

Prediction of Genes Related to Positive Selection Using Whole-Genome Resequencing in Three Commercial Pig Breeds

HyoYoung Kim¹, Kelsey Caetano-Anolles², Minseok Seo³, Young-jun Kwon³,
Seoae Cho⁴, Kangseok Seo^{5*}, Heebal Kim^{1,4,6**}

¹Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea,

²Department of Animal Sciences, University of Illinois, Urbana, IL 61801, USA,

³Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea,

⁴C&K Genomics Inc., Seoul National University Research Park, Seoul 08826, Korea,

⁵Department of Animal Science and Technology, College of Life Science and Natural Resources,
Sunchon National University, Suncheon 57922, Korea,

⁶Department of Agricultural Biotechnology, Animal Biotechnology Major, and Research Institute
for Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

Selective sweep can cause genetic differentiation across populations, which allows for the identification of possible causative regions/genes underlying important traits. The pig has experienced a long history of allele frequency changes through artificial selection in the domestication process. We obtained an average of 329,482,871 sequence reads for 24 pigs from three pig breeds: Yorkshire (n = 5), Landrace (n = 13), and Duroc (n = 6). An average read depth of 11.7× was obtained using whole-genome resequencing on an Illumina HiSeq2000 platform. In this study, cross-population extended haplotype homozygosity and cross-population composite likelihood ratio tests were implemented to detect genes experiencing positive selection for the genome-wide resequencing data generated from three commercial pig breeds. In our results, 26, 7, and 14 genes from Yorkshire, Landrace, and Duroc, respectively were detected by two kinds of statistical tests. Significant evidence for positive selection was identified on genes *ST6GALNAC2* and *EPHX1* in Yorkshire, *PARK2* in Landrace, and *BMP6*, *SLA-DQA1*, and *PRKG1* in Duroc. These genes are reportedly relevant to lactation, reproduction, meat quality, and growth traits. To understand how these single nucleotide polymorphisms (SNPs) related positive selection affect protein function, we analyzed the effect of non-synonymous SNPs. Three SNPs (rs324509622, rs80931851, and rs80937718) in the *SLA-DQA1* gene were significant in the enrichment tests, indicating strong evidence for positive selection in Duroc. Our analyses identified genes under positive selection for lactation, reproduction, and meat-quality and growth traits in Yorkshire, Landrace, and Duroc, respectively.

Keywords: non-synonymous, swine, positive selection, re-sequencing, single nucleotide polymorphism

Introduction

Artificial selection is an aspect of the process of domestication development [1] that is reflected in the evolution of domestic animals. Positive selection increases the fitness of adaptive traits [2]. Previous research has demonstrated that the selection and mating of domestic

animals is determined by human-generated pressures, and the genetic differences resulting from artificial selection have led to economic development [3]. Among domestic animals, the pig has experienced a particularly long history of haplotype changes through artificial selection in the process of domestication [2, 4]. For most of their history, studies have found positive selection in Yorkshire, Landrace, and Duroc to be associated with specific genes related to

Received July 29, 2015; Revised November 21, 2015; Accepted November 21, 2015

*Corresponding author: Tel: +82-61-750-3232, Fax: +82-61-750-3230, E-mail: sks@scnu.kr

**Corresponding author: Tel: +82-2-880-4803, Fax: +82-2-883-8812, E-mail: heebal@snu.ac.kr

Copyright © 2015 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

lactation [5, 6], reproduction [7], meat quality [8], and growth [3] trait. Given the importance of these traits to the pig farming industry, it is necessary to detect the signals of positive selection in pigs.

Identifying genomic selection using high-density single nucleotide polymorphism (SNP) arrays or sequencing data is useful for the discovery of putative trait-related genes. A variant under selection pressure between and/or within populations will show an increase of the frequency of particular alleles or linkage disequilibrium (LD) of SNPs. Selective sweep can cause genetic differentiation across populations [9]. The cross-population extended haplotype homozygosity (XPEHH) [10] and cross-population composite likelihood ratio (XPCLR) [9] tests are widely used to search for signals of selection between the populations. The XPEHH test detects the occurrence of selection based on measuring LD, while the XPCLR test considers the spatial patterns of allele frequencies of SNPs [9]. The XPCLR test is a likelihood method that uses differentiation of multi-locus allele frequencies between two populations to detect selective sweep. Selection with significant signals can break up the haplotype structures more rapidly than mutation or recombination processes [11].

In this study, the XPEHH and XPCLR between-population methods were implemented to detect positive selection in three pig breeds through resequencing using the Illumina HiSeq2000 platform (Illumina, Inc., San Diego, CA, USA). To further explain the biological implications of positively selected genes, functional enrichment analysis was performed.

