



# Evolutionary and Functional Analysis of Korean Native Pig Using Single Nucleotide Polymorphisms

Jongin Lee<sup>1,2</sup>, Nayoung Park<sup>1,2</sup>, Daehwan Lee<sup>1</sup>, and Jaebum Kim<sup>1,\*</sup>

<sup>1</sup>Department of Biomedical Science and Engineering, Konkuk University, Seoul 05029, Korea, <sup>2</sup>These authors contributed equally to this work.

\*Correspondence: [jbkim@konkuk.ac.kr](mailto:jbkim@konkuk.ac.kr)  
<https://doi.org/10.14348/molcells.2020.0040>  
[www.molcells.org](http://www.molcells.org)

**Time and cost-effective production of next-generation sequencing data has enabled the performance of population-scale comparative and evolutionary studies for various species, which are essential for obtaining the comprehensive insight into molecular mechanisms underlying species- or breed-specific traits. In this study, the evolutionary and functional analysis of Korean native pig (KNP) was performed using single nucleotide polymorphism (SNP) data by comparative and population genomic approaches with six different mammalian species and five pig breeds. We examined the evolutionary history of KNP SNPs, and the specific genes of KNP based on the uniqueness of non-synonymous SNPs among the used species and pig breeds. We discovered the evolutionary trajectory of KNP SNPs within the used mammalian species as well as pig breeds. We also found olfaction-associated functions that have been characterized and diversified during evolution, and quantitative trait loci associated with the unique traits of KNP. Our study provides new insight into the evolution of KNP and serves as a good example for a better understanding of domestic animals in terms of evolution and domestication using the combined approaches of comparative and population genomics.**

**Keywords:** domestication, evolution, Korean native pig, single nucleotide polymorphism

## INTRODUCTION

Advances in next-generation sequencing (NGS) technologies have significantly reduced the time and cost of producing population-scale genome sequences of various species. Many population-scale genome projects, such as the 1,000 genomes project (Auton et al., 2015), 1,000 bull genomes project (Daetwyler et al., 2014) and 100,000 genomes project (Turnbull et al., 2018), have been launched to understand genetic diversity in a population as well as their phenotypic consequences. Genome projects of domestic animals have been of particular interest due to their importance as a food resource for humans. For example, variants associated with milk production and curly coat were identified by the sequencing of 234 cattle genomes in the 1,000 bull genomes project (Daetwyler et al., 2014), and genes related to lipid metabolism were identified in the sheep genome project (Naval-Sanchez et al., 2018). Evolutionary studies of the various distinctive traits of domesticated mammals have also been performed to acquire comprehensive insights into molecular mechanisms underlying species-specific traits (Chen et al., 2019; Gordon et al., 2007; Lee et al., 2017; Vieira and Rozas, 2011).

Pigs are a member of the order Artiodactyla, which represents even-toed ungulates such as pigs, deer, giraffes, sheep, and goats. They diverged from rodents and primates approximately 94 million years ago (Groenen et al., 2012), and were first domesticated from wild boar populations in

Received 5 February, 2020; revised 21 July, 2020; accepted 24 July, 2020; published online 18 August, 2020

eISSN: 0219-1032

©The Korean Society for Molecular and Cellular Biology. All rights reserved.

©This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

Europe and Asia 10,000 years ago (Giuffra et al., 2000). Since then, pigs have played an important role as a major source of meat for humans, and subsequent stages of selective breeding have significantly improved economically important traits, such as meat quality, backfat depth, and growth rate. A number of population-scale studies have been performed on various pig breeds to understand genomic features, such as genes, regulatory elements and variants, and the underlying molecular mechanisms affecting such beneficial traits (Davoli and Braglia, 2007; Sachs and Galli, 2009; Zhu et al., 2017).

The Korean native pig (KNP) is a domestic pig breed in Korea that originated from northern China approximately 2,000 years ago (Kim and Choi, 2002). KNP has higher fat content in its muscle, and more chewy meat texture, but also has a lower growth rate and poorer productivity than other pig breeds (Choi et al., 2015). To uncover the genomic determinants of KNP-specific traits, it is necessary to compare the genomes of many different pig breeds as well as other related species. To this end, there have been genome-wide association studies aimed at identifying genomic elements, such as genes and single nucleotide polymorphisms (SNPs), that have relevance to economically important traits of KNP, such as meat quality and productivity (Choi et al., 2015; Edea and Kim, 2014; Lee et al., 2012a; 2012b). These genomic bases for specific traits have gained their functions as a result of natural or artificial selection (selective breeding) during evolution. Therefore, studies designed to explore evolutionary perspectives can contribute to extending our understanding of KNP-specific genomic features and their functional consequences. This type of study was recently performed for Korean cattle (Lee et al., 2016) and the Korean bob-tailed native dog (Lee et al., 2017) to discover evolutionary changes of SNPs, genes, and interactions, which were associated with their characteristic traits. These studies also led to the development of a web interface for easily identifying the rewiring of protein-protein interactions that have occurred during evolution (Kwon et al., 2018).

In this study, we identified SNPs and genes relevant to KNP-specific traits, and discovered the mode and extent of their evolutionary changes using NGS data of six mammalian species (cattle, sheep, horse, dog, cat, and human) and six pig breeds (Korean wild boars [KWb], Duroc [DUR], Landrace [LAN], Yorkshire [YOR], Minipig [MINI], and KNP). We predicted the evolutionary trajectory of KNP SNPs within the used species as well as pig breeds. In addition, inter-species KNP-specific genes were identified based on the unique presence of SNPs in KNP, and their association with the olfactory function was discovered by enrichment tests using the Gene Ontology (GO) and pathway databases. Finally, the potential effects of intra-species KNP-specific nsSNPs within the different pig breeds on KNP-specific traits were examined by using pig quantitative trait loci regions. This study provides new insight into the evolution of the KNP and serves as a good example of the use of both population and comparative genomic approaches for a better understanding of domestic animals in terms of evolution and domestication.

## MATERIALS AND METHODS

### Species selection

The six mammalian species, *Bos Taurus* (cattle), *Canis lupus familiaris* (dog), *Equus caballus* (horse), *Felis catus* (cat), *Ovis aries* (sheep), and *Homo sapiens* (human), were used for inter-species evolutionary analyses. These species were selected based on several criteria, including relatedness to the pig, the existence of chromosome-level assembly, and the availability of the dbSNP data. Additionally, human was selected to be used as an outgroup species for evolutionary analyses. Six pig breeds, KWb, KNP, DUR, YOR, LAN, and MINI, were used for intra-species population analyses.

