Comparative population genomic analysis uncovers novel genomic footprints and genes associated with small body size in Chinese pony

CURRENT STATUS: UNDER REVIEW

Hojjat Asadollahpour Nanaei
Shahid Bahonar University of kerman

Ali Esmailizadeh
Shahid Bahonar University of Kerman

Ahmad Ayatollahi Mehrgardi
Shahid Bahonar University of Kerman

Han Jianlin
International Livestock Research Institute

Dong-Dong Wu
Kunming Institute of Zoology Chinese Academy of Sciences

Yan Li
Yunnan University

Ya-Ping Zhang
Kunming Institute of Zoology Chinese Academy of Sciences

DOI: 10.21203/rs.3.rs-21864/v1

SUBJECT AREAS
Epigenetics & Genomics Animal Science

KEYWORDS
Body size, Artificial selection, Population genomics, Horse, NELL1
Abstract

Background Body size is considered as one of the most fundamental properties of an organism. Due to intensive breeding and artificial selection throughout the domestication history, horses exhibit striking variations for heights at withers and body sizes. Debao pony (DBP), a famous Chinese horse, is known for its small body size and lives in Guangxi mountains of southern China. In this study, we employed comparative population genomics to study the genetic basis underlying the small body size of DBP breed based on the whole genome sequencing data. To detect genomic signatures of positive selection, we applied three methods based on population comparison, fixation index ($F_{ST}$), cross population composite likelihood ratio (XP-CLR) and nucleotide diversity (Pi), and further analyzed the results to find genomic regions under selection for body size-related traits.

Results A number of protein-coding genes with significant ($P$-value < 0.01) higher $F_{ST}$ values (367 genes), XP-CLR scores (681 genes), and a lower value for nucleotide diversity (332 genes) were identified. The most significant signal of positive selection was mapped to the $NELL1$ gene, probably underlies the body size and development traits, and may also have independently been selected for short stature in the DBP population. In addition, some other loci on different chromosomes were identified to be potentially involved in the development of body size.

Conclusions Results of our study identified some positively selected genes across the horse genome, which are possibly involved in body size traits. These novel candidate genes may be useful targets for clarifying our understanding of the molecular basis of body size and as such they should be of great interest for future research into the genetic architecture of relevant traits in horse breeding program.

Background

For thousands of years, enormous variety of domestic breeds with different morphological, physiological and behavioral characters have been domesticated and raised in every corner of the world. They therefore offer a powerful source of biological models for the biology studies and have played significant role in developmental, evolutionary and biomedical research [1, 2, 3].

Due to intensive breeding and artificial selection throughout the domestication history, horses exhibit striking variation of height at withers and body size. Today, body size is one of the most important
criteria for the evaluation and classification of different breeds, and also is an essential parameter for breeding programmers to improve marketability and performance [4]. According to this criterion, most horse breeds can be divided into three main categories including high stature/heavy horses (draft breeds), light horse (riding breeds) and small stature/low weight (pony breeds) [5, 6]. The pony breed is defined as a group of horses with a common height less than 14.2 hands (147 cm) at the withers [7]. There are several pony breeds in the world, though varying in body size, skin color and geographic origins, all of them share some basic traits that make them different from other horse breeds. Debao pony (DBP), a famous Chinese horse breed, is known for its small body size and lives in Guangxi mountains of southern China. This breed is strictly protected by the Chinese government, therefore it has the largest population compared with other local pony breeds. DBPs exhibit peculiar morphoanatomical adaptation to facilitate work in the mountainous regions, for example their average adult height is around 94 cm and 98 cm for male and females, respectively [8, 9]. Although previous studies reported a few candidate genes such as LCORL, HMG2, ZFAT, NCAPG [6, 10], TBX3 [9] and ANKRD1 [11] with major effects on height and body size variations in several horse breeds, these studies were mostly carried out using SNP genotyping involving Illumina EquineSNP50 Genotyping Beadchip or GeneSeek Equine SNP70 Beadchip, which certainly have some limitation due to factors such as low SNP density in many genomic regions, differential probe affinity as well as ascertainment biases that prevent them from detecting novel variation in entire genome [12]. In the present study, whole genome sequencing (WGS) data were used for comparative population genomics to identify the genetic basis underlying the small stature of DBP. We identified some novel loci that are hypothesized to be involved in body size variation. Our results indicate that WGS technique provides useful information in the horse to identify novel variants of functional importance and to further elucidate how historical process of natural selection has shaped the equine genome. 

Results

Sequencing and read alignment

Individual genomes of 17 DBPs were sequenced on an Illumina Hiseq 2000 platform with a read length of 125 bp. To facilitate comparisons with other horse breeds, the sequenced horse data were
jointly analyzed with publicly available WGS data of 15 different breeds (n= 68) (Additional file 1: Table S1 and Additional file 2: Fig. S1). The mean sequence depth was ~13.5X per sample (Additional file 2: Figs. S2). We detected a total of 18,384,176 SNPs in all individuals.

