



OPEN ACCESS

EDITED BY

Li Ma,
University of Maryland, College Park,
United States

REVIEWED BY

Guiguiggba-Kossigan Dayo,
Centre International de Recherche-
Développement sur l'Élevage en Zone
Subhumide, Burkina Faso
Kefei Chen,
Curtin University, Australia

*CORRESPONDENCE

Mirte Bosse,
✉ mirte.bosse@wur.nl

SPECIALTY SECTION

This article was submitted to Livestock
Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 15 October 2022

ACCEPTED 19 December 2022

PUBLISHED 13 January 2023

CITATION

Peng Y, Derks MFL, Groenen MAM, Zhao Y
and Bosse M (2023), Distinct traces of
mixed ancestry in western commercial pig
genomes following gene flow from
Chinese indigenous breeds.
Front. Genet. 13:1070783.
doi: 10.3389/fgene.2022.1070783

COPYRIGHT

© 2023 Peng, Derks, Groenen, Zhao and
Bosse. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Distinct traces of mixed ancestry in western commercial pig genomes following gene flow from Chinese indigenous breeds

Yebo Peng¹, Martijn FL Derks^{2,3}, Martien AM Groenen²,
Yiqiang Zhao¹ and Mirte Bosse^{2,4*}

¹State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China, ²Animal Breeding and Genomics, Wageningen University & Research, Wageningen, Netherlands, ³Topigs Norsvin Research Center, Beuningen, Netherlands, ⁴Amsterdam Institute of Life and Environment (A-Life), VU University Amsterdam, Amsterdam, Netherlands

Studying gene flow between different livestock breeds will benefit the discovery of genes related to production traits and provide insight into human historical breeding. Chinese pigs have played an indispensable role in the breeding of Western commercial pigs. However, the differences in the timing and volume of the contribution of pigs from different Chinese regions to Western pigs are not yet apparent. In this paper, we combine the whole-genome sequencing data of 592 pigs from different studies and illustrate patterns of gene flow from Chinese pigs into Western commercial pigs. We describe introgression patterns from four distinct Chinese indigenous groups into five Western commercial groups. There were considerable differences in the number and length of the putative introgressed segments from Chinese pig groups that contributed to Western commercial pig breeds. The contribution of pigs from different Chinese geographical locations to a given western commercial breed varied more than that from a specific Chinese pig group to different Western commercial breeds, implying admixture within Europe after introgression. Within different Western commercial lines from the same breed, the introgression patterns from a given Chinese pig group seemed highly conserved, suggesting that introgression of Chinese pigs into Western commercial pig breeds mainly occurred at an early stage of breed formation. Finally, based on analyses of introgression signals, allele frequencies, and selection footprints, we identified a ~2.65 Mb Chinese-derived haplotype under selection in Duroc pigs (CHR14: 95.68–98.33 Mb). Functional and phenotypic studies demonstrate that this *PRKG1* haplotype is related to backfat and loin depth in Duroc pigs. Overall, we demonstrate that the introgression history of domestic pigs is complex and that Western commercial pigs contain distinct traces of mixed ancestry, likely derived from various Chinese pig breeds.

KEYWORDS

introgression, hybridization, selection, commercial pigs, gene flow

1 Introduction

Introgression and hybridization played a distinct role in the evolutionary diversification of plants and animals (Dowling and Secor, 1997; Mallet, 2005; Arnold et al., 2008; Stukenbrock, 2016; Grant and Grant, 2019). Genetic material introgressed from sister lineages has often been adaptive in plant and animal evolution (Dowling et al., 2016; Burgarella et al., 2019; Janzen et al., 2019; Cao et al., 2021). In wild animals and plants, adaptive introgression played an essential

role in disease resistance and environmental adaptation. Examples entail introgression in *P. trichocarpa* (Suarez-Gonzalez et al., 2016; Suarez-Gonzalez et al., 2018), *Zea mays* (Hufford et al., 2013), and sheep (Cao et al., 2021). Sometimes morphological characteristics changed, for example, wing patterning in *Heliconius* butterflies (Pardo-Diaz et al., 2012; Enciso-Romero et al., 2017). In modern humans, a variant of the *EPAS1* gene was introduced from Denisovans into Tibetans, which has proven beneficial to the adaptation of Tibetans to high altitudes (Huerta-Sanchez et al., 2014; Zhang W. et al., 2020). However, introgressed haplotypes can also have adverse effects. Examples are risk factors for type 2 diabetes, lupus, biliary cirrhosis (Sankararaman et al., 2014), and even COVID-19 inherited from Neanderthals (Zeberg and Pääbo, 2020).

Human activities have impacted over 75% of the global land area over the past ten thousand years (Venter et al., 2016; Bullock et al., 2018). Domestication and dispersal of pets, plants, and livestock have strongly altered the worldwide distribution of flora and fauna (Wichmann et al., 2009; Ottoni et al., 2013; Koch et al., 2015; Bullock et al., 2018). During the first industrial revolution, humans deliberately promoted crossbreeding of local animal and plant breeds to accelerate the process of breeding. Human-mediated hybridization between different breeds has been an important factor in shaping domestic plants and animals' genomic and phenotypic diversity (Larson and Burger, 2013; Meng et al., 2018). The hybridization from bovine ancestors improved Mongolian yak management and breeding (Medugorac et al., 2017). Likewise, haplotypes introgressed from Holstein and Brown Swiss affect protein and fat content of milk, calving traits, body conformation, feed efficiency, carcass, and fertility traits (Zhang et al., 2018).

Pigs have a long history of admixture. In the genus *Sus*, post-divergence interspecific admixture occurred before the domestication of *Sus scrofa* (Frantz L. A. F. et al., 2013; Frantz et al., 2014; Frantz et al., 2016; Liu et al., 2019). For *Sus. scrofa*, *Sus. cebifrons*, and *Sus. verrucosus*, around 23% of their genomes have been affected by admixture during the later Pleistocene climatic transition (Frantz et al., 2014). Gene flow also happened extensively between domesticated pigs to their wild ancestors during the domestication process (Giuffra et al., 2000; Frantz A. C. et al., 2013; Zhu et al., 2017). Hybridization between China and Western animals may date back to the 1st–fourth century AD (Wang et al., 2011). Historical records report that Chinese pigs were repeatedly introduced into Europe to improve the local pig breeds from the 18th century (Giuffra et al., 2000; Wang et al., 2011; White, 2011), followed by introduction into America from the 19th century onwards (Wang et al., 2011; White, 2011). Vice versa, Western commercial pigs were introduced into China since the start of the 20th century (Wang et al., 2011; White, 2011). The complex hybrid history between China and Western pigs has shaped the present genomic landscape in pigs.

There are 118 native pig breeds in China (Megens et al., 2007) with diverse phenotypic characteristics. Characteristic for Eastern Chinese pigs is early sexual maturity, higher ovulation number, and higher litters size (>15 for some breeds) (Wang et al., 2011). South Chinese pigs have inferior reproductive performance (8–10 piglets per parity for Luchan pigs), thinner skin, and excellent heat resistance (Wang et al., 2011; Chen et al., 2020).

In recent years, genomic studies have revealed some Chinese haplotypes in Western pig breeds that were likely introgressed and selected for. *AHR* is a toxicity- and fertility-related gene (Denison et al., 2011; Onteru et al., 2012). It is introgressed from a Chinese breed

into Dutch Large White pigs (Bosse et al., 2014). Based on the Illumina Porcine 60 K SNP Beadchip dataset of Erhualian, White Duroc × Erhualian F2 population, Duroc and Landrace pigs. Yang et al. found a mutation in *VRTN* that increased vertebra number, carcass length, and teat number in Western pigs and was inherited from Chinese Erhualian pigs (Yang et al., 2016). The meat quality-related genes (*SALI*, *ME1*) and fertility-related genes (*GNRHR*, *GNRH1*), are reported as being introgressed into Duroc from Meishan pigs by Zhao et al., using whole-genome re-sequencing data of 32 Chinese Meishan and 31 Duroc pigs (Zhao et al., 2018). Recently, Chen et al. also explored whole-genome sequencing data from 266 Eurasian wild boars and domestic pigs. They found that the *GOLM1-NAA35*, a gene that is responsible for cytokine interleukin 6 (IL-6) production in human immune cells (Li et al., 2016), is inherited from south Chinese pigs (SCN) in French Large White (LWHFR) (Chen et al., 2020). They also found a haplotype spanning *KATNAL1* that originated from east Chinese pigs (ECN) pigs and has been selected to increase the fertility in LWHFR pigs. Although only LWHFR and two Chinese native pig groups were included, their study provided the novel perspective that introgression from Chinese pigs to commercial breeds may vary considerably. Therefore, in the current study we extensively explore source of introgression and genomic regions that contained introgressed segments on a large scale, including multiple Western pig breeds and a broad sampling of Asian breeds.

Thus, many genomic segments from local Chinese pigs that contributed to favorable characteristics of Western commercial breeds have been identified (Bosse et al., 2014; Frantz et al., 2015; Chen et al., 2017; Chen et al., 2020; Wang et al., 2020), but these records are sporadic, and no systematic survey has been conducted. How extensive these episodes of introgression and improvement of Western domesticated pigs with animals from Asia have been, and where in China these pigs originated, are still unanswered questions. Although Chinese pigs are highly polymorphic (Amaral et al., 2008; Frantz A. C. et al., 2013; Zhao et al., 2019), they form a close genetic group, and there has been an extensive genetic exchange between (local) breeds (Huang et al., 2020). Disentangling the sources of the introgressed haplotypes will shed new light on historical breeding practices, help understand the molecular mechanisms underlying phenotype change, and be of great significance to future breeding.

Even though the overall level of introgression from Chinese pigs to Western commercial breeds seems relatively stable across breeds, the underlying haplotypes, genomic loci, and breed origins may vary (Bosse et al., 2014). In this paper, we present a comprehensive study of the gene flow of pigs from different Chinese origins into five distinct Western commercial lines and illustrate the difference of global haplotype introgression patterns between donor-recipient combinations.

2 Materials and methods

2.1 SNP calling, phasing, and imputation

The datasets analyzed during the current study are available from the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>) under project PRJEB1683 (Groenen et al., 2012), PRJEB29465 (Grahofer et al., 2019), PRJEB9922 (Frantz et al., 2015), PRJNA186497 (Li et al., 2013), PRJNA213179 (Ai et al., 2015), PRJNA231897, PRJNA238851 (Wang et al., 2015), PRJNA254936,

PRJNA255085 (Ramirez et al., 2015), PRJNA260763 (Choi et al., 2015), PRJNA273907, PRJNA305081, PRJNA305975, PRJNA309108 (Li et al., 2017), PRJNA314580, PRJNA320525 (Bianco et al., 2015), PRJNA320526, PRJNA320527, PRJNA322309, PRJNA369600, PRJNA378496 (Zhao et al., 2018), PRJNA398176 (Zhu et al., 2017), PRJNA438040, PRJNA488327 (Yan et al., 2018), PRJNA488960 (Zhang Y. et al., 2020), PRJNA524263 (Zhang W. et al., 2020), and PRJNA550237 (Chen et al., 2020).

