

Genome-Wide Detection of Selection Signatures in Duroc Revealed Candidate Genes Relating to Growth and Meat Quality

Jian Yu,¹ Pengju Zhao,¹ Xianrui Zheng, Lei Zhou, Chuduan Wang, and Jian-Feng Liu²

National Engineering Laboratory for Animal Breeding; Key Laboratory of Animal Genetics, Breeding, and Reproduction, Ministry of Agriculture; College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China
ORCID ID: 0000-0002-5766-7864 (J.-F.L.)

ABSTRACT With the development of high-throughput genotyping techniques, selection signatures in the genome of domestic pigs have been extensively interrogated in the last decade. The Duroc, a major commercial pig breed famous for its fast growth rate and high lean ratio, has not been extensively studied focusing on footprints of intensively artificial selection in their genomes by a lot of re-sequencing data. The goal of this study was to investigate genomic regions under artificial selection and their contribution to the unique phenotypic traits of the Duroc using whole-genome resequencing data from 97 pigs. Three complementary methods (*d_i*, CLR, and iHH12) were implemented for selection signature detection. In Total, 464 significant candidate regions were identified, which covered 46.4 Mb of the pig genome. Within the identified regions, 709 genes were annotated, including 600 candidate protein-coding genes (486 functionally annotated genes) and 109 lncRNA genes. Genes undergoing selective pressure were significantly enriched in the insulin resistance signaling pathway, which may partly explain the difference between the Duroc and other breeds in terms of growth rate. The selection signatures identified in the Duroc population demonstrated positive pressures on a set of important genes with potential functions that are involved in many biological processes. The results provide new insights into the genetic mechanisms of fast growth rate and high lean mass, and further facilitate follow-up studies on functional genes that contribute to the Duroc's excellent phenotypic traits.

KEYWORDS

Whole-genome sequencing
Duroc pig
Selection signatures

Pigs (*Sus scrofa*) play an important role economically, supplying energy and abundant animal protein for humans. In addition, they also can serve as an ideal animal model for research on human diseases, because they have similar physiological characteristics to humans (Lunney 2007). Pigs were domesticated in Europe and Asia approximately 9000 years ago from the Eurasian wild boar

(Larson *et al.* 2005). Since then, the pig has experienced natural and artificial selection to meet the increased human demand for food, which has reduced genetic diversity and increased linkage disequilibrium in selected regions (Ai *et al.* 2013; Akey *et al.* 2010). Intense artificial selection has led to different pig breeds with unique, distinct phenotypic and physiological features (Ai *et al.* 2015; Bosse *et al.* 2012; Frantz *et al.* 2015). With the development of genome-wide genotyping and sequencing technology, researchers can readily detect variations in genomic distribution and use state-of-the-art statistical methods to detect selection signatures in order to explore the genetic mechanisms underlying the unique Duroc phenotype caused by selection (Mokhber *et al.* 2018; Moon *et al.* 2015).

There are at least 16 species in the Suidae family around the world, and most of them are distributed in Europe and Asia (Giuffra *et al.* 2000; Li *et al.* 2012). Of the pig breeds, the Duroc is a well-known domesticated pig breed because of its superior performance in growth, carcass, and feed conversion efficiency due to intense artificial selection for more than 100 years ago (Cameron and Enser 1991). Therefore, the

Copyright © 2020 Yu *et al.*

doi: <https://doi.org/10.1534/g3.120.401628>

Manuscript received March 5, 2020; accepted for publication August 10, 2020; published Early Online August 28, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.11944890>.

¹These authors contributed equally to this work.

²Corresponding author: College of Animal Science and Technology, Room 455, China Agricultural University (West Campus), No.2 Yuanmingyuan West Road, Beijing 100193, China. E-mail: liujf@cau.edu.cn

Duroc is often used as the terminal sires, *i.e.*, used as parental sires in the three-way crossbreeding system, to improve the lean percentage and growth rate of hybrid pigs in pig industry. These selection processes have left a unique footprint in genomic regions causing the excellent production traits of the Duroc. So far, a few studies have focused on the detection of selection signatures in the Duroc genome, and there are limited indications of potential functional genes for its superior growth and carcass characteristics (Edea *et al.* 2017; Yang *et al.* 2014).

We suspected that the difference in phenotype between Duroc and other pig breeds may be due to different selection pressure in the Duroc genome, which causing genetic mutation in some important genes associated with growth and carcass quality. Detecting the mutations would facilitate future breeding of Durocs to improve these traits through genomic selection. Consequently, large-scale resequencing was performed in 31 unrelated Duroc and other domesticated pigs to detect genomic evidence for the excellent productive phenotype of the Duroc and to identify promising genes for future application in pig breeding.

MATERIALS AND METHODS

Sample collection and sequencing

In this study, 63 samples were collected from different areas in China, which included 32 unrelated Meishan samples from Kunshan, Jiangsu Province, and 31 unrelated Duroc samples from Yancheng, Jiangsu Province. DNA was extracted from ear tissue using the Qiagen DNeasy Tissue kit (Qiagen, Germany), and then analyzed by spectrophotometry and agarose gel electrophoresis to ensure high quality. Subsequently, all samples were constructed from a library (paired-end, 2×125 bp) and sequenced using an Illumina HiSeq 2000 sequencing system at Novogene (Beijing, China). Finally, a total of 1,403.35 Gb of sequence data were generated, which were used for subsequent analysis. In addition, genomic data of 4 Duroc and 30 Tibetan wild boars (Li *et al.* 2013; Zhao *et al.* 2016) were downloaded from NCBI with accession numbers ERP001813 and SRA065461, respectively. In total, there was whole-genome sequencing data for 32 Meishan samples, 35 Duroc samples, and 30 Tibetan pig samples in this study.

Reads quality control and mapping

For raw reads, three quality control criteria were checked to ensure high accuracy in reads mapping using the NGSQC Toolkit (Patel and Jain 2012): First, trimming reads contained adapter sequences; second, undesirable reads contained no less than 30% low-quality bases were discarded (quality value ≤ 20 , or N bases); and third, low-quality 3' end reads whose base quality score was lower than 20 were modified. Subsequently, the clean reads (including all the autosomes and the X chromosome) were aligned to the swine reference genome (*Sus scrofa* 11.1ftp://ftp.ensembl.org/pub/release-99/fasta/sus_scrofa/dna/) using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) with default parameters. Duplicated reads were removed using Picard tools (http://picard.sourceforge.net). In addition, base quality recalibration and mapping statistics (such as sequence depth) were implemented for subsequent analyses using Genome Analysis Toolkit (GATK) (McKenna *et al.* 2010) and SAMtools (Li *et al.* 2009) respectively.