### Sequence and annotation data preparation

The genome sequences of the seven mammalian species were obtained from the National Center for Biotechnology Information (NCBI) RefSeq database (O'Leary et al., 2016). Genome sequence identifiers are GCF\_000003055.6 (cattle, bosTau8), GCF\_000002285.3 (dog, canFam3), GCF\_000002285.3 (horse, equCab2), GCF\_000181335.2 (cat, felCat8), GCF\_000298735.2 (sheep, oviAri3), GCF\_000003025.6 (pig, susScr11), and GCF\_000001405.33 (human, hg38). Next generation sequencing (NGS) data of six pig breeds was obtained from previous studies (Choi et al., 2015; Kim et al., 2015). The NGS data was generated from a total of 59 samples, including 10 of KNP, 10 of KWb, 6 of DUR, 14 of LAN, 15 of YOR, and 4 of MINI. Also, the coding DNA sequences and amino acid sequences of pig (assembly version: Sscrofa11.1), which are required for gene annotation and functional enrichment analysis, were downloaded from the Ensembl database (Zerbino et al., 2018). Gene annotation information which is formatted as the GTF format was downloaded from the Ensembl database. And, for the functional enrichment analyses, GO and orthologous gene information were obtained by the Ensembl Genes 95 version of the Ensembl biomaRT (Kinsella et al., 2011).

### Single nucleotide polymorphism identification

SNPs were detected for each pig breed separately by integrating NGS data from multiple samples. The SNP calling and filtration were performed by the guidelines of the GATK Best Practice pipelines (Van der Auwera et al., 2013). SNPs were called using BWA-mem v0.7.15-r1140 with an additional -M option (Li, 2013) for NGS read alignment on the reference pig genome (assembly version: Sscrofa11.1), Samtools v1.3.1 (Li, 2011) and Picard v2.17.11 (<https://broadinstitute.github.io/picard/>) for preprocessing of alignments, and GATK v3.8 (DePristo et al., 2011) for SNP calling and filtering. To obtain high-quality SNPs, we filtered the called SNPs by several criteria: QD (Variant Confidence/Quality by Depth) < 2.0, MQ (RMS Mapping Quality) < 40.0, FS (Phred-scaled P value using Fisher's exact test to detect strand bias) > 60.0, HaplotypeScore (Consistency of the site with at most two segregating haplotypes) > 13.0, MQRankSum (Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities) < -2.5, and ReadPosRankSum (Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias) < -8.0. Then, the transition to transversion (Ti/Tv) ratio was calculated to ensure the qual-

ity of the called SNPs by vcftools v0.1.17 with --TsTv-summary option (Danecek et al., 2011). For seven mammalian species, SNP datasets were downloaded from the dbSNP database (build 148 for cat; 150 for cattle and pig; 151 for dog, horse, sheep and human) (Sherry et al., 2001).

### Identification of orthologous genomic positions for SNP evolutionary analysis

Pairwise whole-genome sequence alignments of five species (cattle, sheep, horse, dog, and cat) against pig were built by using the LASTZ program (Harris, 2007) with E = 30, H = 2000, K = 3000, L = 2200, M = 50, O = 400, T = 1, and Y = 9400 options. The pairwise whole-genome sequence alignment between pig and human was downloaded from the UCSC genome browser (Haeussler et al., 2019). By using the whole-genome sequence alignments and the liftOver tool (Kent et al., 2002), the orthologous genomic positions of KNP SNPs in other mammalian species were identified.

### Phylogenetic information preparation

The phylogenetic information among the seven mammalian species (pig, cattle, sheep, horse, dog, cat, and human) was downloaded from the TimeTree website (Kumar et al., 2017). The phylogenetic relationships among the six pig breeds were estimated using the SNPPhylo program (Lee et al., 2014) based on their SNP data. In the case of the six pig breeds, we used only the topology, not branch lengths, of the estimated phylogenetic tree.

### Inference of SNP changes during evolution

The evolutionary changes, such as appearance and disappearance of KNP SNPs during evolution, were inferred by using the maximum parsimony algorithm which minimizes the total number of changes during the evolutionary process (Farris, 1970; Fitch, 1971). First, the existence or absence of SNPs at the orthologous genomic positions in other species were identified. Then, the existence or absence of SNPs at each SNP position in the common ancestors was predicted by considering the existence of SNPs at orthologous genomic positions in all species using the maximum parsimony algorithm. Human was used as an outgroup species to resolve ambiguous cases. Next, we inferred the appearance or disappearance of them at each branch by comparing the SNP existence between all ancestor-descendant pairs. The SNP appearance and disappearance were defined as SNP absence in ancestor and existence in the descendant, and SNP existence in ancestor and absence in the descendant, respectively. In the case of evolutionary analysis within six pig breeds, a similar analysis was performed by using the other six mammalian species (except pig) as outgroup species. The maximum parsimony algorithm and inference of evolutionary changes were implemented by in-house scripts using the Perl programming language.

### Definition of the evolutionary speed in inter- and intra-species evolutionary analysis

In the inter-species evolutionary analysis, for comparing the evolutionary speed of each branch, the rate of SNP changes was defined as the number of appeared and disappeared

SNPs per billion year. For intra-species evolutionary analysis, the ratio of SNPs changes was used, which is the fraction of SNP changes against the number of SNPs in the immediate common ancestor nodes.

### Identification of positively selected genes

To identify positively selected genes (PSGs) in KNP, we generated multiple protein sequence alignments for the inter-species KNP-specific genes and all corresponding orthologous genes (Ensembl Genes 95) in other mammalian species using MUSCLE v3.8.425 (Edgar, 2004). Out of 1,747 inter-species KNP-specific genes, a total of 252 genes with one or more corresponding orthologous genes in each of the other mammalian species were used in this analysis. Gaps in the multiple protein alignment results were removed and the alignments were transformed to DNA sequence alignments by using trimAl v1.4. with -phylip\_paml -nogaps -splitbystop-codon options (Capella-Gutierrez et al., 2009). Also, the DNA sequence alignments were filtered out by the length cutoff (<150 bp). The PSGs were identified by using the codeml program in the PAML package (Yang, 1997) with branch-site models which were specified by the following model parameters: model = 2, NSsites = 2, omega = 1, and fix\_omega = 1 for a null model and fix\_omega = 0 for an alternative model. The alternative and null models were used for modeling the dN/dS ratios on branches, and the dN/dS ratio was fixed to 1 for all branches in the null model. The Chi-square test was used to check the statistical significance of each gene, and q-values (False Discovery Rate-adjusted *P* value) were calculated for multiple test correction. For some KNP genes, there were multiple orthologous genes in other species. However, only one of the orthologous genes per species can be used in the test for PSGs. Therefore, the codeml program was run for each of different combinations of orthologous genes (one for each species) with the q-value cutoff 0.05, and the KNP genes with at least one of the combinations passing the q-value cutoff were considered as the PSG.

### SNP annotation, and functional enrichment and pathway analysis

The KNP-specific SNP sets were annotated, and their functional effects were predicted by the SnpEff program v4.3t (Cingolani et al., 2012). The pig reference genome, CDS sequences, and gene annotation information of pig (assembly version: Sscrofa11.1) were used for SNP annotation. Non-synonymous SNPs (nsSNPs) were identified by selecting the SNPs which were classified into the four sequence ontology terms: missense variant, start loss, stop gained and stop loss. Functional enrichment analyses of the annotated genes were performed by enrichment test with GO information. The enrichment tests were conducted by the hypergeometric test, and q-values were calculated for multiple test correction with 0.05 as a cutoff. The statistical tests and corrections were performed with R (<https://www.R-project.org>). Additionally, we performed the pathway analysis using the additional pathway databases, KEGG (Kanehisa et al., 2019) and REACTOME (Fabregat et al., 2018), by g:GOST, which is the functional profiling module of g:Profiler (Reimand et al., 2016), with default parameters.