Methods

Sampling and whole-genome sequencing

We used genomic DNA samples gathered from 12 males and 12 females of three pig breeds: Yorkshire, 2 males and 3 females; Landrace, 7 males and 6 females; and Duroc, 3 males and 3 females. Blood samples were collected for DNA extraction. Sample collections and DNA quality check procedures were performed according to the manufacturer's instructions. Next, we constructed genomic DNA libraries for each sample using TruSeq DNA Library kits (Illumina). The paired-end library was sequenced on an Illumina HiSeq 2000 sequencing platform.

The pair-end sequence reads were aligned to the reference pig genome sequence from University of California, Santa Cruz (UCSC; <http://genome.ucsc.edu/>; susScr3) using Bowtie2 with the default setting. We used the following open-source software: Bowtie2, Picard tools 1.94 (<http://picard.sourceforge.net>), Samtools 0.1.19 [12], Genome Analysis Toolkit (GATK) 2.6.4 [13], and VCFtools 4.0 [14]

for resequencing data processing and SNP calling. Substitution calling was performed using GATK Unified-Genotyper. We phased the haplotypes for the entire pig populations using BEAGLE [15]. Picard tools was used for duplicate read removal and all mate-pair information confirmation. Samtools was used for indexing the results from bam files and calculating the mapped reads using the flagstat option. GATK was used for realignment and SNP calling from resequencing data, and VCFtools was used when VCF files were handled. After filtering, non-biallelic SNPs were excluded.

Detection of selective sweep

To detect selective sweep, we implemented two between-populations methods (XPEHH and XPCLR) between each pairwise breed contrast. These statistics used the phased data based on the SNP genotypes obtained from the whole-genome resequencing. For each contrast, XPEHH and XPCLR were implemented across all SNPs between the three pig breeds to identify potential positive selection. We used a p-value of <0.01 for XPEHH and the top 1,000 genes in XPCLR as the cutoff criteria. These tests revealed that some SNPs might have been under positive selection in the Yorkshire, Landrace, and Duroc breeds, respectively.

Bioinformatic analysis of genes under positive selection

Based on the findings of positive selection, enrichment analysis was performed to examine the biological functions of genes in detected regions. In this study, regions with SNPs under positive selection were extended approximately 10 kb upstream and downstream [16]. We assembled genes located within the extended region using the RefGene from the UCSC Genome Browser (<http://genome.ucsc.edu/>; ver. hg19). Gene enrichment analysis was performed GO terms [17], including biological process, molecular function, cellular component, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [18] using the DAVID (<http://david.abcc.ncifcrf.gov/>) tool [19].

Analysis of SNP effects

To identify non-synonymous coding SNPs (nsSNPs), we analyzed SNP effects using the SNPeffect database, which detects phenotypic effects of variation (<http://snpeffect.vib.be/>) [20]. For more exact results, a Fisher's exact test was performed to identify breed-specific amino acids in genes with nsSNPs. The statistical test was used to analyze 2×2 contingency tables composed with two factors. Considering our purpose, one factor was breed information, and the other was amino acid information. From the 2×2 contingency tables, we expected to observe enrichment of specific amino

acids in specific breeds at each position. All data parsing and calculations were performed using Python (ver. 2.5) and R (ver. 3.1.2).

Results

Sequencing and alignments

Genome-wide positive selection in pigs was identified using whole-genome resequencing, in which we aligned whole genome reads from three pig breeds, Yorkshire, Landrace, and Duroc, to the pig reference sequence. These breeds have undergone positive selection with strong signals for lactation, reproduction, meat quality, and growth. On average, ~329,482,871 sequences were mapped to the pig reference genome from the University of California, Santa Cruz (UCSC) Genome Browser (assembly ID: susScr3) using Bowtie2. Table 1 provides a summary of the alignments of 24 pig samples. Over 92.8% of all reads were aligned to the pig reference genome. Pig resequencing data with an average read depth of 11.7× were generated using the Illumina HiSeq2000 platform. A total of 18,440,443 SNPs were identified using the GATK.

Selective sweep detected by between-population methods

To detect signals of positive selection in the three pig breeds, two between-population methods (XPEHH and XPCLR) were implemented with phased haplotypes based on pairwise populations of Yorkshire-Landrace (Y-L), Yorkshire-Duroc (Y-D), Landrace-Yorkshire (L-Y), Landrace-Duroc (L-D), Duroc-Yorkshire (D-Y), and Duroc-Landrace (D-L). Here, for Y-L, Yorkshire was regarded as the observed population and Landrace as the reference population. XPEHH and XPCLR scores denote selection that happened in the observed population. As a result, several regions of positive selection were identified at a significance level of p-value < 0.01 [21] in XPEHH, and the top 1,000 genes in XPCLR. Table 2 summarizes the numbers of genes under significant positive selection, which were identified in the six breed pairs (Y-L, Y-D, L-Y, L-D, D-L, and D-Y). For example, for the Y-L breed pair, which detects selection that occurred in the Yorkshire population when compared to the Landrace population, 376 and 1,869 genes were under positive selection based on the XPEHH and XPCLR tests, respec-

Table 1. Summary of alignment rates in 24 pig samples

Sample	Total raw sequences	Total filtered sequences	Average sequence depth
S_1 (Y)	317,653,962	304,227,661	11.31
S_2 (Y)	339,820,584	326,552,972	12.09
S_3 (Y)	351,288,112	335,300,941	12.50
S_4 (Y)	341,544,022	326,045,928	12.16
S_5 (Y)	333,812,260	318,386,421	11.88
S_7 (L)	327,963,242	310,633,736	11.67
S_8 (L)	326,204,388	308,126,065	11.61
S_9 (L)	332,203,902	316,685,253	11.82
S_10 (L)	345,708,096	33,135,123	12.30
S_11 (L)	355,525,994	335,692,509	12.65
S_12 (L)	335,544,768	319,080,587	11.94
S_13 (L)	323,931,824	310,404,686	11.53
S_14 (L)	338,310,130	268,717,357	12.04
S_15 (L)	328,367,780	241,735,861	11.69
S_17 (L)	301,513,518	259,832,740	10.73
S_18 (L)	301,322,414	261,545,335	10.72
S_19 (L)	332,943,476	303,693,608	11.85
S_20 (L)	322,384,718	289,496,700	11.47
S_21 (D)	308,284,332	296,013,130	10.97
S_22 (D)	310,837,424	298,848,776	11.06
S_23 (D)	351,342,118	334,179,965	12.50
S_24 (D)	327,445,612	315,483,972	11.65
S_25 (D)	327,904,008	316,956,022	11.67
S_26 (D)	325,732,212	312,978,523	11.59
Average	329,482,871	305,915,453	11.73

Samples S_6 and S_16 were excluded after DNA quality checks. Y, Yorkshire; L, Landrace; D, Duroc.

Table 2. Summary of genes under significant positive selection

	Breed pair	XPEHH	XPCLR	Overlapping genes	Gene names for breed specific genes
Yorkshire (Y)	Y_L	376 (26,458)	1,869 (4,848)	119	26 (<i>ST6GALNAC2^a</i> , <i>TBC1D13</i> , <i>ARHGAP26</i> , <i>TRAPPC12</i> , <i>RPS7</i> , <i>COLEC11</i> , <i>DEPTOR</i> , <i>ENPP2</i> , <i>OLFM3</i> , <i>EXTL2</i> , <i>CLEC2D</i> , <i>DUS2</i> , <i>CHST9</i> , <i>HOOK1</i> , <i>PDLIM5</i> , <i>TMEM63A</i> , <i>EPHX1^a</i> , <i>MGAT5B</i> , <i>OSBPL10</i> , <i>COL13A1</i> , <i>ADAMTS14</i> , <i>TBATA</i> , <i>NUP155</i> , <i>SEMA5A^a</i> , <i>PHACTR3</i> , <i>LRGUK</i>)
	Y_D	336 (24,887)	1,498 (3,978)	99	
Landrace (L)	L_Y	756 (47,167)	1,739 (4,751)	69	7 (<i>MAS1</i> , <i>ZC3H12D</i> , <i>EEA1</i> , <i>GLG1</i> , <i>ER^a</i> , <i>PARK2^a</i> , <i>IDO1</i>)
	L_D	408 (26,698)	1,733 (4,539)	143	
Duroc (D)	D_L	495 (76,009)	2,108 (5,327)	60	14 (<i>TGFB1</i> , <i>LMTK2</i> , <i>PRDM14</i> , <i>BMP6^a</i> , <i>SLA-DQB2</i> , <i>SLA-DQA1^a</i> , <i>DST</i> , <i>TBC1D14</i> , <i>CWH43</i> , <i>ENOPH1</i> , <i>GOLT1A</i> , <i>PRKG1^a</i> , <i>LIPM</i> , <i>ANKRD22</i>)
	D_Y	587 (68,590)	2,032 (5,364)	64	

The single nucleotide polymorphisms under positive selection were extended by approximately ± 10 kb.

Parentheses indicate the number of single nucleotide polymorphisms at the significance level.

XPEHH, cross-population extended haplotype homozygosity; XPCLR, cross-population composite likelihood ratio.

^aBreed specific genes related to lactation, reproduction, meat quality, and growth traits in previous studies.

tively. For the Y-D breed pair, when using the XPCLR test, 1,498 genes were under positive selection with 3,978 SNPs detected in Yorkshire, with Duroc as the reference population. By the consensus set of genes under positive selection identified with both the XPEHH and XPCLR tests, 26, 7, and 14 genes were identified in Yorkshire, Landrace, and Duroc, respectively (Fig. 1). For example, the 26 Yorkshire genes identified by both the XPCLR and XPEHH tests in both the Y-L and Y-D comparisons, and so effectively detected four times.

The biological functions of genes under positive selection

We performed functional enrichment analysis on the following genes under positive selection using DAVID ver. 6.7: *ST6GALNAC2* and *EPHX1* in Yorkshire; *PARK2* in Landrace; and *BMP6*, *SLA-DQA1*, and *PRKG1* in Duroc. The results revealed enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway terms related to metabolism, diseases, genetic or environmental information processing, and organismal systems (Table 3). The gene *BMP6* acts as a hedgehog signaling pathway; the hedgehog protein is an important regulator of slow oxidative fiber clustering and is distributed throughout the tissue in pigs. Previous research has reported that *BMP6* plays an important role in meat quality [8]. Among the 26 positively selected genes in Yorkshire, the *ST6GALNAC2* gene is expressed during lactation, contributing to immune-related functions [5]. Up-regulation of the *EPHX1* (epoxide hydrolase 1) gene occurs during mammary gland involution. Nakamura *et al.* [6] suggested that *EPHX1* is involved in the initial stage of involution. The *SEMA5A* (semaphoring 5A) gene for Yorkshire reported strong candidate gene which affects the milk production traits for economic traits in cattle. The

estrogen receptor (*ER*) gene for Landrace reported that mice lacking *ER α* show severe reproduction [7]. This implies that the *ER* gene might be a positively selected lactation trait in pigs. The Gene Ontology (GO) terms included immune system process (*SLA-DQA1*) and metabolic process (*PRKG1*) for Duroc (Supplementary Table 2). Williams *et al.* [3] reported that a low level of chronic immune system resulted in greater growth in pigs. These results suggest that the 26, 7, and 14 population-specific genes with significant signals may result from potential positive selection for lactation in Yorkshire, reproduction in Landrace, and meat-quality and growth traits in Duroc.

SNP effects for population-specific genes

To understand how SNPs related positive selection affect protein function, we analyzed SNP effects using the SNPeffect database. We identified two genes, *SLA-DQA1* and *EPHX1*, with nsSNPs showing evidence of positive selection in D-Y, respectively (Fig. 2, Supplementary Table 1). To obtain accurate results, Fisher's exact test was performed on these two genes. The total numbers of statistical tests performed on the non-synonymous genes were $20 \times 3 \times 13$ (number of cases of amino acids \times number of breeds \times number of non-synonymous amino acid positions) for *SLA-DQA1*, and $20 \times 3 \times 4$ for *EPHX1*. From the enrichment tests, we identified three positions (rs324509622, rs80931851, and rs80937718) that were significant under a Bonferroni-adjusted p-value of <0.05 with an alternative hypothesis: the odds ratio is greater than 1 in the *SLA-DQA1* gene. The p-values of the three non-synonymous positions were 5.20e-05, 5.20e-05, and 7.43e-06, respectively. As a note, the above three significant enrichment results were observed only in the Duroc breed. The non-synonymous positions showed strong evidence for positive selection

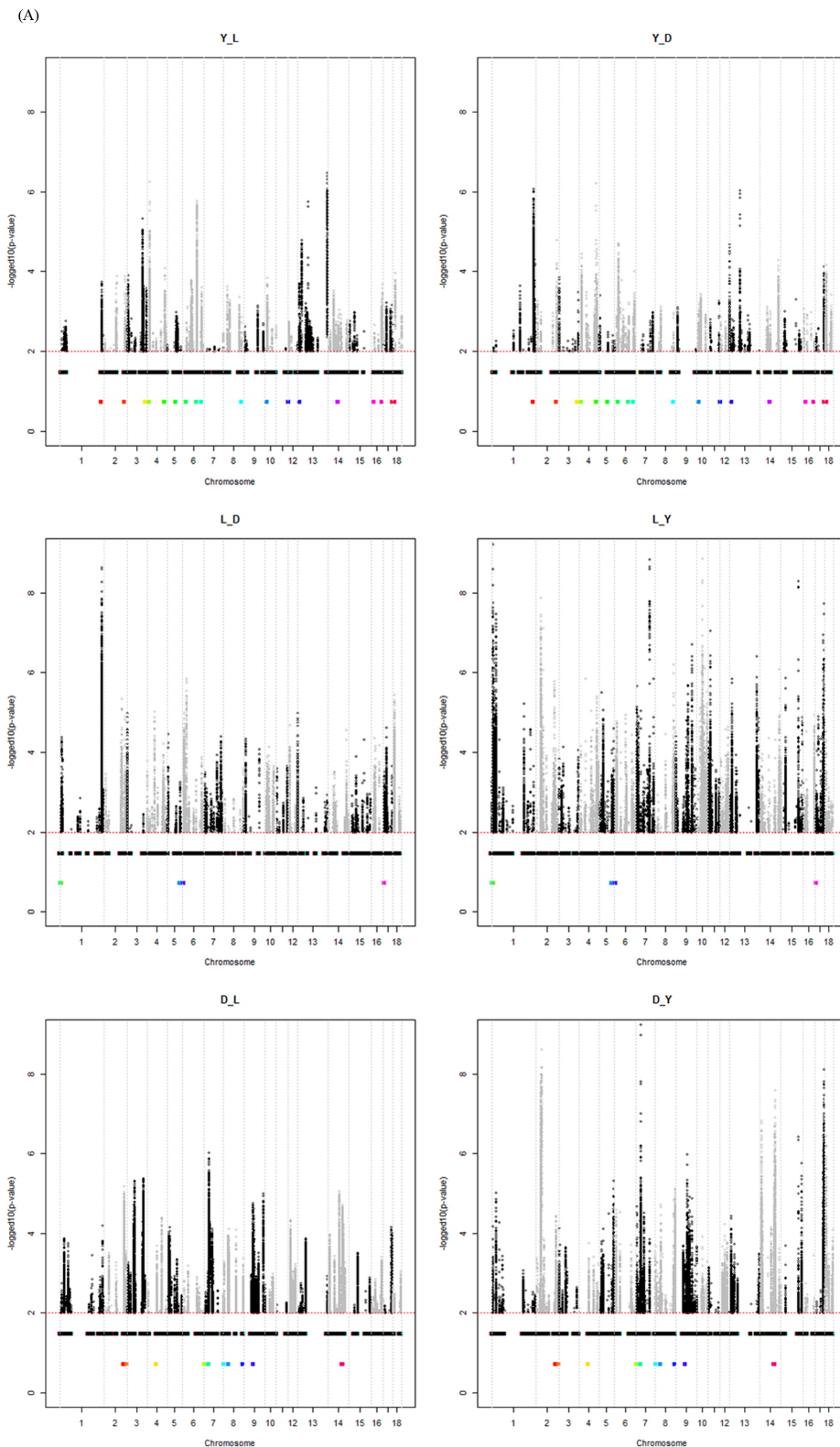


Fig. 1. Visualization of genome-wide positive selection. Manhattan plots of the degree of selection signals based on pairwise populations of three pig breeds using cross-population extended haplotype homozygosity (A) and cross-population composite likelihood ratio (B). Red lines indicate the significance level and colored dots indicate the genes under significant positive selection. Y-L, Yorkshire-Landrace; Y-D, Yorkshire-Duroc; L-D, Landrace-Duroc; L-Y, Landrace-Yorkshire; D-L, Duroc-Landrace; D-Y, Duroc-Yorkshire.

between D-Y based on the XPEHH test. As the result, we validated the Duroc-specific enrichment of amino acids affected by non-synonymous SNPs using statistical tests.

This result will contribute to improving the meat quality of pigs.