A total of 730 samples were included with Asian, Western, domestic and wild backgrounds (See Table S1). Raw reads were aligned to the Sscrofa11.1 reference genome (Warr et al., 2020) using the bwa-mem algorithm (Li and Durbin, 2009). Samtools-v1.8 (Li, 2011) was used for sorting, merging, and marking potential PCR duplications. Finally, haplotype-based variant detection was conducted with freeBayes-v1.1 (--min-base-quality 10 --min-mapping-quality 20 --min-alternate-fraction 0.2 --haplotype-length 0 --pooled-continuous--ploidy 2 --min-alternate-count 2) (Garrison and Marth, 2012). After SNP calling, SNP loci were screened and retaining with a quality value greater than 20 (vcffilter -f "QUAL > 20"). Further quality control was conducted with the following criteria: minor allele frequency (MAF) > 0.01, missing rate < 0.01, call rate > 90%, sequencing depth of sample > 4. Individuals and loci satisfying the above criteria were retained for further analyses, and assigned to their specific background (193 Chinese indigenous pigs, 30 Asian wild boars, 13 Yucatan mini-pigs, 40 Western wild boars, 298 Western commercial pigs and 18 wild suidae; Supplementary Table S1). Finally, phasing and imputation were performed based on this data set with Beagle 5.1 (Browning et al., 2018; Browning et al., 2021) (window = 20 overlap = 4 gp = true ap = true).

2.2 Genetic structure analysis

T-SNE dimensionality reduction was first conducted by *sklearn.manifold.TSNE* ($n_components = 2$, $perplexity = 24$) in scikit-learn-0.23.1 python package on high-quality *Sus. Scrofa* samples. To construct the Neighbor-joining tree (NJ-tree), we calculated the IBS distance matrix by plink-1.9 on phased SNP data with default parameters. Then the NJ-tree was constructed by fastME-v2.15 (-D 1 -m N -b 10000 -T 10 -s -I) (Lefort et al., 2015) with *Sus cebifrons* as the outgroup. The tree was plotted by the iTOL-v5 online tool (Letunic and Bork, 2021). Model-based global ancestry estimation was conducted with Admixture-1.3 (-B10 -c10) (Alexander et al., 2009) with cross-validation to assess the best fitting K-value.

2.3 Local introgression detection

Western wild boars and Yucatan minipigs were combined as the Western haplotypes background for the introgression study. Chinese groups were set as the donor population for every commercial line. Then putative introgression segments were detected for every donor-recipient combination with relative Identity-by-descent (rIBD) method using whole-genome sequencing data (Bosse et al., 2014). Identity-by-descent (IBD) detection was performed with the refinedIBD algorithm (length = 0.1 trim = 0.01 lod = 1) (Browning and Browning, 2013). These parameters were adjusted to detect not only segments that are identical, but segments with similar origins

(i.e., Western or Asian) that show higher similarity than expected between Chinese and European ancestries.

The rIBD values were calculated on non-overlapping bins of 10 kb along the genome. For every bin, we calculated rIBD values with the following formula: $rIBD = nIBD_{R,D} - nIBD_{R,B}$. $nIBD_{R,D}$ denotes the normalized IBD (nIBD) value of the recipient-donor pair, $nIBD_{R,B}$ denotes the nIBD value of the recipient-background pair. $nIBD = \frac{Count_IBD}{Total_IBD}$, $Count_IBD$ = shared IBD counts between group1 and group2, $Total_IBD = N_1 * N_2$. N_1 and N_2 are the sample size of group1 and group2, respectively. That way, $rIBD > 0$ indicates that commercial breeds (recipient) shares more IBD traces with Chinese indigenous (donor) than Western background, and thus denotes introgression from the donor into the recipient population. In contrast, a negative rIBD value indicates that the number of haplotypes shared by the recipient and the background population is greater than that shared with the donor at that locus. We then performed a Z-transformation of the rIBD values with the mean and standard deviation values of overall IBD from all donor-recipient pairs. We independently selected the presumed introgression bins with a Z-rIBD threshold of $\mu + 2\sigma$ for every pair, where μ and σ are the mean and standard deviation of Z-rIBD values. Positive significant Z-rIBD values are thus indicative of the presumed introgression from Chinese breeds into the Western commercial pig.

2.4 Overlapping ratio of Z-rIBD segments

To measure the coincidence of significant positive/negative Z-rIBD fragments between different donor recipients, we calculated the overlapping ratio by the following formula:

$$Overlapping\ ratio_{C,A,B} = 1 - \left(\frac{Noneoverlapped\ Counts_{C,A} + Noneoverlapped\ Counts_{C,B}}{Total\ Counts_{C,A} + Total\ Counts_{C,B}} \right)$$

Where A, B, C denote the three populations. When we compare the overlapping level between "C - > A" and "C - > B", population C denotes one donor population while A and B denote two different recipients. To compare the overlapping level between "A - > C" and "B - > C", population C denote on recipient while A and B denote two different donors. $Noneoverlapped\ Counts_{C,A}$: the number of positive/negative fragments/bins shared between C and A but not shared with B. $Total\ Counts_{C,A}$: the total number of positive/negative fragments shared by C and A.

2.5 Selective sweep analysis

We performed a genome scan to detect recent adaptive introgression events using polymorphism data from the recipient populations only, using the VolcanoFinder-v1.0 tool (Setter et al., 2020). The ancestral genome was constructed with 16-way Enredo-Pecan-Ortheu multiple alignments files, downloaded from the Ensembl v.103 databases (<https://www.ensembl.org/>). We obtained the allele frequency and the unnormalized site frequency spectrum files required for Volcanofinder, and performed the analysis according to the standard workflow from VolcanoFinder (<https://doi.org/10.5061/dryad.7h44j0zr7>) (Setter et al., 2020). Finally, variants were polarized by the ancestor alleles status and used as the input for VolcanoFinder-v1.0 (-big 30000, -1 1 1) (Setter et al., 2020).

2.6 Selection of introgression segments for further analysis

To locate important introgressed segments, we merged the consecutive significant ZrIBD bins into one introgression segment. We ranked introgression segments by segment length as the first criterion and average rIBD value as the second. Then, we selected the segments with a length larger than 11 Kb and Log (10) likelihood ratio of selective sweep footprint >11 (the 0.95 quantile). After that, we computed the length of the introgressed segment, selective sweep footprints, average ZrIBD value and average minor allele frequency for every introgression segment and selected the segments that matched all criteria as top candidates for further analysis.

2.7 Haplotype origin tracing

To trace the origin of haplotypes, alleles were first joined into a “FASTA” format sequence from the phased VCF file by an in-house python script. For a genomic region of interest, variants belonging to the same haplotype were joined to a sequence. These haplotypes thus consist of a string of variants derived from the phased VCF. Subsequently, the SNP distance matrix between haplotypes was calculated with SNP-dists v0.7.0 (<https://github.com/tseemann/snp-dists>). Finally, hierarchical clustering was conducted in R using the gplots package. Paterson’s D-statistics (Patterson et al., 2012) were computed by Dtrios (-j100) of the Dsuite v0.4 (Malinsky et al., 2021) tool package.

2.8 Determination of Chinese-derived alleles

We refer to an allele as a “Chinese-derived allele” when it occurs in Duroc and Chinese pigs, but is nearly absent in European wild boars. So the “Chinese-derived allele” should match the following criteria: 1) allele frequency in Duroc pigs $\geq 0.1.2$. Allele frequency in any of the Chinese local pig groups $\geq 0.1.3$ allele frequency in European wild boars ≤ 0.0125 (i.e., only one European wild boar among 40 boars has that allele and is heterozygous).

2.9 Candidate variants selection and LD calculation

Chinese-derived variants were annotated by snpEff-v5.0 (Cingolani et al., 2012). To pinpoint potential causal variants with a high effect on the phenotype, the variants were then ranked using pCADD’s PHRED score. Briefly, the pCADD is the “pig combined annotation dependent depletion”, a model to score single nucleotide variants in pig genomes in terms of their putative deleteriousness, or effect on phenotypes, based on a combination of annotations, see (Gross et al., 2020). The pCADD model is a pig-specific variant of the original CADD model that was developed for human aims to discriminate neutral variants from variants with high impact. Then, “Candidate variants” were selected by the following principles: 1) PHRED score >4.3 (the whole genome mean value). 2) missense variant, 3’UTR variant, or 5’UTR variant.

The LD level of the proxy SNPs with other variants from the sequence data was calculated by plink v1.90b6 (--ld-snp new14_97387849 --ld-

window 3000 --ld-window-kb 3000 --r2 --ld-window-r2 0) in the Duroc population. The mean r2 values between proxy SNP and other variants in every block were used as the LD level of the proxy SNP and that block.

2.10 SNP selection from the illumina 50 K SNP array dataset

To be able to test phenotypic effects of the Chinese-derived introgressed haplotypes on chromosome 14, we wanted to expand our sample size by incorporating genotype data from commercial Duroc pigs. The genotype data was obtained from routinely screened pigs from Topigs Norsvin pigs that were genotyped by the (Illumina) Geneseek custom 50 K SNP chip with 50,689 SNPs (50 K) (Lincoln, NE, USA). The chromosomal positions are based on the *Sscrofa11.1* reference assembly. In our set of re-sequenced Duroc pigs, SNPs were filtered using the following requirements: Each marker had a MAF greater than 0.01, a call rate greater than 0.85, and an animal call rate >0.7. SNPs with a *p*-value below 1×10^{-5} for the Hardy-Weinberg equilibrium exact test were also discarded. All pre-processing steps were performed using plink v1.90b3.30 (Chang et al., 2015). SNPs on chromosome 14 were retained for further LD analyses to identify the SNP in highest LD with the candidate variants on the introgressed haplotypes. We tested LD between the candidate SNPs from the sequence in the Asian derived haplotype and SNPs on the 50 K chip by using Plink-1.9 (--ld-snp new14_97387849 --ld-window 3000 --ld-window-kb 3000 --r2 --ld-window-r2 0).

2.11 Phenotype-genotype association

To estimate the impact of genotypes on production traits, we used the genotype data for the candidate SNP from 11,255 Duroc animals (not all animals have all phenotypes) to test the association of our introgressed allele with the following traits: daily gain from birth to Tstart (25 kg) for 9,921 animals, daily gain from Tstart to the Tend (25–120 kg) in 10,986 animals, backfat at 120 kg (Tend) in 7,192 animals, lean meat percentage and loin depth at the end (120 kg) in 7,688 and 7,192 animals respectively. The corrected phenotypes for all traits of each animal were obtained from the routine genetic evaluation by Topigs Norsvin. Then, for each trait, we conducted a Welch’s *t*-test (significance threshold $p < 0.05$) to test for differences in phenotypes of the different genotypes at our candidate SNP, that were assigned to either European or Asian background.

3 Results

3.1 Data collection

We collected 730 samples (Supplementary Table S1) from NCBI (<https://www.ncbi.nlm.nih.gov/>) and conducted SNP Calling with freeBayes-v1.1 (Garrison and Marth, 2012). After strict quality control, 19, 656, 271 SNPs and 592 samples were retained for further analyses, including 193 Chinese indigenous pigs, 30 Asian wild boars, 13 Western local pigs (Yucatan mini-pigs), 40 Western wild boars, 298 Western commercial pigs, and 18 samples from suidae in Southeast Asian islands (Supplementary Table S1).

