In silico approaches to discover the functional impact of non-synonymous single nucleotide polymorphisms in selective sweep regions of the Landrace genome

Donghyun Shin¹, Kyung-Hye Won¹, and Ki-Duk Song¹,²,*

Objective: The aim of this study was to discover the functional impact of non-synonymous single nucleotide polymorphisms (nsSNPs) that were found in selective sweep regions of the Landrace genome.

Methods: Whole-genome re-sequencing data were obtained from 40 pigs, including 14 Landrace, 16 Yorkshire, and 10 wild boars, which were generated with the Illumina HiSeq 2000 platform. The nsSNPs in the selective sweep regions of the Landrace genome were identified, and the impacts of these variations on protein function were predicted to reveal their potential association with traits of the Landrace breed, such as reproductive capacity.

Results: Total of 53,998 nsSNPs in the mapped regions of pigs were identified, and among them, 345 nsSNPs were found in the selective sweep regions of the Landrace genome which were reported previously. The genes featuring these nsSNPs fell into various functional categories, such as reproductive capacity or growth and development during the perinatal period. The impacts of amino acid sequence changes by nsSNPs on protein function were predicted using two in silico SNP prediction algorithms, i.e., sorting intolerant from tolerant and polymorphism phenotyping v2, to reveal their potential roles in biological processes that might be associated with the reproductive capacity of the Landrace breed.

Conclusion: The findings elucidated the domestication history of the Landrace breed and illustrated how Landrace domestication led to patterns of genetic variation related to superior reproductive capacity. Our novel findings will help understand the process of Landrace domestication at the genome level and provide SNPs that are informative for breeding.

Keywords: Landrace; Next-generation Sequencing; Non-synonymous Single Nucleotide Polymorphism; Reproductive Capacity; Selective Sweep

INTRODUCTION

The recently developed high-throughput and cost-effective genotyping techniques allow the thorough exploration of genetic variation in domestic animals. In particular, whole-genome sequencing is a powerful approach for detecting massive amounts of single nucleotide polymorphisms (SNPs) in genome-wide sequence data. One of the strategies for studying genetic variation is to detect the selective sweep signatures based on patterns of linkage disequilibrium (LD) [1], which was proposed by Smith and Haigh [2], and other researchers have expanded and applied it [3-6]. Wang et al [7] performed a relative extended haplotype homozygosity (REHH) test to detect selective sweep regions of the Landrace genome using genotyping by genome sequencing. The genetic signature for selection of body size investigated by estimating the XP-EHH statistic in the Yucatan miniature pig [8]. Whole-genome re-sequencing of Jeju black pig (JBP) and Korean native pigs (which live on the Korean peninsula) were
performed to identify signatures of positive selection in JBP, the true and pure Korean native pigs [9]. Studies of selective sweeps in pigs have revealed strong selection signatures associated with genes underlying economic traits such as the body length, disease resistance, pork yield, muscle development, and fertility [10,11].

Diverse types of variants, e.g. copy number variations, insertion/deletion (InDel) and structural variations, have been identified in the selective sweep regions of the Landrace genome [7]. Unlike many SNPs are phenotypically neutral, non-synonymous SNPs (nsSNPs) that are located in protein-coding regions and lead to amino acid substitutions in the corresponding protein product might have functional impacts and play a role in biological processes through altering the protein structure, stability, or function; these variations are often strongly associated with several phenotypes [12]. In the case of pigs, previous studies reported the different polymorphic patterns of nsSNPs in the Toll-like receptor genes between European wild boars and domestic pigs [13].

In this study, we aimed to identify nsSNPs in the selective sweep regions of the Landrace genome that might be related to superior reproductive capacity or growth and development during the perinatal period, and gene networks that were enriched in Landrace genome. Finally, impact of amino acid changes by nsSNPs on protein function was also investigated using in silico bioinformatic tools.