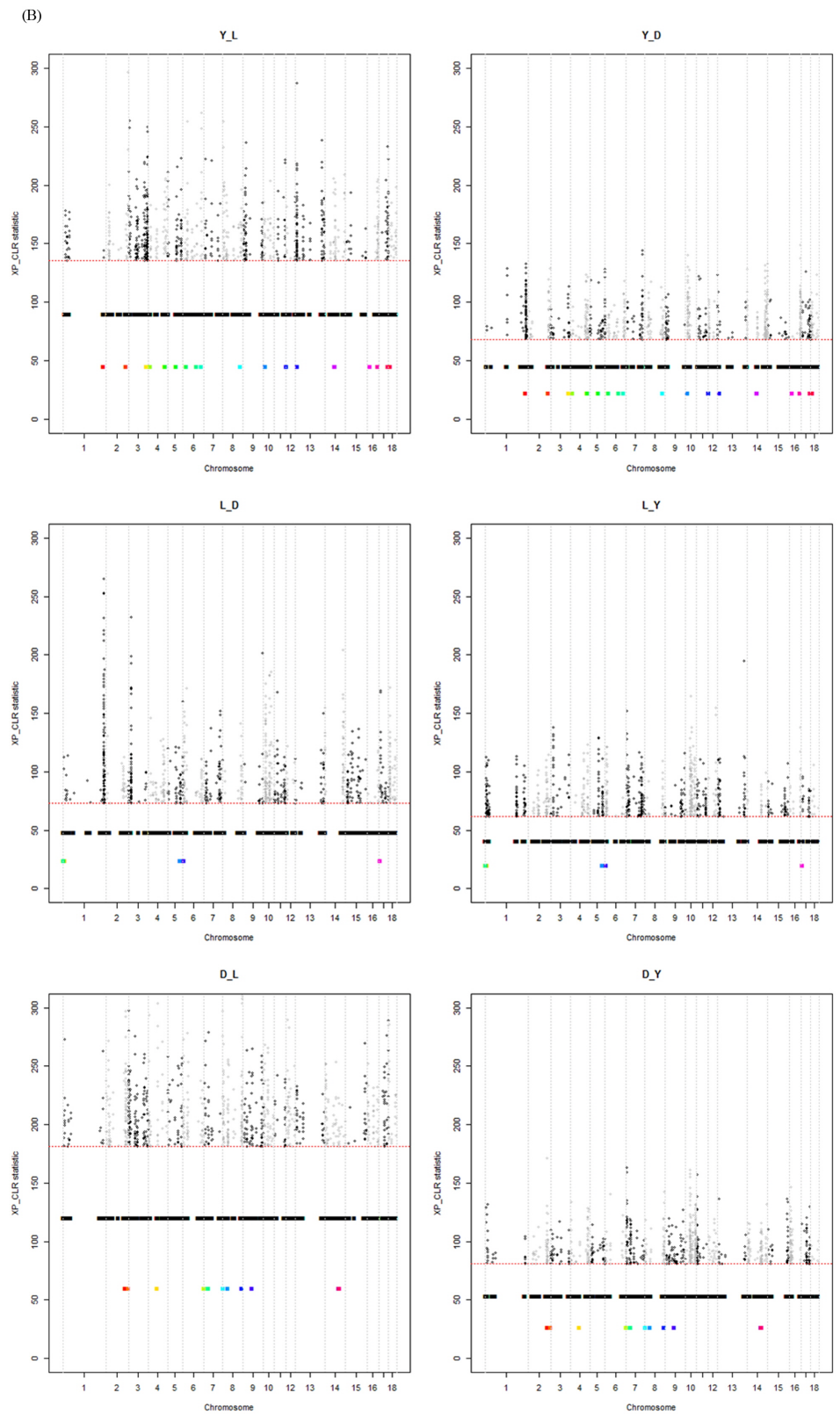


Fig. 1. Continued.

Discussion

Identifying positive selection assists in the exploration of

the genetic mechanisms of phenotypic diversity [2]. Selective sweep causes changes in the allele frequencies of SNPs, and the resulting changes in genetic variation can be

Table 3. Enriched KEGG pathways associated with positively selected genes

Breed	Gene symbol	KEGG pathway	Higher-level KEGG pathway	The highest level of KEGG pathway
Yorkshire	<i>ST6GALNAC2</i>	Glycosphingolipid biosynthesis	Glycan biosynthesis and metabolism	Metabolism
	<i>EPHX1</i>	Metabolism of xenobiotics by cytochrome P450	Xenobiotics biodegradation and metabolism	
Landrace	<i>PARK2</i>	Parkinson's disease, Ubiquitin mediated proteolysis	Neurodegenerative diseases	Human diseases
			Folding, sorting and degradation	Genetic information processing
Duroc	<i>BMP6</i>	Hedgehog signaling pathway [8, 22]	Signal transduction	Environmental information processing
		TGF-beta signaling pathway		
	<i>SLA-DQA1</i>	Cell adhesion molecules	Signaling molecules and interaction	Environmental information processing
		Antigen processing and presentation, intestinal immune network for IgA production	Immune system	Organismal systems
		Type I diabetes mellitus	Endocrine and metabolic diseases	Human diseases
	<i>PRKG1</i>	Asthma, autoimmune thyroid disease, systemic lupus erythematosus, allograft rejection, graft-versus-host disease	Immune diseases	Immune diseases
		Viral myocarditis	Cardiovascular diseases	Human diseases
		Long-term depression	Nervous system	Organismal systems
		Olfactory transduction	Sensory system	
		Gap junction	Cell communication	Cellular processes

KEGG, kyoto encyclopedia of genes and genomes.

useful in detecting selective pressure [9]. Although several studies have reported selective sweeps identified using chips such as the Illumina Porcine 60K SNP BeadChip [22] in pigs, few have identified selective sweeps using whole-genome sequencing. Therefore, detecting positive selection in the pig genome appears worthwhile.

Here we implemented two representative between-population methods, XPEHH and XPCLR, to explore positive selection. XPEHH is designed to detect near fixation or fixed selective sweeps by comparing haplotypes between two populations [10]. XPCLR is more robust for SNP discovery bias than allele frequency-based methods [9]. XPEHH tests whether the SNP site is homozygous in one population and polymorphic in another by comparing the extended haplotype homozygosity score of two populations on a core SNP. A negative XPEHH value indicates that selection happened in the reference population [15]. XPCLR was designed as a multi-locus composite likelihood ratio method to avoid the bias influence in SNP discovery. XPEHH and XPCLR find one core SNP and grid window using multi-locus information, and thus these tests identify more significant selection regions than F_{st} or Tajima's D [23]. Identifying significant genome-wide selection regions with multiple loci using a single-marker F_{st} test is difficult. Therefore, for multiple SNP methods, we carried out XPEHH and XPCLR and excluded

F_{st} and Tajima's D. XPEHH assumes an approximately normal distribution. Park *et al.* [16] determined empirical p-values using the top 1% of the significance, and our results show that these significance cutoffs are more reasonable. In addition, positive selection identified by multiple test methods is more convincing.

