Whole-Genome Resequencing Reveals Genetic Structure and Introgression in Chinese Pudong White Pigs

Min Huang  
South China Agricultural University

Zhongping Wu  
Zhongkai University of Agriculture and Engineering

Xiaopeng Wang  
South China Agricultural University

Desen Li  
South China Agricultural University

Shaojuan Liu  
South China Agricultural University

Sumei Zheng  
South China Agricultural University

Lijuan Yang  
South China Agricultural University

Bingbing Liu  
South China Agricultural University

Guixin Li  
South China Agricultural University

Yongchuang Jiang  
South China Agricultural University

Hui Zhang (mailto: zanghui520627@hotmail.com)  
South China Agricultural University

Hao Chen  
Jiangxi Science and Technology Normal University

Jun Ren  
South China Agricultural University

Research

Keywords: Pudong White pig, Whole-genome resequencing, Genetic structure, Introgression, KIT, White coat color
Abstract

Background: Pudong white (PDW) pigs originating from Shanghai, are the only Chinese indigenous pigs with white coat color except Rongchang (RC) pigs. However, there is limited information about its overall genetic structure, relationship with other breeds especially the East Chinese (ECN) and European pig due to the white coat of PDW. Whole-genome sequencing provides the effective approach to get the unique information of genome. The high-depth whole-genome sequencing data of 26 global pig breeds, European Wild boars (EWB), Chinese Wild boars (CWB) and out group (OUT) were implemented to detect the genetic structure, signature of selection and potential exotic introgression in PDW pigs.

Results: The PDW pigs belonging to ECN pigs based on genetic relationship, and harbor lower genetic diversity and higher inbreeding coefficient compared to other Chinese indigenous pigs. Both the $f^3$ and $D$-statistics analysis demonstrated that PDW pigs shared apparent alleles with Large White (LW) pigs. Then, two statistics, haplotype heat-map, copy number variation (CNV) and rIBD analysis further revealed that PDW pigs carry the same KIT genotype and share haplotypes at PARG-MARCHF8 locus with LW pigs, suggesting that the lineage of European (EUR) pigs in PDW originated from LW pigs. After detecting the KIT mutations in different pig breeds, PDW was confirmed to be same with LW at DUP1, DUP2 and the splicing mutation on intron 17 of KIT which determine the white coat color phenotype in European white pigs.

Conclusions: This study shows that ECN pigs crossed with LW pigs after introduced to China about 110-164 years ago, where the offspring carrying KIT genotype that caused white coat color phenotype, and then were selected due to the rare white coat color in Chinese indigenous pigs, gradually forming PDW pig breed. To our knowledge, this study gives the first thorough description of the genetic structure of PDW pig via whole-genome resequencing data. This study not only advances our understanding of genetic structure, molecular phylogeny, and molecular origin of PDW pigs, but also provides a basis for facilitating the development of a national project for the conservation and utilization of this unique Chinese local population.

Background

The domestication of farm animals is one of the most important events in human history. It has deeply influenced human history allowing a transition from hunting and gathering to more settled lifestyles. Pigs (Sus scrofa) are important domestic animals that domesticated largely in Europe and China approximately 10,000 years ago [1–3]. Since then, under the different forces of natural and human-mediated artificial selection, distinct phenotypes of different pig breeds emerged in the world [4] with diverse phenotypes in appearance, growth, palatability, fertility and local fitness. Due to different geographic origins, a rich genetic resource of pig breeds have been formed in China which occupied more than one-third (~ 100) of the total number of global pig breeds [5]. Chinese domestic pigs are important genetic resource according to the good adaptability to local environment [6], high prolificacy, desirable meat quality [7, 8] and high resistance to disease [9]. These excellent phenotypes marked them to be
good genetic materials for the improvement of economic phenotypes and support the sustainable
development of the pig industry not only in China but also in western countries [10, 11].

Pudong White (PDW) pigs are primarily distributed in Pudong New District, Shanghai. They are
geographically adjacent to Taihu pig breeds (also called ECN pigs, including Meishan, Fengjing,
Shawutou, Erhualian, Jiaxing Black and Mi pigs) which are mainly distributed in a narrow region around
Taihu Lake in the lower Yangtze River Valley of China. Of note, PDW pigs have similar economic traits to
Taihu pigs including high fecundity, high resistance to crude feed, satisfactory adaptability to the local
environment, and superior meat quality [12]. For these reasons, it have been included in the conservation
list of China's livestock and poultry genetic resource by the Ministry of Agriculture of China [12].
Noteworthy, the completely different in coat color phenotypes of PDW has raised the debate of this
genetic resource and its relationship to Taihu and European pig breeds. Some studies have shown that
PDW is a distinctive genetic resource with a unique genetic structure separate from other pig breeds
through mitochondrial DNA-based genetic structure analysis [13] or genetic structure analysis based on
low density of whole-genome single nucleotide polymorphism (SNP) data [14]. Since the Opium War with
the opening of Shanghai, EUR pigs such as white pig, Yorkshire, and miniature pigs with erect ears were
then introduced to the Shanghai in succession to cross with local breeds [15]. It hence arises the
possibility that the PDW pig could have genetic components introgressed from European breeds
considering its coat color. Therefore, it is worth to study whether there is introgression from European
breeds into PDW pigs, and the genomic distribution of introgressed European haplotypes and the
molecular mechanism of white coat color in PDW pigs should be elusived.

Whole-genome sequencing provides the effective approach to get the unique information of genome
including data on whole-exome sequencing, translocations, complex rearrangements and genome-wide
mutational patterns. Based on these genome information, it is proved to be efficiently characterize the
identify causative mutations underlying monogenic defects [16, 17], estimate the number of loci affecting
the complex traits [18], and detect the regions shaped by selection [19–22] in domestic animals.
Moreover, in pigs, the genomic signatures of selection contributing to the domestication of EUR pigs, the
local adaptation of the Chinese Tibetan wild boars [6, 23, 24] and the evolutionary history of pigs and the
role of introgression in adaptation were identified by whole-genome sequencing and selective sweep
analysis [25].

To elucidate the genetic structure, introgression and the regions shaped by selection for white coat color
in PDW pigs, we sequenced the genomes of 15 PDW pigs at high genomic coverage (20 ×) and obtained
whole-genome resequencing data of 320 global pigs from public database. After performing the
population genetic structure analysis, the $f_3$, D-statistics analysis and rIBD in 26 global pig breeds,
European Wild boars and Chinese Wild boars, PDW pigs were proved to belong to ECN ecotype pigs, and
with the introgression of EUR pigs lineage at the KIT and PARG-MARCHF8 locus which share the same
haplotypes with LW pigs, and the mutations in KIT probably determine the white coat color in PDW pigs.

Materials And Methods
Sample collection and DNA extraction

All procedures involving animals complied with guidelines for the care and utility of experimental animals prescribed by the Ministry of Agriculture of China. The ear tissues of PDW pigs were collected from Zhejiang Qinglian Food Company. Genomic DNA was extracted from ear tissues using a routine phenol/chloroform protocol and diluted to a final concentration of 100 ng/µl. The integrity of genomic DNA was verified with agarose gel electrophoresis.

Data collection and whole-genome sequencing

A total of 15 PDW pigs with distant genetic relationships were sequenced at 20× coverage depth using a whole-genome shotgun strategy. 150 bp paired-end libraries reads of these 15 individuals were sequenced on Illumina Hiseq 2000 instrument (Illumina, United States) to produce the raw data. The whole-genome resequencing data of 320 individuals from 27 populations, including 18 Chinese indigenous breeds (Jinhua (JH), Erhualian (EHL), Meishan (MS) Wannan Black (WNB), Wuzhishan (WZS), Luchuan (LUC), Bamaxiang (BMX), Baoshan (BS), RC, Neijiang (NJ), Hetao (HT), Laiwu Black (LWH), Min (MIN), Bamei (BAM), Tibet Tibetan (TT), Sichuan Tibetan (SCT), Yunnan Tibetan (YNT), Gansu Tibetan (GST), n = 181), CWB (n = 13), six European domestic breeds (Duroc (DU), French Large White (FLW), English Large White (ELW), Korean Landrace (KLR), Danish Landrace (DLR), Pietrain (PI), n = 113), EWB (n = 21) and OUT (Phacochoerus africanus, Sus cebifrons, Sus celebensis, Sus verrucosus, Sus barbatus, Sumatras wild boars, n = 7) were downloaded from NCBI public database (Access No. PRJNA398176 [26], PRJNA213179 [25], PRJNA488327 [27], PRJNA550237 [28], PRJEB9922 [29], PRJEB1683 [23], PRJNA553106 [30], PRJNA231897 [31], PRJNA239399 [32], PRJNA506339 [33] and PRJNA260763 [34], Table S1). The whole-genome sequencing raw data of 335 pigs was filtered by the following steps: (i) removing the paired reads if the content of N contained in a single-end sequencing read exceeded 10% of the length of the read; (ii) filtering the paired reads if the number of low-quality (Q ≤ 5) bases contained in a single-end sequencing read exceeded 50% of the length of the read; (iii) removal of the adapter paired-end reads.

Variants detection

First, a reference genome index of Sus scrofa 11.1 [35] was established by means of the index function from BWA v0.7.17 [36]. Then, the paired-end reads were aligned to the genome index using mem algorithm from BWA and obtained sorted binary bam files from sam files via SAMtools v1.9 [37]. LocusCollectora and Realigner functions from sentieon (https://www.sentieon.com/) were used to remove duplicate reads and realigned indels, and QualCal function was used for Base Quality Score Recalibration (BQSR). Afterward, the Haplotyper method from sentieon was used to generate gvcf file of each individual and GVCFtyper was used to perform joint calling. Thereafter, hard filtering was implemented under conditions of QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 3.0 via VariantFiltration method from GATK v4.0.2.1 (https://gatk.broadinstitute.org/hc/en-us). Finally, PLINK v1.9 [38] was used to reserve variants with
major allele frequency (MAF) > 0.01 and call rate > 0.9, where 30,987,305 single nucleotide polymorphisms (SNPs) were obtained for subsequent statistical analysis.

**Population genetics and genetic diversity analysis**

The obtained SNPs data set was filtered with MAF > 0.05 and call rate > 0.9 to acquire a new data set which was used for the following analysis. The command --distance-matrix from PLINK was used to calculate the average share allele (1-Dst) distance matrix between individuals. An unrooted neighbor-joining (NJ) tree was constructed by phylip v3.69 [39] and visualized with Figtree v1.42 (http://tree.bio.ed.ac.uk/software/figtree/). Genetic differentiation between populations was calculated using VCFtools v0.1.16 [40] with no-overlap 40-kb sliding windows. An interbreed NJ-tree was constructed by phylip using average WEIGHTED_FST values and visualized with Figtree as well. Then, --pca command from gcta v1.92.4 [41] was used to estimate the first 4 principal components (PC) of each individual without OUT. Subsequently, --r2 command in PLINK was used to calculate linkage disequilibrium (LD) decay of each population with a maximum distance of 1000 kb between SNPs. Run of homozygosity (ROH) was estimated through command --homozyg in PLINK. Each ROH region adhered to the following conditions: (i) number of SNPs more than 200; (ii) numbers of heterozygote SNPs less than 2; (iii) missing genotype less than 5; and (iv) the length more than 50 kb. The number of ROH length < 100 kb, 100–500 kb, 500–1000 kb, 1000–5000 kb, and > 5000 kb respectively were counted. Finally, the coefficient of inbreeding (F) and heterozygosity of each population were estimated by commands --het and --hard from PLINK.

**Admixture analysis**

The whole-genome $f_3$(Source1, Source2; Target) statistic was estimated via qp3Pop procedure from Admixtools v6.0 [42] under the default parameters. The $f_3$-statistic measures shared genetic drift between the two source populations and the target population [43]. Larger values indicated greater shared genetic drift between two source populations, hence a closer relationship between the two sources. Further, $f_3(X, Y; outgroup)$ was used to measure the shared genetic drift between PDW and the other pig breeds, where Y referred to PDW, ECN, South Chinese (SCN), North Chinese (NCN) and Southwest (SWCN) pigs, and X referred to the other pig groups. In the D-statistics, $D(Pop1,Pop2, Pop3, OUT)$, where $D > 0$ indicates that Pop1 and Pop3 shared more alleles than Pop2 and Pop3 share, while $D < 0$ means that Pop2 and Pop3 share more alleles than Pop1 and Pop3 share. The model of $D(pop1, pop2, PDW, OUT)$ was used, where pop1 and pop2 referred to LW pigs (including FLW and ELW) and the other pig breeds. Then, --indep-pairphase 1000 100 0.5 command in PLINK was used to prune the dataset, where three values referred to window size in kb, step size in variants number, and LD decay ($r^2$). To reduce the impact of strong linkage, a new data set including SNPs without $r^2$ values greater than 0.5 was used to reckon the ancestral lineage composition under default parameters via ADMIXTURE v1.3.0 [44]. In this analysis, we removed OUT and randomly selected 6 individuals from the remaining 27 groups. Treemix software [45] used a Gaussian approximation of genetic drift and allele frequency data to estimate a maximum likelihood (ML) tree and allowed the addition of admixture events in its framework. Under conditions of 100
bootstrap replicates and 1000 SNPs in one block, we rooted the tree using OUT and inferred the ML tree under several different migration events.

**Selective signal analysis**

SNPs with MAF > 0.05 and call rate > 0.9 in 15 PDW and 51 ECN pigs with black coat color (ECNB, including EHL, MS and WNB pigs) were extracted to conduct selective signal analysis. The autosomal genetic differentiation index $F_{st}$ between PDW and ECNB pigs was estimated using VCFtools with no-overlap 40-kb sliding windows and selected the empirical top 1% of $F_{st}$ as the threshold ($F_{st_{top1%}} = 0.64$). $F_{st}$, nucleotide polymorphism ($\pi$) and Cross Population Extended Haplotype Homozogosity (XP-EHH) on $KIT$ locus and $PARG$-$MARCHF8$ locus with no-overlap 10-kb sliding windows were performed using VCFtools and –xpehh function from selscan [46].

**Introgression analysis**

Whole-genome $D(LW, ECNB, PDW, OUT)$ analysis was computed in non-overlap 40-kb windows using script ABBABABAwindows.py [47], where LW included ELW and FLW pigs. We estimated identical by decent (IBD) between PDW and two different groups (LW and ECNB) for 10 times using BEAGLE with a 40-kb bin. Normalized IBD (nIBD) was acquired by formula nIBD = clIBD/tIBD, where clIBD indicated the count of haplotypes IBD between PDW pigs and one group, while tIBD meant total pairwise comparisons between PDW pigs and one group. The distribution of nIBD ranged between 0 (no IBD detected) and 1 (all individuals in the group share IBD). Relative IBD (rIBD) between PDW and LW pigs was calculated by formula rIBD = nIBD$_{PDW_LW}$ - nIBD$_{PDW_ECNB}$ [10]. The empirical top 1% windows was selected as the threshold both in D-statistic ($D_{top1%} = -0.979$) and rIBD (rIBD$_{top1%} = 0.042$). Also, the $D(LR, ECNB, PDW, OUT)$ and the rIBD were calculated between PDW and another two groups (LR and ECNB), where LR included KLR and DLR pigs.

**Copy number variation (CNV) prediction and PCR Amplification**

A Gaussian hidden Markov model method described by Miles was used to predict the diploid copy number state of DUP1 region (chr8: 41,223,212 – 41,783,660 bp) in the $KIT$ locus [10, 48]. The CNV prediction was conducted with a non-overlap 800 bp window size for each alignment. The primers for DUP1-4 duplication breakpoints amplification were designed by Wu, et al [48]. PCR amplifications were run on a Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA) at 95°C for 5 min, 30 × (95°C for 30 s, 57°C for 30 s, 72°C 1–2 min), 72°C 10 min and 12°C forever. PCR products were visualized through 1% agarose gel electrophoresis. Here, mixed genome DNA from ten individuals of PDW, LW, Jiaxing black (JXH) and Yushan black (YSH) pigs respectively and performed the PCR amplifications of each mix pool for three times.

**Results**
Population genetic structure analysis verified PDW pigs belong to the ecotype of ECN pigs

We conducted variants detection in 335 pigs, including 320 individuals from the public database (Table S1) and 15 PDW pigs, notably detected 30.98 million SNPs, were used for subsequent analysis. To examine phylogenetic relationships of the 335 pigs, an individual-based NJ-tree analysis was conducted using the SNPs set (Fig. 1A). The 335 pigs were subdivided into 3 main branches, representing Chinese pigs, EUR pigs and OUT respectively. Individuals from different populations formed different groups, among which, PDW pigs were close to the Chinese pigs, while KLR and DLR were in similar Landrace (LR) pig strain. Herein, Chinese indigenous pigs were divided into four ecotypes, i.e., (i) ECN including EHL, JH, MS, and WNB pigs; (ii) SCN including BMX, LUC and WZS pigs; (iii) SWCN including BS, NJ, RC, GST, YNT, SCT and TT pigs; and (iv) NCN including LWH, HT, BAM and MIN pigs, which was consistent with previous findings [5].
<table>
<thead>
<tr>
<th>Breed</th>
<th>Abbrev</th>
<th>Ecotype</th>
<th>Ho</th>
<th>He</th>
<th>Pn</th>
<th>r^{2}_{0.3}(kb)</th>
<th>F</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese Wild boar</td>
<td>CWB</td>
<td>CWB</td>
<td>0.31</td>
<td>0.35</td>
<td>0.58</td>
<td>1.76</td>
<td>0.25</td>
<td>1.96e-03</td>
</tr>
<tr>
<td>Pudong White</td>
<td>PDW</td>
<td>ECN</td>
<td>0.38</td>
<td>0.36</td>
<td>0.46</td>
<td>11.10</td>
<td>0.41</td>
<td>1.74e-03</td>
</tr>
<tr>
<td>Erhualian</td>
<td>EHL</td>
<td>ECN</td>
<td>0.36</td>
<td>0.34</td>
<td>0.51</td>
<td>2.51</td>
<td>0.26</td>
<td>1.70e-03</td>
</tr>
<tr>
<td>Meishan</td>
<td>MS</td>
<td>ECN</td>
<td>0.35</td>
<td>0.35</td>
<td>0.41</td>
<td>2.56</td>
<td>0.33</td>
<td>1.57e-03</td>
</tr>
<tr>
<td>Jinhua</td>
<td>JH</td>
<td>ECN</td>
<td>0.37</td>
<td>0.34</td>
<td>0.49</td>
<td>12.22</td>
<td>0.38</td>
<td>1.74e-03</td>
</tr>
<tr>
<td>Wannan Black</td>
<td>WNB</td>
<td>ECN</td>
<td>0.27</td>
<td>0.32</td>
<td>0.37</td>
<td>1.51</td>
<td>0.42</td>
<td>1.14e-03</td>
</tr>
<tr>
<td>Bamei</td>
<td>BAM</td>
<td>NCN</td>
<td>0.36</td>
<td>0.34</td>
<td>0.65</td>
<td>6.04</td>
<td>0.21</td>
<td>2.31e-03</td>
</tr>
<tr>
<td>Hetao</td>
<td>HT</td>
<td>NCN</td>
<td>0.39</td>
<td>0.35</td>
<td>0.69</td>
<td>4.61</td>
<td>0.10</td>
<td>2.48e-03</td>
</tr>
<tr>
<td>Laiwu Black</td>
<td>LWH</td>
<td>NCN</td>
<td>0.36</td>
<td>0.35</td>
<td>0.60</td>
<td>4.83</td>
<td>0.27</td>
<td>2.21e-03</td>
</tr>
<tr>
<td>Min</td>
<td>MIN</td>
<td>NCN</td>
<td>0.38</td>
<td>0.36</td>
<td>0.62</td>
<td>4.98</td>
<td>0.23</td>
<td>2.26e-03</td>
</tr>
<tr>
<td>Bamaxiang</td>
<td>BMX</td>
<td>SCN</td>
<td>0.38</td>
<td>0.35</td>
<td>0.64</td>
<td>3.22</td>
<td>0.21</td>
<td>2.26e-03</td>
</tr>
<tr>
<td>Luchuan</td>
<td>LUC</td>
<td>SCN</td>
<td>0.39</td>
<td>0.35</td>
<td>0.52</td>
<td>5.11</td>
<td>0.33</td>
<td>1.93e-03</td>
</tr>
<tr>
<td>Wuzhishan</td>
<td>WZS</td>
<td>SCN</td>
<td>0.37</td>
<td>0.34</td>
<td>0.71</td>
<td>2.46</td>
<td>0.14</td>
<td>2.49e-03</td>
</tr>
<tr>
<td>Neijiang</td>
<td>NJ</td>
<td>SWCN</td>
<td>0.37</td>
<td>0.35</td>
<td>0.57</td>
<td>4.57</td>
<td>0.29</td>
<td>2.09e-03</td>
</tr>
<tr>
<td>Rongchang</td>
<td>RC</td>
<td>SWCN</td>
<td>0.30</td>
<td>0.35</td>
<td>0.53</td>
<td>1.76</td>
<td>0.39</td>
<td>1.91e-03</td>
</tr>
<tr>
<td>Baoshan</td>
<td>BS</td>
<td>SWCN</td>
<td>0.39</td>
<td>0.34</td>
<td>0.69</td>
<td>2.71</td>
<td>0.10</td>
<td>2.41e-03</td>
</tr>
<tr>
<td>Sichuan Tibetan</td>
<td>SCT</td>
<td>SWCN</td>
<td>0.34</td>
<td>0.36</td>
<td>0.64</td>
<td>3.05</td>
<td>0.24</td>
<td>2.29e-03</td>
</tr>
<tr>
<td>Tibetan</td>
<td>TT</td>
<td>SWCN</td>
<td>0.33</td>
<td>0.34</td>
<td>0.75</td>
<td>2.43</td>
<td>0.14</td>
<td>2.50e-03</td>
</tr>
<tr>
<td>Gansu Tibetan</td>
<td>GST</td>
<td>SWCN</td>
<td>0.34</td>
<td>0.33</td>
<td>0.70</td>
<td>2.23</td>
<td>0.21</td>
<td>2.41e-03</td>
</tr>
<tr>
<td>Yunnan Tibetan</td>
<td>YNT</td>
<td>SWCN</td>
<td>0.37</td>
<td>0.36</td>
<td>0.69</td>
<td>2.51</td>
<td>0.12</td>
<td>2.41e-03</td>
</tr>
<tr>
<td>French Large White</td>
<td>LW</td>
<td>EUR</td>
<td>0.33</td>
<td>0.33</td>
<td>0.38</td>
<td>4.72</td>
<td>0.45</td>
<td>1.46e-03</td>
</tr>
</tbody>
</table>

1 Abbreviations of breeds. 2 ECN, East China; SCN, South China, SWCN, Southwest China, NCN, North China. EUR, European

3 Observed heterozygosity. 4 Expected heterozygosity.

5 The proportion of polymorphic markers. 6 Linkage disequilibrium values of r^2 = 0.3. 7 Inbreeding coefficient.

8 Average nucleotide polymorphism in 10-kb windows.
The OUT was eliminated and PCA analysis was performed where the result revealed that the first principal component (PC1) divided the remaining 27 groups into Chinese and European pig populations, the PC2 divided Chinese pigs into multiple groups, and PDW pigs were still close to the Chinese indigenous pigs (Fig. 1B). Then, an NJ-tree between populations was constructed based on the $\text{Fst}$ (Fig. 1C). PDW and ECN pigs were on the same branch which indicated that genetic differentiation between PDW and ECN pigs was smaller than other pigs.

PDW was affected by the lineage of European pigs

ADMIXTURE analysis was performed using the remaining 27 groups after removing OUT to further observe the lineage composition of PDW pigs. The result demonstrated that when the assumed ancestor number K was 2–4, PDW pigs primarily comprised the lineage of ECN pigs but contained a certain proportion of EUR pigs’ lineage with the average proportions of 29.52%, 16.93%, and 13.5% respectively, indicating that PDW pigs belong to the ECN ecotype in genetic relationship but influenced by EUR pigs at the same time (Fig. 1D). However, when the K was more than 5, an independent lineage was formed in PDW pigs. Additionally, the NCN pigs contained an apparent lineage of EUR pigs, this was consistent with the results of previous studies [5, 28]. We conducted Treemix analysis with OUT as root and the result revealed that Chinese indigenous pigs were divided into ECN, SCN, and SWCN pig groups. However, NCN and PDW pigs were located between Chinese and EUR pigs due to the effect by lineage of EUR pigs (Fig. S1). When the migration event was assumed to be 8, an apparent gene flow (migration weights = 0.23) between PDW and LW pigs was observed (Fig. S2). In line with other findings, the results above suggest that PDW pigs belong to the ecotype of ECN pigs but are affected by EUR pigs.
To further observe the genetic diversity of PDW pigs in the context of global pig breeds, we estimated the expected heterozygosity (He), observed heterozygosity (Ho), SNPs polymorphism ratio (Pn), LD \((r^2 = 0.3)\) fragment length, inbreeding coefficient, and Run of homozygosity (ROH) of each pig group (Table 1, Table S2). As a result, PDW pigs had similar heterozygosity (He and Ho), lower SNPs polymorphism, longest LD \((r^2 = 0.3)\) fragment length and higher inbreeding coefficient compared to other Chinese indigenous pigs (Table 1, Fig. S3). We counted the number of ROHs of different fragment lengths in 28 pig groups. When the length was less than 500 kb, the number of ROHs in PDW pigs was similar to Chinese indigenous pigs. Nevertheless, when the ROH length was more than 500 kb, its number in PDW pigs was significantly higher than other China indigenous pigs and close to European domestic pigs (Table S2). Considering that the ROHs in PDW pigs were between China indigenous pigs and European domestic pigs, it further indicates the possibility of introgression from EUR pigs into PDW pigs.

### The gene flow, selected signals and introgression regions at the genome of PDW pigs from LW

We estimated the \(f_3\) test of each ecotype and found Chinese indigenous pigs have low genetic similarity with EUR pigs (Table S3-S6). Among these, NCN groups have higher genetic similarity with EUR pigs compared with other ecotypes (Table S6). Besides, each ecotype had higher genetic similarity with Chinese Wild boar (CWB) compared with European domesticated pigs. The \(f_3(PDW, X; OUT)\) was estimated to further evaluate the shared genetic drift between PDW and the other pig breeds, and the results revealed that PDW pigs shared the most genetic similarity with ECN pigs (Table S7). Noteworthy, the genetic similarity between PDW and European domesticated pigs was slightly higher than SCN pigs or CWB. The calculation results of \(D(LW, X, PDW, OUT)\) showed that PDW pigs shared the most alleles with ECN pigs (Table S8). Also, PDW shared the most alleles with LW pigs compared with other European domestic pigs which were even higher than SCN pigs and CWB.

Although PDW pig breed is a type of ECN pigs in genetic relationships, its coat color is completely different from the latter. We considered that PDW pigs have been strongly selected through human-mediated, making it possible to fix the genes causing the white coat phenotype in PDW pigs. To reduce the signal noise caused by different genetic backgrounds, \(Fst\) was performed between ECNB and PDW pigs where 18 genes were identified on the top windows of autosome including \(ECT2L, CCDC28A, ELMOD3, MMP19, PYM1, FANCA, ZNF276, VPS9D1, PRSS16, KIT, CCDC126, TAF3, EFCAB3, C3orf14, IDE, CDH18, RALGAPB\) and \(KMT2C\) (Fig. 2A). Among these, the \(KIT\) is a famous gene causing white coat color phenotype in LW pigs \([48]\). Total of 251 genes were identified from 623 windows with \(Fst\) value higher than \(Fst_{top1%}\). The Gene Ontology (GO) terms and KEGG pathways analysis were performed online website Metascape (http://metascape.org), which showed these genes mainly participate in regulating import growing development process, such as hormone metabolic process and labyrinthine layer morphogenesis (Fig. 2B).

Above results suggested a significant gene flow between PDW and LW pigs so that we performed whole-genome \(rIBD\) analysis between PDW and LW (ELW and FLW) pigs with ECNB as control group (Fig. 3A). Several significant introgression regions from LW pigs into PDW pigs were identified on chromosome 1, 2,
5, 6, 8, 9, 10, 14 and 15. Among these regions, two on chromosome 8 (Region2) and 14 (Region1) overlapped with the whole-genome ABBA-BABA test (Fig. S4). We examined Region1 carefully and found that a chr14:90.3–91.5 Mb region (Fig. 3B) including PARG, NCOA4, MSMB, ZFAND4, MARCHF8, ALOX5, RASSF4, ZNF22, DEPP1 and TMEM72 genes were identified. These genes mainly involved in growth-related biological processes, bone development and disease-resistance processes. The haplotypes heat-map showed that PDW pigs shared similar haplotypes with KLW, ELW, DU and sectional PI pigs (Fig. 3C). However, KLR and DLR pigs did not carry the shared haplotypes. Interestingly, sectional Chinese indigenous pigs (WZS, BS, BAM, HT and LWH pigs) also carried a small amount of the shared haplotypes, further implying that EUR pigs have introgressed into Chinese indigenous pigs.

The Fst and XP-EHH of PDW and the two groups (LW and ECNB pigs) in the area showed that the genetic differentiation and the degree of selection between PDW and LW pigs were lower compared to those of ECNB pigs (Fig. 3D and 3E).

The causality of splicing mutation, DUP1 and DUP2 of KIT in relation to the coat color phenotype of Pudong pigs

We found the import white coat color related KIT gene in European white pigs in Region2 (Fig. 3A). Previous studies showed that the KIT gene harbors various alleles. Among them, LW and LR pigs carry the I alleles, such as I1 and I2 [48], while Chinese indigenous pigs carry the recessive wild-type i allele [49–51], implying that the KIT identified in PDW pigs might be from introgression of European white pigs. Further, we made a close examination of IBD in the KIT locus and its surrounding area (8 Mb on both sides), and found that the frequency of shared IBD haplotypes between PDW and LW pigs on the chr8:40.85–41.60 Mb region was significantly higher compared to those surrounding regions (Fig. 4A). The haplotype heat-map constructed using the region revealed that the haplotypes of PDW pigs were completely different from Chinese indigenous pigs but highly similar to European domestic pigs, specifically LW and LR pigs (Fig. 4B). Also, we calculated the IBD between PDW and the other two groups (LR and ECNB pigs), where the shared IBD at the KIT locus was not found (Fig. S5). PDW and ECNB pigs were highly differentiated in a 400-kb region on both sides of the KIT locus, which was consistent with the degree of differentiation between ECNB and LW pigs. However, PDW and LW pigs had extremely low differentiation (Fig. 4C top). Besides, PDW and LW pigs had lower nucleotide polymorphisms in the KIT locus compared to ECNB pigs (Fig. 4C bottom). The genetic relationship of the KIT gene showed that Chinese indigenous pigs and EUR pigs formed two distinct branches, where PDW pigs were clustered with FLW, ELW and LR pigs (Fig. 4D).

Previous studies confirmed four duplications (DUP1-DUP4) on the KIT locus, and of them, DUP1, DUP2 and the splicing G > A mutation on intron 17 of KIT are necessary for the manifestation of a solid white coat color in European domestic white pigs [23, 48]. We counted the genotype frequency of the KIT splicing mutation (chr8:41486012 bp) of pigs used in this study and found that Chinese indigenous pigs and Wild boars did not carry the mutation. Nonetheless, 15 PDW and 13 LR pigs and 93% of LW pigs carried the mutation. Also, other 51 PDW pigs were genotyped using primers designed by Wu, et al. [48]
where we found that all 51 PDW pigs carried the mutation. We predicted the CNV within the 560-kb DUP1 region using the Hidden Markov Model (HMM), which showed PDW, LW and LR pigs had 3–6 copies of DUP1 whereas Chinese indigenous pigs, CWB, DU and EWB carried normal copies (Fig. 5A). PCR was performed using the primers designed on the breakpoint of each DUP (Fig. 5B) [48] in PDW, LW, JXH and YSH pigs. Each breakpoint sequence was identified in the amplified of PDW and LW pigs, while JXH and YSH pigs did not (Fig. 5C). The above results confirmed that PDW pigs carried DUP1, DUP2 and the splice mutation on intron 17 of KIT, which caused the white coat color phenotype in PDW pigs.

Discussion

PDW pigs and RC pigs are white coat color phenotype distinguish with other Chinese indigenous pigs that covered with black or spotted hair. As for this reason, debate was raised of PDW pigs’ genetic resource and its relationship to Chinese indigenous pigs and EUR pigs. Although the PDW was identified to be a distinctive genetic resource with a unique genetic structure [14, 48], there is limited information about its overall genetic structure and the molecular mechanism of the white coat color compared with global domesticated pigs. In this study, we used high-coverage whole-genome sequencing to generate a comprehensive catalog of genetic variants of PDW pigs. Population genetic analyses were performed with the global swine including 26 domesticated pig breeds, CWB, EWB and OUT. We then conducted in-depth population genetics analyses for the PDW pigs from the global pigs which based on the whole-genome sequencing data. To the best of our knowledge, this is the first population genomics analyses to use high-coverage whole-genome sequencing to dissect the genetic structure of PDW, uncover the introgression in PDW from other pig breeds and bring a comprehensive understanding of mechanism of white coat color in PDW pigs.

Our study supports PDW pigs belong to the ecotype of ECN pigs but carry EUR pigs’ lineage with the following reasons: (1) NJ-tree and PCA analysis classified PDW pigs to the ecotype of ECN pigs; (2) PDW pigs were located between Chinese and EUR pigs due to its genetic components of European origin by ADMIXTURE analysis; and (3) PDW pigs had similarity ROHs distribution with EUR domestic pigs; (4) f3 and D-statistics analysis demonstrated that PDW pigs shared apparent alleles with LW pigs; besides, (5) rIBD further discovered that PDW pigs carry the same KIT genotype and share haplotypes at PARG-MARCHF8 locus with LW pigs, confirming that the lineage of EUR pigs in PDW originated from LW pigs; (6) The DUP1, DUP2 and splicing mutation on intron 17 of KIT determine the white coat color phenotype of PDW pigs that same with LW pigs.

At the previous studies, PDW was identified to be a distinctive genetic resource with a unique genetic structure separate from other pig breeds [14], and genetic distances between PDW and European domesticated pigs were shorter than EUR pigs and other Chinese indigenous pigs. In view of this phenomenon, the source of the investigated European pig samples and the limit of the scale of the breeding farm were considered in terms of the potentially results of inbreeding and genetic drift. However, the results in this study revealed that there was introgression form European domesticated pigs in history occurred in Chinese indigenous pig, PDW. First, in this study, NCN and PDW pigs were located between
Chinese and EUR pigs in the population structure analysis. Interestingly, LWH pig breed is one of Chinese local pig breed classified to NCN also investigated here, which proved to have genetic components introgressed from EUR pigs [52], which supported by genetic evidence and historical evidence that EUR pigs were then introduced to the Shandong province and crossed with local breeds during the colonization of Yantai and Qingdao cities of Shandong Province by Britain and Germany at the early 20th century. Similarly, for PDW, EUR pigs such as Yorkshire were introduced to the Shanghai and crossed with local breeds subsequently since the Opium War with the opening of Shanghai [15]. This provides the historical evidence of introgression from European white pigs into PDW pigs. Moreover, we found that the number of ROHs in PDW pigs was similar to Chinese indigenous pigs when the length was less than 500 kb, otherwise they are closer to European domestic pigs with significantly higher than that in Chinese indigenous pigs. For the European domestic pigs, they undergone a long period of selective pressure to meet needs of consumers and then be a strong degree of inbreeding, which leads to more long homozygous segments in the genome. However, the intensity and degree of human-mediated breeding was lower in Chinese local pigs than that in European commercial pig, then resulted in a higher genetic diversity and recombination degree of Chinese local pigs than European commercial pigs [53]. For these reasons, the ROHs in European commercial pigs were longer than Chinese local pigs. When European domesticated pigs were introgressed into PDW pigs, it turned to be normal that ROHs in PDW pigs were similar to European domestic pig breeds when the length was more than 500 kb. It also proved that there is introgression from European domesticated pigs in PDW.

The results of $rIBD$ analysis not only further verified that European domesticated pigs’ lineage in PDW pigs, but also revealed the significant introgression at the $KIT$ locus and $PARG$-$MARCHF8$ locus from LW pigs. The PDW pigs shared similar haplotypes at these loci with LW, LR, DU and sectional PI pigs. Noteworthy, PI pigs were crosses of French Bayeux pigs $\times$ British Berkshire pigs $\times$ LW pigs in the 1950s, suggesting that shared similarity haplotypes of PI pigs with LW pigs was normal [12]. In the late 1970s, China successively introduced DU and PI pigs, which meant that the European pig lineage in PDW pigs did not originate from these two pig breeds according to the breeding history of PDW pigs. These results indicated that the lineage of EUR pigs in PDW pigs originated from the introgression of LW or LR pigs. Among the $KIT$ and $PARG$-$MARCHF8$ locus, the $MARCHF8$ at $PARG$-$MARCHF8$ locus was identified to be recruited by BST2 (bone marrow stromal cell antigen 2) to catalyze the ubiquitination of the PEDV (porcine epidemic diarrhea virus) N protein. The Porcine epidemic diarrhea (PED) was first recognized in England in 1971 [54], and causing acute diarrhea, vomiting, and high mortality rates in neonatal piglets. This shows that EUR pigs, especially the LW originated in England, may carry the haplotype at $MARCHF8$ susceptible to PEDV. After being introduced into China, LW crossed with Chinese local pigs and result in the same haplotype at $MARCHF8$ appeared in the offspring of hybrid pig lines, such as PDW. It further proves the possibility that the introgression in PDW origin from LW pigs. Further investigations are required to test this assumption.

The coat color phenotypes was one of the most intuitive breed features of domestic pigs. Notably, China has the largest number of local pig breeds in the world with various coat color phenotypes, including solid black, spotted, brown, belted, two-end black, white, etc. [55–57]. So far, several major genes have been
reported to affect the coat color in Chinese indigenous pigs, such as \textit{MC1R} gene for the solid black coat color [58], \textit{TYRP1} gene for the brown coloration phenotype in Chinese Kele, Dahe and Tibetan pigs [59], \textit{EDNRB} gene for the two-end black coat color phenotype in BMX pigs [5] and \textit{MITF} gene for the white coat color in RC pigs [60]. However, there is no novel mutations in these genes of PDW pigs compared with other Chinese domestic pigs. Interestingly, the \textit{KIT} gene at the \textit{KIT} locus was responsible for white coat color related in European white pigs, LW and LR. In addition, genetic relationship at this locus revealed that PDW pigs were clustered with LW and LR pigs. There are four duplications (DUP1-DUP4) on the \textit{KIT} locus, among these DUP1, DUP2 and the splicing mutation on intron 17 of \textit{KIT} are necessary for the manifestation of a solid white coat color in European domestic pigs [23, 48]. The experimental results showed that PDW pigs carried DUP1, DUP2 and the splicing mutation which is the same with LW and LR pigs. In combination with above results provide convincing evidence for the \textit{KIT} caused the white coat color phenotype in PDW pigs.

Historically, LW pig is one of ancient pig breed from Britain and was introduced to Chinese coastal areas along with human activities and subsequent crossed with the ECN pigs, especially Taihu pigs. Further, ALDER software [61] was used to estimate the time of admixture between ECN and FLW or ELW pigs, which yielded average estimate of 27.86 \pm 4.98 (139.31 \pm 24.91 years) and 27.12 \pm 5.21 (135.58 \pm 26.03 years) generations ago (Table S9), respectively, and coincided with the introduction of LW into China. Considering the nearly 200 years of history of PDW, the lineage of EUR pigs identified in PDW pigs might be from LW pigs, but not LR pigs bred after 1890s. Hence, we believe that LW pigs were crossed with local pigs after being introduced into the Shanghai area with human activities in the mid-19th century. Individuals carrying \textit{KIT} locus in the hybrid offspring were selected because of the rare white coat color in Chinese indigenous pigs [12] and to be unique genetic resource of PDW. Therefore, the introgression of the \textit{KIT} haplotype from LW pigs was the most likely causal gene for the white coat color of PDW.

**Conclusions**

In conclusion, this work used high-depth sequencing data of 335 global pig breeds, and based on genetic relationship, we confirmed that PDW pigs belong to ECN ecotype pigs and carry a certain proportion of EUR pigs’ lineage. Nonetheless, PDW pigs had a lower genetic diversity and higher inbreeding coefficient compared to other Chinese indigenous pigs. Based on \textit{rIBD} analysis and human activities in history, the peculiar white coat color phenotype in PDW pigs might be attributed to introgression from LW pigs in the mid-19th century. Further, we ascertained that PDW pigs carried the same DUP1, DUP2 and splicing mutation of \textit{KIT} with LW pigs, which determine the white coat color in PDW pigs.

**Abbreviations**

MAF: Major allele frequency; LD: Linkage disequilibrium; \( r^2_{0.3} \): Linkage disequilibrium values of \( r^2=0.3 \); BAM: Bamei; BMX: Bamaxiang; BS: Baoshan; CNV: Copy number variation; CWB: Chinese wild boars; DLR: Danish Landrace; DU: Duroc; DUP: Duplications; ECN: East Chinese; ECNB: Eastern Chinese black; EHL: Erhualian; ELW: English Large White; EWB: European wild boars; EUR: European; F: Inbreeding
coefficient; FLW: French Large White; Fst: Genetic differentiation index; GST: Gansu Tibetan; He: Expected heterozygosity; Ho: Observed heterozygosity; HT: Hetao; IBD: Identical By Descent; cIBD: Count IBD; nIBD: Normalized IBD; tIBD: Total IBD; JH: Jinhua; KLR: Korean Landrace; LR: Landrace; LUC: Luchuan; LW: Large White pigs; LWH: Laiwu black; MIN: Min; ML-tree: Maximum likelihood tree; MS: Meishan; NCN: North Chinese pigs; NJ: Neijiang; NJ-tree: Neighbor-joining tree; OUT: Outgroup; PCA: Principal components analysis; PDW: Pudong white; PI: Pietrain; Pn: he proportion of polymorphic markers; RC: Rongchang; ROH: Run of homozygosity; SCN: South Chinese pigs; SCT: Sichuan Tibetan; SNP: single nucleotide polymorphism; SWCN: Southwest Chinese pigs; TT: Tibetan; WNB: Wannan Black; WZS: Wuzhishan; XP-EHH: Cross Population Extended Haplotype Homozogysity; YNT: Yunnan Tibetan; π: nucleotide polymorphism.

Declarations

Acknowledgements

We are grateful to Zhejiang Qinglian Food Company provided the samples and the researchers who shared the whole genome sequencing data all over the world.

Authors’ Contributions

JR proposed the idea for the study. HC and HZ designed the study. MH and ZW collected and analyzed the data. MH wrote the manuscript. MH and ZW performed the bioinformatics analyses. HC and HZ revised the article. XW, DL, XW, SL, SZ, LY, BL GL and YC performed genotyping experiments.

Funding

This research was supported by the Science and Technology Research Project of Jiangxi Provincial Education Department (Grant Number: GJJ180627) and the Youth Science Foundation of Jiangxi Province (CN) (Grant Number: 20202BABL215021).

Availability of data and materials

On top of the public data sets used, unpublished raw sequencing data in this research is available via National Genomics Data Center (https://bigd.big.ac.cn/) Bioproject PRJCA003970.

Ethics approval and consent to participate

This study does not involve direct handling of animals so not applicable.
Consent for publication

Not applicable.

Competing interests

All authors declare that they have no conflict of interest.

References


55. Lai FJ, Jun R, Ai HS, Ding NS, Ma JW, D.Q. Z, et al. Chinese white Rongchang pig does not have the dominant white allele of KIT but has the dominant black allele of MC1R. J Hered. 2007;98:84.
Population relationship and structure. (A) Neighbor-joining (NJ) tree based on an identical-by-state matrix among 335 pigs tested in this study. Each abbreviation represents a pig breed and each color represents an ecotype (Table 1). (B) Principal component analysis of 27 populations without OUT. Each figured point represents the eigenvector of one individual. The first (PC1) and second components (PC2) are shown. Color codes for different ecotypes are as in Fig. 1. (C) NJ-tree of 28 populations based on Fst. (D)
Population structures were inferred using ADMIXTURE with the assuming number of ancestral cluster K from 2 to 8.

Figure 2

Genome-wide selection signals between PDW and Eastern China black pigs (ECNB, including Erhualian, Meishan and Wannan black pigs). (A) Autosomal manhattan plot of Fst between the two pig populations. Each point refers to one non-overlap 40 kb window and the gray dotted line represents threshold value defined by the top 1% windows (Fsttop1% = 0.64). (B) The Gene Ontology (GO) terms and KEGG pathways of the 251 genes identified from Fst analysis.
Figure 3

Introgressed European domestic haplotypes in PDW pigs. (A) Manhattan plot of rIBD values between Pudong White (PDW) and Large White (LW) pigs (positive value) or Eastern Chinese black (ECNB) (negative value). The grey dashed line indicates the top 1% significance threshold. (B) rIBD values in an 8-Mb region (chr14:90.3-91.5 Mb) harboring the MARCHF8 gene. The region between grey dashed lines including PARG, NCOA4, MSMB, ZFAND4, MARCHF8, ALOX5, RASSF4, ZNF22, DEPP1 and TMEM72I...
(C) Haplotype heat map in region indicated in the grey dashed lines in (B). (D) The intergroup Fst and (E) XP-EHH in a 400-kb region harboring MARCHF8 gene.

Figure 4

(A) rIBD values in an 8-Mb region harboring the KIT gene. The brown dashed line indicates the 1% threshold line (rIBDtop1%= 0.04) and the region between grey dashed lines including GSX2, PDGFRA and KIT genes. (B) Haplotype heat map in region indicated in the grey dashed lines in (A). Major and minor
alleles in PDW pigs are indicated by beige and light blue respectively. (C) The intergroup Fst (top) and intragroup nucleotide polymorphism ($\pi$, bottom) in a 400-kb region harboring KIT gene. (D) NJ-tree was constructed based on IBS matrix using variants in KIT region.

Figure 5

Copy number variation in the KIT locus. (A) Heat-map of copy number prediction along the KIT locus for 63 pigs. For each individual, diploid copy numbers were predicted in 800 bp non-overlapping bins by normalized coverage of DUP1 and its 500 kb flanking region on each side. The rectangles represent KIT and DUP2-4, which make up DUP1. The abbreviations in the figure correspond to Table S1 and CDOM means Chinese domestic pigs. (B) