyclic adenosine monophosphate-dependent *PRKAR1A* is an important regulator of osteogenic differentiation. The level of *PRKAR1A* increased significantly during osteoblast differentiation. After knocking down *PRKAR1A* in mouse and human cells, the formation of bone nodules was obstructed and osteoblast markers were significantly reduced, suggesting that *PRKAR1A* was associated with osteogenesis (Zhang *et al.*, 2014). Changes in cartilage and bone in mice with bone tumor have shown that it was caused by incomplete differentiation of osteoblasts from tumor cells due to the reduction in *PRKAR1A* protein in tumor cells (Pavel *et al.*, 2008).

The *TPO* is a main growth factor of megakaryocytes and can enhance the immune factors of peripheral blood mononuclear cells and promote the resorption of osteoclasts in progenitor cells, thus causing osteoporosis or osteitis (Bethel *et al.*, 2015). In addition, *RHEB* can promote expression of SRY(sex determining region Y)-box transcription factor 9 (*SOX9*) and the synthesis of glycosaminoglycan and collagen II, which can promote adipose mesenchymal stem cells to differentiate into chondrocytes, enhance the function of RUNX family transcription factor 2 (*RUNX2*) to promote bone formation and increase the content of calcium and phosphate. And serine protease 7 (*SP7*) induced by *RUNX2* is an essential transcription factor of osteoblast differentiation and chondrocyte maturation (Komori, 2018). These previous studies reflect the reliability of our results, and these genes can be used as candidate genes for studying foot-and-limb firmness.

Our measurement method for BMD is objective and the GWAS approach for detecting candidate genomic region underlying BMD was efficient, though the small sample size was one possible potential limitation of this study to some extent. Future investigations with a larger sample size in other pig breeds and higher throughput sequencing of the candidate genomic region are necessary to confirm and extend our findings.

2449

Conclusion

In our study, it was found that breed, parity and the interaction effect between breed and parity had very significant effect on the metatarsal BMD of sow. The GWAS identified a total of nine SNPs significantly associated with BMD trait, and *CUL7*, *PTK7*, *SRF*, *VEGFA*, *RHEB*, *PRKAR1A* and *TPO* were selected as the plausible candidate genes related to BMD trait by enrichment analysis and functional annotation.

Declaration of interest

The authors declare to have no conflicts of interest.

Ethics statement

All animals used in this study were kept according to the Chinese legislation for pig production. The collection of samples and care of the animals used in this study followed the guidelines for experimental animals established by the Animal Care Committee and were approved by the Ethics Committee of Huazhong Agricultural University, China.

Software and data repository resources

None of the data were deposited in an official repository. Data may be available upon request by contacting the corresponding author.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731120001111

References

Bethel M, Barnes CLT, Taylor AF, Cheng YH, Chitteti BR, Horowitz MC, Bruzzaniti A, Srour EF and Kacena MA 2015. A novel role for thrombopoietin in regulating osteoclast development in humans and mice. Journal of Cellular Physiology 230, 2142–2151.

Blott SC, Swinburne JE, Sibbons C, Fox-Clipsham LY, Helwegen M, Hillyer L, Parkin TD, Newton J and Vaudin M 2014. A genome-wide association study demonstrates significant genetic variation for fracture risk in Thoroughbred racehorses. BMC Genomics 15, 147–152.

Browning SR and Browning BL 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81, 1084–1097.

Chen H, Wang C, Conomos MP, Stilp AM and Lin X 2016. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. American Journal of Human Genetics 98, 653–666.

Dan H, O'Sullivan J, Urquhart J, Daly S, Bhaskar SS, Biesecker LG, Cole T, Kirk J, Chandler K and Donnai D 2011. Exome sequencing identifies CCDC8 mutations in 3-M syndrome, suggesting that CCDC8 contributes in a pathway with CUL7 and OBSL1 to control human growth. American Journal of Human Genetics 89, 148–153.

de Sevilla XF, Fabrega E, Tibau J and Casellas J 2009. Genetic background and phenotypic characterization over two farrowings of leg conformation defects in Landrace and Large White sows. Journal of Animal Science 87, 1606–1612.

Duan X, Murata Y, Liu Y, Nicolae C, Olsen BR and Berendsen AD 2015. Vegfa regulates perichondrial vascularity and osteoblast differentiation in bone development. Development 142, 1984–1991.

Fan B, Onteru SK, Mote BE, Serenius T, Stalder KJ and Rothschild MF 2009. Large-scale association study for structural soundness and leg locomotion traits in the pig. Genetics Selection Evolution 41, 14–18.

Fukawa K and Kusuhara S 2001. The genetic and non-genetic aspects of leg weakness and osteochondrosis in pigs. Journal of Animal Science 14, 114–122.

Guo YM, Ai HS, Ren J, Wang GJ, Wen Y, Mao HR, Lan LT, Ma JW, Brenig B and Rothschild MF 2009. A whole genome scan for quantitative trait loci for leg weakness and its related traits in a large F2 intercross population between White Duroc and Erhualian. Journal of Animal Science 87, 1569–1575.

Gurgul A, Mikszacybulska A, Szmatoła T, Jasielczuk I, Piestrzyńskakajtoch A, Fornal A, Semikgurgul E and Bugnoponiewierska M 2018. Genotoping-by-sequencing performance in selected livestock species. Genomics 43, 620–628.

Hayes M, Gao X, Yu LX, Paria N, Henkelman RM, Wise CA and Ciruna B 2014. ptk7 mutant zebrafish models of congenital and idiopathic scoliosis implicate dysregulated Wnt signalling in disease. Nature Communications 5, 4777–4780.

Hu K, Olsen BR, Hu K and Olsen BR 2016. Osteoblast-derived VEGF regulates osteoblast differentiation and bone formation during bone repair. Journal of Clinical Investigation 126, 509.