### Analysis with quantitative trait loci

Pig quantitative trait loci (QTL) data was obtained from the Animal QTL database (Release 39; pig assembly version SS\_11.1) (Hu et al., 2019). To estimate the degree of association between the intra-species KNP-specific nsSNPs and QTLs, the number of intra-species KNP-specific nsSNPs which were placed in the QTL regions was normalized by million base pair. The example of intra-species KNP-specific nsSNPs was visualized by using the UCSC genome browser track hub (Raney et al., 2014) with pig RefSeq genes (O'Leary et al., 2016).

### Data access

All SNPs of the six pig breeds detected in this study have been submitted to the European Variation Archive database (<https://www.ebi.ac.uk/eva/>; project ID: PRJEB34412). The intra-species KNP-specific nsSNP data within the pig QTL regions is available on My Hubs page on the UCSC track hub website (<https://genome.ucsc.edu/cgi-bin/hgHubConnect>) using the hub URL ([http://bioinfo.konkuk.ac.kr/KNP\\_nsSNPs/hub.txt](http://bioinfo.konkuk.ac.kr/KNP_nsSNPs/hub.txt)).

## RESULTS

### Evolutionary analysis of KNP SNPs among related species

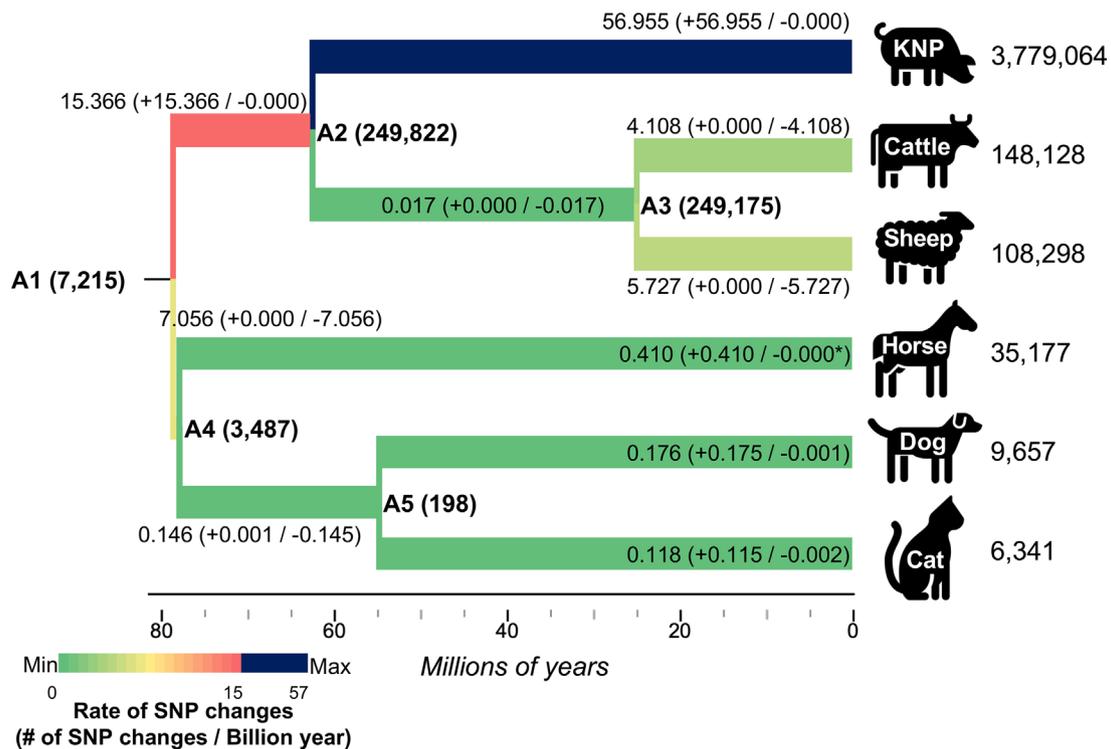
A total of 36,283,251 SNPs was called from six pig breeds:

KNP, KWB, DUR, LAN, YOR, and MINI (Table 1; Materials and Methods section). Among them, more than 80% of SNPs in each pig breed were found in the dbSNP database (build 150). The transition to transversion (Ti/Tv) ratio was ranged from 2.31 to 2.36, which is similar to the numbers found in other studies of pigs (Choi et al., 2015; Li et al., 2017). For evolutionary analysis, additional SNPs of other six mammalian species (cattle, sheep, horse, dog, cat, and human) were downloaded from the dbSNP database (Supplementary Table S1; Materials and Methods section).

**Table 1.** Statistics of SNPs of six pig breeds

Pig breeds	No. of identified SNPs	No. of SNPs in dbSNP <sup>a</sup>	Ti/Tv ratio
KNP	16,545,631	13,930,964 (0.84)	2.32
KWB	23,528,922	19,587,524 (0.83)	2.36
DUR	11,518,093	9,677,013 (0.84)	2.32
LAN	19,717,820	16,346,309 (0.83)	2.29
YOR	18,422,733	15,161,229 (0.82)	2.27
MINI	10,293,949	8,600,902 (0.84)	2.31
Total	36,283,251	29,077,408 (0.80)	

<sup>a</sup>The fractions of identified SNPs found in dbSNP (build 150; total 58,431,079 pig SNPs) are shown in parentheses. For calling SNPs, the pig reference assembly version susScr11 was used.



**Fig. 1. Evolution of KNP SNPs.** The number next to KNP represents the number of KNP SNPs used in this analysis, and the numbers next to the other descendant species show the number of SNPs found at orthologous genomic positions of KNP SNPs in those species. The numbers next to the names of ancestors (A1 to A5) are the number of KNP SNPs predicted to be present in those ancestors. The numbers on branches represent the rate of SNP changes which is the number of appearing and disappearing events per billion year on each branch. The rate of appearance and disappearance events are separately shown in parentheses with + and - sign respectively. The zero rate in parentheses with an asterisk (\*) symbol means a very small rate that cannot be shown with a current precision of numbers.