Single nucleotide variants (SNVs) calling and annotation

After alignment, SAMtools was then used to convert SAM into the best alignments in BAM format. The 'SortSam', 'MarkDuplicates',

'BaseRecalibrator' of GATK were used separately to sorting, to remove potential PCR duplications and for base quality score recalibrations. The BAM files were indexed by SAMtools. The 'HaplotypeCaller', 'GenomicsDBImport' and 'GenotypeGVCFs' of GATK with default parameters were used to call SNPs, which generated genotype calls in Variant Call Format (VCF). We further removed those INDELS in the VCF file using 'SelectVariants' of GATK. Finally, the GATK 'VariantFiltration' module was used to exclude SNP calling errors in according to the following criteria (McKenna *et al.* 2010). High-quality SNPs with Quality score > 30 , MQ RMS mapping quality > 20 , DP > 5 , coverage $> 30\%$, and MAF > 0.01 were kept for further analysis. In addition, the dbSNP database (Sherry *et al.* 2001) was used to identify novel genetic variations. Finally, All filtered SNPs were functionally annotated using the gene-based annotation modules of ANNOVAR software (Wang *et al.* 2010), and the corresponding gene annotation file was downloaded from the Ensembl database (https://hgdownload.soe.ucsc.edu/goldenPath/susScr11/bigZips/genes/susScr11.ensGene.gtf.gz). In the annotation step, SNVs were classified into eight categories based on their genome locations, including exonic regions (synonymous, nonsynonymous, stop gain and stop loss), splicing sites, intronic regions, 5' and 3' untranslated regions (UTRs), upstream and downstream regions, and intergenic regions.

Selection signature detection of the Duroc population

To detect the genomic region under selection in the Duroc, three complementary methods were implemented to ascertain the selection signatures. First, the population differentiation coefficient (F_{ST}) (Holsinger and Weir 2009) was used to measure population differentiation between the Duroc and the other two populations (Meishan and Tibetan wild boars) using non-overlapping 100kb windows. The F_{ST} was further translated into the d_i statistic (Akey *et al.* 2010) for each SNP, *i.e.*, $d_i = \sum_{i \neq j} \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{sd[F_{ST}^{ij}]}$, where $E[F_{ST}^{ij}]$ and $sd[F_{ST}^{ij}]$ represent

the expected value and standard deviation of F_{ST} between breeds i (Duroc) and j (Meishan or Tibetan wild boars) calculated from 18 autosomes. The d_i values for Duroc vs. Meishan and Duroc vs. Tibetan for each locus were summed, and the d_i values were averaged over SNVs in non-overlapping 100-kb windows. We empirically selected extremely high d_i values in the 5% right-tail as potential candidate regions under positive selection.

Second, a recently reported method, Integrated Haplotype Homozygosity Pooled (iHH12)(Torres *et al.* 2018), was used to detect selection signatures in the Duroc genome. Beagle software (Browning and Browning 2007) was used to impute and phase the missing genotypes of the Duroc (with default settings) before iHH12 was calculated. The iHH12 method was developed based on a window-based statistic (H12) (Garud *et al.* 2015) with good power to detect soft sweeps by adapting H12 into an integrated haplotype homozygosity framework. For this statistic, EHH12 of the entire sample of haplotypes from different loci was first calculated. Based on that, iHH12 was calculated. And We calculated iHH12 values for each loci using Selscan software (Szpiech and Hernandez 2014). Finally, the iHH12 values are normalized across the whole genome, *i.e.*, $iHH12 = \frac{iHH12 - E[iHH12]}{sd[iHH12]}$. The genome was divided into non-overlapping 100kb windows. We empirically selected extremely high iHH12 values in the 5% right-tail as potential candidate regions under positive selection.

Lastly, the composite likelihood ratio (CLR) test (Pavlidis *et al.* 2013) was implemented to search for a recent selection sweep caused by a change in allele frequency at linked sites, which generates

an excess of derived alleles at high frequencies and reducing genetic diversity (Fay and Wu 2000; Kim and Stephan 2002). CLR values were calculated for sites every 1kb across the genome using SweepFinder2 software (DeGiorgio *et al.* 2016), and the genome was divided into non-overlapping 100kb windows. We empirically selected extremely high CLR values in the 5% right-tail as potential candidate regions under positive selection. To further control the false positive rates of the detection of selection signatures, genomic regions identified by at least two methods were used in further analyses.

Gene annotation and enrichment analysis

All identified candidate regions were annotated to further identify potential candidate genes based on the Ensembl Genes 100 Database (<http://asia.ensembl.org/>). Subsequently, these candidate genes were transformed into homologous identical genes in the human genome using the BioMart database (<http://asia.ensembl.org/biomart/martview/>) (Durinck *et al.* 2005), due to incomplete gene function annotations in the pig genome. We further investigated the function of annotated genes undergoing positive selection by Gene Ontology (GO) term analysis and KEGG pathway analysis using KOBAS (Wu *et al.* 2006) to reveal the potential biological function of candidate genes harbored in selected regions. Furthermore, we downloaded the published pig QTL from the Pig QTLdb (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=bedSS_11.1) to identify QTL overlapping with candidate selected regions in the current study based on the putative QTL.

Data availability

A total 63 pig samples with 1,403.35Gbases were uploaded to NCBI with BioProject ID: PRJNA378496. Illumina paired-end sequences for the other 34 pigs used in this study were downloaded from NCBI with accession numbers ERP001813 and SRA065461. Supplemental material available at figshare: <https://doi.org/10.25387/g3.11944890>.

RESULTS

Whole-genomic resequencing and SNV detection

Whole-genome resequencing of 63 samples generated 1,403.35 G raw paired-end reads. After quality control, the clean reads were aligned to the *sus scrofa* 11.1 reference genome (ftp://ftp.ensembl.org/pub/release-99/fasta/sus_scrofa/dna/) using BWA. The sequence depth for each sample ranged from 3.92X to 13.43X, with an average of > 9.35X (Table S1). Subsequently, we called SNVs on a population level. Finally, 37,438,998 SNVs were we identified in Duroc genome. After filtration of the raw variants, 10,398,223 high-quality SNVs were kept for subsequent analyses. Approximately 87% of the variants (8,969,913 SNVs) were found in the dbSNP database, whereas more than 13% (1,428,310 SNVs) of the variants that we identified were absent from the SNPs database. These novel SNVs substantially expand the database of Duroc genetic variants. Genome-wide distribution of detected SNVs for Duroc were showed in Figure 1.

All detected SNVs in the Duroc population were annotated using a gene annotation file(<https://hgdownload.soe.ucsc.edu/goldenPath/susScr11/bigZips/genes/susScr11.ensGene.gtf.gz>) downloaded from