**MATERIALS AND METHODS**

**Sample preparation and whole-genome re-sequencing**

In this study, a whole-genome sequence data set consisting of 14 Landrace (Danish), 16 Yorkshire (Large White) pigs, and 10 wild boars, were obtained from the NCBI Sequence Read Archive database (SRP047260). FastQC software [14] were used to perform a quality check on raw sequence data. Using Trimmomatic-0.32 [15], potential adapter sequences were removed before sequence alignment. Paired-end sequence reads were mapped to the pig reference genome (*Sus scrofa* 10.2.75) from the Ensembl database using Bowtie2 [16] with the default settings. For downstream processing and variant calling, following software packages were used: Picard tools (http://broadinstitute.github.io/picard/), SAMtools [17], and Genome Analysis Toolkit (GATK) [18]. “CreateSequenceDictionary” and “MarkDuplicates” Picard command-line tools were used to read reference FASTA sequences for writing bam files with only a sequence dictionary and to filter potential polymerase chain reaction duplicates, respectively. Using SAMtools, index files were created for the reference and bam files. Local realignment of sequence reads was performed to correct misalignment due to the presence of small insertions and deletions using GATK “Realigner-TargetCreator” and “IndelRealigner” arguments. In addition, base quality score recalibration was performed to obtain accurate quality scores and to correct the variation in quality with machine cycle and sequence context. For calling variants, GATK “UnifiedGenotyper” and “SelectVariants” arguments were used with the following filtering criteria. All variants with i) a Phred-scaled quality score of less than 30; ii) read depth less than 5; iii) MQ0 (total count across all samples of mapping quality zero reads) >4; or iv) a Phred-scaled p-value using Fisher's exact test of more than 200 were filtered out to reduce false-positive calls due to strand bias. “vcf-merge” tools of VCFtools [19] were used to merge all of the variants calling format files for the 40 samples. Additionally, tri-allelic SNPs were excluded, and all filtered SNPs on autosomes (a total of 26,240,429 SNPs) were annotated using an SNP annotation tool, SnpEff version 4.1a and the Ensemble *Sus scrofa* gene set version 75 (Sscrofa10.2.75). 53,998 nsSNPs (missense variant) were identified on autosomes from 40 sets of pig whole-genome data (Figure 1). Then, certain SNPs due to poor genotyping quality were removed; 4,174 SNPs were excluded based on Hardy-Weinberg equilibrium testing (p ≤ 0.000001). In addition, a total of 19,002 SNPs with a minor allele frequency of <0.05 were excluded. After genomic data quality control, there were 30,822 SNPs for downstream analysis.

**Population structure analysis**

Population structure analysis was performed to infer the population structure of the 40 pigs with whole-genome sequence data. The program STRUCTURE (https://web.stanford.edu/group/pritchardlab/structure.html) was used to evaluate the extent of substructure among the 40 individuals belonging to three pig breeds. Bayesian clustering analysis implemented in STRUCTURE (version 2.3.4) was used to estimate the population structure using 30,822 nsSNPs from the whole-genome sequencing data of the 40 pigs [20]. An initial burn-in of 10,000 iterations were followed by 10,000 iterations for parameter estimation was sufficient to ensure the convergence of parameter estimates. To estimate the number of populations (the K parameter of STRUCTURE), the dataset was analyzed by allowing for the values of K = 3 (Figure 2).

**Identify nsSNPs in Landrace selective sweep regions**

A previous study identified 269 selective sweep regions of the Landrace genome using the REHH test (p-value ≤ 0.01), which was used to detect the recent positive selection signatures by evaluating how LD decays across the genome 7. A total of 261 of 269 selective sweep regions of the Landrace genome were on autosomes, and 345 nsSNPs belonged to 55 Landrace selective sweep regions were identified (Figure 3). Overall, 345 nsSNPs in 55 selective sweep regions of the Landrace genome belonged to 90 genes, and gene function 64 of total 90 genes were discovered. Gene ontology (GO) network analysis was performed using ClueGO [21] to infer the biological mean-
Predicting damaging amino acid substitutions of non-synonymous SNPs specific to the Landrace breed

In this study, the functional effects of nsSNPs were predicted using the following in silico algorithms: sorting intolerant from tolerant (SIFT) [22] and polymorphism phenotyping v2 (Polyphen-2) [23]. Total 345 nsSNPs in 55 selective sweep regions of the Landrace genome were analyzed using SIFT. NsSNPs with less than 0.05 of SIFT score, which was regarded as deleterious, were used for PolyPhen-2 ver. 2.2.2 (http://genetics.bwh.harvard.edu/pph2/) analysis to predict the influence of an amino acid change on the structure and function of a protein by using specific empirical rules [23]. From the results of Polyphen-2 analysis, nsSNPs were classified into probably damaging, possibly damaging, and benign based on their scores (ranging from 0 to 1); if Polyphen-2 score for nsSNPs was more than 0.95, nsSNPs were considered to be “probably damaging”; while for values between 0.5 and 0.95, they were regarded as “possibly damaging”. The scores below 0.5 were classified as “benign”. In this study, probably damaging and possibly damaging SNPs were judged as to have strong effects on protein function.