To understand the evolutionary patterns of KNP SNPs, we traced their evolutionary history based on SNP information in other six mammalian species shown above, orthologous genomic positions of KNP SNPs in the used species, and the predicted SNPs in the common ancestors of the used species by the maximum parsimony algorithm (Materials and Methods section). In this analysis, only 3,779,064 KNP SNPs, which have their orthologous genomic positions in all other six mammalian species, were used. As shown in Fig. 1, less than 4% out total of 3,779,064 KNP SNPs were observed at orthologous genomic positions in other species, and the percentage was decreased as a function of divergence from KNP. A total of 7,215 KNP SNPs in A1, 249,822 KNP SNPs in A2, 249,175 KNP SNPs in A3, 3,487 KNP SNPs in A4, and 198 KNP SNPs in A5 were predicted as being present in each ancestor, which also showed a similar decreasing pattern of the number of KNP SNPs as divergence time increases. Those ancestral SNPs were further compared to discover the evolutionary changes of KNP SNPs by calculating the rate of SNP changes, which was defined as the number of appeared and disappeared SNPs per billion year. For branches from A1 to KNP, a total of 15.366 KNP SNPs per billion year appeared from A1 to A2, and this rate increased to 56.955 from A2 to KNP. Among KNP, cattle, and sheep, the number of ancestral SNPs was almost preserved from the ancestor A2 to A3, and more than 40% of SNPs disappeared during evolution from A3 to cattle and sheep. As expected, a very small number of SNPs was predicted to be present in A4 compared with A2. A4 is the common ancestor of horse, dog, and cat, which are more divergent species than cattle and sheep from KNP. Interestingly, the number of SNPs was increased from A4 to horse, and from A5 to dog and cat, which was opposite to the pattern observed from A2 to KNP, cattle, and sheep. This was because compared with the number of orthologous genomic positions with SNPs in each species (horse, dog, and

cat), only a small number of genomic positions have common SNPs in all three species, which led to the prediction of a small number of ancestral SNPs in A4 and A5.

### Functional enrichment and pathway analysis of KNP genes with SNPs only in KNP

The SNPs in six mammalian species (KNP, cattle, sheep, horse, dog, and cat) were used to identify KNP genes with SNPs not found at orthologous genome positions in the other five species (hereafter “inter-species KNP-specific genes”; 1,747 genes found). The GO enrichment test was then performed with q-value cutoff 0.05 (Materials and Methods section). The inter-species KNP-specific genes were enriched in three major categories of function; olfaction, immune system, and translation (Supplementary Table S2).

The aforementioned 1,747 inter-species KNP-specific genes were further reduced to genes with only nsSNPs (total 1,016 genes found). The GO enrichment test was also performed as above, and a total of 34 functions in the three GO categories were significantly enriched for those genes (Table 2, Supplementary Table S3). In terms of the biological process category, olfaction-associated GO terms, such as G protein-coupled receptor signaling pathway (GO:0007186), signal transduction (GO:0007165), detection of chemical stimulus involved in sensory perception of smell (GO:0050911), response to stimulus (GO:0050896), and sensory perception of smell (GO:0007608), were significantly enriched. Pathway analysis using g:Profiler (Reimand et al., 2016) also confirmed that those genes are significantly related to olfactory function (Materials and Methods section; Supplementary Table S4). For example, olfaction-related pathways, such as olfactory transduction (KEGG:04740) in the KEGG database (Kanehisa et al., 2019), and olfactory signaling pathway (REAC:R-SSC-381753), G alpha (s) signaling events (REAC:R-SSC-418555), GPCR downstream signaling (REAC:R-

**Table 2.** Significantly enriched functions (q-value cutoff 0.05) of 1,016 inter-species KNP-unique genes containing non-synonymous SNPs only in KNP not in the five other mammalian species

GO ID	GO term	No. of associated genes	q-value
GO:0007186	G protein-coupled receptor signaling pathway	523	< 1.0E-05
GO:0007165	Signal transduction	515	< 1.0E-05
GO:0050911	Detection of chemical stimulus involved in sensory perception of smell	514	< 1.0E-05
GO:0050896	Response to stimulus	490	< 1.0E-05
GO:0007608	Sensory perception of smell	489	< 1.0E-05
GO:0006412	Translation	105	< 1.0E-05
GO:0002323	Natural killer cell activation involved in immune response	9	4.0E-04
GO:0002286	T cell activation involved in immune response	9	6.9E-04
GO:0033141	Positive regulation of peptidyl-serine phosphorylation of STAT protein	9	6.9E-04
GO:0022900	Electron transport chain	15	8.5E-04
GO:0006952	Defense response	16	2.7E-03
GO:0042100	B cell proliferation	9	7.9E-03
GO:0006826	Iron ion transport	9	1.5E-02
GO:0043330	Response to exogenous dsRNA	9	1.5E-02
GO:0006959	Humoral immune response	9	1.7E-02
GO:0098664	G protein-coupled serotonin receptor signaling pathway	7	2.7E-02

The functions in the biological process category are shown (others are in Supplementary Table S3).

SSC-388396), signaling by GPCR (REAC:R-SSC-372790), signal transduction (REAC:R-SSC-162582), regulation of IFNA signaling (REAC:R-SSC-912694), and interferon alpha/beta signaling (REAC:R-SSC-909733) in the Reactome database (Fabregat et al., 2018), were found to be significantly enriched. The 1,016 genes were also significantly enriched for the GO terms related to the immune system, such as natural killer cell activation involved in immune response (GO:0002323), T cell activation involved in immune response (GO:0002286), positive regulation of peptidyl-serine phosphorylation of STAT protein (GO:0033141), B cell proliferation (GO:0042100), defense response (GO:0006952), response to exogenous dsRNA (GO:0043330), and humoral immune response (GO:0006959).

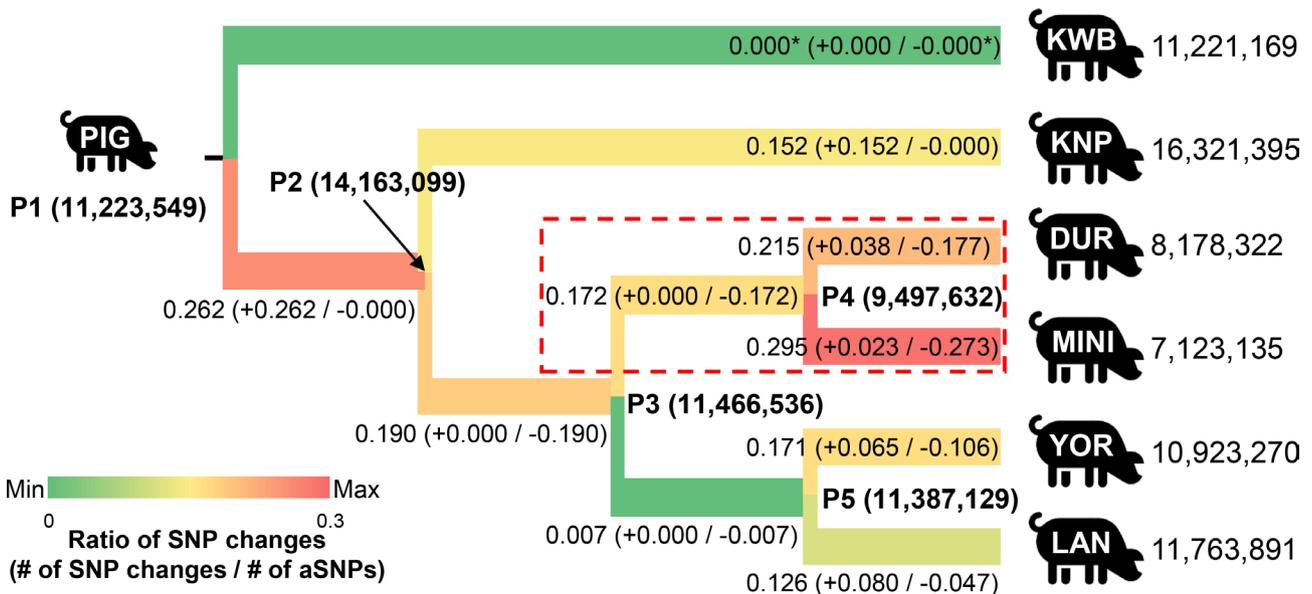
To find KNP genes strongly affected by selective pressure during evolution, PSGs in KNP were identified from 252 out of 1,747 inter-species KNP-specific genes, which have orthologous genes in all of the other six mammalian species (cattle, sheep, horse, dog, cat, and human; Materials and Methods section). A total of 48 PSGs were discovered (Supplementary Table S5), and the GO enrichment test showed that most of the PSGs were enriched in olfaction-related GO terms, such as sensory perception of smell (GO:0007608), response to stimulus (GO:0050896), and olfactory receptor activity (GO:0004984) (Supplementary Table S6).