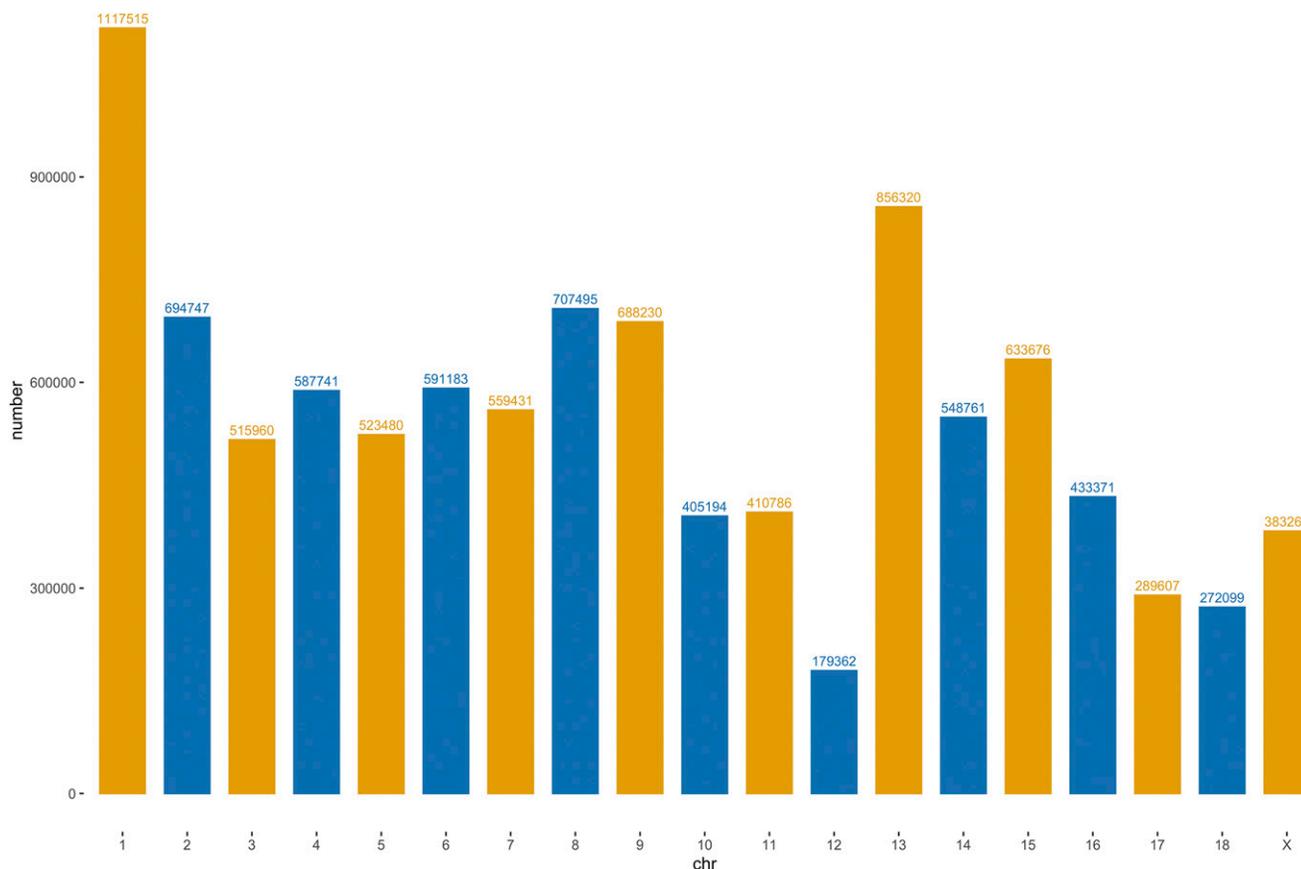


Figure 1 Genome-wide distribution of detected SNVs on 19 chromosomes for the Duroc. X-axis represents 18 autosomes and X chromosome. Y-axis represents the number of SNVs.

the Ensembl database. Most of the SNVs were located in intergenic regions (~51.06%), followed by introns (~44.76%), upstream and downstream regions (~1.32%), exons (~1.59%), untranslated regions (UTRs) (~1.24%), and splicing site regions (~0.004%) (Figure 2). Further research on exonic regions detected 15,198 synonymous mutations, 14,177 non-synonymous mutations, 303 stop-gain mutations, and 49 stop-loss mutations. In addition, the results of functional annotation were used to identify the causal mutants in these candidate genes. Studying the functional variations provides an opportunity to explain the genetic mechanism of the features of the Duroc due to artificial selection.

Detection of selection signatures

Our previous study showed that strong linkage disequilibrium (LD) levels in the Duroc compared to the other breeds because of intense artificial selection (Zhao *et al.* 2018). Based on these findings, it was speculated that there should be specific selection signatures in the Duroc population arising from long-term artificial selection after domestication. Therefore, positive selection signature detection was performed on the genome of Duroc pigs with large-scale resequencing data combining the three complementary methods with non-overlapping windows. Across the genome, d_i values for Duroc to Meishan and Tibetan wild boar populations with non-overlapping 100-kb windows ranged from -1.89 to 4.51, with an average of 0.087. The genome-wide distribution of the averaged d_i values for 100-kb non-overlapping windows is shown in Figure 3. Extremely high d_i values ($d_i > 1.45$) located at the 5% right tail of the empirical distribution were selected as potential candidate regions of artificial selection in the genome of the Duroc.

Furthermore, two within-population methods (CLR and iHH12) were implemented to detect the selection signatures of Duroc. The genome-wide distribution of CLR values for 100-kb non-overlapping windows is shown in Figure 4. The CLR values for these windows ranged from ~0 to 4,080, with an average of 28.27. Extremely high CLR values (CLR > 81.85) at the 5% right tail of the empirical distribution were selected as potential candidate signals in the Duroc. For iHH12 detection, extremely high iHH12 values (iHH12 > 2.81) at the 5% right tail of the empirical distribution were selected

as potential candidate signals for positive selection in the Duroc. Accordingly, the iHH12 values for windows ranged from 0 to 10.31, with an average value of 0.27. The iHH12 further confirmed some strong selection signatures on SCC6, SSC14, and SSC15, which were identified in CLR (Figure 5).

In order to reduce false positive candidate regions, potential candidate regions identified by more than two methods were chosen as the final candidate regions for positive selection. The genome distribution of candidate regions detected by different methods is shown in Figure 6. A total of 464 candidate regions (Figure 7) were identified by more than two methods, which were mainly distributed on SSC1, SSC2, and SSC9, covering 46.4 Mb genomic regions. Meanwhile, 27 candidate regions were identified by all three methods.

Gene annotation and functional analysis

All identified candidate regions were annotated to identify functionally important genes relevant to the excellent phenotype of the Duroc, such as high daily gain, meat content in the carcass and feed conversion ratio. Seven hundred and nine genes were identified in candidate regions, including 600 candidate protein code genes (486 functionally annotated genes) (Table S6) and 109 lncRNA genes (Table S7). There were many genes involved in important biological processes, such as *insulin receptor (INSR)*, *insulin-like growth factor 1 receptor (IGF1R)*, and *insulin-like growth factor 2 receptor (IGF2R)* for muscle growth and protein synthesis. Further analysis of these genes revealed that the *INSR* gene (identified by all three methods) has undergone intense selection ($d_i = 3.25$, CLR = 935.59, iHH12 = 4.68). Functional annotation showed that the *INSR* gene contains 114 SNVs, most of which are intronic variations. In particular, a synonymous SNV (ENSSSC-G00000013566:ENSSSCT00000014817:exon12:c.T2640C:p.G880G; ENSSSCG00000013566:ENSSSCT00000014815:exon13:c.T2676C:p.G892G; rs338663742 in the dbSNP database) was found within the *INSR* gene. Intriguingly, this mutation was significantly different between the Duroc with Meishan and Tibetan populations (Table S8).