If the SIFT score of each SNP was less than 0.05, the SNP was regarded as being deleterious, which could strongly affect protein function. Additionally, we performed PolyPhen-2 (version 2.2.2) analysis to predict the influence of an amino acid change on the structure and function of a protein by using specific empirical rules [23]. Amino acid sequences corre-
RESULTS

DNA sequencing, data preprocessing, and genetic variant calling

A total of 26,240,429 SNPs were extracted on autosomes from the whole-genome sequences of the 40 pigs, including 14 Landrace individuals, and annotated all extracted SNPs using SnpEff version 4.1a (http://snpeff.sourceforge.net/SnpSift.html) [24]. Through this SNP annotation, all SNPs were divided into 31 functional classes, including nsSNPs (Figure 1). Most of the SNPs were located in intergenic or intronic regions; finally, we identified 53,998 nsSNPs (0.205% of the total SNPs). After quality control for all of the nsSNPs, there were 30,822 nsSNPs. Population structure analysis using the genotypic information on these SNPs provided the genetic relationship among breeds. The results from analyzing the population structure clearly distinguished Landrace, Yorkshire, and wild boar (Figure 2).

nsSNPs in Landrace selective sweep regions

A total of 269 selective sweep regions were obtained from a previous study on the Landrace breed to identify nsSNPs related to selective sweeps [7], and a total of 345 nsSNPs were identified from 55 Landrace selective sweep regions (Figure 3) by re-analyzing the data of previous study resequencing data of Landrace and Yorkshire [7]. Information of 345 nsSNPs in the selective sweep regions of the Landrace genome belonged to 90 genes were shown in Table 1. The average number of nsSNPs per gene was 3.83, and the gene length was not correlated to the number of nsSNPs (Figure 4). The deleted in malignant brain tumors 1 (DMBT1) gene consisted of 18 exons harboring 26 nsSNPs that were evenly distributed; this gene had the highest number of nsSNPs among the 90 genes. Moreover, there were considerable frequency differences between Landrace and other breeds (Yorkshire and wild boar) in nsSNPs of the DMBT1 gene (Figure 5). This suggests that...
DMBT1 is significantly affected by many nsSNPs in Landrace breed establishment. Previous studies strongly suggested an important role of DMBT1 in the process of fertilization in pigs; it was shown to be seconed in the oviduct and involved in the mechanism of fertilization in porcine species [25,26]. In particular, Ambruosi et al [25] reported that oviduct fluid containing DMBT1 protein was strongly related to the preparation of gametes for fertilization, fertilization itself, and subsequent embryonic development. Therefore, we assumed that nsSNPs of DMBT1 of Landrace might correlate with the fertilization capacity that was acquired during artificial selection, making the reproductive capacity of Landrace pigs superior to that of other breeds [27].

Among 90 genes, the function of 64 genes were predicted, and we performed GO network analysis of these 64 genes using ClueGO [21] to draw inferences on the biological effects of nsSNPs in Landrace selective sweep regions. The information on these networks is shown in Figure 6 and Table 2. The GO network analysis revealed that 19 of the total of 64 genes were associated with five major GO terms, and these major terms were closely related to the reproductive capacity or growth and development of the Landrace breed during the perinatal period. In the GO network, seven genes (C-C motif chemokine ligand 1 [CCL1], CCL23, hemopexin, mucolipin 1, leucine zipper and EF-hand containing transmembrane protein 2, phospholipase A2 group VI [PLA2G6]), and protein tyrosine phosphatase, receptor type, C [PTPRC]) were related to cellular metal ion homeostasis in seven major GO terms, and this cluster was the largest in this network. Moreover, these terms were similar to the GO results of a positively selected region identified in Wang's study of Landrace selective sweeps [7]. Metal ions are one major group of minerals; since components of follicular fluid such as Ca, Cu, and Fe significantly increase as the follicles increase in size, some minerals appear to play an important role in pig reproduction [28]. Five genes (ATPase phospholipid transporting 8A1 [ATP8A1], CCL1,
kinesin family member 20B, plasminogen, and PTPRC) were shown to be involved in the positive regulation of locomotion, and its network consisted of four GO terms (positive regulation of locomotion, positive regulation of cellular component movement, positive regulation of cell motility, and positive regulation of cell migration). This cellular movement is a central process in the development and maintenance of multicellular organisms. In addition, tissue formation during embryonic development requires the orchestrated movement of cells in a particular direction. It is reasonable to assume that several genes of these four significant GO terms in the selective sweep regions of the Landrace genome might be related to the superior growth and development of Landrace during the perinatal period. Ten genes (ATP8A1, bridging integrator 2, CD93 mole-