**Phylogenetic analysis, linkage disequilibrium decay and runs of homozygosity**

We firstly performed different classical analyses including phylogenetic tree, principle component analysis (PCA), Bayesian model-based analysis and haplotype population structure based on chromosome painting from ChromoPainter and fineSTRUCTURE analyses to reveal genetic relationships among different horses (Fig. 1). Based on the phylogenetic tree results (Fig. 1A), all DBPs were separated from Middle East, European and American horse breeds. The topological pattern found in the tree was also supported by both PCA (Fig. 1B) and Bayesian model-based analysis (Fig. 1C). The results of painting algorithm from ChromoPainter and fineSTRUCTURE analyses simplified their relationships into a co-ancestry matrix, which presents expected number of chunks from donors (column) to recipients (row), and visualized as a heat map plot (Fig. 2A).

In addition, runs of homozygosity (ROH) along the whole genome have been applied to quantify individual autozygosity to improve the understanding of inbreeding depression in populations. The mean number of ROHs longer than 100 and 1000 Kb per each population is plotted in Figure 2B. Also, to investigate the effects caused by selection and genetic bottleneck on population history and demography, we determined the scale of linkage disequilibrium (LD) decay for each breed. The average level of LD, measured by r^2, between adjacent single-nucleotide polymorphism (SNPs) across the complete genome was estimated (Fig. 2C).

**Scans for signatures of selection**

Based on the results obtained from phylogenetic analysis and to avoid bias from unequal sample sizes, we compared the whole genome of DBP individuals, as a pony-sized (genetically homogeneous) breed with Thoroughbred (THB) horses, as a large well-defined (genetically homogeneous) breed, to identify signatures of positive selection related to body size traits. Here, we examined three different parameters for a greater statistical power to localize the source of selection signals. Population
differentiation ($F_{ST}$) was calculated for each SNP between DBP and THB horses as described in Akey et al. (2002) [13]. A sliding window analysis was employed with 50 kb window size and 25 kb step size. A total of 367 protein-coding genes were identified by the outlier approach in the top 1% windows with high $F_{ST}$ values (Additional file 1: Table S2). To get a clearer insight into the genetic mechanisms related to these candidate genes, further downstream analyses were conducted. Among the candidate genes, five ($NELL1$, $FGFR1$, $SNTG1$, $BMP2$ and $TBX15$) are involved in body size related traits (Table 1). Gene set enrichment analysis (GSEA) identified some significantly enriched categories related with skeletal development, such as “broad hallux” (HP:0010055), “broad phalanges of the hand” (HP:0009768) and “broad toe” (HP:0001837) (Additional file 1: Table S3). We further calculated nucleotide diversity ($\pi$), to trace the regions that might underlie selective sweeps. A sliding window analysis yielded 332 candidate genes within regions of relatively low $\pi$ estimates in DBPs (Additional file 1: Table S4). For these outlier genes, some enriched gene ontology (GO) terms were found to be related with skeletal development categories, including “aplasia/hypoplasia of the tibia” (HP:0005772) and “Short femur” (HP:0003097) (Additional file 1: Table S5). In addition, we found some size-related genes such as, $NELL1$, $FGFR1$ and $CNN3$ (Table 1). Finally, to confirm the selection signals detected by $F_{ST}$ and $\pi$ methods, we adopted a cross-population composite likelihood ratio test (XP-CLR) to evaluate historical selections based on the comparison of allele frequency spectrum. A total of 681 candidate genes were identified using the outlier approach (top 1% cutoff) (Additional file 1: Table S6). The results of functional enrichment analysis from all these genes showed some categories related with both muscle and skeletal development such as “small hand” (HP:0200055), “increased body weight” (HP:0004324), “large forehead” (HP:0002003), “animal organ development” (GO:0048513), “anatomical structure development” (GO:0048856) and “tissue development” (GO:0009888) (Table 1; Additional file 1: Table S7).

In addition, to confirm these results, when we compared the genome of DBP, used as a test population, with all other horse studied breeds (mixed-breed and purebred) used as a reference population, we also observed that $NELL1$ signals in the high values (the top 1% in both $F_{ST}$ and XP-
CLR methods, Additional file 2: Fig. S3).

Discussion

Before to the application of methods for detecting signatures of selection, we first assessed the phylogeny of the horse breeds to evaluate the phylogenetic position of DBP within the species. In agreement with previous studies, DBP was phylogenetically distant from other horse breeds while being relatively close to the Mongolian horse (MNG) [8, 9]. The distance patterns between horse breeds was also identified by PCA. However, THB and DBP had the greatest distance from each other, two German Warmblood breeds, Hanoverian (HNV) and Holsteiner (HLS), were not clearly distinguishable in either the phylogenetic tree and PCA, indicating the possibility of shared genetic components between the two breeds [14]. Similar to the results from PCA, admixture analysis at \( K = 2 \) separated the DBP, Yakutian (YKT), MNG and THB horses from other populations, while at \( K = 3 \) the Standardbred (STB) horses separate from the remaining horse populations. Consistent with previous studies, we found that the DBP and THB populations are genetically homogeneous for all \( K \) values (\( K = 2 \) to \( K = 8 \)) [4, 9].

In addition, LD decay analysis revealed a markedly lower level of LD across all genomic distances in DBPs than other breeds. The high LD in commercial breeds, especially for THB horses, could be a consequence of artificial selection for specific abilities, e.g. racing performance, in the breeding programs [15] while DBPs clearly have an ancient origin following a long-term natural selection. Also, computing the ROH per each breed showed the lower level in the DBP population compared to other horses. Here, we found high level of ROH for Sorraia (SOR) and THB populations, that is concordant with previous study in the six different horse breeds [16].

Potential independent of positive selection in the DBP population

Because of the small body size of DBP, that is notably less than average horse breeds, we focused specifically on the loci that may play more important roles in the rapid evolution of body size during the domestication process. Here, we used comparative genome analysis between DBP and THB breeds to identifying the genetic basis underlying the size variation among DBPs. In our broad spectrum analysis, we identified several previously reported genes that are associated with body size
related traits. Highly significant candidate genes that may be potentially involved in body size traits are listed in Table 1. Within the regions showing extremely high values (top 0.01), both $F_{ST}$ and XP-CLR methods showed $BMP2$ gene shared overlapped selection signatures as positively selected genes (PSGs) (Figs. 3A and 3B). $BMP2$, a bone formation-related gene, was found as one of the candidates on ECA5. This protein belongs to the TGF-β superfamily, which has diverse biological activities related to bone physiology and metabolism [17, 18]. Previous studies have found associations of the $BMP-2$ variants with bone and cardiac development, bone mineral density, as well as body size traits [19, 20]. Also in human, BMP-2 appears to be the most important BMP affecting the adult skeleton [21].

Another possible candidate gene, $FGFR1$, was found in one of the selection regions on ECA27 (top 1% cutoff for $F_{st}$ and $Pi$ methods) (Fig. 3A). $FGFR1$ is an important candidate gene that influences bone growth and skeletal development. Previous studies found that FGFR1 protein plays a critical role in formation of muscle and bone tissues [22, 23, 24, 25]. Considering the important function of $FGFR1$ in skeletal development, this gene is an important candidate for body size variation in mammals.

Results from the detection of selection signatures revealed consistently high signal values in $F_{ST}$ and XP-CLR analyses as well as high differences in allele frequencies for $NELL1$ gene, which is overlapped among candidate PSGs (Figs. 3A, 3B and 4). $NELL1$, encodes a mammalian cell-signaling protein (protein kinase C-b1, PKC-b1) that has been shown to regulate skeletal ossification [28, 29].

Overexpression of this gene in both human and mice induces craniosynostosis, the premature fusion of cranial sutures [30]. Previous studies have shown that absence of $Nell1$ leads to decreased cell differentiation and cell proliferation in several organs such as heart, bone and cartilage tissue [30, 31]. Recently, an interstitial 11p14.1-p15.3 deletion involving the $Nell-1$ gene was also reported in associated with short stature in children [32]. Moreover, it was demonstrated that the $NELL-1$ has potential roles as a bone-forming growth factor in sheep [33].

Body size is recognized as one of the most fundamental properties of an organism, affecting nearly all biological aspects. In the last decades, new insights from the genetic and physiological studies have refined our understanding of genetic basis of body size, as the target of positive selection in human
and domesticated animals. Human body size is a polygenic trait affected by variants of numerous genes and their interactions with environmental factors. For example, hundreds of genetic variants, in at least 180 loci, with small effects, have impact on final human adult height [34]. In contrast, several independent studies in domesticated animals have shown that changes in body size can be controlled by a few genes with large effects. For instance, it has been demonstrated that one specific haplotype defined by 20 SNPs spanning the recent selection sweep covering IGF-1 gene has a major effect on body size within all small dogs [35]. A similar study has shown that one SNP within the strong linkage region of BMP10 gene explained around 22% of the overall body weight variance in five chicken lines [2]. Also, one study on dairy and beef cattle revealed the variation in the average height can be controlled by only 10 genes in eight genomic regions [36].