Ibeagha-Awemu EM, Peters SO, Akwanji KA, Imumorin IG and Zhao X 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Scientific Reports 6, 31109–31115.

Jensen TB, Kristensen HH and Toft N 2012. Quantifying the impact of lameness on welfare and profitability of finisher pigs using expert opinions. Livestock Science 149, 209–214.

Joshi NA and Fass JN 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Retrieved from https://github.com/najoshi/sickle

Komori T 2018. Runx2, an inducer of osteoblast and chondrocyte differentiation. Histochemistry and Cell Biology 149, 313–323.

Laenoi W, Uddin MJ, Cinar MU, Großebrinkhaus C, Tesfaye D, Jonas E, Scholz AM, Tholen E, Looft C and Wimmers K 2011. Quantitative trait loci analysis for leg weakness-related traits in a Duroc × Pietrain crossbred population. Genetics Selection Evolution 43, 1–7.

Li H and Durbin R 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li X, Masinde G, Gu W, Wergedal J, Mohan S and Baylink DJ 2002. Genetic dissection of femur breaking strength in a large population (MRL/MpJ×SJL/J) of F2 mice: single QTL Effects, epistasis, and pleiotropy. Genomics 79, 734–740.

Li Z, Pei XH, Yan J, Yan F, Cappell KM, Whitehurst AW and Xiong Y 2014. CUL9 mediates the functions of the 3M complex and ubiquitylates survivin to maintain genome integrity. Molecular Cell 54, 805–819.

Liu X, Huang M, Fan B, Buckler ES and Zhang Z 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. Plos Genetics 12, e1005767.

Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S and Daly M 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20, 1297–1303.

Meazza C, Lausch E and Pagani S 2013. 3-M syndrome associated with growth hormone deficiency: 18. Italian Journal of Pediatrics 39, 1–6.

Moris M, Peretz A, Tjeka R, Negaban N and Bergmann P 1995. Quantitative ultrasound bone measurements: normal values and comparison with bone mineral density by dual X-ray absorptiometry. Calcified Tissue International 57, 6–10.

Pavel E, Nadella K, Towns WH and Kirschner LS 2008. Mutation of Prkar1a causes osteoblast neoplasia driven by dysregulation of protein kinase A. Molecular Endocrinology 22, 430–435.

Pértille F, Guerrero-Bosagna C, Silva VHD, Boschiero C, Nunes JDRDS, Ledur MNC, Jensen P and Coutinho LL 2016. High-throughput and cost-effective chicken genotyping using next-generation sequencing. Scientific Reports 6, 26929–26935.

Peter P, Charles YC, Gonzales J, Anderson J, Sakhaee K and Rubin C 1993. Measurement of intrinsic bone quality *in vivo* by reflection ultrasound: correction of impaired quality with slow-release sodium fluoride and calcium citrate. Journal of Bone and Mineral Research 8, 234–240.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIWD and Daly MJ 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

Racine JS 2012. RStudio: a platform-independent IDE for R and sweave. Journal of Applied Econometrics 27, 167–172.

Rangkasenee N, Murani E, Brunner RM, Schellander K, Cinar MU, Luther H, Hofer A, Stoll M, Witten A, Ponsuksili S and Wimmers K 2013. Genome-wide association identifies TBX5 as candidate gene for osteochondrosis providing a functional link to cartilage perfusion as initial factor. Frontiers in Genetics 4, 78–82.

Rothammer S, Bernau M, Kremer-Rücker PV, Medugorac I and Scholz AM 2017. Genome-wide QTL mapping results for regional DXA body composition and bone mineral density traits in pigs. Archives Animal Breeding 60, 51–59.

Shaikh S, Shettigar SKG, Kumar S, Kantharia S, Kurva J and Cherian S 2019. Novel mutation in Cul7 gene in a family diagnosed with 3M syndrome. Journal of Genetics 98, 21–26.

Shanks 2016. A cross-sectional study of radial and tibial ultrasonography between different ethnic and physically active groups. Master thesis, Essex University, Kochester, Britain.

Tan C, Wu Z, Ren J and Huang Z 2017. Genome-wide association study and accuracy of genomic prediction for teat number in Duroc pigs using genotyping-by-sequencing. Genetics Selection Evolution 49, 35–40.

Wallner C, Schira J, Wagner JM, Schulte M, Fischer S, Hirsch T, Richter W, Abraham S, Kneser U and Lehnhardt M 2015. Application of VEGFA and FGF-9 enhances angiogenesis, osteogenesis and bone remodeling in type 2 diabetic long bone regeneration. PLoS ONE 10, e0118823.

Wang Y, Xuemin C, Yiqiang Z, Jing F, Xiaoxiang H, Ning L and Peng X 2017. Optimized double-digest genotyping by sequencing (ddGBS) method with high-density SNP markers and high genotyping accuracy for chickens. PLoS ONE 12, e0179073.

Watts NB, Lewiecki EM, Miller PD and Baim S 2008. National Osteoporosis Foundation 2008 Clinician's guide to prevention and treatment of osteoporosis and the World Health Organization Fracture Risk Assessment Tool (FRAX): what they mean to the bone densitometrist and bone technologist. Journal of Clinical Densitometry the Official Journal of the International Society for Clinical Densitometry 11, 473–477.

Zhang M, Manchanda PK, Wu D, Wang Q and Kirschner LS 2014. Knockdown of PRKAR1A, the gene responsible for Carney complex, interferes with differentiation in osteoblastic cells. Molecular Endocrinology 28, 295–300.

Zhou Z, Liu G and Su L 2016. A new approach to detect epistasis utilizing parallel implementation of ant colony optimization by MapReduce framework. International Journal of Computer Mathematics 93, 511–523.