The functions of these annotated genes undergoing positive selection was investigated further by GO term analysis and KEGG pathway analysis using KOBAS. Considering insufficiency in pig genome gene function annotations, the candidate genes were transformed into

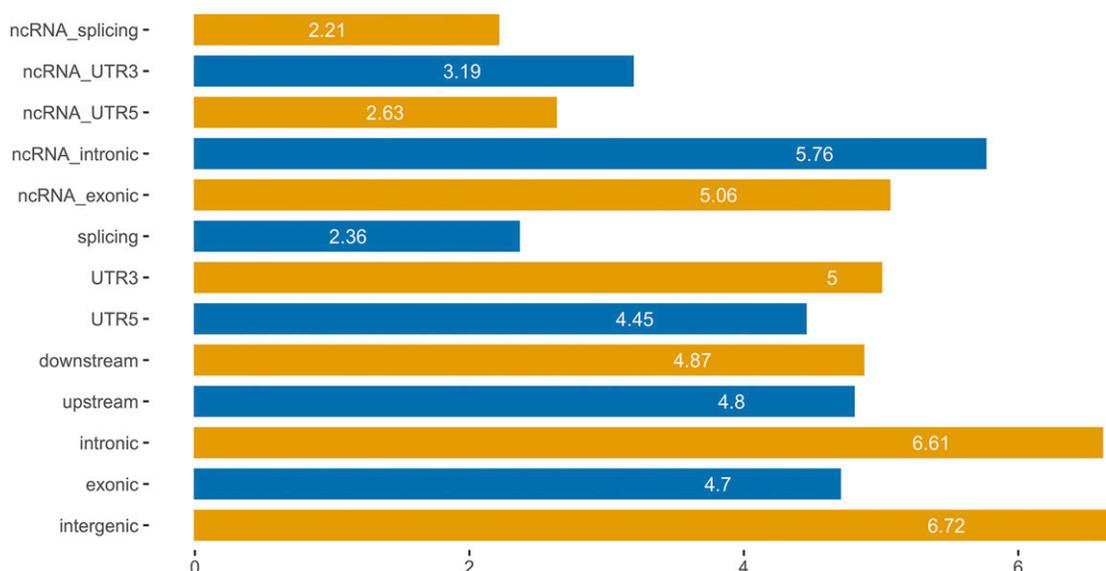


Figure 2 Genome-wide annotation of genetic variations. X-axis represents the number of genetic variations (log10) within various functional regions. Y-axis represents various functional regions in genes.

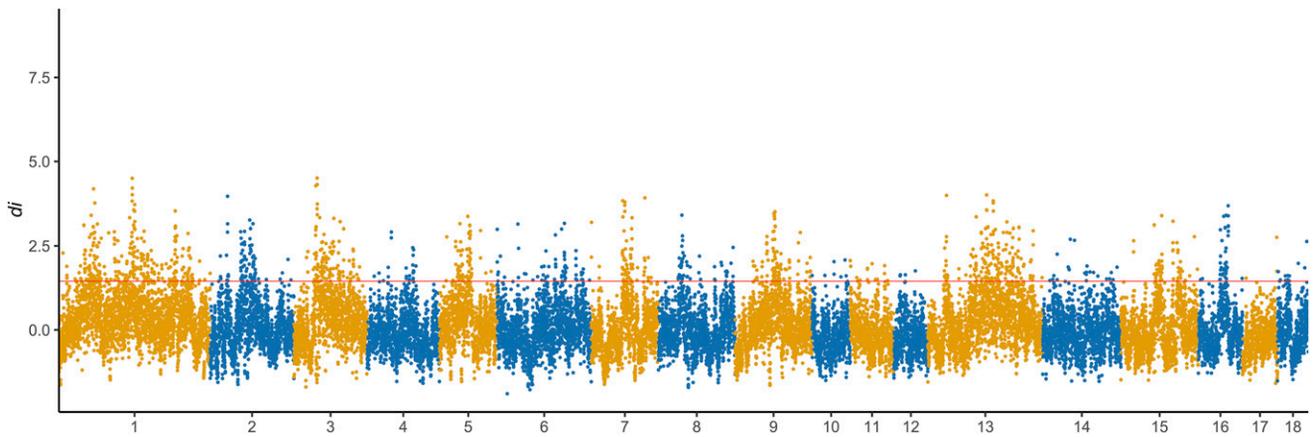


Figure 3 Genome-wide distribution of selection signatures detected by d_i on 18 chromosomes. X-axis represents 18 autosomes, and Y-axis represents d_i values for non-overlapping 100 kb windows. Red line displays the threshold level of 5%.

homologous identical genes in the human genome using the BioMart database (<http://asia.ensembl.org/biomart/martview/>). Finally, a total of 271 significantly enriched GO terms with P-values < 0.05 were observed (Table S9) that were associated with growth, protein synthesis, and metabolism, such as insulin-like growth factor receptor signaling pathway (GO:0048009), protein modification process (GO:0036211), regulation of biological process (GO:0050789), and regulation of protein metabolic process (GO:0051246). In addition, some of these significantly enriched terms were involved in energy metabolism and the immune response.

In addition, 53 significantly enriched KEGG pathway terms were observed (Table S10), including regulation of actin cytoskeleton (hsa04810), insulin resistance signaling pathway (hsa04931), and PI3K-Akt signaling pathway (hsa04151). Seven genes (*INSR*, *PDPK1*, *OGA*, *GFPT2*, *CREB3*, *CREB3L3*, and *PPP1R3E*) were significantly enriched in the insulin resistance signaling pathway, which plays an important role in energy metabolism, protein synthesis, and metabolism in muscle cells. Previous studies have reported that knockdown of *CREB3* in mice results in decreased estradiol and progesterone synthesis and promotes cell proliferation (Zhao *et al.* 2016).

The Pig QTLdb database (Release 41, Apr 26, 2020), which contain 30,580 QTL of 691 different traits reported in the past decades, was downloaded to identify quantitative trait loci (QTL) overlapping with candidate selected regions. In total, 3,404 porcine

QTL (Table S11) were identified to be located within or overlapping with the candidate regions. Of these, 2,665 (~78.29%) were associated with production and carcass traits, suggesting selection for production traits during the breeding of Duroc pigs.

DISCUSSION

Natural and artificial selection has left distinct imprints in the genomes of domestic animals, such as reduced genetic diversity and increased linkage disequilibrium in regions under selection (Ai, Huang, and Ren 2013; Akey *et al.* 2010). Identifying these regions in domesticated animals can help identify causal variants contributing to economic traits. In the current study, a number of regions with signatures of positive selection were identified, and further functional analyses on these regions identified a list of genes involved in fast growth and high lean deposition, supporting that these regions have been under selection. These findings begin to elucidate the genetic mechanism of the excellent phenotypic traits of the Duroc.