So far, previous genome-wide association studies (GWAS) have revealed several QTLs for body size and stature in different horse breeds. Based on the standard additive model, Makvandi-Nejad et al. (2012) [6] identified four loci on the ECA3, 6, 9 and 11 that explained 83% of size variance in 48 horses, three each of eight large and eight small horse breeds. Using the same dataset of these 48 horses, a recent GWAS study involving both dominant and recessive mixed-model approaches as well as a genome-wide scan for signatures of selection based on the $F_{ST}$ genetic differentiation and XP-CLR test, ANKRD1 gene was identified and validated as a novel candidate, explaining 7.98% of the genetic variance in body size of the American Miniature horse (AMH). Compared with the fixed status of all four loci identified by Makvandi-Nejad et al. (2012) [6], ANKRD1 gene could be applied in effective genotype-assisted selection for body size in AMH [11]. In other independent studies, the differential SNPs in LCORL gene on ECA3 [10, 37, 38, 39, 40], ZFAT gene on ECA9 [37], TBX3 gene on ECA8 [9] and LASP1 gene on ECA11 [41] have also been shown to be strongly associated with body size traits in horses.

In this study, we have investigated the genetic basis underlying the body size variation in DBP. In our broad spectrum of analyses by three methods, we did not find any significant selection signal within or near genes which were previously identified as horse body size-related candidates. Instead, we observe that NELL1 gene likely played an important role in the evolution of the small stature of DBP,
an ancient small pony that was evolved in the mountainous areas in southwestern China. In addition, some other loci on different chromosomes were also identified to be potentially involved in body development process. No evidence of directional selection for the detected genes in this study has been reported to date in other horse populations, suggesting that these genes have been probably selected independently for the short stature of DBPs.

Conclusions
In this study, using next-generation sequencing analysis, we identified some novel candidate genes under selection for body size traits in DBP population. This results suggest that the imaging a common evolutionary mechanism that influences patterns of genetic variation in all horse breeds is misconception as many of them were adapted for different environments and/or for various goals. These novel candidate genes may be useful target for clarifying our understanding of the molecular basis of body size and they should be of great interest for future research addressing the genetic architecture of relevant traits in horse breeding program.

Methods
Sample collection and sequencing
Here, we included totally 86 whole genomes of different horse breeds. A total of 17 DBPs were genotyped using the Illumina HiSeq 2000. In addition, previously published genome sequence data from 15 other breeds of horses (n=69) were obtained from the Sequence Read Archive (NCBI) (Additional file 1: Table S1 and Additional file 2: Fig. S1).

Alignments and variant identification
High-quality reads in the present study and previous studies were aligned against the reference horse genome, ENSEMBL (version 94), using Burrows-Wheeler Aligner (BWA) with the default parameters (https://github.com/lh3/bwa) [42]. Bam files were imported into SAMtools [43] for sorting/merging and into Picard tools (https://broadinstitute.github.io/picard/, latest release 2.18.26) to mark duplicated reads. We then used RealignerTargetCreator, IndelRealigner, and BaseRecalibrator tools in the Genome Analysis Toolkit (GATK) 4.0.12.0 [44] for realigning around indels and base quality recalibration. The variant calling of sequence data were handled using UnifiedGenotyper and
VariantFiltration command in GATK. Final variants were filtered to be supported by a minimum mapping quality of 25, minimum genotype quality of 40, and also, loci with >2 alleles and within clusters (>3 SNP in a 10-bp window) were removed.

**Population variation and population genetic analyses**

The phylogenetic tree for the individual genome sequences was constructed from the SNP data by using the neighbor-joining method among the horses. PCA was performed using the software GCTA [45], after pruning the SNPs for short-range linkage disequilibrium (--indep-pairwise 50 10 0.1 option). In order to visualize population structure between different breeds, ADMIXTURE [46] was used with an ancestor population size ranging from 2 to 8 and 10,000 iterations for each run, based on the pruned data. In addition, we used a haplotype-based approach to explore patterns of haplotype sharing using ChromoPainter and fineSTRUCTURE [47]. ROH and LD were calculated by PLINK [48].

**Identification of candidate genes under positive selection in the DBP**

Three different test statistics were used to identify regions of the genome that showed the strongest signals of selective sweeps in DBPs. To characterize genetic differentiation, $F_{ST}$ values for each SNP were calculated between DBP and THB horses [13]. The absolute differences in allele frequency ($\Delta \pi = \pi(\text{other}) - \pi(\text{DBP})$) for each SNP were estimated between DBP and THB groups. Sliding window analyses were performed upon a window size of 50 kb and a step size of 25 kb across of the genome. The average $F_{ST}$ and Pi values of SNPs in each window were calculated. We additionally performed a multilocus allele frequency differentiation scan using an updated cross-population composite likelihood ratio (CLR), to detect regions in the genome where the change in allele frequency at the locus almost-occurred recently. The scores were estimated using the code provided by Hua Chen (Department of Genetics, Harvard Medical School) [49]. We used the following options: sliding windows of 0.1 cM, grid size 10 k, maximum number of SNPs within each window as 300, and correlation value for 2 SNPs weighted with a cutoff of 0.99. Here, DBPs was taken to be the object population and other horses chosen as the reference population. The regions with the XP-CLR values in the top 0.01 of the empirical distribution were designated as signal of positive selection.
**Gene set enrichment and pathway analysis**

To search the possible pathways involved in the outlier regions ($F_{ST}$, Pi and XP-CLR), candidate selective regions were annotated using the Variant Effect Predictor available at (http://asia.ensembl.org/info/docs/tools/vep/index.html). Functional enrichment analysis was performed using the ‘g:Profiler’ enrichment analysis tool [50]. And the $P$-value of the gene enrichment was corrected by Benjamini–Hochberg FDR (false discovery rate).