**Figure 4.** Correlation between length and number of single nucleotide polymorphisms (SNPs) in genes related to non-synonymous SNPs (nsSNPs) in Landrace selective sweep regions.

**Figure 5.** Frequency difference of non-synonymous single nucleotide polymorphisms (nsSNPs) in deleted in malignant brain tumors 1 genes between Landrace and other breeds (Yorkshire and wild boar).
cule [CD93], exophilin 5, GRB2 associated binding protein 2, n-ethylmaleimide-sensitive factor attachment protein, beta, PLA2G6, PTPRC, and vesicle associated membrane protein 1 [VAMP1]) were associated with exocytosis, and five genes (ATP8A1, CD93, DMBT1, PTPRC, and VAMP1) were classified under the secretory granule membrane term in the GO network. The acrosome contains a single secretory granule and is located in the head of mammalian sperm; secretion from this granule is an absolute requirement for fertilization [29]. Acrosome exocytosis is a synchronized and tightly regulated all-or-nothing process, which provides a unique model for studying the multiple steps of the membrane fusion cascade [29]. Therefore, we assumed that these genes containing nsSNPs in the selective sweep region, which are related to exocytosis and the secretory granule membrane, might have been influenced by artificial selection, considering the distinctive reproductive capacity of the Landrace breed [27].

Figure 6. Gene ontology (GO) network analysis of genes related to non-synonymous single nucleotide polymorphisms (SNPs) in Landrace selective sweep regions. Significant results of GO analysis using genes related to non-synonymous SNPs in the selective sweep regions of the Landrace genome with our criteria in ClueGO packages of Cytoscape (number of genes = 4, sharing group percentage = 40.0). These results are largely divided into eight clusters as follows.

Table 2. Information of gene ontology (GO) network analysis of genes related to non-synonymous SNPs in Landrace selective sweep regions