Previous studies have shown that it is more powerful to utilize complementation of different methods to detect selection signatures for livestock and poultry (Ma *et al.* 2015a; Qanbari *et al.* 2014). Therefore, three methods based on different sources of information were used to construct a comprehensive selection map for the Duroc genome in our study. Regions identified by at least two methods

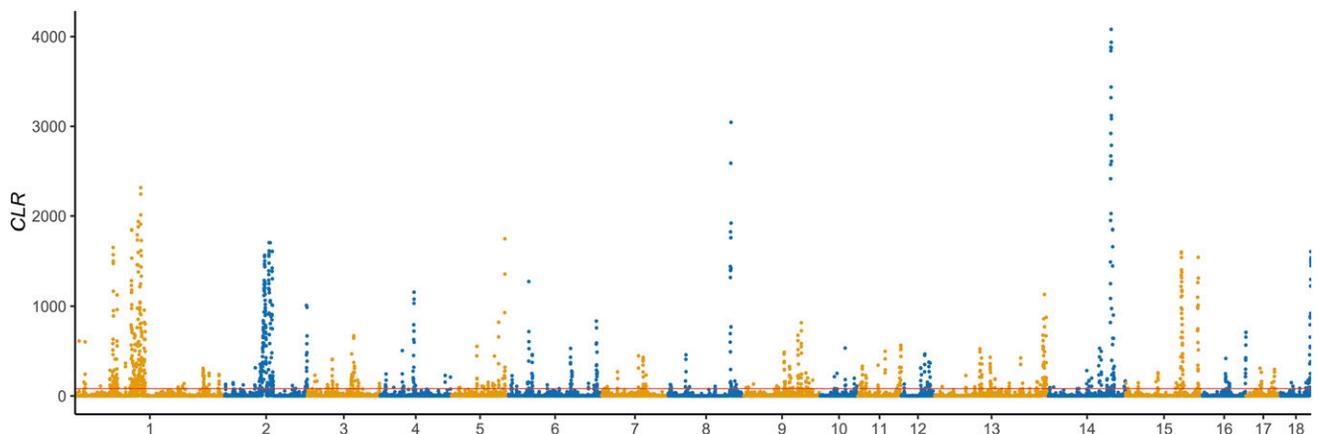


Figure 4 Genome-wide distribution of selection signatures detected by CLR on 18 chromosomes. X-axis represents 18 autosomes, and Y-axis represents CLR values for non-overlapping 100 kb windows. Red line displays the threshold level of 5%.

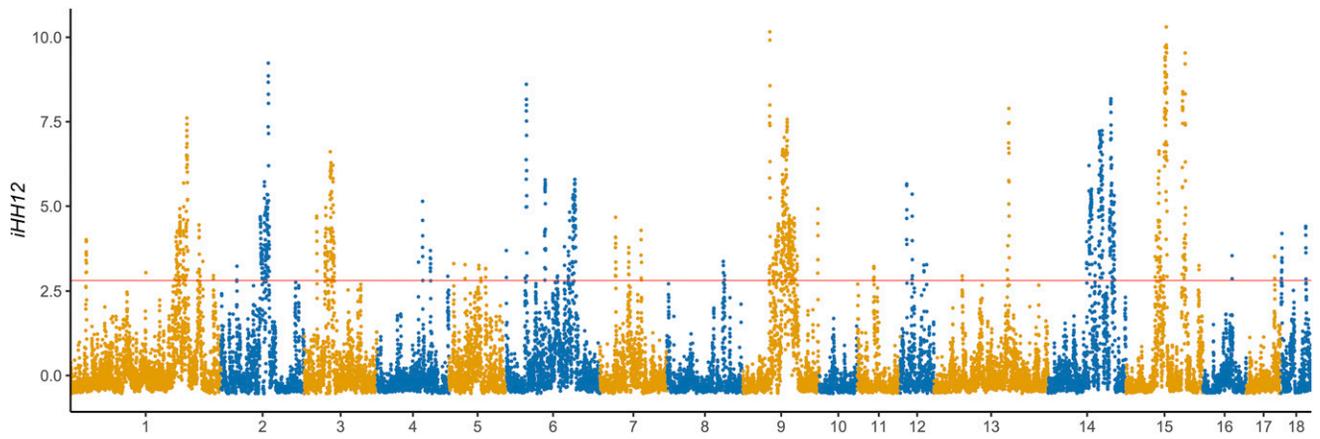


Figure 5 Genome-wide distribution of selection signatures detected by iHH12 on 18 chromosomes. X-axis represents 18 autosomes, and Y-axis represents iHH12 values for non-overlapping 100 kb windows. Red line displays the threshold level of 5%.

were identified as candidate regions of positive selection to reduce false positives. Although some older selection signatures could have been lost in some way, it does make the results more reliable, which was well supported by the QTL identified as overlapping with candidate selected regions. Most QTL (~78.29%) were associated

with production and carcass traits. In addition, most of the regions identified by all three methods had an extremely high signal value in the three methods, whereas the regions identified by only two methods tended to have a high signal value in one particular method. This means that regions identified by only two methods may have higher