**Abbreviations**

BAM: Binary Alignment MAP; BWA: Burrows-Wheeler Aligner program; LD: linkage disequilibrium; GATK: Genome Analysis Toolkit; GO: Gene ontology; GWAS: Genome-wide association studies; PCA: Principle component analysis; PSGs: positively selected genes; ROH: Runs of homozygosity.

**Declarations**

**Acknowledgments**

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA2004010302), National Natural Science Foundation of China (91731304), the Sino-Africa Joint Research Center, Chinese Academy of Sciences (SAJC201611), and Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences. YL was supported by the Young Academic and Technical Leader Raising Foundation of Yunnan Province (to Yan Li). Samples used in this study were provided by the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding).

**Funding**

This study was funded by Strategic Priority Research Program of the Chinese Academy of Sciences (XDA2004010302), National Natural Science Foundation of China (91731304), the Sino-Africa Joint Research Center, Chinese Academy of Sciences (SAJC201611), and Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences

**Availability of data and materials**

The datasets generated and/or analyzed during the current study are available in the XXXXXXXX.

**Authors’ contributions**
A.E., Y.L., and Y.-P.Z., lead the project, and designed and conceived the study. H.A.N., A.E., and Y.L., provided samples. H.A.N., Y.L., and D.D.W., analyzed and interpreted the data. H.A.N. drafted the manuscript. A.E., HAN, Y.L., A.A.M., and J.H., revised the manuscript. All authors revised and approved the final manuscript.

**Ethics approval and consent to participate**

The animal procedures were approved by Animal Care and Use Committee of Kunming Institute of Zoology (SMKX2017007)

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Darwin C. The variation of animals and plants under domestication. London: Cambridge University Press. 1868.

2. Wang MS, Huo YX, Li Y, Otecko NO, Su LY, Xu HB, Wu SF, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. J Mol Cell Biol. 2016;8:542-552.

3. Andersson L. Domestic animals as models for biomedical research. Ups J Med Sci. 2016;121:1-11.

4. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. Plos One. 2013;8:e29013.

5. Brooks SA, Makvandi-Nejad S, Chu E, Allen JJ, Streeter C, Gu E, McCleery B, Murphy BA, Bellone R, Sutter NB. Morphological variation in the horse: defining complex traits of body size and shape. Anim Genet. 2010;41: 159–165.

6. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, Loredo Al, Bellone
RR, Mezey JG, Brooks SA, Sutter NB. Four loci explain 83% of size variation in the horse. PLoS ONE. 2012;7:e39929.

7. Mattern J. Horses on the farm. Rourke Pub Group. 2010.

8. Jiang Q, Wei Y, Huang Y, Jiang H, Guo Y, Lan G, Liao J. The complete mitochondrial genome and phylogenetic analysis of the Debao pony (Equus caballus). Mol Biol Rep. 2011;38:593-599.

9. Kader A, Li Y, Dong K, Irwan DM, et al. Population variation reveals independent selection toward small body size in Chinese Debao pony. Genome Biol Evol. 2016;8:42-50.

10. Metzger J, Schrimpf R, Philipp U, Distl U. Expression levels of LCORL are associated with body size in horses. Plos One. 2013;8:e56497.

11. Al-Abri MA, Posbergh C, Palermo K, Sutter NB, Eberth J, Hoffman GE, Brooks SA. Genome-wide scans reveal a quantitative trait locus for withers height in horses near the ANKRD1 gene. J Equine Vet Sci. 2018;60:67-73.

12. Dreger DL, Rimbault M, Davis BW, Bhatnagar A, Parker HG, Ostrander EA. Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. Dis Model Mech. 2016;9:1445-1460.

13. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 2002:12;1805-1814.

14. Nolte W, Thaller G, Kuehn C. Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. PLoS One. 2019;14(4):e0215913.

15. Moon S, Lee JW, Shin D, Shin KY, Kim J, Choi IY, et al. A Genome-wide Scan for Selective Sweeps in Racing Horses. Asian Austral J Anim. 2015;28(11):1525-1531.
16. Metzger J, Karwath M, Tonda R, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics. 2015;16:764.