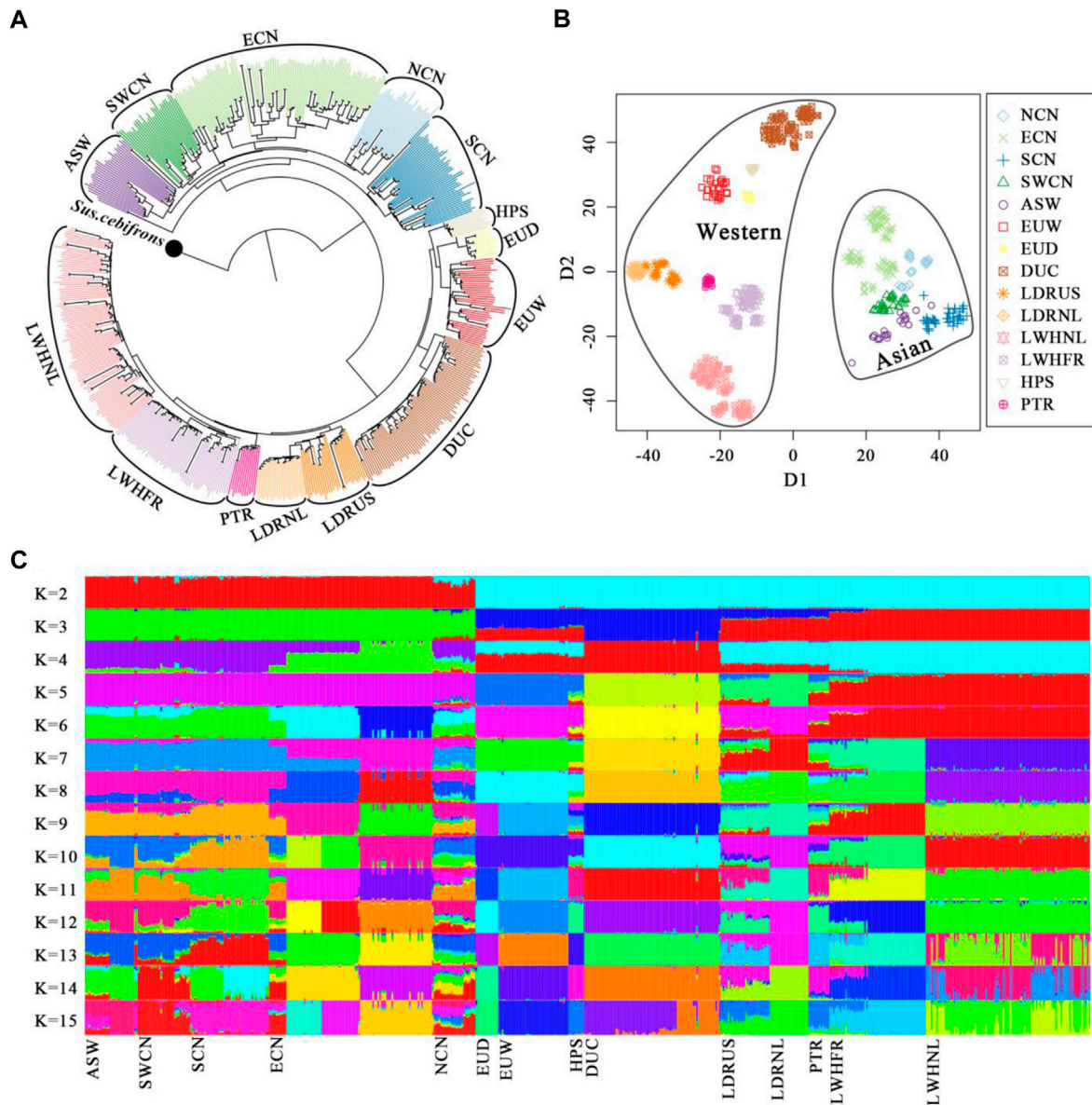


FIGURE 1
 Genetic structure of pigs in this study. **(A)** The Neighbor-joining tree was constructed by fastME-v2.1.5 based on the IBS-distance matrix and set *Sus. cebifrons* from Southeast Asian islands as the outgroup. **(B)** Dimensionality reduction of whole-genome SNPs with the t-SNE algorithm. **(C)** Global ancestry inference of Chinese and Western pigs conducted with ADMIXTURE-v1.3.0. NCN, North Chinese pigs; ECN, East Chinese pigs; SCN, South Chinese pigs; SWCN, Southwest Chinese pigs; ASW, Asian Wild boars; EUW, European Wild boars; EUD, European local pigs; DUC, Duroc pigs; LDRUS, American Landrace pigs; LDRNL, Dutch Landrace pigs; LWHNL, Dutch Large White pigs; LWHFR, French Large White pigs; HPS, Hampshire pigs; PTR, Pietrain pigs.

3.2 Genetic structure analysis

We predefined the groups of Chinese pigs according to our previous analysis (Peng et al., 2022) and their geographical origins (Wang et al., 2011) (Supplementary Figure S1). The NJ-tree, t-SNE dimensionality and global ancestry analysis were used to dissect the genetic structure of our samples. The NJ-tree and ancestry inference separate Western and Chinese-derived pigs (Figures 1A–C). Chinese animals clustered into a monophyletic clade, and different Chinese origins clustered into sub-clades except for Chinese Northern pigs (Figure 1A). For Western pigs, every commercial line clustered into a monophyletic clade (Figure 1A) and breeds were clearly distinguished

in the t-SNE plot (Figure 1B). Admixture analysis was consistent with this pattern, with increasing values of K above six indicating local ancestry for European pigs, and Asian substructure was best captured with K = 14, when the cross-validation reached a plateau (Figure 1C and Supplementary Figure S2). After we removed a few of the eastern Chinese samples that showed hybrid ancestral components in the admixture result, we assigned the China and Western pig breeds to specific clusters according to geographical sources and genetic relationships.

Finally, Duroc (DUC), Dutch Large White (LWHNL), French Large White (LWHFR), Dutch Landrace (LDRNL), and American Landrace (LDRUS) were recognized as five distinct Western

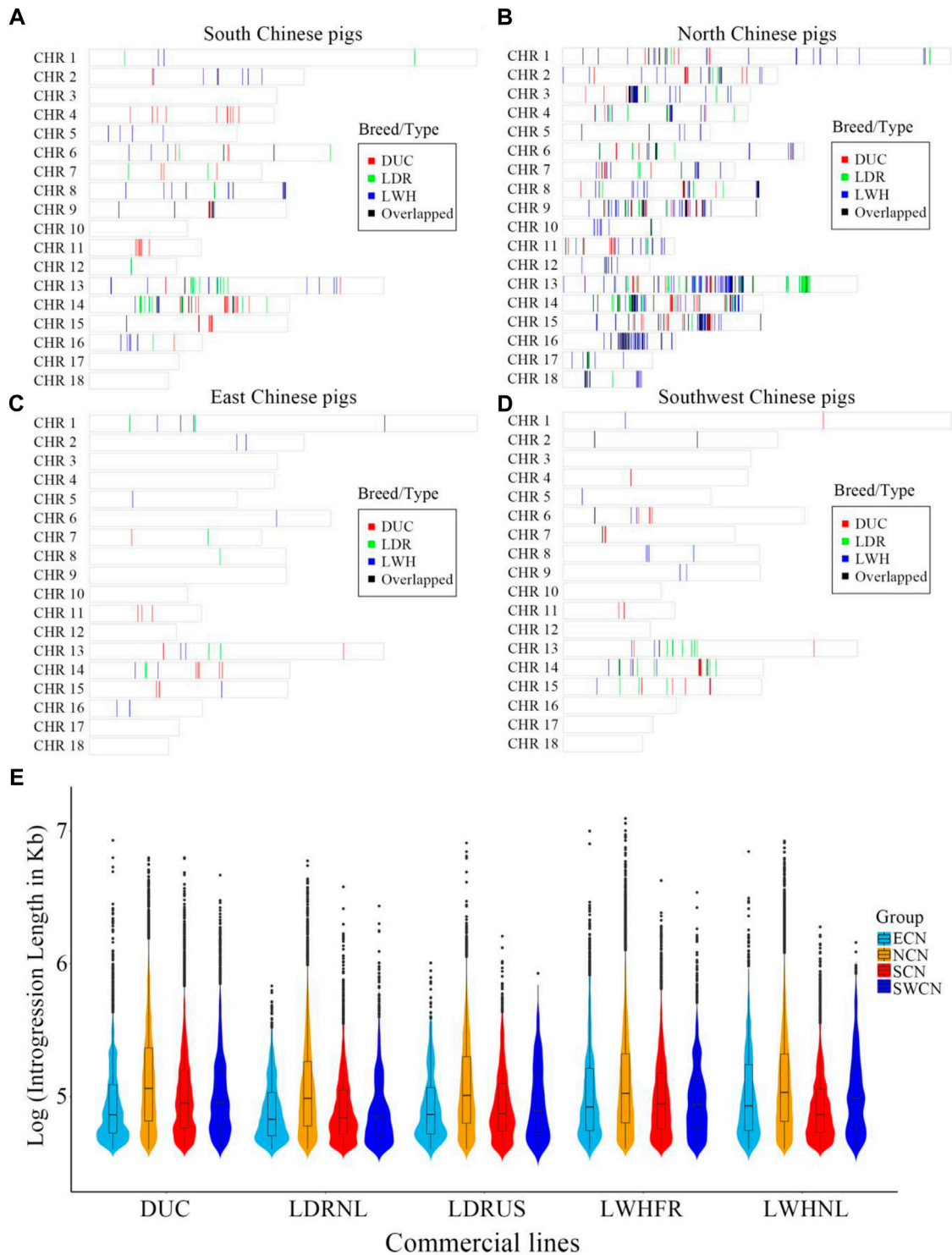
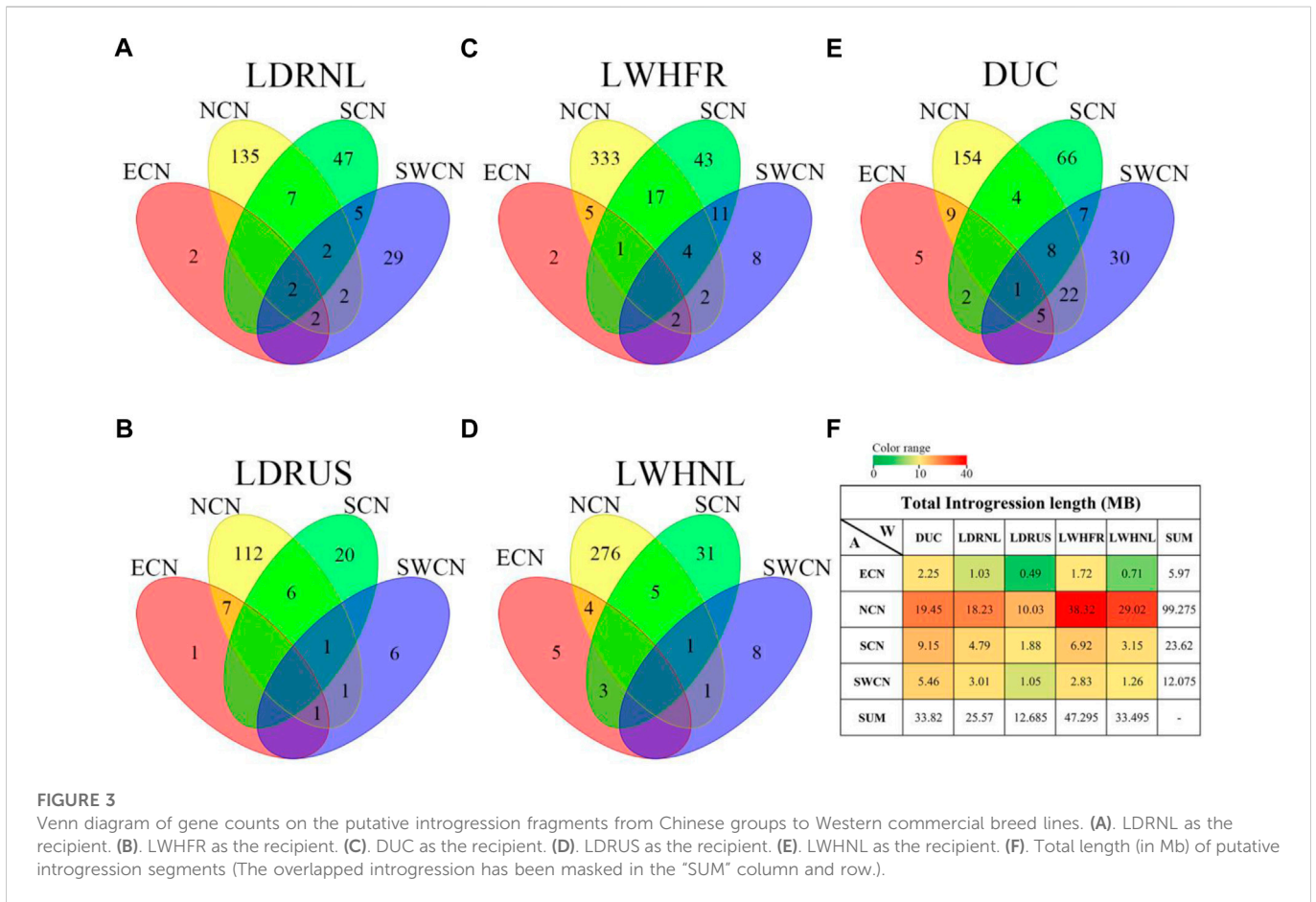