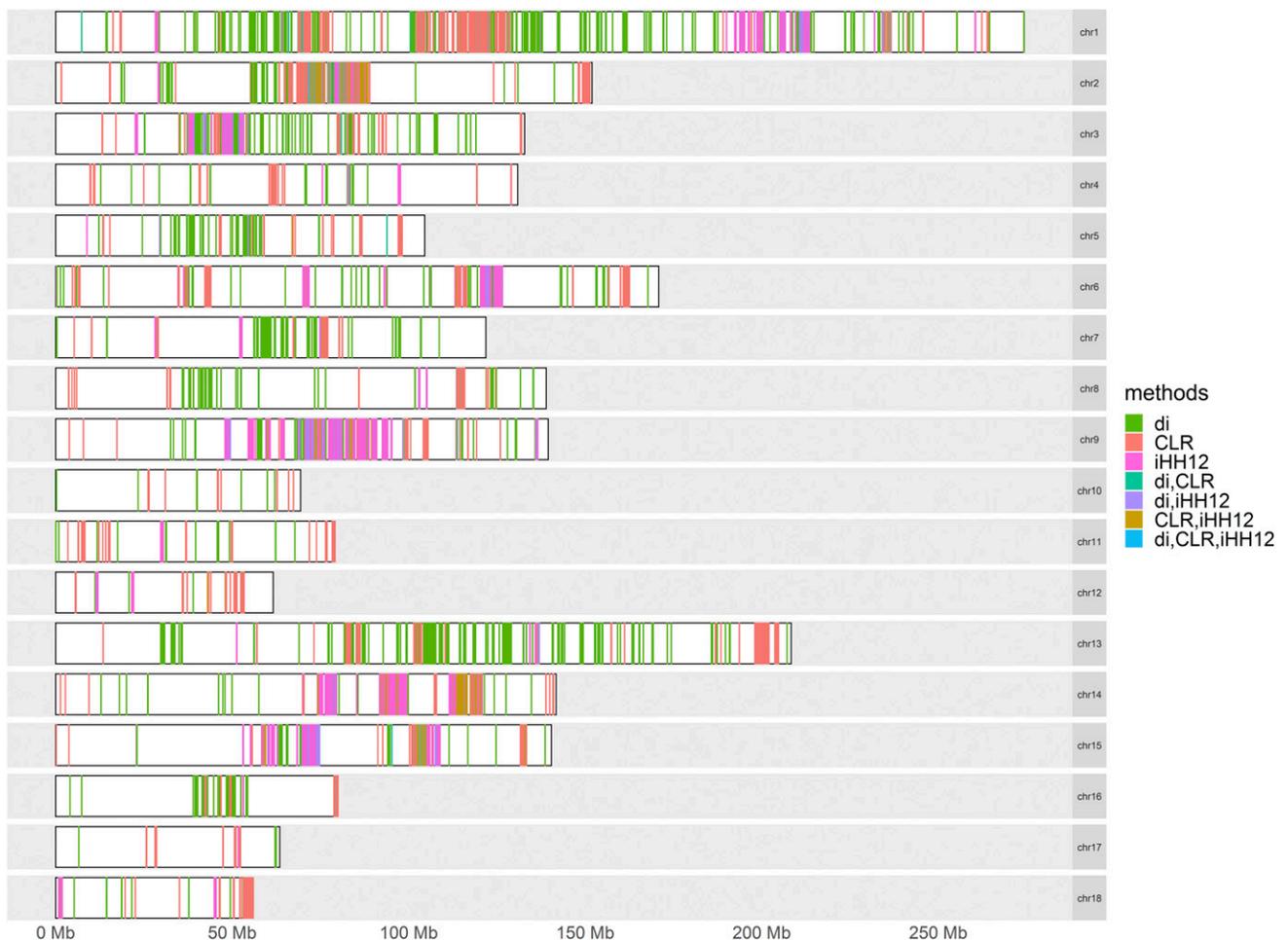


Figure 6 Genomic distribution of selection signatures detected by d_i , CLR and iHH12 on 18 chromosomes. In the figure, " d_i ," "CLR" and "iHH12" represent the regions detected by only one method; " d_i ,CLR", " d_i ,iHH12" and "CLR,iHH12" represent the regions detected by two different methods; " d_i ,CLR,iHH12" represents the regions detected by all three methods.

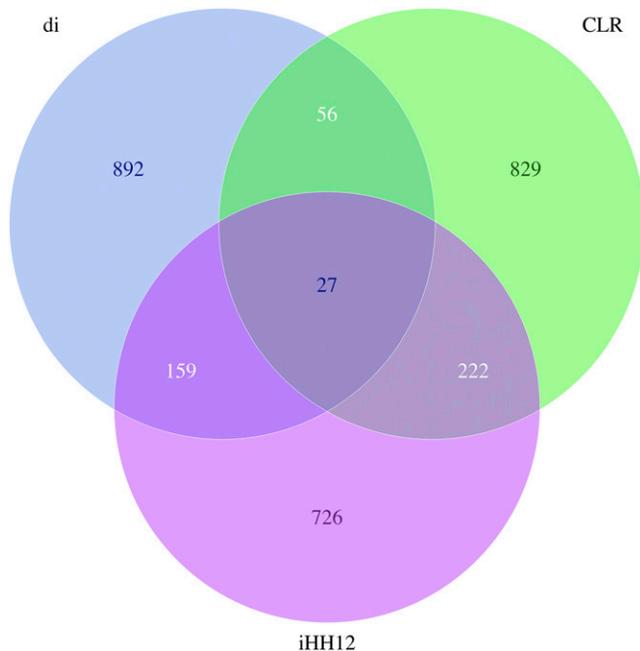


Figure 7 Venn diagram for numbers of candidate regions detected by different methods. Blue represents the numbers of candidate regions detected by di; green represents the numbers of candidate regions detected by CLR; orchid represents the numbers of candidate regions detected by iHH12.

false positive rates than regions identified by all three methods. Previous study also suggested that combining different selection signature statistics can reduce those selective sweeps merely derived from independent domestication processes (Ma *et al.* 2015a). Compared with previous genome-wide scans for selection signatures in Duroc pigs (Edea *et al.* 2017; Diao *et al.* 2018; Ma *et al.* 2018), a number of overlapping signatures were found with previous studies, such as 7.3–7.4 Mb on SSC1, and 88.1–88.4 Mb on SSC2. These regions contain some important genes, such as *IGF2R*, *HOMER1* and *CMYA5*, which are associated with muscle development, carcass traits and meat quality (Hao *et al.* 2017; Xu *et al.* 2011). Some novel candidate regions containing genes associated with growth and carcass quality, such as *INSR* and *IGF1R*, were also found in the current study. These findings may have led to a gap in detecting selective sweeps in Duroc pigs.

Compared with the Duroc, the Meishan and Tibetan wild boar have an extremely slow growth rate and low lean meat ratio because of various breeding goals during long-term domestication, which has made a huge difference in their genomes. Therefore, this research design using Meishan and Tibetan wild pigs as reference populations has facilitated the accurate identification of genetic mechanisms under great phenotypic difference. Of the candidate genes, *INSR*, *IGF1R*, and *IGF2R*, are of particular significance, as they are synonymous with growth regulation and protein metabolism also seen in the Duroc. In particular, a synonymous exonic SNV within *INSR* gene was found, which was identified by all three methods with extremely high signal values. In the meanwhile, a novel heterozygous nonsense mutation in exon 2 of *INSR* was found in humans in a previous report (Saito-Hakoda *et al.* 2018), which causes type A insulin resistance. Interestingly, all Duroc pigs were homozygous wild type genotypes, and heterozygous mutations existed only in the Meishan and Tibetan pig populations. Therefore, it is suspected that the *INSR* gene may be

a positional candidate gene explaining why the Duroc has excellent production traits. Moreover, *IGF1R* and *IGF2R* also play a crucial role in regulating cell proliferation and muscle growth. Previous studies have shown that *IGF1R* affects muscle growth and protein turnover (Schiaffino and Mammucari 2011; Meek *et al.* 1998) by regulating insulin and *IGF-1* to stimulate muscle protein synthesis (Fulks *et al.* 1975; Rommel *et al.* 2001) and inhibit protein degradation via the ubiquitin-proteasome and autophagy-lysosome pathways (Mammucari *et al.* 2007; Sandri *et al.* 2004). It is notable that *IGF2R* has previously been implicated in distinct patterns of selection in domesticated Duroc pigs and Tibetan wild boars (Li *et al.* 2013) and as a possible positional candidate gene detected by F_{ST} and CLR when comparing Duroc pigs with Asian wild boars and European wild boars (Ma *et al.* 2018). It has been reported that a new quantitative trait nucleotide is located in the regulatory sequence of the imprinted *IGF-2* gene, which explains 15–30% of the phenotypic variation in muscle mass and 10–20% of the variation in back-fat thickness (Van Laere *et al.* 2003). Therefore, *INSR*, *IGF1R* and *IGF2R* may be possible positional candidate genes explaining why Duroc pigs have excellent production traits.