17. Bhatia M, Bonnet D, Wu D, Murdoch B, Wrana J, Gallacher L, Dick JE. Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. J Exp Med. 1999;189:1139-1148.

18. Seib FP, Lanfer B, Bornhauser B, Werner C. Biological activity of extracellular matrix-associated BMP-2. J Tissue Eng Regen Med 2010; 4: 324-327.

19. Kugimiya F, Kawaguchi H, Kamekura S, Chikuda H, Ohba S, Yano F, Ogata N, Katagiri T, Harada Y, Azuma Y, Nakamura K, Chung UI. Involvement of endogenous bone morphogenetic protein (BMP) 2 and BMP6 in bone formation. J Biol Chem. 2005; 280(42):35704-12.

20. Tan TY, Gonzaga-Jauregui C, Bhoj EJ, et al. Monoallelic BMP2 Variants Predicted to Result in Haploinsufficiency Cause Craniofacial, Skeletal, and Cardiac Features Overlapping Those of 20p12 Deletions. Am J Hum Genet. 2017;101(6):985–994.

21. Gori F, Thomas T, Hicok KC, Spelsberg TC, Riggs BL. Differentiation of human marrow stromal precursor cells: bone morphogenetic protein-2 increases OSF2/CBFA1, enhances osteoblast commitment, and inhibits late adipocyte maturation. J Bone Miner Res. 1999;14(9):1522-35.

22. Yamaguchi TP, Harpal K, Henkemeyer M, Rossant J. FGFR-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. Genes Dev. 1994;8: 3032-3044.

23. Burke D, Wilkes D, Blundell TL, Malcolm S. Fibroblast growth factor receptors: lessons from the genes. Trends Biochem Sci. 1998;23: 59-62.

24. Wang Q, Green RP, Zhao G, Ornitz DM. Differential regulation of endochondral bone
growth and joint development by FGFR1 and FGFR3 tyrosine kinase domains.

Development. 2001;128: 3867-3876.

25. Su N, Jin M, Chen L. Role of FGF/FGFR signaling in skeletal development and homeostasis: learning from mouse models. Bone Res. 2014;2:14003.

26. Sabeti PC, Schaffner SF, Fry B, et al. Positive natural selection in the human lineage. Science. 2006;312:1614-1620.

27. Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here?. Genome Res. 2009;19:711-722.

28. Kuroda S, Tanizawa K. Involvement of Epidermal Growth Factor-like Domain of NELL Proteins in the Novel Protein-Protein Interaction with Protein Kinase C. Biochem Biophys Res Commun. 1999;265(30):752-757.

29. Qi H, Kim JK, Ha P, Chen X, Chen E, Chen Y, et al. Inactivation of Nell-1 in Chondrocytes Significantly Impedes Appendicular Skeletogenesis. J Bone Miner Res. 2019;34(3):533-546.

30. Desai J, Shannon ME, Johnson MD, Ruff DW, et al. Nell1-deficient mice have reduced expression of extracellular matrix proteins causing cranial and vertebral defects. Hum Mol Gen. 2006;15:8.

31. Shen J, James AW, Chung J, Lee K, Zhang JB, Ho S, Lee KS, Kim TM, Niimi T, Kuroda S, Ting K, Soo C. NELL-1 promotes cell adhesion and differentiation via Integrinβ1. J Cell Biochem. 2012;113(12):3620-8.

32. Dateki S, Watanabe S, Kinoshita F, Yoshiura KI, Moriuchi H. Identification of 11p14.1-p15.3 deletion probably associated with short stature, relative macrocephaly, and delayed closure of the fontanelles. Am J Med Genet A.2017;173(1):217-220.

33. James AW, Chiang M, Asatrian G, et al. Vertebral Implantation of NELL-1 Enhances Bone Formation in an Osteoporotic Sheep Model. Tissue Eng Part A. 2016;22(11-
34. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832-8.

35. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, et al. A single IGF1 allele is a major determinant of small size in dogs. Science. 2007;316:112-115.

36. Pryce JE, Hayes BJ, Bolormaa S, Goddard ME. Polymorphic regions affecting human height also control stature in cattle. Genetics. 2011;187:981-984.

37. Signer-Hasler H, Flury C, B Haase, Burger D, Simianer H, Leeb T, Rieder S. A Genome-Wide Association Study Reveals Loci Influencing Height and Other Conformation Traits in Horses. PLoS ONE. 2012;7(5): e37282.

38. Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. Anim Genet. 2013;44(4):467-71.

39. Staiger EA, Al Abri MA, Pflug KM, Kalla SE, Ainsworth DM, Miller D, Raudsepp T, Sutter NB, Brooks SA. Skeletal variation in Tennessee Walking Horses maps to the LCORL/NCAPG gene region. Physiol Genomics. 2016;48: 325-335.