FIGURE 2

The distribution of genomic regions with introgression signature from (A) South Chinese pigs, (B) North Chinese pigs, (C) East Chinese pigs, and (D) Southwest Chinese pigs to different Western commercial breeds. DUC: Duroc, LDR: American and Dutch Landrace pigs, LWH: French and Dutch Large White pigs, Overlapped: the overlapped introgressed region between any two pairs. (E). The features of natural logarithms transformed introgressed fragment lengths (in Kb) from China to Western pigs. ECN, East Chinese pigs; NCN, North Chinese pigs; SCN, South Chinese pigs; SWCN, Southwest Chinese pigs. DUC, Duroc; LDRUS, American Landrace pigs; LDRNL, Dutch Landrace pigs; LWHFR, French Large White pigs; LWHNL, Dutch Large White pigs.

commercial lines (WS). European wild boars (EUW) plus local European pigs (EUD) were defined as the Western background population (WB). Moreover, Southern (SCN), Eastern (ECN), Northern (NCN), and Southwestern (SWCN) Chinese pigs were

defined as the four Chinese local groups (AB). We combined Chinese wild boars (CNW), Korean wild boars (KRW), and Thai wild boars (THW) into the Chinese background population (AB) (Supplementary Table S1).



3.3 Introgression landscape from Chinese to western pigs

We assessed local signatures of introgression in the Western pig genomes using an IBD haplotype sharing method. There are large introgressed fragments and introgression clusters from China to Western pigs (Figures 2A–D). On a genome-wide scale, the proportion and local regions of putative introgression are highly diverse between different donor-recipient pairs. The highest proportion of introgression into Western commercial breeds is NCN, followed by SCN (Figures 2A–D). The amount of introgression varied between the European breeds, with most putative introgression segments from Chinese pigs found in Large White breeds and the French Large White line in particular (Total length of Chinese-derived segments is 33.82 Mb for DUC, 25.57 Mb for LDRNL, 12.68 Mb for LDRUS, 47.29 Mb for LWHFR, 33.49 Mb for LWHNL. Figures 2A–D and Figure 3F). The putative introgression fragments also varied in length and number (Figures 2E, 3E). The longest introgression fragments reach ~1.2 Mb between LWHFR and NCN (Total length: ~38 Mb), but only ~0.34 Mb between LDRNL and ECN (Total length: ~1 Mb). For any Western commercial line, the average introgressed segment length from NCN is longer than from other Chinese populations (Figure 2E), suggesting a relatively recent genetic exchange between NCN and Western pigs.

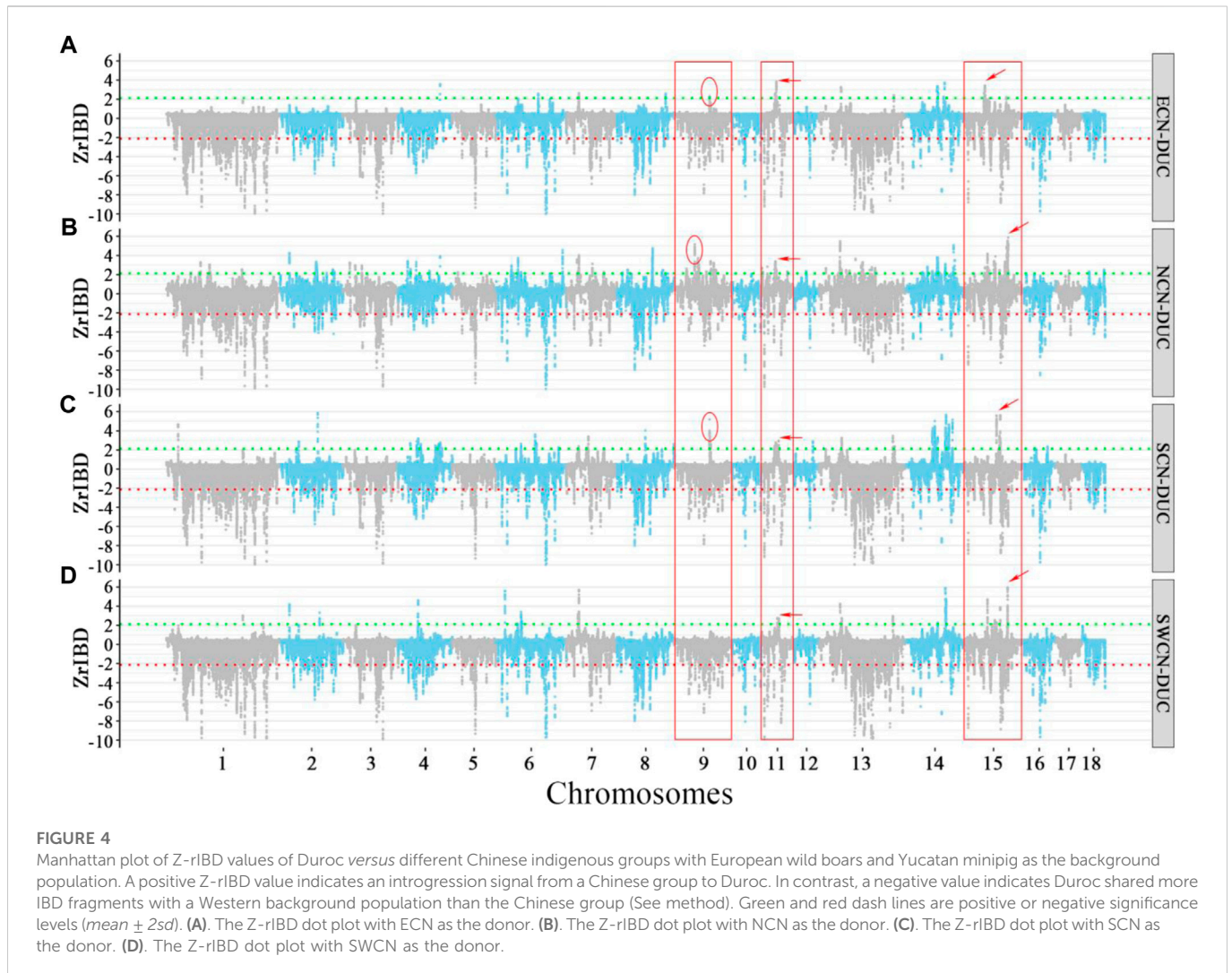
We studied the number of genes affected by the introgression fragments for different donor-recipient pairs. Results are consistent with the total introgression length (Figure 3F). Most of the genes affected by introgression from local Chinese pigs into Western

commercial pigs are specific for every donor-recipient pair (Figures 3A–E). Furthermore, most introgressed genes are from NCN to Large White (LWH), especially LWHFR (428 genes, Figure 3C).

3.4 Various segments from different Chinese groups are introgressed into specific western breeds

We further compared the degree of overlap of positive/negative Z-rIBD segments among donor-recipient combinations to study the global introgression differences. We found that the overall positive Z-rIBD (introgression footprint, see method) patterns are less similar than negative Z-rIBD patterns in different donor-recipient pairs (Figure 4, And Supplementary Figure S3). For donor groups from different sources in China, the degree of overlap significant positive Z-rIBD segments between donor-recipient pairs range from 3.3% to 20.56% (Supplementary Table S2), but for negative Z-rIBD segments, it is 69.73%–93.51% (Supplementary Table S2). The degree of overlap of positive Z-rIBD segments is much lower than the negative Z-rIBD segments (Supplementary Table S2), suggesting specific introgression.

Here we take Duroc as an example. The difference in the positive peaks of Z-rIBD is noticeable (Figure 4). There are peaks at different locations or heights on chromosome nine for ECN, NCN, and SCN (Figures 4A–C), but there is no significant Z-rIBD peak for the SWCN-DUC pair (Figure 4D). On chromosome 11, there are peaks located at 34–39 Mb with different heights or widths (Figures 4A–D). Likewise, on



chromosome 15, the highest peak is located at different positions for the Chinese groups (Figures 4A–D) except for NCN and SWCN (Figures 4B–D). Suggesting that pigs from different regions in China contributed differently to Western commercial pig breeds.