Bioinformatic analysis of genes under positive selection in this study revealed interesting biological information related to lactation, reproduction, meat quality, and growth. We identified several GO terms using the DAVID functional annotation tool. Immune system process (GO:0002376) was found to be significant in Duroc. Williams *et al.* [3] reported that a low level of chronic immune system resulted in greater growth in pigs. This GO term suggests that genes with some growth traits may have been under selection in the domestication process. Growth traits are referred to as important indicators of the immune system. This suggests that *SLA-DQA1* might play a role in the pig immune system. In addition, *BMP6* and *PRKG1* in Duroc had biological pathways related to meat-quality traits. Fonseca *et al.* [8] reported that the *BMP6* gene plays a role in increasing the quality of meat juiciness and tenderness. Lee *et al.* [22] identified the significant SNPs located in the *BMP6* gene related to meat quality through a genome-wide association study using the Porcine 60K BeadChip. Therefore, the

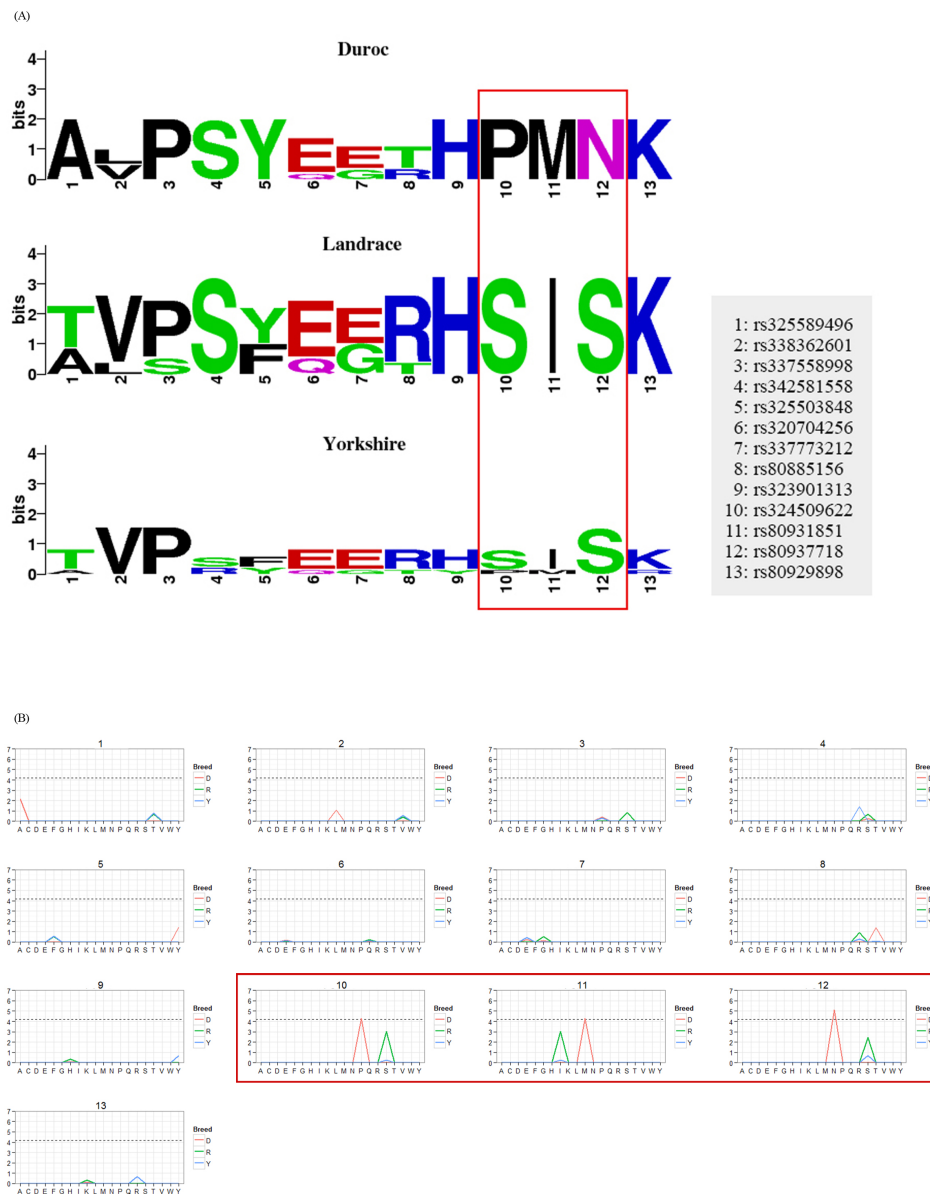


Fig. 2. Visualization of single nucleotide polymorphism (SNP) effects on positively selected genes. (A) Visualization of the sequence logo for the *SLA-DQA1* gene with 13 non-synonymous SNPs (nsSNPs) showing evidence of positive selection in Duroc. (B) Visualization of enrichment tests for the 13 nsSNPs. The red box indicates the significant positive selection regions under the Bonferroni adjusted p-value < 0.05.