40. Sevane N, Dunner S, Boado A, Cañon J. Polymorphisms in ten candidate genes are associated with conformational and locomotive traits in Spanish Purebred horses. J Appl Genetics. 2016;58(3):355-361.

41. Junior AB, Quirino CR, Vega WHO, Rua MAS, David CMG, Jardim JG. Polymorphisms in the LASP1 gene allow selection for smaller stature in ponies. Livest Sci. 2018;160-164.

42. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler...
transform. Bioinformatics. 2009;15: 1754-60.

43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-2079.

44. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-1303.

45. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88: 76-82.

46. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655-1664.

47. Lawson, DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. Plos Genet. 2012;8:e1002453.

48. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-575.

49. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010;20:393-402.

50. Reimand J, Arak T, Vilo J. g: Profiler--a web server for functional interpretation of gene lists (2011 update), Nucleic Acids Res. 2011;39:W307–315.

51. An B, Xia J, Chang T, Wang X, Xu L, Zhang L, Gao X, Chen Y, Li J, Gao H. Genome-wide association study reveals candidate genes associated with body measurement traits in Chinese Wagyu beef cattle. Anim Genet. 2019;50(4):386-390.

52. Lee KY, Singh MK, Ussar S, et al. Tbx15 controls skeletal muscle fibre-type determination and muscle metabolism. Nat Commun. 2015;6:8054.
53. Zhang L, Ma X, Xuan J, Wang H, Yuan Z, Wu M, Liu R, Zhu C, Wei C, Zhao F, Du L. Identification of MEF2B and TRHDE Gene Polymorphisms Related to Growth Traits in a New Ujumqin Sheep Population. PLoS One. 2016;11(7):e0159504.

54. Yang X-Q, Guo L-J, Zhai C-Y, Yu H, Liu H, Liu D. Expression, characterization, and variation of the porcine calpain 7 gene. Acta Agric Scand Sect A- Anim Sci. 2009;59:204-210.

55. Monda KL, Chen GK, Taylor KC, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat Genet. 2013;45(6):690-696.

56. Shin S, Heo J, Yeo J, Lee C, Chung E. Genetic association of phosphodiesterase 1B (PDE1B) with carcass traits in Korean cattle. Mol Biol Rep.2012;39(4):4869-74

57. Zhang X, Chu Q, Guo G, et al. Genome-wide association studies identified multiple genetic loci for body size at four growth stages in Chinese Holstein cattle. PLoS One. 2017;12(4):e0175971.

58. Tanaka KI, Xue Y, Nguyen-Yamamoto L, Morris JA, Kanazawa I, Sugimoto T, Wing SS, Richards JB, Goltzman D. FAM210A is a novel determinant of bone and muscle structure and strength. Proc Natl Acad Sci U S A.2018;115(16):E3759-E3768.

59. Tang Z, Liang R, Zhao S, Wang R, Huang R, Li K. CNN3 is regulated by microRNA-1 during muscle development in pigs. Int J Biol Sci. 2014;10(4):377-385.

60. Edea Z, Jeoung YH, Shin SS, et al. Genome-wide association study of carcass weight in commercial Hanwoo cattle. Asian-Australas J Anim Sci. 2018;31(3):327-334.

61. Wang J, Li ZJ, Lan XY, Hua LS, Huai YT, Huang YZ, Ma L, Zhao M, Jing YJ, Chen H, Wang JQ. Two novel SNPs in the coding region of the bovine PRDM16 gene and its associations with growth traits. Mol Biol Rep.2010;37(1):571-7.

62. Wood TL, Streck RD, Pintar JE, Expression of the IGFBP-2 gene in post-implantation
rat embryos. Development 1992;114: 59-66;

63. James E. Koltes, Bishnu P. Mishra, et al. A nonsense mutation in cGMP-dependent type II protein kinase (PRKG2) causes dwarfism in American Angus cattle. Proc Natl Acad Sci U S A. 2009;106(46)19250-19255.

64. Gudbjartsson DF, Walters GB, Thorleifsson G, et al. Many sequence variants affecting diversity of adult human height. Nat Genet. 2008; 40(5):609-15.

65. Metzger J, Rau J, Naccache F, Bas Conn L, Lindgren G, Distl O. Genome data uncover four synergistic key regulators for extremely small body size in horses. BMC Genomics. 2018;19(1):492.

66. Yang M, Fu J, Lan X, Sun Y, Lei C, Zhang C, Chen H. Effect of genetic variations within the SH2B2 gene on the growth of Chinese cattle. Gene. 2013;528(2):314-9.

67. Deng MT, Zhu F, Yang YZ, Yang FX, Hao JP, Chen SR, Hou ZC. Genome-wide association study reveals novel loci associated with body size and carcass yields in Pekin ducks. BMC Genomics. 2019; 20(1):1.