| GO ID   | GO Term                              | Term p-value | Group p-value | #Genes | Associated genes found                      |
|---------|--------------------------------------|--------------|---------------|--------|--------------------------------------------|
| GO:0002274 | Myeloid leukocyte activation         | 0.005        | 0.005         | 7      | ATP8A1, BIN2, CD93, GAB2, MAPT, PTPRC, SHPK |
| GO:0006887 | Exocytosis                          | 0.001        | 0.001         | 10     | ATP8A1, BIN2, CD93, EXPH5, GAB2, NAPB, PLA2G6, PLG, PTPRC, VAMP1 |
| GO:0030667 | Secretory granule membrane           | 0.003        | 0.003         | 5      | ATP8A1, CD93, DMBT1, PTPRC, VAMP1         |
| GO:0040017 | Positive regulation of locomotion    | 0.016        | 0.017         | 5      | ATP8A1, CCL1, KIF20B, PLG, PTPRC          |
| GO:0051272 | Positive regulation of cellular component movement | 0.013 | 5 | ATP8A1, CCL1, KIF20B, PLG, PTPRC |
| GO:2000147 | Positive regulation of cell motility | 0.012        | 5              | 5      | ATP8A1, CCL1, KIF20B, PLG, PTPRC         |
| GO:0030335 | Positive regulation of cell migration | 0.010        | 5              | 5      | ATP8A1, CCL1, KIF20B, PLG, PTPRC         |
| GO:0006873 | Cellular ion homeostasis             | 0.003        | 0.006         | 7      | CCL1, CCL23, HPX, LETM2, MCOLN1, PLA2G6, PTPRC |
| GO:0055080 | Cation homeostasis                   | 0.005        | 7              | 7      | CCL1, CCL23, HPX, LETM2, MCOLN1, PLA2G6, PTPRC |
| GO:0030003 | Cellular cation homeostasis          | 0.003        | 7              | 7      | CCL1, CCL23, HPX, LETM2, MCOLN1, PLA2G6, PTPRC |
| GO:0055065 | Metal ion homeostasis                | 0.003        | 7              | 7      | CCL1, CCL23, HPX, LETM2, MCOLN1, PLA2G6, PTPRC |
| GO:0072507 | Divalent inorganic cation homeostasis| 0.015        | 5              | 5      | CCL1, CCL23, MCOLN1, PLA2G6, PTPRC       |
| GO:0006875 | Cellular metal ion homeostasis       | 0.001        | 7              | 7      | CCL1, CCL23, HPX, LETM2, MCOLN1, PLA2G6, PTPRC |
| GO:0072503 | Cellular divalent inorganic cation homeostasis | 0.013 | 5 | CCL1, CCL23, MCOLN1, PLA2G6, PTPRC |
| GO:0055074 | Calcium ion homeostasis              | 0.011        | 5              | 5      | CCL1, CCL23, MCOLN1, PLA2G6, PTPRC       |
| GO:0006874 | Cellular calcium ion homeostasis     | 0.010        | 5              | 5      | CCL1, CCL23, MCOLN1, PLA2G6, PTPRC       |

SNPs, single nucleotide polymorphisms.

Significant results of GO analysis using genes related to non-synonymous SNPs in the selective sweep regions of the Landrace genome with our criteria in ClueGO packages of Cytoscape (number of genes = 4, sharing group percentage = 40.0). These results are largely divided into eight clusters as follows.
Predicting strong effects of nsSNPs on amino acid substitutions in Landrace selective sweep region

Two in silico SNP prediction algorithms, SIFT [22] and PolyPhen-2 [23], were applied to estimate the possible effects of the stabilizing residues on protein functions for 345 nsSNPs in Landrace selective sweep regions. The results of SIFT and Polyphen-2 for 345 non-synonymous SNPs are shown in Tables 3, 4.

According to the SIFT analysis, 75 of 345 nsSNPs were classified as being deleterious (for some SNPs, there was low confidence in the accuracy of classification as deleterious).

### Table 4. Forty-six non-synonymous SNPs with strong effects on protein functions based on SIFT and Polyphen-2

| SNP         | CHR | POS      | A1 | A2 | SIFT prediction | SIFT score | Polyphen-2 prediction | Polyphen-2 score | Gene       | Selective sweep |
|-------------|-----|----------|----|----|----------------|------------|------------------------|------------------|------------|-----------------|
| rs326699354 | 2   | 62,624,837 | G  | A  | deleterious     | 0.017      | probably damaging       | 0.993            | ENSSSCG00000013821 | 2:62355986:62756249 |
| rs340857214 | 2   | 62,625,107 | G  | A  | deleterious     | 0.021      | probably damaging       | 0.934            | Effector1      | 11:82112057:8213672 |
| rs326644986 | 2   | 62,625,190 | A  | T  | deleterious     | 0.028      | probably damaging       | 0.999            | ENSSSCG00000013822 | 2:62355986:62756249 |
| rs335820735 | 2   | 62,644,986 | A  | T  | deleterious     | 0.008      | probably damaging       | 0.934            | Effector1      | 11:82112057:8213672 |
| rs343007761 | 2   | 62,645,014 | T  | G  | deleterious     | 0.018      | probably damaging       | 0.934            | ENSSSCG00000013822 | 2:62355986:62756249 |

SIFT, sorting intolerant from tolerant; Polyphen-2, polymorphism phenotyping v2.