#### Evolutionary analysis of KNP SNPs within pig breeds

To study the evolution of KNP SNPs during the domestication process of six pig breeds (KNP, KWB, DUR, LAN, YOR, and MINI), the presence and absence of KNP SNPs in ancestors

were predicted using the maximum parsimony algorithm based on the phylogenetic tree obtained from the SNP data (Materials and Methods section; Fig. 2). In the phylogenetic tree, a subtree was created with KNP and the other four pig breeds (DUR, MINI, YOR, and LAN), and KWB was placed outside of the subtree. From 44% (MINI) to 72% (LAN) of KNP SNPs were shared with the other pig breeds. The existence information of KNP SNPs in other pig breeds was used to predict the presence of KNP SNPs in ancestors, and 11,223,549 KNP SNPs in P1, 14,163,099 KNP SNPs in P2, 11,466,536 KNP SNPs in P3, 9,497,632 KNP SNPs in P4, and 11,387,129 KNP SNPs in P5 were predicted as present. Note that the number of KNP SNPs in this analysis was smaller than the number in Table 1 because some of KNP SNPs which failed to produce unambiguous ancestral SNPs were filtered out.

For profiling and comparing the evolutionary changes of KNP SNPs in the used pig breeds, we calculated the ratio of SNP changes, which is defined as the number of SNP changes against the number of SNPs in an immediate ancestor on each branch. Note that the rate per billion year, which was used in the inter-species evolution study, was not used here because the exact time of domestication of each breed is hard to obtain. For branches from P1 to KNP, the gradual increase of SNPs was observed (ratio +0.262 from P1 to P2, +0.152 from P2 to KNP). However, the number of ancestral SNPs was decreased from P2 to four pig breeds (DUR, MINI, YOR, and LAN), and the ratio of SNP disappearance from P3 to P4 (ratio -0.172) was almost 25 folds greater than the ratio from P3 to P5 (ratio -0.007). The biggest and smallest



**Fig. 2. Changes of SNPs during pig domestication.** The numbers next to the names of ancestors (P1 to P5) and descendant pig breeds represent the numbers of existing SNPs. The numbers on branches represent the ratio of SNP changes which is the number of appeared and disappeared SNPs against the number of present SNPs of an immediate ancestor (aSNPs). The ratio of SNP appearance and disappearance are separately shown in parentheses with + and - sign, respectively. The numbers which are higher in disappeared SNPs than appeared SNPs are represented in bold, and the red-dashed rectangle highlights the group that has undergone strong negative selection pressure. The tree topology was predicted by SNPhylo. The zero rate in parentheses with an asterisk (\*) symbol means a very small rate that cannot be shown with a current precision of numbers.

**Table 3.** Significantly enriched functions of 2,608 KNP genes with unique non-synonymous SNPs

GO ID	GO term	GO domain	No. of associated genes	q-value
GO:0007165	Signal transduction	BP	295	< 1.0E-05
GO:0007186	G protein-coupled receptor signaling pathway	BP	307	< 1.0E-05
GO:0007608	Sensory perception of smell	BP	277	< 1.0E-05
GO:0050896	Response to stimulus	BP	278	< 1.0E-05
GO:0050911	Detection of chemical stimulus involved in sensory perception of smell	BP	294	< 1.0E-05
GO:0005886	Plasma membrane	CC	310	< 1.0E-05
GO:0016020	Membrane	CC	454	< 1.0E-05
GO:0016021	Integral component of membrane	CC	451	< 1.0E-05
GO:0004930	G protein-coupled receptor activity	MF	289	< 1.0E-05
GO:0004984	Olfactory receptor activity	MF	294	< 1.0E-05
GO:0003735	Structural constituent of ribosome	MF	38	4.4E-02

BP, biological process; CC, cellular component; MF, molecular function.

SNP changes were observed on branches from P4 to MINI (ratio 0.295) and from P1 to KWB (ratio ~0), respectively.

### Functional enrichment and QTL analysis of KNP genes with SNPs absent in other pig breeds

From the comparison of SNP positions in six pig breeds, a total of 510,474 SNPs was identified as being present only in KNP. Among them, 4,304 were nsSNPs (hereafter “intra-species KNP-specific nsSNPs”). To examine the effect of these intra-species KNP-specific nsSNPs, we performed a functional enrichment analysis of 2,608 genes containing at least one of the 4,304 intra-species KNP-specific nsSNPs. Among the 2,608 genes, a total of 503 genes were associated with the 11 significantly enriched GO terms, and most of them were enriched in functions associated with olfaction (Table 3).

To further investigate the association with interesting traits, the intra-species KNP-specific nsSNPs were compared with pig QTL based on the Pig QTLdb (Hu et al., 2019). We obtained a total of 4,241 intra-species KNP-specific nsSNPs positions within 4,864 QTLs (Supplementary Table S7). Based on the normalized number of intra-species KNP-specific nsSNPs per million base pairs in each QTL, the top-ranked QTLs, including HDL/LDL ratio QTL, feed conversion ratio QTL, intramuscular fat content QTL, marbling QTL, and muscle moisture percentage QTL, were examined. We found that the intra-species KNP-specific nsSNPs are highly related to meat quality, especially, marbling and muscle moisture level. The intramuscular fat content QTL region, which contains 21 intra-species KNP-specific nsSNPs and the *TAS1R2* gene associated with dietary adaptation (da Silva et al., 2014), is visualized in Fig. 3 by using the UCSC genome browser track hub (Raney et al., 2014). In this region, there are 18 pig genes and 21 intra-species KNP-specific nsSNPs (Fig. 3A). Also, the example of the *TAS1R2* gene shows an amino acid change caused by the intra-species KNP-specific nsSNP (Fig. 3B).