68. Yasukochi Y, Sakuma J, Takeuchi I, Kato K, Oguri M, Fujimaki T, Horibe H, Yamada Y. Identification of three genetic variants as novel susceptibility loci for body mass index in a Japanese population. Physiol Genomics. 2018;50(3):179-189.

Tables
Table 1 Candidate genes putatively selected by three statistical methods ($F_{ST}$, Pi and XP-CLR) affecting body size traits in DBP.
| Method  | Gene   | Chr. | Ensembl ID       | Summary of gene function                                                                 |
|---------|--------|------|------------------|-----------------------------------------------------------------------------------------|
| F<sub>ST</sub> (top 1%) | NELL1  | 7    | ENSECAG00000024835 | Cell differentiation and cell proliferation [32]                                         |
|         | FGFR1  | 27   | ENSECAG00000015006 | Bone growth and skeletal development                                                     |
|         | SNTG1  | 9    | ENSECAG00000000087 | Body measurement traits [51]                                                            |
|         | BMP2   | 22   | ENSECAG00000021201 | Bone physiology and metabolism                                                          |
|         | TBX15  | 5    | ENSECAG00000023325 | Skeletal muscle and muscle metabolism                                                   |
| Pi (top 1%) | NELL1  | 7    | ENSECAG00000024835 | Cell differentiation and cell proliferation [32]                                         |
|         | FGFR1  | 27   | ENSECAG00000015006 | Bone growth and skeletal development                                                     |
|         | TRHDE  | 28   | ENSECAG00000009284 | Growth Traits [53]                                                                     |
|         | CAPN7  | 16   | ENSECAG00000004932 | Growth Traits [54]                                                                     |
|         | GALNT10| 14   | ENSECAG00000006979 | Body mass index [55]                                                                   |
|         | PDE1B  | 6    | ENSECAG00000022552 | Muscle growth [56]                                                                     |
|         | ARPP21 | 16   | ENSECAG00000022872 | Body size traits [57]                                                                  |
|         | FAM210A| 8    | ENSECAG00000007436 | Bone and muscle structure [58]                                                          |
|         | CNN3   | 5    | ENSECAG00000020188 | Skeletal muscle development [59]                                                        |
|         | SWT1   | 5    | ENSECAG00000012085 | Carcass weight [60]                                                                    |
|         | PRDM16 | 2    | ENSECAG00000017045 | Body weight [61]                                                                       |
| XP-CLR (top 1%) | NELL1  | 7    | ENSECAG00000024835 | Cell differentiation and cell proliferation [32]                                         |
|         | SNTG1  | 9    | ENSECAG00000000087 | Body measurement traits [51]                                                            |
|         | BMP2   | 22   | ENSECAG00000021201 | Bone physiology and metabolism                                                          |
|         | TRHDE  | 28   | ENSECAG00000009284 | Growth Traits [53]                                                                     |
|         | IGF2BP2| 19   | ENSECAG00000020685 | Embryonic development [62]                                                              |
|         | PRKG2  | 3    | ENSECAG00000024387 | Dwarfism [63]                                                                           |
|         | ADAMTS17| 1   | ENSECAG00000000579 | Human height [64], Development [65]                                                      |
|         | SH2B2  | 13   | ENSECAG00000024201 | Growth performance [66]                                                                 |
|         | PLXDC2 | 29   | ENSECAG00000017520 | Body size traits [67]                                                                   |
|         | TN53   | 4    | ENSECAG00000020052 | Bone length [67]                                                                        |
|         | AGTPBP1| 23   | ENSECAG00000019812 | Body mass index [68]                                                                    |

Figures
Figure 1

Phylogenetic analyses. (A) Phylogenetic tree was built based on weighted neighboring-joining (NJ) method, using Mongolian horse as outgroup. (B) PCA. (C). Population structure by Admixture program with K=2 to 8 (K=2 best). AKH, Akhal-Teke; ARB, Arabian; AMH, American Miniature Horse; DBP, Debao pony; FCH, Franches-Montagnes; HLS, Holsteiner; HNV, Hanoverian; MNG, Mongolian; MNM, Mangalarga Marchador; PCH, Percheron; QRT, Quarter Horse; SOR, Sorraia; STB, Standardbred; THB, Thoroughbred; TWH, Tennessee Walking Horse and Yakutian, YKT.
Figure 2

(A) Finestructure. A heat map of a co-ancestry matrix generated by chromosome painting with fineSTRUCTURE. The color of each cell represents the expected number of ‘chunks’ imported from a donor genome (column) to a recipient genome (row). (B) Runs of homozygosity (ROH). (C) Linkage disequilibrium (LD) decay.
Figure 3

Genomic landscape of population differentiation by FST (A) and XP-CLR (B).
Figure 4

Positive selection on NELL1 gene, (A) haplotype across NELL1 gene, (B) FST and (C) pi.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Additional file 1.xlsx
Additional file 2.docx