Enrichment analysis showed that some important genes were significantly enriched in insulin resistance, which is a common pathophysiological mechanism of obesity and type 2 diabetes mellitus in humans (DeFronzo 1997; DeFronzo *et al.* 1985). Skeletal muscle is one of the major target organs of insulin-mediated glucose uptake, metabolism and utilization, and it is the earliest and most important site of insulin resistance. Recent studies have reported that the impairment of glucose uptake, insulin signaling pathway and mitochondrial biosynthesis are closely related to skeletal muscle insulin resistance (Asmann *et al.* 2006; Chavez-Velazquez *et al.* 2007; Toledo *et al.* 2006; Abdul-Ghani *et al.* 2008). Seven candidate genes (*INSR*, *PDPK1*, *OGA*, *GFPT2*, *CREB3*, *CREB3L3*, and *PPP1R3E*) are involved in this biological process, which is closely related to cell proliferation and metabolism. These candidate genes can elucidate domestication history in Duroc populations, and potentially offer new ways to explore evolutionary mechanisms.

In addition to these findings associated with growth and carcass quality, some important functional genes associated with immunity and transcription (such as *PALM*, *PTBP1*, and *CRHBP*) are also detected under selection pressure between European pigs and Asian pigs (Ma *et al.* 2015b; Moon *et al.* 2015), which may explain the difference in adaptability between Duroc and Asian pigs.

CONCLUSION

In summary, we identified selection signatures in the genome of Duroc pig, a well-known European domestic pig breed. We found some promising genes associated with fast growth rate and high lean ratio of Duroc. These results provide novel insights into the genome evolution and selection mechanisms in domesticated pig breeds.

ACKNOWLEDGMENTS

The whole sample collection and treatment were conducted in strict accordance with the protocol approved by the Institutional Animal Care and Use Committee (IACUC) of China Agricultural University (permit number: DK996). This work was supported by the National Natural Science Foundation of China (31661143013, 31790414-1), the Natural Science Foundation of Beijing, China (6192010). J-F.L. conceived and designed the experiments. P.Z. and J.Y. performed SNVs prediction and population analyses. X.Z. collected samples and prepared for sequencing. J.Y., P.Z., C.W. and J-F.L. wrote and revised the paper. All authors read and approved the final manuscript.

LITERATURE CITED

- Abdul-Ghani, M., R. Jani, M. Molina, A. Chaves, D. Tripathy *et al.*, 2008 Decreased mitochondrial ATP synthesis in obese non-diabetic subjects. *Diabetes* 57: A365–A366.
- Ai, H., X. Fang, B. Yang, Z. Huang, H. Chen *et al.*, 2015 Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47: 217–225. <https://doi.org/10.1038/ng.3199>
- Ai, H. S., L. S. Huang, and J. Ren, 2013 Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. *PLoS One* 8: e56001. <https://doi.org/10.1371/journal.pone.0056001>
- Akey, J. M., A. L. Ruhe, D. T. Akey, A. K. Wong, C. F. Connelly *et al.*, 2010 Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. USA* 107: 1160–1165. <https://doi.org/10.1073/pnas.0909918107>
- Asmann, Y. W., C. S. Stump, K. R. Short, J. M. Coenen-Schimke, Z. Guo *et al.*, 2006 Skeletal muscle mitochondrial functions, mitochondrial DNA copy numbers, and gene transcript profiles in type 2 diabetic and nondiabetic subjects at equal levels of low or high insulin and euglycemia. *Diabetes* 55: 3309–3319. <https://doi.org/10.2337/db05-1230>
- Bosse, M., H. J. Megens, O. Madsen, Y. Paudel, L. A. Frantz *et al.*, 2012 Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 8: e1003100. <https://doi.org/10.1371/journal.pgen.1003100>
- Browning, S. R., and B. L. Browning, 2007 Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81: 1084–1097. <https://doi.org/10.1086/521987>
- Cameron, N. D., and M. B. Enser, 1991 Fatty acid composition of lipid in Longissimus dorsi muscle of Duroc and British Landrace pigs and its relationship with eating quality. *Meat Sci.* 29: 295–307. [https://doi.org/10.1016/0309-1740\(91\)90009-F](https://doi.org/10.1016/0309-1740(91)90009-F)
- Chavez-Velazquez, A., R. Jani, M. Abdul-Ghani, V. Frohlich, P. Sathya-Narayana *et al.*, 2007 Short-term elevation of plasma free fatty acids (FFA) decreases skeletal muscle mitochondrial membrane potential in healthy glucose tolerant subjects. *Diabetes* 56: A63.
- DeFronzo, R. A., 1997 Pathogenesis of type 2 diabetes: metabolic and molecular implications for identifying diabetes genes. *Diabetes Rev. (Alex.)* 5: 177–269.
- DeFronzo, R. A., R. Gunnarsson, O. Bjorkman, M. Olsson, and J. Wahren, 1985 Effects of Insulin on Peripheral and Splanchnic Glucose-Metabolism in Noninsulin-Dependent (Type-II) Diabetes-Mellitus. *J. Clin. Invest.* 76: 149–155. <https://doi.org/10.1172/JCI111938>
- DeGiorgio, M., C. D. Huber, M. J. Hubisz, I. Hellmann, and R. Nielsen, 2016 SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics* 32: 1895–1897. <https://doi.org/10.1093/bioinformatics/btw051>
- Diao, S., Y.-Y. Luo, Y.-L. Ma, X. Deng, Y.-T. He *et al.*, 2018 Genome-wide detection of selective signatures in a Duroc pig population. *Journal of Integrative Agriculture* 17: 2528–2535. [https://doi.org/10.1016/S2095-3119\(18\)61984-7](https://doi.org/10.1016/S2095-3119(18)61984-7)
- Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor *et al.*, 2005 BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21: 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>
- Edea, Z., J. K. Hong, J. H. Jung, D. W. Kim, Y. M. Kim *et al.*, 2017 Detecting selection signatures between Duroc and Duroc synthetic pig populations using high-density SNP chip. *Anim. Genet.* 48: 473–477. <https://doi.org/10.1111/age.12559>
- Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
- Frantz, L. A., J. G. Schraiber, O. Madsen, H. J. Megens, A. Cagan *et al.*, 2015 Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47: 1141–1148. <https://doi.org/10.1038/ng.3394>
- Fulks, R. M., J. B. Li, and A. L. Goldberg, 1975 Effects of insulin, glucose, and amino acids on protein turnover in rat diaphragm. *J. Biol. Chem.* 250: 290–298.
- Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent Selective Sweeps in North American *Drosophila melanogaster* Show Signatures of Soft Sweeps. *PLoS Genet.* 11: e1005004. <https://doi.org/10.1371/journal.pgen.1005004>
- Giuffra, E., J. M. Kijas, V. Amarger, O. Carlborg, J. T. Jeon *et al.*, 2000 The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* 154: 1785–1791.
- Hao, X., G. Plastow, C. Zhang, S. Xu, Z. Hu *et al.*, 2017 Genome-wide association study identifies candidate genes for piglet splay leg syndrome in different populations. *BMC Genet.* 18: 64. <https://doi.org/10.1186/s12863-017-0532-4>
- Holsinger, K. E., and B. S. Weir, 2009 Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat. Rev. Genet.* 10: 639–650. <https://doi.org/10.1038/nrg2611>
- Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160: 765–777.
- Larson, G., K. Dobney, U. Albarella, M. Fang, E. Matisoo-Smith *et al.*, 2005 Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307: 1618–1621. <https://doi.org/10.1126/science.1106927>
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, M., S. Tian, L. Jin, G. Zhou, Y. Li *et al.*, 2013 Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45: 1431–1438. <https://doi.org/10.1038/ng.2811>
- Li, Y., S. Mei, X. Zhang, X. Peng, G. Liu *et al.*, 2012 Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics* 13: 725. <https://doi.org/10.1186/1471-2164-13-725>
- Lunney, J. K., 2007 Advances in swine biomedical model genomics. *Int. J. Biol. Sci.* 3: 179–184. <https://doi.org/10.7150/ijbs.3.179>
- Ma, Y., X. Ding, S. Qanbari, S. Weigend, Q. Zhang *et al.*, 2015a Properties of different selection signature statistics and a new strategy for combining them. *Heredity* 115: 426–436. <https://doi.org/10.1038/hdy.2015.42>
- Ma, Y. L., J. L. Wei, Q. Zhang, L. Chen, J. Y. Wang *et al.*, 2015b A Genome Scan for Selection Signatures in Pigs. *PLoS One* 10: e0116850. <https://doi.org/10.1371/journal.pone.0116850>
- Ma, Y., S. Zhang, K. Zhang, C. Fang, S. Xie *et al.*, 2018 Genomic Analysis To Identify Signatures of Artificial Selection and Loci Associated with Important Economic Traits in Duroc Pigs. *G3 (Bethesda)* 8: 3617–3625 (Bethesda). <https://doi.org/10.1534/g3.118.200665>
- Mammucari, C., G. Milan, V. Romanello, E. Masiero, R. Rudolf *et al.*, 2007 FoxO3 controls autophagy in skeletal muscle in vivo. *Cell Metab.* 6: 458–471. <https://doi.org/10.1016/j.cmet.2007.11.001>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20: 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Meek, S. E., M. Persson, G. C. Ford, and K. S. Nair, 1998 Differential regulation of amino acid exchange and protein dynamics across splanchnic and skeletal muscle beds by insulin in healthy human subjects. *Diabetes* 47: 1824–1835. <https://doi.org/10.2337/diabetes.47.12.1824>
- Mokhber, M., M. Moradi-Shahrbabak, M. Sadeghi, H. Moradi-Shahrbabak, A. Stella *et al.*, 2018 A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. *BMC Genomics* 19: 449. <https://doi.org/10.1186/s12864-018-4759-x>
- Moon, S., T. H. Kim, K. T. Lee, W. Kwak, T. Lee *et al.*, 2015 A genome-wide scan for signatures of directional selection in domesticated pigs. *BMC Genomics* 16: 130. <https://doi.org/10.1186/s12864-015-1330-x>