### Commonality of nsSNPs and genes found in between inter- and intra-species analysis

The KNP-specific nsSNPs and genes obtained from the inter- and intra-species analysis were further examined to find their commonality. Among 7,992 inter-species and 4,304

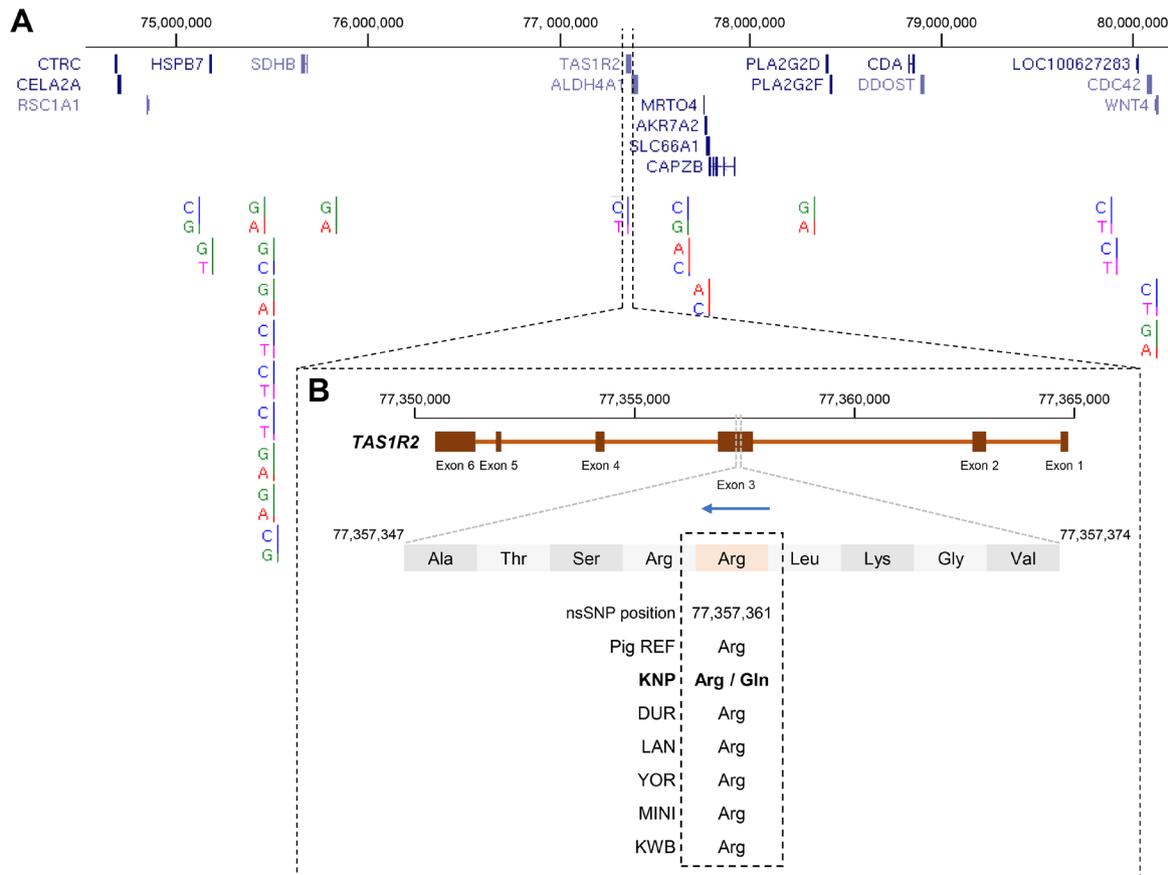
intra-species KNP-specific nsSNPs, 662 common nsSNPs were found (Supplementary Fig. S1A). These common nsSNPs contributed to 331 common genes between 1,747 inter-species and 2,608 intra-species KNP-specific genes (Supplementary Fig. S1B). The 331 common genes include the gene (ENSSSCG00000000615) annotated in the backfat-thickness related QTL region, which is one of the KNP-specific traits (Paszek et al., 2001), and other genes (ENSSSCG00000001827, ENSSSCG000000040964) related to the KNP-specific phenotypes such as carcass length (de Koning et al., 2003; Falker-Gieske et al., 2019).

## DISCUSSION

In this study, the evolutionary and functional analysis of KNP with the six mammalian species (cattle, sheep, horse, cat, dog, and human) and five different pig breeds (KWB, DUR, LAN, YOR, and MINI) were performed by using the SNP data. Through the evolutionary analysis, we discovered how KNP SNPs have generated during evolution within the used species as well as pig breeds. The inter-species KNP-specific genes with the uniquely present SNPs only in KNP and their functions associated with interesting traits were discovered using the GO and pathway databases. In addition, the relationships between intra-species KNP-specific nsSNPs and pig quantitative trait loci regions were examined.

In the evolutionary analysis, the changes of KNP SNPs during evolution were investigated by using the information of shared KNP SNPs in other mammalian species (Fig. 1) as well as other pig breeds (Fig. 2), and they were quantified in terms of divergence time (in the case of inter-species analysis) or the ratio of SNP changes (in the case of inter-pig breed analysis). To reduce confounding effect by the different number of known SNPs in the used species (or pig breeds), we only focused on KNP SNPs and their existence at the orthologous genomic positions in the other species (or pig breeds). If reliable evolutionary trajectories of SNPs of other species (or pig breeds) are obtained, the evolutionary history of KNP SNPs can be more characterized, and distinguishing evolutionary patterns compared with other species (or pig breeds) can be discovered.

### Intramuscular fat content QTL (chr6: 74,531,339 - 80,190,960)



**Fig. 3. Examples of intra-species KNP-specific nsSNPs.** (A) Pig RefSeq genes were represented by blue blocks with their gene symbols and coordinates on the top panel. The intra-species KNP-specific nsSNP information is represented by a vertical line containing a pair of two alleles (top: reference allele, bottom: alternative allele) under the gene annotations, and their relative fractions are shown by vertical lines with different heights for each allele. This track was displayed by the UCSC genome browser track hub. (B) In the *TAS1R2* gene example, the top panel shows gene structure with the direction of transcription (blue arrow), and bottom panel represents the comparison of amino acids in the third exon among six different pig breeds. Two different amino acids caused by the intra-species KNP-specific nsSNP are shown together with a slash delimiter.

A total of 1,016 out of 1,747 inter-species KNP-specific genes contained nsSNPs not found at orthologous genome positions in the other mammalian species (cattle, sheep, horse, dog, and cat). These genes were found in various data resources, such as the GO, KEGG and REACTOME database, as being significantly related to functions associated with olfaction, immune system and translation. These functions have been known as the characteristics of pigs associated with adaptation and domestication (Groenen et al., 2012; Vincent et al., 2012; Zhu et al., 2017). Pigs have the largest olfactory receptor gene families, including more than 1,300 functional olfactory receptor genes located on 16 pig chromosomes (Nguyen et al., 2012). Among the 1,016 inter-species KNP-specific genes, 177 genes were olfactory receptor genes (Supplementary Table S8) related with receptor signaling pathway and the detection of chemical stimulus involved in sensory perception of smell. In addition, immunity-related genes in pigs have highly expanded in the

evolutionary process. For example, pigs have more than 50 type I interferon genes, which are significantly more than in human and mouse (Sang et al., 2014). Immune response is significantly associated with the feed efficiency and product quality of pigs (Horodyska et al., 2018). In our analysis, a total of 12 immunity related genes were observed in the above inter-species KNP-specific genes (Supplementary Table S9).

Out of the 1,747 inter-species KNP-specific genes, a total of 48 genes were predicted as PSGs. Among them, six genes (ENSSSCG00000036715, ENSSSCG00000038050, ENSSSCG00000037486, ENSSSCG00000033023, ENSSSCG00000038781, ENSSSCG00000040604) were classified as the G protein-coupled receptor and annotated with odorant binding GO term (GO:0005549) which was enriched in 48 PSGs. It suggests that KNP have been characterized and diversified the functions associated with olfaction in the evolution process, which is in line with the characteristics of pigs found in previous evolutionary and

comparative studies (Groenen et al., 2012; Zhu et al., 2017). Additional GO terms, such as single-stranded RNA binding (GO:0003727) in the molecular function category and transposition, RNA-mediated (GO:0032197) in the biological process category, were also enriched in the 48 PSGs. Seven genes (ENSSSCG00000032719, ENSSSCG0000003778, ENSSSCG000000364866, ENSSSCG00000036856, ENSSSCG00000040198, ENSSSCG00000040767, and ENSSSCG00000033328) annotated with those GO terms belong to the LINE-1 retrotransposable element ORF1 protein subfamily that is highly related to LINE-1 which is known as main repetitive element groups in pig (Groenen et al., 2012).

We identified 4,241 intra-species KNP-specific nsSNPs positioned within 4,864 pig QTL regions. The associations between the intra-species KNP-specific nsSNPs and QTL regions were evaluated by calculating the normalized number of intra-species KNP-specific nsSNPs per million base pairs. Interestingly, the top-ranked QTL regions were the intramuscular fat content QTL and muscle moisture percentage QTL that are known to having positive correlation with the meat quality-related features such as high-fat content in muscle, meat tenderness, juiciness, and taste (Grindflek et al., 2001). The intramuscular fat content QTL is associated with the fat content and fatty acid composition in muscle (Hocquette et al., 2010; Sanchez et al., 2007), and the muscle moisture percentage QTL is also related to the meat quality in pigs (Kärst et al., 2011; Wang et al., 2016; Watanabe et al., 2018). Those features may be related to the formation of the well-known features of KNP meat, which are high-fat content in muscle and chewy texture (Cho et al., 2015). In addition, QTLs, such as the feed conversion ratio QTL and the total number born alive QTL, were also highly ranked and seemed to be associated with the poor productivity of KNP. Of interest was the example of the *TAS1R2* gene, which contains an amino acid change in KNP. The *TAS1R* gene family is known to be associated with taste and nutrient sensing as well as dietary adaptations (da Silva et al., 2014) (Fig. 3B), and therefore might also be related to the meat quality of KNP.

Lastly, the commonality of nsSNPs and genes between inter- and intra-species analysis was examined, and 662 nsSNPs within 331 genes were found as common. These nsSNPs and genes were observed in the KNP-specific traits-related QTL regions, such as the backfat-thickness related and carcass length QTL.

In summary, SNPs of KNP, other pig breeds, and related species were used to discover the mode and extent of the evolutionary changes of the KNP SNPs, inter- and intra-species KNP-specific nsSNPs, inter-species KNP-specific genes, their functions, and QTLs harboring those intra-species KNP-specific nsSNPs. As sequencing technologies become faster and cheaper, population-scale genomic studies have been actively performed for many species and breeds. Furthermore, comprehensive comparative and evolutionary analysis can help enhance a deeper understanding of species diversity and their history. For domestic animals in particular, it is very important to understand their genomic history and the characteristics which lead to phenotypic consequences, in order to preserve and improve their valuable traits. This study suggests a good example of both population and com-

parative genomic approaches for a better understanding of domestic animals in terms of evolution and domestication.

*Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).*

## ACKNOWLEDGMENTS

This work was supported by Agenda (project No. PJ01334302) of the National Institute of Animal Science funded by the Rural Development Administration (RDA), Republic of Korea, a grant (2014M3C9A3063544) funded by the Ministry of Science and ICT, Republic of Korea, and a grant (2019R1F1A1042018) funded by the Ministry of Education, Republic of Korea.

## AUTHOR CONTRIBUTIONS

J.K. conceived and supervised the study. J.L. and N.P. collected data. J.L., N.P., D.L., and J.K. performed analyses and interpreted results. J.L., N.P., and J.K. wrote the manuscript. All authors reviewed the manuscript.

## CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

## ORCID

Jongin Lee <https://orcid.org/0000-0001-7870-1526>  
Nayoung Park <https://orcid.org/0000-0003-2722-0639>  
Daehwan Lee <https://orcid.org/0000-0002-7210-0786>  
Jaebum Kim <https://orcid.org/0000-0002-2287-9760>

## REFERENCES

- Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973.
- Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., et al. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* 364, eaav6202.
- Cho, I.C., Yoo, C.K., Lee, J.B., Jung, E.J., Han, S.H., Lee, S.S., Ko, M.S., Lim, H.T., and Park, H.B. (2015). Genome-wide QTL analysis of meat quality-related traits in a large F2 intercross between Landrace and Korean native pigs. *Genet. Sel. Evol.* 47, 7.
- Choi, J.W., Chung, W.H., Lee, K.T., Cho, E.S., Lee, S.W., Choi, B.H., Lee, S.H., Lim, W., Lim, D., Lee, Y.G., et al. (2015). Whole-genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. *DNA Res.* 22, 259-267.
- Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). 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 (Austin)* 6, 80-92.
- da Silva, E.C., de Jager, N., Burgos-Paz, W., Reverter, A., Perez-Enciso, M., and Roura, E. (2014). Characterization of the porcine nutrient and taste receptor gene repertoire in domestic and wild populations across the globe. *BMC Genomics* 15, 1057.
- Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brondum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of

- monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858-865.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
- Davoli, R. and Braglia, S. (2007). Molecular approaches in pig breeding to improve meat quality. *Brief. Funct. Genomic. Proteomic.* **6**, 313-321.
- de Koning, D.J., Pong-Wong, R., Varona, L., Evans, G.J., Giuffra, E., Sanchez, A., Plastow, G., Noguera, J.L., Andersson, L., and Haley, C.S. (2003). Full pedigree quantitative trait locus analysis in commercial pigs using variance components. *J. Anim. Sci.* **81**, 2155-2163.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491-498.
- Edea, Z. and Kim, K.S. (2014). A whole genomic scan to detect selection signatures between Berkshire and Korean native pig breeds. *J. Anim. Sci. Technol.* **56**, 23.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649-D655.
- Falker-Gieske, C., Blaj, I., Preuß, S., Bennewitz, J., Thaller, G., and Tetens, J. (2019). GWAS for meat and carcass traits using imputed sequence level genotypes in pooled F2-designs in pigs. *G3 (Bethesda)* **9**, 2823-2834.
- Farris, J.S. (1970). Methods for computing Wagner trees. *Syst. Zool.* **19**, 83-92.
- Fitch, W.M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**, 406-416.
- Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T., and Andersson, L. (2000). The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785-1791.
- Gordon, L., Yang, S., Tran-Gyamfi, M., Baggott, D., Christensen, M., Hamilton, A., Crooijmans, R., Groenen, M., Lucas, S., Ovcharenko, I., et al. (2007). Comparative analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of gene-rich vertebrate regions. *Genome Res.* **17**, 1603-1613.
- Grindflek, E., Szyda, J., Liu, Z., and Lien, S. (2001). Detection of quantitative trait loci for meat quality in a commercial slaughter pig cross. *Mamm. Genome* **12**, 299-304.
- Groenen, M.A., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.J., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398.
- Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853-D858.
- Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. Ph.D. thesis in Computer Science and Engineering (State College: The Pennsylvania State University).
- Hocquette, J.F., Gondret, F., Baéza, E., Médale, F., Jurie, C., and Pethick, D.W. (2010). Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers. *Animal* **4**, 303-319.
- Horodyska, J., Wimmers, K., Reyer, H., Trakooljul, N., Mullen, A.M., Lawlor, P.G., and Hamill, R.M. (2018). RNA-seq of muscle from pigs divergent in feed efficiency and product quality identifies differences in immune response, growth, and macronutrient and connective tissue metabolism. *BMC Genomics* **19**, 791.
- Hu, Z.L., Park, C.A., and Reecy, J.M. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* **47**, D701-D710.
- Kärst, S., Cheng, R., Schmitt, A.O., Yang, H., de Villena, F.P., Palmer, A.A., and Brockmann, G.A. (2011). Genetic determinants for intramuscular fat content and water-holding capacity in mice selected for high muscle mass. *Mamm. Genome* **22**, 530-543.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590-D595.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996-1006.
- Kim, H., Song, K.D., Kim, H.J., Park, W., Kim, J., Lee, T., Shin, D.H., Kwak, W., Kwon, Y.J., Sung, S., et al. (2015). Exploring the genetic signature of body size in Yucatan miniature pig. *PLoS One* **10**, e0121732.
- Kim, K.S. and Choi, C.B. (2002). Genetic structure of Korean native pig using microsatellite markers. *Korean J. Genet.* **24**, 1-7.
- Kinsella, R.J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**, bar030.
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812-1819.
- Kwon, D., Lee, D., Kim, J., Lee, J., Sim, M., and Kim, J. (2018). INTERSPIA: a web application for exploring the dynamics of protein-protein interactions among multiple species. *Nucleic Acids Res.* **46**, W89-W94.
- Lee, D., Cho, M., Hong, W.Y., Lim, D., Kim, H.C., Cho, Y.M., Jeong, J.Y., Choi, B.H., Ko, Y., and Kim, J. (2016). Evolutionary analyses of Hanwoo (Korean cattle)-specific single-nucleotide polymorphisms and genes using whole-genome resequencing data of a Hanwoo population. *Mol. Cells* **39**, 692-698.
- Lee, D., Lim, D., Kwon, D., Kim, J., Lee, J., Sim, M., Choi, B.H., Choi, S.G., and Kim, J. (2017). Functional and evolutionary analysis of Korean bob-tailed native dog using whole-genome sequencing data. *Sci. Rep.* **7**, 17303.
- Lee, K.T., Lee, Y.M., Alam, M., Choi, B.H., Park, M.R., Kim, K.S., Kim, T.H., and Kim, J.J. (2012a). A whole genome association study on meat quality traits using high density SNP chips in a cross between Korean native pig and Landrace. *Asian-Australas. J. Anim. Sci.* **25**, 1529-1539.
- Lee, T.H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162.
- Lee, Y.M., Alam, M., Choi, B.H., Kim, K.S., and Kim, J.J. (2012b). A whole genome association study to detect single nucleotide polymorphisms for blood components (immunity) in a cross between Korean native pig and Yorkshire. *Asian-Australas. J. Anim. Sci.* **25**, 1674-1680.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, <https://arxiv.org/abs/1303.3997v2>
- Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., Li, D., Yeung, C.K.L., Che, T., Jin, L., et al. (2017). Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res.* **27**, 865-874.
- Naval-Sanchez, M., Nguyen, Q., McWilliam, S., Porto-Neto, L.R., Tellam, R., Vuocolo, T., Reverter, A., Perez-Enciso, M., Brauning, R., Clarke, S., et al. (2018). Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat. Commun.* **9**, 859.
- Nguyen, D.T., Lee, K., Choi, H., Choi, M.K., Le, M.T., Song, N., Kim, J.H.,

- Seo, H.G., Oh, J.W., Lee, K., et al. (2012). The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. *BMC Genomics* **13**, 584.
- O'Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733-D745.
- Paszek, A.A., Wilkie, P.J., Flickinger, G.H., Miller, L.M., Louis, C.F., Rohrer, G.A., Alexander, L.J., Beattie, C.W., and Schook, L.B. (2001). Interval mapping of carcass and meat quality traits in a divergent swine cross. *Anim. Biotechnol.* **12**, 155-165.
- Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N., Paten, B., Zweig, A.S., Karolchik, D., et al. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003-1005.
- Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83-W89.
- Sachs, D.H. and Galli, C. (2009). Genetic manipulation in pigs. *Curr. Opin. Organ Transplant.* **14**, 148-153.
- Sanchez, M.P., Iannuccelli, N., Basso, B., Bidanel, J.P., Billon, Y., Gandemer, G., Gilbert, H., Larzul, C., Legault, C., Riquet, J., et al. (2007). Identification of QTL with effects on intramuscular fat content and fatty acid composition in a Duroc x Large White cross. *BMC Genet.* **8**, 55.
- Sang, Y., Bergkamp, J., and Blecha, F. (2014). Molecular evolution of the porcine type I interferon family: subtype-specific expression and antiviral activity. *PLoS One* **9**, e112378.
- Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308-311.
- Turnbull, C., Scott, R.H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F.B., Halai, D., Baple, E., Craig, C., Hamblin, A., et al. (2018). The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687.
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1-11.10.33.
- Vieira, F.G. and Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol. Evol.* **3**, 476-490.
- Vincent, A., Louveau, I., Gondret, F., Lebret, B., and Damon, M. (2012). Mitochondrial function, fatty acid metabolism, and immune system are relevant features of pig adipose tissue development. *Physiol. Genomics* **44**, 1116-1124.
- Wang, J., Yan, X.L., Liu, R., Fu, Q.Q., Zhou, G.H., and Zhang, W.G. (2016). Differences in calpain system, desmin degradation and water holding capacity between commercial Meishan and Duroc x Landrace x Yorkshire crossbred pork. *Anim. Sci. J.* **87**, 109-116.
- Watanabe, G., Motoyama, M., Nakajima, I., and Sasaki, K. (2018). Relationship between water-holding capacity and intramuscular fat content in Japanese commercial pork loin. *Asian-Australas. J. Anim. Sci.* **31**, 914-918.
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555-556.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* **46**, D754-D761.
- Zhu, Y., Li, W., Yang, B., Zhang, Z., Ai, H., Ren, J., and Huang, L. (2017). Signatures of selection and interspecies introgression in the genome of Chinese domestic pigs. *Genome Biol. Evol.* **9**, 2592-2603.