We identified that 46 of 345 non-synonymous SNPs in the selective sweep regions of the Landrace genome had strong effects on protein function as determined with both in silico tools: SIFT and PolyPhen-2.
adipose tissue, which is related to the concentration of lipid
this gene has been reported to be important in porcine backfat
not reveal any particularly important network of
function in this study. Although our GO network analysis did
strong effect among the 90 genes. Five SNPs of the total of 11
was shown to have the greatest number of nsSNPs with a
important roles of these genes during selection. Additionally,
of nsSNPs in genes related to olfactory receptors suggested
reflects the strong reliance of pigs on their sense of smell
number of functional olfactory receptor genes most probably
involved the olfactory receptor gene family [30]. Martien et
of the three genes would be very similar. Because the
orthologous genes consisted of 18 genes from 8 species from
placental mammals and all 18 genes were related to olfactory
receptors, we assumed that ENSSSCG00000013821, ENSSSCG
0000013822, and ENSSSCG0000013819 were inferred as
orthologous genes (Table 5). Because the lists of orthologs
of the three genes were the same, we guessed that the functions
of the three genes would be very similar. Because the
orthologous genes consisted of 18 genes from 8 species from
placental mammals and all 18 genes were related to olfactory
receptors, we assumed that ENSSSCG00000013821, ENSSSCG
0000013822, and ENSSSCG0000013819 were inferred as
orthologous genes. In a previous study of pig evolution, one
of the several significant features of porcine genome expansion
involved the olfactory receptor gene family [30]. Martien et
al [26] reported that there are 1,301 porcine olfactory receptor
genes and 343 partial olfactory receptor genes. This large
number of functional olfactory receptor genes most probably
reflects the strong reliance of pigs on their sense of smell
while scavenging for food. The presence of greater number
of nsSNPs in genes related to olfactory receptors suggested
important roles of these genes during selection. Additionally,
the monoacylglycerol O-acyltransferase 2 (MOGAT2) gene
was shown to have the greatest number of nsSNPs with a
strong effect among the 90 genes. Five SNPs of the total of 11
nsSNPs in the MOGAT2 gene had strong effects on protein
function in this study. Although our GO network analysis did
not reveal any particularly important network of MOGAT2,
this gene has been reported to be important in porcine backfat
adipose tissue, which is related to the concentration of lipid
and lipid synthesis, as revealed by a transcriptome analysis
comparing Landrace and other breeds [31]. In addition, 3 of
26 nsSNPs in the DMBTI gene were considered to have strong
effects on protein function, as revealed by the SIFT and Poly-
phen-2 results.

DISCUSSION

Given the interest of the meat production industry in improv-
ing the meat quality or piglet number, a genetic investigation
focusing on the selective sweep regions of the Landrace ge-
ome was previously performed [7]. This study provided vital
information for domestic pig breeding. In most selective sweep
studies using whole-genome sequencing data, all SNPs, in-
cluding nsSNPs, were used to detect selective sweep regions.
As nsSNPs are mutations that alter the amino acid sequences
of encoded proteins, their presence results in a phenotypic
change in the organism. Such changes are usually subjected to
natural selection. In the case of Landrace, the domestica-
tion process had a shorter generation interval than natural selection.
Therefore, we believe that nsSNPs had a diverse evolutionary
history during the domestication and artificial selection pro-
cesses, and advanced studies are required to achieve an accurate
interpretation of the Landrace genome using nsSNP inform-
ation after exploring Landrace positive selection based on
whole-genome sequence data. In this study, we performed
several analyses of nsSNPs of the Landrace genome to obtain
a better understanding of the whole genome. We assumed that
the information on these nsSNPs might be associated with novel
important biological mechanisms related to particular
traits of the Landrace breed. For the precise analysis of the
characteristics of the Landrace breed from a genomic perspec-
tive, we investigated the biological meaning of nsSNPs in the
selective sweep regions of the Landrace genome used in a pre-
vious study [7]. As a result, there was no correlation between
the number of nsSNPs and gene length per 90 genes contain-
ing an nsSNP within the selective sweep regions of the Landrace
genome (Figure 5), which was contrary to our expectations.
Considering that 22 of 90 genes overlapped with multiple
selective sweep regions while the others belonged to a single
selective sweep region, we assumed that genes containing many
nsSNPs in the selective sweep regions of the Landrace genome
were more meaningful than our expectation. Subsequently,
based on GO network analysis using genes containing 345
nsSNPs in the selective sweep regions of the Landrace genome,
a large proportion of selective sweep regions of the Landrace
genome where strong amino acid sequence changes had oc-
curred, were involved in the superior reproductive capacity
or growth and development of the Landrace breed during the
perinatal period. Some of the GO network results overlapped
with the GO analysis of all the selective sweep regions in a
previous study, while others involved novel interpretations
of the Landrace genome [7].