SLA-DQA1 and *BMP6* genes might be effects of positive selection for growth and meat quality, respectively, in Duroc. In Yorkshire, three genes under significant positive selection were *ST6GALNAC2*, *EPHX1*, and *SEMA5A*; these genes were related to lactation traits in several previous studies. The *ST6GALNAC2* gene is expressed during lactation [5], *EPHX1* is expressed in the early involution of mammary glands [6], and *SEMA5A* is reportedly related to milk production traits in cattle [21]. In Landrace, the *ER* gene was reported to be related to reproduction traits in previous studies. Generally *ER* refers to either *ER α* or *ER β* and estrogens are involved in differentiation and maintenance of reproductive tissue. Krege *et al.* [7] reported that mice lacking *ER α* exhibited severe reproductive problems. This implies that the *ER* gene could be an effect of positive selection for lactation traits in

pigs as well as in mice. Our findings show that several significant SNPs were present in genes found to be related to lactation, reproduction, meat quality, and growth traits in previous studies. These results indirectly support the effects of positive selection. Unfortunately, information on many of the genes with significant signals has not yet been analyzed. Based on SNP analyses for positively selected genes, *SLA-DQA1* and *EPHX1* genes exhibited evidence of positive selection in D-Y, respectively. However, an enrichment test revealed that the three significant amino-acid positions in *SLA-DQA1* and *EPHX1* were not significant.

Our analyses suggested that several genes—such as *EPHX1*, *PARK2*, *BMP6*, *SLA-DQA1*, *PRKG1*, and *ST6GALNAC2*—might have undergone positive selection. These genes are associated with lactation, reproduction, meat

quality, and growth traits. All findings in this study provide useful data for the study of pigs, however our results are preliminary and need to validate these genes in another set of samples.

Supplementary materials

Supplementary data including two tables can be found with this article online at <http://www.genominfo.org/src/sm/gni-13-137-s001.pdf>.

Acknowledgments

This work was supported by a grant from the Next-Generation BioGreen 21 Program (No. PJ008116), Rural Development Administration, Republic of Korea. All authors read and approved the final manuscript.

References

- Hafez ES. *Adaptation of Domestic Animals*. Philadelphia: Lea & Febiger, 1968.
- Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, *et al.* Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci USA* 2007;104:15276-15281.
- Williams NH, Stahly TS, Zimmerman DR. Effect of chronic immune system activation on the rate, efficiency, and composition of growth and lysine needs of pigs fed from 6 to 27 kg. *J Anim Sci* 1997;75:2463-2471.
- Wilkinson S, Lu ZH, Megens HJ, Archibald AL, Haley C, Jackson IJ, *et al.* Signatures of diversifying selection in European pig breeds. *PLoS Genet* 2013;9:e1003453.
- Church PC, Goscinski A, Lefevre C. EXP-PAC: providing comparative analysis and storage of next generation gene expression data. *Genomics* 2012;100:8-13.
- Nakamura M, Tomita A, Nakatani H, Matsuda T, Nadano D. Antioxidant and antibacterial genes are upregulated in early involution of the mouse mammary gland: sharp increase of ceruloplasmin and lactoferrin in accumulating breast milk. *DNA Cell Biol* 2006;25:491-500.
- Krege JH, Hodgin JB, Couse JF, Enmark E, Warner M, Mahler JF, *et al.* Generation and reproductive phenotypes of mice lacking estrogen receptor beta. *Proc Natl Acad Sci USA* 1998; 95:15677-15682.
- Fonseca S, Wilsons IJ, Horgan GW, Maltin CA. Slow fiber cluster pattern in pig longissimus thoracis muscle: implications for myogenesis. *J Anim Sci* 2003;81:973-983.
- Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res* 2010;20:393-402.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; 419:832-837.
- Szpiech ZA, Hernandez RD. selscan: an efficient multi-threaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 2014;31:2824-2827.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078-2079.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-1303.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, *et al.* The variant call format and VCFtools. *Bioinformatics* 2011;27:2156-2158.
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009; 84:210-223.
- Park W, Kim J, Kim HJ, Choi J, Park JW, Cho HW, *et al.* Investigation of de novo unique differentially expressed genes related to evolution in exercise response during domestication in Thoroughbred race horses. *PLoS One* 2014;9: e91418.
- Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004;32:D258-D261.
- Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res* 2002;30:42-46.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, *et al.* DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:P3.
- Reumers J, Schymkowitz J, Ferkinghoff-Borg J, Stricher F, Serrano L, Rousseau F. SNPeff: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. *Nucleic Acids Res* 2005;33:D527-D532.
- Pareek CS. Emerging trends on functional genomics in farm animals. In: *Advances in Applied Biotechnology* (Parihar P, ed.). Jodhpur: Agrobios, 2010. pp. 23-34.
- Lee T, Shin DH, Cho S, Kang HS, Kim SH, Lee HK, *et al.* Genome-wide association study of integrated meat quality-related traits of the duroc pig breed. *Asian-Australas J Anim Sci* 2014;27:303-309.
- Ma Y, Zhang H, Zhang Q, Ding X. Identification of selection footprints on the X chromosome in pig. *PLoS One* 2014; 9:e94911.