- Patel, R. K., and M. Jain, 2012 NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7: e30619. <https://doi.org/10.1371/journal.pone.0030619>
- Pavlidis, P., D. Zivkovic, A. Stamatakis, and N. Alachiotis, 2013 SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol. Biol. Evol.* 30: 2224–2234. <https://doi.org/10.1093/molbev/mst112>
- Qanbari, S., H. Pausch, S. Jansen, M. Somel, T. M. Strom *et al.*, 2014 Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.* 10: e10004148. <https://doi.org/10.1371/journal.pgen.1004148>
- Rommel, C., S. C. Bodine, B. A. Clarke, R. Rossman, L. Nunez *et al.*, 2001 Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. *Nat. Cell Biol.* 3: 1009–1013. <https://doi.org/10.1038/ncb1101-1009>
- Saito-Hakoda, A., A. Nishii, T. Uchida, A. Kikuchi, J. Kanno *et al.*, 2018 A follow-up during puberty in a Japanese girl with type A insulin resistance due to a novel mutation in INSR. *Clin. Pediatr. Endocrinol.* 27: 53–57. <https://doi.org/10.1297/cpe.27.53>
- Sandri, M., C. Sandri, A. Gilbert, C. Skurk, E. Calabria *et al.*, 2004 Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. *Cell* 117: 399–412. [https://doi.org/10.1016/S0092-8674\(04\)00400-3](https://doi.org/10.1016/S0092-8674(04)00400-3)
- Schiaffino, S., and C. Mammucari, 2011 Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models. *Skelet. Muscle* 1: 4. <https://doi.org/10.1186/2044-5040-1-4>
- Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan *et al.*, 2001 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29: 308–311. <https://doi.org/10.1093/nar/29.1.308>
- Szpiech, Z. A., and R. D. Hernandez, 2014 selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Mol. Biol. Evol.* 31: 2824–2827. <https://doi.org/10.1093/molbev/msu211>
- Toledo, F. G. S., S. Watkins, and D. E. Kelley, 2006 Changes induced by physical activity and weight loss in the morphology of intermyofibrillar mitochondria in obese men and women. *J. Clin. Endocrinol. Metab.* 91: 3224–3227. <https://doi.org/10.1210/jc.2006-0002>
- Torres, R., Z. A. Szpiech, and R. D. Hernandez, 2018 Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* 14: e1007387. <https://doi.org/10.1371/journal.pgen.1007387>
- Van Laere, A. S., M. Nguyen, M. Braunschweig, C. Nezer, C. Collette *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425: 832–836. <https://doi.org/10.1038/nature02064>
- Wang, K., M. Li, and H. Hakonarson, 2010 ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38: e164. <https://doi.org/10.1093/nar/gkq603>
- Wu, J., X. Mao, T. Cai, J. Luo, and L. Wei, 2006 KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 34: W720–W724. <https://doi.org/10.1093/nar/gkl167>
- Xu, X., X. Xu, Q. Yin, L. Sun, B. Liu *et al.*, 2011 The molecular characterization and associations of porcine cardiomyopathy associated 5 (CMYA5) gene with carcass trait and meat quality. *Mol. Biol. Rep.* 38: 2085–2090. <https://doi.org/10.1007/s11033-010-0334-5>
- Yang, S., X. Li, K. Li, B. Fan, and Z. Tang, 2014 A genome-wide scan for signatures of selection in Chinese indigenous and commercial pig breeds. *BMC Genet.* 15: 7. <https://doi.org/10.1186/1471-2156-15-7>
- Zhao, F., N. Wang, Y. L. Yi, P. F. Lin, K. Q. Tang *et al.*, 2016 Knockdown of CREB3/Luman by shRNA in Mouse Granulosa Cells Results in Decreased Estradiol and Progesterone Synthesis and Promotes Cell Proliferation. *PLoS One* 11: e0168246. <https://doi.org/10.1371/journal.pone.0168246>
- Zhao, P., J. Li, H. Kang, H. Wang, Z. Fan *et al.*, 2016 Structural Variant Detection by Large-scale Sequencing Reveals New Evolutionary Evidence on Breed Divergence between Chinese and European Pigs. *Sci. Rep.* 6: 18501. <https://doi.org/10.1038/srep18501>
- Zhao, P., Y. Yu, W. Feng, H. Du, J. Yu *et al.*, 2018 Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *Gigascience* 7: giy058. <https://doi.org/10.1093/gigascience/giy058>

Communicating Editor: D.-J. de Koning