3.5 Hybridization occurred in the early breeding process of commercial pigs

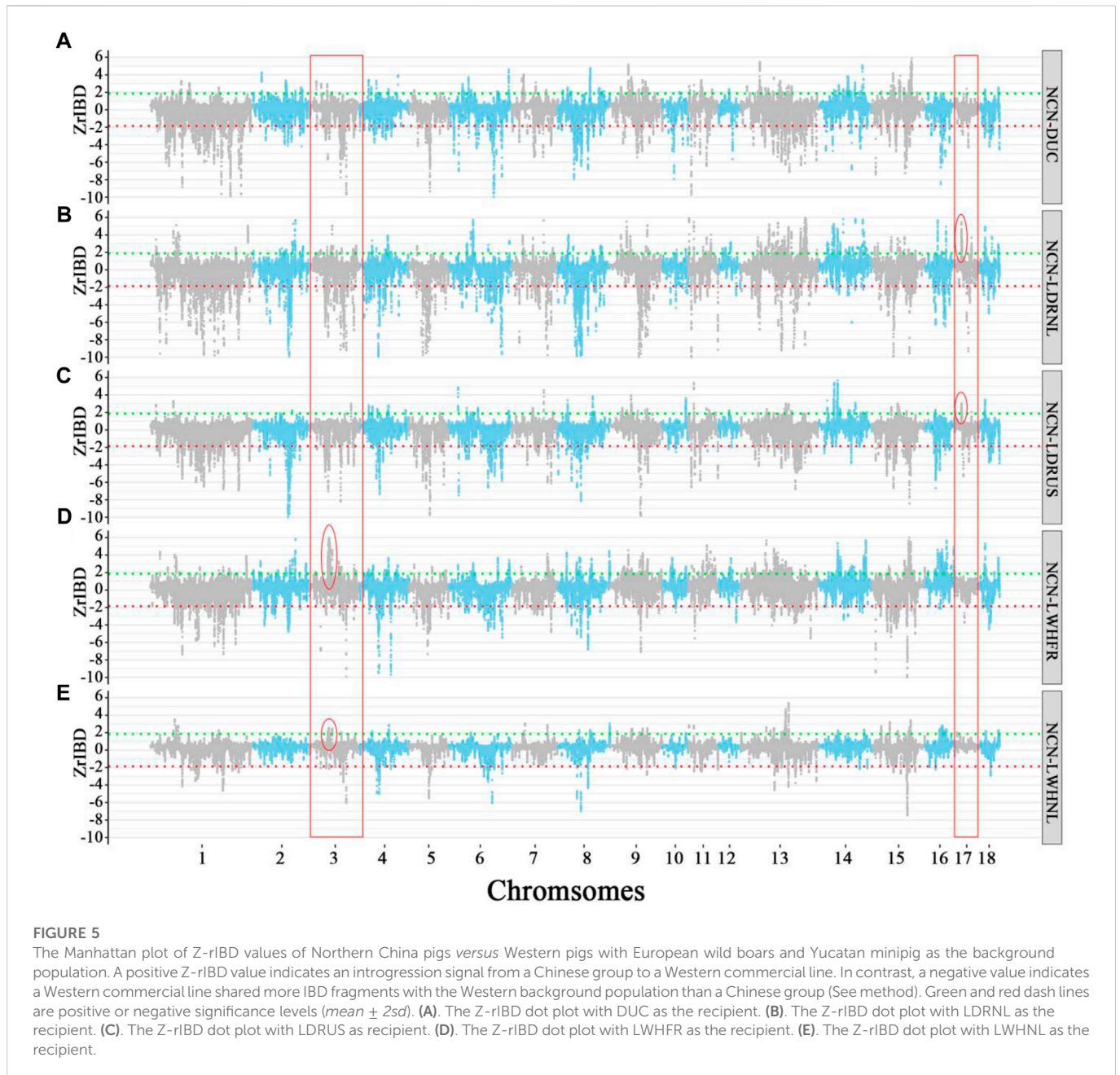
There is a large difference in the introgression patterns between specific Chinese groups and Western commercial lines (Figure 5, And Supplementary Figure S4). The degree of overlap of positive Z-rIBD segments for the specific Chinese local pigs to different European commercial pigs ranges from 0% to 34.83% (Supplementary Table S3). In contrast, for the negative Z-rIBD segments, it ranges from 27.48% to 49.81% (Supplementary Table S3). However, the differences in the introgression patterns within related breeds from a specific Chinese group are smaller. Dutch and French Large White breeds show a similar introgression pattern from North Chinese pigs compared with other commercial pigs (Figures 5D,E, and Supplementary Table S3). The degree of overlap of positive Z-rIBD for these breeds is as high as 34.83% (Supplementary Table S3). In contrast, this is only around 10% compared to the other Western commercial lines

(Supplementary Table S3). A broad peak on chromosome 3 (chr3:48–52 Mb) was found in Dutch Large White and French Large White (Figures 5D,E), but not in the other breeds (Figures 5A–C).

In the Landrace breed, the degree of overlap of positive Z-rIBD is as high as 26.55% (Figures 5B,C, and Supplementary Table S3). A significant introgression signal on chromosome 17 (CHR17: 17–18 Mb) is observed in both Dutch and American Landrace (Figures 5B,C) but not seen in the other breeds (Figures 5A,D,E). Besides, the Z-rIBD peaks mentioned above are different in the different pig lines. The observed difference in introgression signal from specific Chinese groups to related Western commercial lines reflects a difference in the extent of introgression. This suggests that gene flow occurred mainly in the early stages of commercial pig breeding rather than after the differentiation of the lines. However, the tendency of artificial selection caused changes in signal strength.

3.6 A Chinese-derived haplotype introgressed into duroc genomes

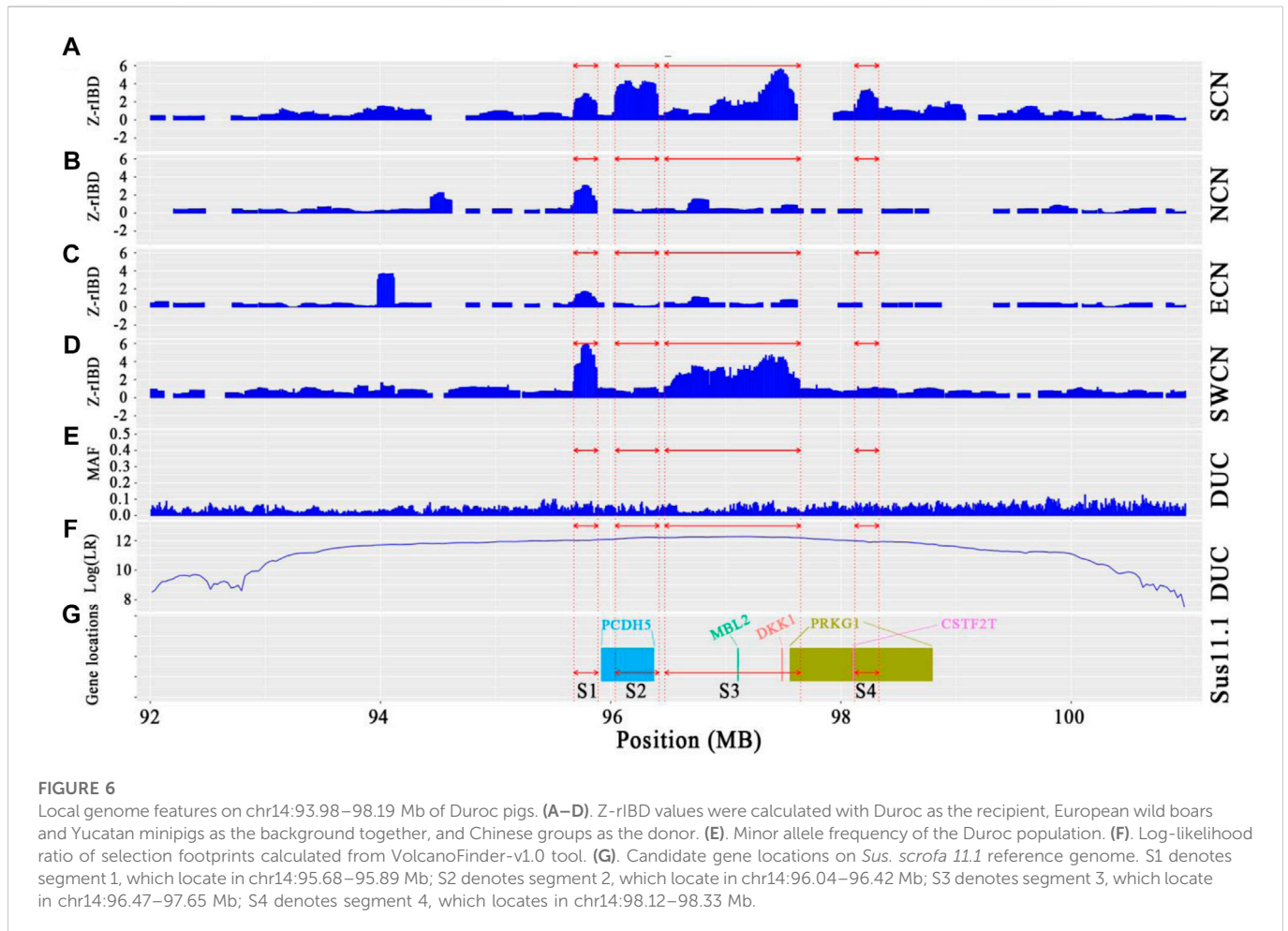
We observed a cluster of Duroc-specific introgression signatures spanning ~2.65 Mb on chromosome 14 (chr14: 95.68–98.33 Mb)



(Figure 6). Such a strong introgression and selection signal is not seen for the other commercial pigs at that region (Supplementary Figures S5, S6). This introgressed region appears to be a set of segments derived from Chinese pigs in the Duroc population. The Z-rIBD value for SCN-DUC is up to 5.65 for segment 3 (the mean Z-rIBD value is 2.48 for segment 1, 3.90 for segment 2, 2.19 for segment 3 and 2.77 for segment 4, Figure 6A). Except for SWCN-DUC in segment 1 (mean Z-rIBD = 5.89, Figure 6D) and segment 3 (mean Z-rIBD = 4.74, Figure 6D), the mean Z-rIBD values is highest in the SCN-DUC pair (Figures 6A–D). We also observed a lower minor allele frequency than expected by chance (0.04 for this region but 0.12 for whole-genome) in Duroc (Figure 6E). These signatures are located within a strong adaptive selection region (chr14: 92–101 Mb) on the Duroc genome (Figure 6F). Moreover, there are five candidate genes in this region: *PCDH15*, *MBL2*, *DKK1*, *PRKG1*, and *CSTF2T* (Figure 6G).

Additionally, PCA plots of Duroc and Chinese pigs from SNPs across the full genome and local SNPs in this region display a strong discordancy (Supplementary Figure S7). The clustering of the Duroc and Chinese pigs in this region hint at introgression, evident from the big difference between the global and local PCA analyses. Combining the above results, we suspect that this haplotype in the Duroc genome was inherited from SCN or SWCN pigs.

To trace the sources of the haplotype, we then calculated a distance matrix between individuals for every segment by SNP-dists v0.7.0 followed by hierarchical clustering in R-4.0 using the gplots package (Warnes, 2020). The results (Figure 7) show that most Duroc pigs clustered together with Chinese pigs (especially with ECN, SCN, and SWCN), in sharp contrast to LWH and Landrace (LDR). The LWH and LDR clustered with Western background populations on segment 1 and segment 4 (Figures 7A,D). In the clustering of segments



2 and 4, more SCN pigs are located within Duroc clades (Figures 7B,D), suggesting that fragment 2 is more likely derived from SCN pigs. The results of the ABBA-BABA test (D-statistics) highlights that Duroc shares more derived alleles with SCN than other Chinese pigs for segment 2, segment 3, and segment 4 (Table 1). Moreover, a high degree of linkage disequilibrium (LD) in this region ($r^2 = 0.56$, Supplementary Figure S8) was found. According to the dist trees, D-statistics and the degree of LD, we believe that the haplotype (chr14:95.68–98.33 Mb) in the Duroc genome is derived from SCN pigs.