CONCLUSION

Our results strongly suggested that Landrace genetic variants, which could give rise to changes in amino acid sequences, might be important factors for the superior reproductive capacity of this breed. We aimed to perform analyses of the Landrace genome using nsSNPs in selective sweep regions. Our results showed that most of the genes affected by nsSNPs in the selective sweep regions may be closely related to the superior reproductive capacity or growth and development of the Landrace breed during the perinatal period. Furthermore, there were indications that nsSNPs in selection had impacted in Landrace breed establishment. This study will provide insights into the impact of the process of domestication on the Landrace genome.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.
ACKNOWLEDGMENTS

This study was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ01110901, PJ01315101), Rural Development Administration, Republic of Korea. The authors are grateful to this organization.

REFERENCES

1. Ennis S. Linkage disequilibrium as a tool for detecting signatures of natural selection. In: Collins AR, editor. Linkage disequilibrium and association mapping. Totowa, NJ, USA: Humana Press; 2007. p. 59-70.

2. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res 1974;23:23-35.

3. Barton NH. Linkage and the limits to natural selection. Genetics 1995;140:821-41.

4. Durrett R, Schweinsberg J. Approximating selective sweeps. Theor Popul Biol 2004;66:129-38.

5. Pennings PS, Hermisson J. Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet 2006;2:e186.

6. Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. Evolution 2005;59:2312-23.

7. Wang Z, Chen Q, Yang Y, et al. A genome-wide scan for selection signatures in Yorkshire and Landrace pigs based on sequencing data. Anim Genet 2014;45:808-16.

8. Kim H, Song KD, Kim HJ, et al. Exploring the genetic signature of body size in Yucatan miniature pig. PloS One 2015;10:e0121732.

9. Kim J, Cho S, Caetano-Anolles K, Kim H, Ryu Y-C. Genome-wide detection and characterization of positive selection in Korean Native Black Pig from Jeju Island. BMC Genet 2015;16:3.

10. Amaral AJ, Ferretti L, Megens H-J, et al. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PloS One 2011;6:e14782.

11. Li M, Tian S, Yeung CK, et al. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. Sci Rep 2014;4:4678.

12. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894-900.

13. Bergman I-M, Rosengren JK, Edman K, Edfors I. European wild boars and domestic pigs display different polymorphic patterns in the Toll-like receptor (TLR) 1, TLR2, and TLR6 genes. Immunogenetics 2010;62:49-58.

14. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114-20.

16. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.

17. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078-9.

18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

19. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156-8.

20. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945-59.

21. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25:1091-3.

22. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812-4.

23. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;76:7.20.1-41.

24. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 2012;6:80-92.

25. Ambrusoi B, Accogli G, Douet C, et al. Deleted in malignant brain tumor 1 is secreted in the oviduct and involved in the mechanism of fertilization in equine and porcine species. Reproduction 2013;146:119-33.

26. Teijeiro JM, Roldán ML, Marini PE. Molecular identification of the sperm selection involved porcine sperm binding glycoprotein (SBG) as deleted in malignant brain tumors 1 (DMBT1). Biochimie 2012;94:263-7.

27. Chen P. Genetic improvement of lean growth rate and reproductive traits in pigs [Retrospective Theses and Dissertations]. Ames, IA, USA: Iowa State University; 2002.

28. Subha G. Role of biochemical factors and mineral supplementation in livestock ration for maintenance of their fertility and healthy reproductive status: a review. Res J Chem Sci ISSN 2013;3:102-6.

29. Mayorga LS, Tomes CN, Belmonte SA. Acrosomal exocytosis, a special type of regulated secretion. IUBMB Life 2007;59:286-90.

30. Groenen MA, Archibald AL, Uenishi H, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 2012;491:393-8.

31. Corominas J, Ramayo-Caldas Y, Puig-Oliveras A, et al. Analysis of porcine adipose tissue transcriptome reveals differences in de novo fatty acid synthesis in pigs with divergent muscle fatty acid composition. BMC Genomics 2013;14:843.