3.7 Prioritizing causal variants within introgressed haplotypes

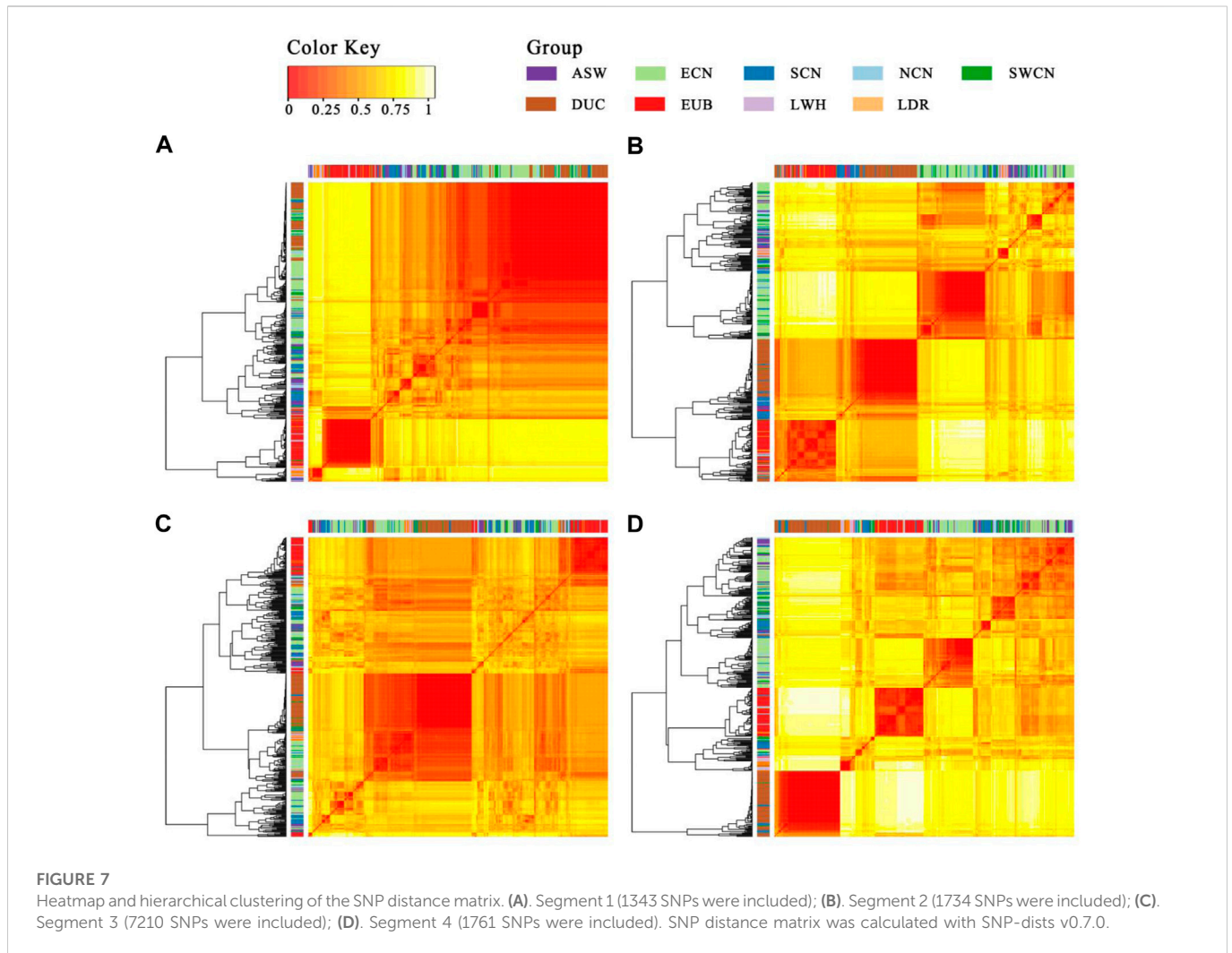
We further investigated SNP allele frequencies in the four segments in different pig populations (Table 2, And Supplementary Table S4). There are many alleles with low (≤ 0.0125) frequencies in Western wild boars but high frequencies in Duroc pigs for each of the segments (123 variants in segment 1, 118 variants in segment 2, 436 variants in segment 3, 383 variants in segment 4, Supplementary Table S4). Furthermore, the derived alleles in Duroc pigs at these loci seem to have undergone strong selection (Figure 6F). We believe these are candidate alleles derived from Chinese pigs due to their moderate allele frequencies in Chinese pigs (Table 2, And

Supplementary Table S4). Seven candidate mutations (Supplementary Table S5) were selected from the putative Chinese-derived set of alleles that potentially have a high functional impact (see methods). These variants are likely to have a strong impact on the phenotype as derived from the pCADD model, with the strongest located within the three prime UTR region of *PRKG1*.

3.8 Association of *PRKG1*-haplotype with production traits

We analyzed genotype and phenotype data of 11,255 animals from a commercial Duroc population to assess the potential phenotypic impact of the introgressed haplotypes. We screened the (Illumina) Geneseek custom 50 K SNP array for SNPs in highest LD with the introgressed haplotypes, and a SNP (INRA0045978) was selected as a proxy for the introgressed segment due to its high LD (r^2 range from 0.65 to 0.73) with the seven candidate alleles in the Duroc population (Supplementary Table S6, Supplementary Figure S9). Next, we used the genotypes for this selected SNP from 11,255 Duroc animals from the same commercial breed to test the association of INRA0045978 with a set of production traits (See methods).

We found a significant association with backfat (genotype “0/0” versus “1/1; *t*-test *p*-value 0.016; Figure 8C) with a and with loin



depth (genotype “0/1” versus “1/1; *t*-test *p*-value 0.028; Figure 8E and Supplementary Table S7). The INRA0045978 SNP has a low Duroc reference allele frequency in Western wild boar (0.0125) but higher in Chinese pigs (0.5517) and Duroc (0.8782). These results suggest that the *PRKG1*-haplotype may decrease backfat (mean difference of 2.3 mm) and increase loin depth (mean difference of 6.1 mm) in Duroc pigs.

4 Discussion

We conducted a comprehensive analysis of the introgression from China to Western commercial pigs. The complexity of the commercial pig breeding process caused unforeseen scenarios. Our findings reveal the distribution and quantity of Chinese pig genetic components in major Western commercial pig breeds.

Interestingly, we found that the overall positive introgression patterns across breeds are less similar than negative patterns. The high degree of overlap for negative Z-rIBD segments was caused by the close genetic relationship among local Chinese pigs. The lower degree of overlap for the positive Z-rIBD segments indicates specific contributions from different Chinese local pigs into Western pigs. This could indicate that some genomic regions in

Western pigs do not allow introgression from such distantly related pig populations and that purifying selection is at play. By contrast, breed-specific traits requirements could promote introgression reserved at specific loci, wherein other breeds, these Chinese-derived haplotypes, are undesired. Therefore, we hypothesize that genomic regions lacking Chinese introgression in all Western pigs contain genes that contribute to traits shared across all Western pigs and identify this as an exciting avenue for future research.

Introgressed sequences from different Chinese pig groups were found for a given Western breed. This may have been influenced by the opening of foreign trade ports in China hundreds of years ago and by the traits of pigs in different places (Chen et al., 2020). Western commercial breeds have retained different proportions and different specific loci of introgression. We believe different Chinese pig breeds were introduced for crossbreeding before current Western breeds were established. After establishing Western commercial breeds, these breeds were selected in different directions. We show introgression signals at the same genomic positions but with different introgression intensities for different lines from the same breed. This suggests the influence of directional selection on the gene flow. These results show that the variation in

TABLE 1 D-statistics result of four segments.

Segment	P1	P2	P3	D-statistic	Z-score	p-value
S1 (14:95.68–95.89 Mb)	EAS	DUC	EUW	0.322572	3.0973	0.001
	DUC	NCN	EUW	0.189092	2.7052	0.0034
	SCN	DUC	EUW	0.160682	1.1969	0.1157
	SWCN	DUC	EUW	0.158532	2.2859	0.0111
S2 (14:96.04–96.42 Mb)	EUW	DUC	ECN	0.132401	1.1937	0.1163
	DUC	EUW	NCN	0.0728634	0.8777	0.1901
	EUW	DUC	SCN	0.447761	6.6222	2E-11
	EUW	DUC	SWCN	0.187129	1.8819	0.0299
S3 (14:96.47–97.65 Mb)	ECN	DUC	EUW	0.266232	3.7459	9E-05
	DUC	NCN	EUW	0.0031214	0.0414	0.4835
	EUW	DUC	SCN	0.354104	9.3692	0
	SWCN	DUC	EUW	0.180207	2.7965	0.0026
S4 (14:98.12–98.33 Mb)	EUW	ECN	DUC	0.0861798	0.8438	0.1994
	EUW	NCN	DUC	0.0833274	0.9529	0.1703
	EUW	SCN	DUC	0.291517	3.3918	0.0003
	EUW	SWCN	DUC	0.114714	1.3727	0.0849

Note: S1, S2, S3, and S4 indicate segments 1–4 introgressed from Chinese pigs into commercial pigs. $D = (ABBA - BABA) / (ABBA + BABA)$, with closely related *Sus* species from Southeast Asian islands as the outgroup. P1-P3, are the combination of EUW, and Chinese native pig groups (There are four valid combinations according to the formula of D-statistics).

TABLE 2 Average allele frequency of the Chinese-derived alleles within the four segments, in every population.

Segment	DUC	LDRNL	LDRUS	LWHFR	LWHNL	EUD	EUW	ECN	NCN	SCN	Secn	ASW
S1	0.8894	0.0022	0.0172	0.0264	0.0372	0.0003	0.0122	0.7238	0.5020	0.4289	0.7300	0.3062
S2	0.8827	0.0016	0.0058	0.0576	0.1149	0.0007	0.0123	0.1060	0.1123	0.4429	0.1874	0.2032
S3	0.8961	0.3528	0.3401	0.2664	0.2177	0.0004	0.0081	0.3453	0.3090	0.3500	0.5444	0.2241
S4	0.8818	0.0551	0.0602	0.0675	0.0448	0.0006	0.0046	0.1958	0.2003	0.3029	0.2138	0.1974

Note: S1, S2, S3, and S4 indicate segments 1–4 introgressed from Chinese pigs into commercial pigs. In the four segments, there are high allele frequencies in Duroc but low allele frequencies in Landrace and Large White, while these allele frequencies levels in Chinese pigs are high or moderate.

phenotypes of Western commercial breeds is caused by i) their initial variety, ii) different Chinese pigs used for introgression, iii) different directions and strength of selection after introgression. For different commercial lines of the same breed, the variation in phenotypes was most likely mainly caused by variation in the strength of selection. An illustration is the identified novel introgression haplotype from Southern China to Duroc pigs on chromosome 14 harboring the *PRKG1* gene. The *PRKG1* gene straddles the two introgressed segments (segment 3 and segment 4). Considering the high degree of LD in this region, it is very likely that they are derived from a single gene flow event. *PRKG1* has previously been reported to have undergone positive selection in Duroc (Kim et al., 2015) and is related to fatty acid composition. The gene showed copy number variation in Iberian - Landrace crosses (Revilla et al., 2017) and is related to average daily gain in Large White pigs (Wu et al., 2019). Furthermore, we showed that this introgressed *PRKG1*-haplotype significantly affects the thickness of the pig backfat

and loin depth (Figures 8C–E), indicating its relevance for commercial breeding.

We also found other genes with essential functions in this region (Table 3). *PCDH15* is related to backfat thickness according to a GWAS result of Landrace and Yorkshire population (Lee and Shin, 2018). Porcine *MBL2* is one of the mannose-binding lectins; it is the central component of innate immunity, facilitating phagocytosis and inducing the lectin activation pathway of the complement system (Phatsara et al., 2007; Bergman et al., 2014). *DKK1* is one of the Wnt signaling inhibitors. Upregulation of *DKK1* expression can be observed in the endometrium in pigs during the pre-implantation period (Zeng et al., 2019). *CSTF2T* plays a potential role in infertility as a mutation in this gene caused male infertility in humans (Gorukmez and Gorukmez, 2020). In conclusion, the introgressed segment contains a set of genes with potential impact on backfat thickness, immunity, daily gain and reproduction.

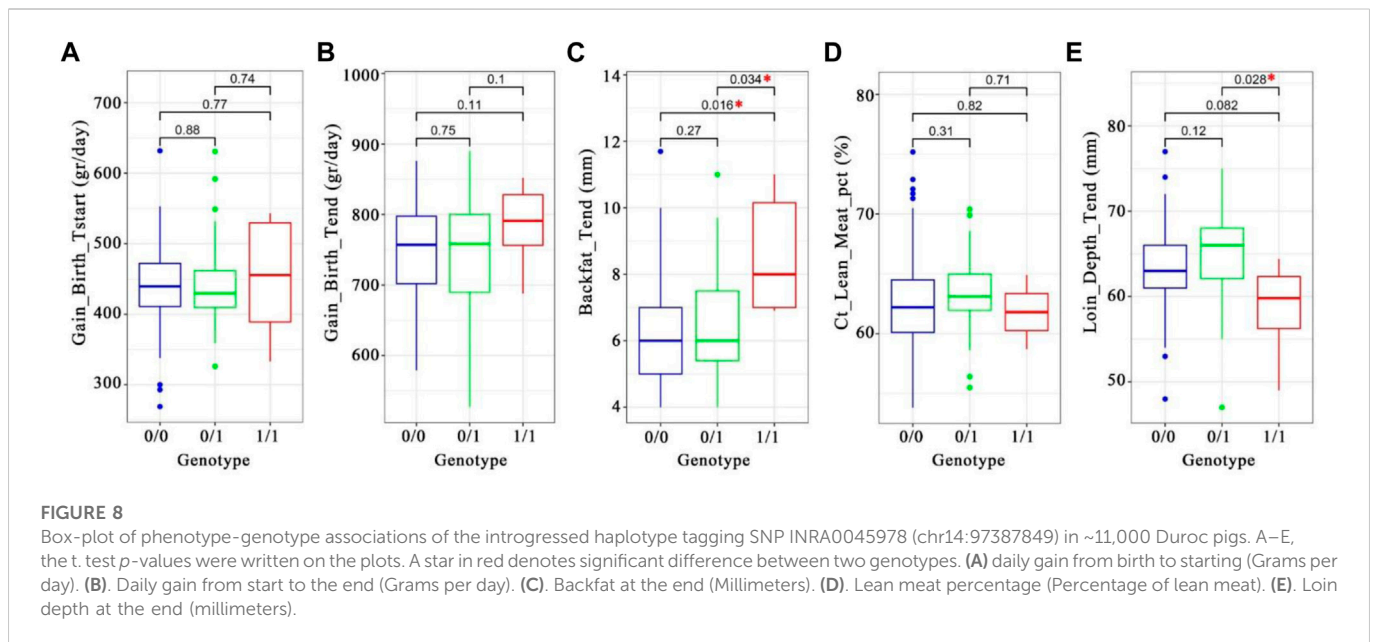


TABLE 3 Genes overlapping with the four segments within the introgressed region.

Segment lable	Position (BP)	Name	Description
S1 & S2	chr14:95,920,700–96,372,532	PCDH15	Protocadherin related 15
S 3	chr14:97,103,926–97,107,635	MBL2	<i>Sus scrofa</i> mannose-binding lectin 2
S 3	chr14:97,487,117–97,490,450	DKK1	Dickkopf WNT Signaling Pathway Inhibitor 1
S 3 & S 4	chr14:97,558,535–98,793,356	PRKG1	Protein Kinase CGMP-Dependent 1
S 4	chr14:98,105,772–98,110,358	CSTF2T	Cleavage Stimulation Factor Subunit 2 Tau Variant

Note: S1, S2, S3, and S4 indicate segments 1–4 introgressed from Chinese pigs into commercial pigs. The name of the gene is GeneCards (<https://www.genecards.org/>) Symbol. Description information is from GeneCards.

We also observed a large number of introgressed haplotypes in commercial Western pig breeds derived from NCN. However, we did not find any relevant written records of such an introduction of NCN into Europe or America. A general view is that ECN/SCN has been introduced to Europe to improve Western commercial pig breeds (Chen et al., 2017; Zhao et al., 2018; Chen et al., 2020). We, therefore, assume that NCN did not participate in the crossbreeding with Western commercial pigs directly but that the haplotypes introgressed and retained in Western pigs are more conserved in NCN than SCN/ECN. This suggests that current NCN pigs resemble the local breeds introduced centuries ago. This assumption should, however, be confirmed in future studies. Furthermore, it is known that Western commercial pigs contributed to NCN after the 20th century. Ai et al. (2015) found an extreme divergence between the northern and southern Chinese pig haplotypes in the 14-Mb region on the X chromosome. These haplotypes found in NCN were also found in European pigs. Therefore, a reciprocal introgression from European-related boars to NCN and *vice versa* cannot be ruled out. Therefore, care should be taken when

assessing the direction of selection and interpretation of the results.

5 Conclusion

A comprehensive analysis of the genetic introgression from Chinese pigs of different regions into different Western commercial lines was studied with 592 re-sequencing pigs. Our analysis revealed different Chinese pig haplotypes' complex introgression patterns and characteristics into Western commercial pig breeds. The results showed that the amount and origin of haplotypes introgressed from different Chinese pig sources to specific Western pigs vary greatly. The impact of Chinese haplotypes from specific sources on different commercial breeds is very different. The introgression likely occurred in the early stages of breed development. Breeding selection tendency experienced by different lines likely led to the observed differences in gene introgression. LWH pigs are most affected by Chinese haplotypes and the haplotypes were better retained in LWHFR. We also found that a ~2.65 Mb Chinese-derived

haplotype in Duroc pigs significantly affects the thickness of the pig backfat and the increase of loin depth.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

MB and YZ conceived the idea, YB and MD performed analyses, MG, YZ, and MB provided supervision, YP wrote the manuscript with input from all authors.

Funding

This work was funded by China Scholarship Council (CSC) File No. 201906350013.

References

- Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., et al. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47, 217–225. doi:10.1038/ng.3199
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109
- Amaral, A. J., Megens, H. J., Crooijmans, R. P., Heuven, H. C., and Groenen, M. A. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179, 569–579. doi:10.1534/genetics.107.084277
- Arnold, M. L., Sapir, Y., and Martin, N. H. (2008). Review. Genetic exchange and the origin of adaptations: Prokaryotes to primates. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 363, 2813–2820. doi:10.1098/rstb.2008.0021
- Bergman, I. M., Edman, K., van As, P., Huisman, A., and Juul-Madsen, H. R. (2014). A two-nucleotide deletion renders the mannose-binding lectin 2 (MBL2) gene nonfunctional in Danish Landrace and Duroc pigs. *Immunogenetics* 66, 171–184. doi:10.1007/s00251-014-0758-5
- Bianco, E., Soto, H. W., Vargas, L., and Pérez-Enciso, M. (2015). The chimerical genome of Isla del Coco feral pigs (Costa Rica), an isolated population since 1793 but with remarkable levels of diversity. *Mol. Ecol.* 24, 2364–2378. doi:10.1111/mec.13182
- Bosse, M., Megens, H. J., Frantz, L. A., Madsen, O., Larson, G., Paudel, Y., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5, 4392. doi:10.1038/ncomms5392
- Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi:10.1534/genetics.113.150029
- Browning, B. L., Tian, X., Zhou, Y., and Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* 108, 1880–1890. doi:10.1016/j.ajhg.2021.08.005
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015
- Bullock, J. M., Bonte, D., Pufal, G., da Silva Carvalho, C., Chapman, D. S., García, C., et al. (2018). Human-mediated dispersal and the rewiring of spatial networks. *Trends Ecol. Evol.* 33, 958–970. doi:10.1016/j.tree.2018.09.008
- Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., et al. (2019). Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. *Front. Plant Sci.* 10, 4. doi:10.3389/fpls.2019.00004
- Cao, Y. H., Xu, S. S., Shen, M., Chen, Z. H., Gao, L., Lv, F. H., et al. (2021). Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. *Mol. Biol. Evol.* 38, 838–855. doi:10.1093/molbev/msaa236
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8
- Chen, H., Huang, M., Yang, B., Wu, Z., Deng, Z., Hou, Y., et al. (2020). Introgression of Eastern Chinese and Southern Chinese haplotypes contributes to the improvement of

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1070783/full#supplementary-material>

fertility and immunity in European modern pigs. *Gigascience* 9, gaa014–13. doi:10.1093/gigascience/giaa014

Chen, M., Su, G., Fu, J., Zhang, Q., Wang, A., Sandø Lund, M., et al. (2017). Population admixture in Chinese and European *Sus scrofa*. *Sci. Rep.* 7, 13178. doi:10.1038/s41598-017-13127-3

Choi, J. W., Chung, W. H., Lee, K. T., Cho, E. S., Lee, S. W., Choi, B. H., 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. doi:10.1093/dnares/dsv011

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (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* 6, 80–92. doi:10.4161/fly.19695

Denison, M. S., Soshilov, A. A., He, G., DeGroot, D. E., and Zhao, B. (2011). Exactly the same but different: Promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci. official J. Soc. Toxicol.* 124, 1–22. doi:10.1093/toxsci/kfr218

Dowling, T. E., Markle, D. F., Tranah, G. J., Carson, E. W., Wagman, D. W., and May, B. P. (2016). Introgressive hybridization and the evolution of lake-adapted catostomid fishes. *PLoS One* 11, e0149884. doi:10.1371/journal.pone.0149884

Dowling, T. E., and Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* 28, 593–619. doi:10.1146/annurev.ecolsys.28.1.593

Enciso-Romero, J., Pardo-Diaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., et al. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Mol. Ecol.* 26, 5160–5172. doi:10.1111/mec.14277

Frantz, A. C., Zachos, F. E., Kirschning, J., Cellina, S., Bertouille, S., Mamuris, Z., et al. (2013a). Genetic evidence for introgression between domestic pigs and wild boars (*Sus scrofa*) in Belgium and Luxembourg: A comparative approach with multiple marker systems. *Biol. J. Linn. Soc.* 110, 104–115. doi:10.1111/bj.12111

Frantz, L. A. F., Madsen, O., Megens, H. J., Groenen, M. A., and Lohse, K. (2014). Testing models of speciation from genome sequences: Divergence and asymmetric admixture in island south-east asian *Sus* species during the plio-pleistocene climatic fluctuations. *Mol. Ecol.* 23, 5566–5574. doi:10.1111/mec.12958

Frantz, L. A. F., Meijaard, E., Gongora, J., Haile, J., Groenen, M. A., and Larson, G. (2016). The evolution of suidae. *Annu. Rev. animal Biosci.* 4, 61–85. doi:10.1146/annurev-animal-021815-111155

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H. J., Bosse, M., Paudel, Y., et al. (2013b). Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol.* 14, R107. doi:10.1186/gb-2013-14-9-r107

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H. J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148. doi:10.1038/ng.3394

- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Available at: <https://arxiv.org/abs/12073907> (Accessed December 16, 2022).
- Giuffra, E., Kijas, J. M., Amarger, V., Carlborg, Ö., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* 154, 1785–1791. doi:10.1093/genetics/154.4.1785
- Gorukmez, O., and Gorukmez, O. (2020). First infertile case with CSTF2TGene mutation. *Mol. Syndromol.* 11, 228–231. doi:10.1159/000509686
- Grahofer, A., Letko, A., Hafliker, I. M., Jagannathan, V., Ducos, A., Richard, O., et al. (2019). Chromosomal imbalance in pigs showing a syndromic form of cleft palate. *BMC Genomics* 20, 349. doi:10.1186/s12864-019-5711-4
- Grant, P. R., and Grant, B. R. (2019). Hybridization increases population variation during adaptive radiation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 23216–23224. doi:10.1073/pnas.1913534116
- Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398. doi:10.1038/nature11622
- Gross, C., Derks, M., Megens, H. J., Bosse, M., Groenen, M. A. M., Reinders, M., et al. (2020). pCADD: SNV prioritisation in *Sus scrofa*. *Genet. Sel. Evol.* 52, 4. doi:10.1186/s12711-020-0528-9
- Huang, M., Yang, B., Chen, H., Zhang, H., Wu, Z., Ai, H., et al. (2020). The fine-scale genetic structure and selection signals of Chinese indigenous pigs. *Evol. Appl.* 13, 458–475. doi:10.1111/eva.12887
- Hufford, M. B., Lubinsky, P., Pyhajarvi, T., Devengenzo, M. T., Ellstrand, N. C., and Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genet.* 9, e1003477. doi:10.1371/journal.pgen.1003477
- Janzen, G. M., Wang, L., and Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytol.* 221, 1279–1288. doi:10.1111/nph.15457
- Kim, H., Caetano-Anolles, K., Seo, M., Kwon, Y. J., Cho, S., Seo, K., et al. (2015). Prediction of genes related to Positive selection using whole-genome resequencing in three commercial Pig breeds. *Genomics Inf.* 13, 137–145. doi:10.5808/GI.2015.13.4.137
- Koch, K., Algar, D., Searle, J. B., Pfenninger, M., and Schwenk, K. (2015). A voyage to terra australis: Human-mediated dispersal of cats. *BMC Evol. Biol.* 15, 262. doi:10.1186/s12862-015-0542-7
- Larson, G., and Burger, J. (2013). A population genetics view of animal domestication. *Trends Genet.* 29, 197–205. doi:10.1016/j.tig.2013.01.003
- Lee, Y. S., and Shin, D. (2018). Genome-wide association studies associated with backfat thickness in Landrace and Yorkshire Pigs. *Genomics Inf.* 16, 59–64. doi:10.5808/GI.2018.16.3.59
- Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference Program. *Mol. Biol. Evol.* 32, 2798–2800. doi:10.1093/molbev/msv150
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi:10.1093/nar/gkab301
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi:10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., 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. doi:10.1101/gr.207456.116
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi:10.1038/ng.2811
- Li, Y., Oosting, M., Deelen, P., Ricano-Ponce, I., Smeekens, S., Jaeger, M., et al. (2016). Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat. Med.* 22, 952–960. doi:10.1038/nm.4139
- Liu, L., Bosse, M., Megens, H. J., Frantz, L. A. F., Lee, Y. L., Irving-Pease, E. K., et al. (2019). Addendum: Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat. Commun.* 10, 6306. doi:10.1038/s41467-020-20106-2
- Malinsky, M., Matschiner, M., and Svardal, H. (2021). Dsuite - fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21, 584–595. doi:10.1111/1755-0998.13265
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends Ecol. Evol.* 20, 229–237. doi:10.1016/j.tree.2005.02.010
- Medugorac, I., Graf, A., Grohs, C., Rothhammer, S., Zagdsuren, Y., Gladyr, E., et al. (2017). Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* 49, 470–475. doi:10.1038/ng.3775
- Megens, H.-J., Crooijmans, R. P. M. A., San Cristobal, M., Hui, X., Li, N., and Groenen, M. A. M. (2007). Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: Differences in microsatellite variation between two areas of domestication. *Genet. Sel. Evol.* 40, 103–128. doi:10.1186/1297-9686-40-1-103
- Meng, J. W., He, D. C., Zhu, W., Yang, L. N., Wu, E. J., Xie, J. H., et al. (2018). Human-mediated gene flow contributes to metapopulation genetic structure of the pathogenic fungus *Alternaria alternata* from potato. *Front. Plant Sci.* 9, 198. doi:10.3389/fpls.2018.00198
- Onteru, S. K., Fan, B., Du, Z. Q., Garrick, D. J., Stalder, K. J., and Rothschild, M. F. (2012). A whole-genome association study for pig reproductive traits. *Anim. Genet.* 43, 18–26. doi:10.1111/j.1365-2052.2011.02213.x
- Otonari, C., Flink, L. G., Evin, A., Georg, C., De Cupere, B., Van Neer, W., et al. (2013). Pig domestication and human-mediated dispersal in Western Eurasia revealed through ancient DNA and geometric morphometrics. *Mol. Biol. Evol.* 30, 824–832. doi:10.1093/molbev/mss261
- Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., et al. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* 8, e1002752. doi:10.1371/journal.pgen.1002752
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi:10.1534/genetics.112.145037
- Peng, Y., Cai, X., Wang, Y., Liu, Z., and Zhao, Y. (2022). Genome-wide analysis suggests multiple domestication events of Chinese local pigs. *Anim. Genet.* 53, 293–306. doi:10.1111/age.13183
- Phatsara, C., Jennen, D. G., Ponsuksili, S., Murani, E., Tesfaye, D., Schellander, K., et al. (2007). Molecular genetic analysis of porcine mannose-binding lectin genes, MBL1 and MBL2, and their association with complement activity. *Int. J. Immunogenet* 34, 55–63. doi:10.1111/j.1744-313X.2007.00656.x
- Ramirez, O., Burgos-Paz, W., Casas, E., Ballester, M., Bianco, E., Olalde, I., et al. (2015). Genome data from a sixteenth century pig illuminate modern breed relationships. *Heredity* 114, 175–184. doi:10.1038/hdy.2014.81
- Revilla, M., Puig-Oliveras, A., Castelló, A., Crespo-Piazuelo, D., Paludo, E., Fernández, A. I., et al. (2017). A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. *PLoS One* 12, e0177014. doi:10.1371/journal.pone.0177014
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., et al. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507, 354–357. doi:10.1038/nature12961
- Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., and Hermisson, J. (2020). VolcanoFinder: Genomic scans for adaptive introgression. *PLoS Genet.* 16, e1008867. doi:10.1371/journal.pgen.1008867
- Stukenbrock, E. H. (2016). The role of hybridization in the evolution and emergence of new fungal Plant pathogens. *Phytopathology* 106, 104–112. doi:10.1094/PHYTO-08-15-0184-RVW
- Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C., et al. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol. Ecol.* 25, 2427–2442. doi:10.1111/mec.13539
- Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C. B., and Douglas, C. J. (2018). Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Mol. Ecol.* 27, 1667–1680. doi:10.1111/mec.14561
- Venter, O., Sanderson, E. W., Magrath, A., Allan, J. R., Beher, J., Jones, K. R., et al. (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* 7, 12558. doi:10.1038/ncomms12558
- Wang, C., Wang, H., Zhang, Y., Tang, Z., Li, K., and Liu, B. (2015). Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. *Mol. Ecol. Resour.* 15, 414–424. doi:10.1111/1755-0998.12311
- Wang, J., Liu, C., Chen, J., Bai, Y., Wang, K., Wang, Y., et al. (2020). Genome-wide analysis reveals human-mediated introgression from western Pigs to indigenous Chinese breeds. *Genes (Basel)* 11, 275. doi:10.3390/genes11030275
- Wang, L., Wang, A., Wang, L., Li, K., Yang, G., He, R., et al. (2011). *Animal genetic resources in China: Pigs*. Beijing, China: China Agriculture Press.
- Warnes, G. R. (2020). gplots: Various R programming tools for Plotting data. R package version 3.0. Available at: <https://CRAN.R-project.org/package=gplots> (Accessed December 16, 2022).
- Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., et al. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* 9, giae051. doi:10.1093/gigascience/giae051
- White, S. (2011). From globalized PIG BREEDS TO CAPITALIST PIGS: A study in animal cultures and evolutionary history. *Environ. Hist.* 16, 94–120. doi:10.1093/envhiv/enq143

- Wichmann, M. C., Alexander, M. J., Soons, M. B., Galsworthy, S., Dunne, L., Gould, R., et al. (2009). Human-mediated dispersal of seeds over long distances. *Proc. Biol. Sci.* 276, 523–532. doi:10.1098/rspb.2008.1131
- Wu, P., Wang, K., Yang, Q., Zhou, J., Chen, D., Liu, Y., et al. (2019). Whole-genome re-sequencing association study for direct genetic effects and social genetic effects of six growth traits in Large White pigs. *Sci. Rep.* 9, 9667. doi:10.1038/s41598-019-45919-0
- Yan, G., Guo, T., Xiao, S., Zhang, F., Xin, W., Huang, T., et al. (2018). Imputation-based whole-genome sequence association study reveals constant and novel loci for hematological traits in a large-scale swine F2 resource Population. *Front. Genet.* 9, 401. doi:10.3389/fgene.2018.00401
- Yang, J., Huang, L., Yang, M., Fan, Y., Li, L., Fang, S., et al. (2016). Possible introgression of the VRTN mutation increasing vertebral number, carcass length and teat number from Chinese pigs into European pigs. *Sci. Rep.* 6, 19240. doi:10.1038/srep19240
- Zeberg, H., and Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 587, 610–612. doi:10.1038/s41586-020-2818-3
- Zeng, S., Ulbrich, S. E., and Bauersachs, S. (2019). Spatial organization of endometrial gene expression at the onset of embryo attachment in pigs. *BMC Genomics* 20, 895. doi:10.1186/s12864-019-6264-2
- Zhang, Q., Calus, M. P. L., Bosse, M., Sahana, G., Lund, M. S., and Gulbrandsen, B. (2018). Human-mediated introgression of haplotypes in a modern dairy cattle breed. *Genetics* 209, 1305–1317. doi:10.1534/genetics.118.301143
- Zhang, W., Yang, M., Wang, Y., Wu, X., Zhang, X., Ding, Y., et al. (2020a). Genomic analysis reveals selection signatures of the Wannan Black pig during domestication and breeding. *Asian-Australas J. Anim. Sci.* 33, 712–721. doi:10.5713/ajas.19.0289
- Zhang, Y., Xue, L., Xu, H., Liang, W., Wu, Q., Zhang, Q., et al. (2020b). Global analysis of alternative splicing difference in peripheral immune organs between tongcheng Pigs and large white Pigs artificially infected with PRRSV *in vivo*. *Biomed. Res. Int.* 2020, 4045204. doi:10.1155/2020/4045204
- Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., et al. (2018). Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *Gigascience* 7, giy058. doi:10.1093/gigascience/giy058
- Zhao, Q. B., Sun, H., Zhang, Z., Xu, Z., Olasege, B. S., Ma, P. P., et al. (2019). Exploring the structure of haplotype blocks and genetic diversity in Chinese indigenous Pig Populations for conservation purpose. *Evol. Bioinforma. online* 15, 1176934318825082. doi:10.1177/1176934318825082
- Zhu, Y., Li, W., Yang, B., Zhang, Z., Ai, H., Ren, J., et al. (2017). Signatures of selection and interspecies introgression in the genome of Chinese domestic Pigs. *Genome Biol. Evol.* 9, 2592–2603. doi:10.1093/gbe/evx186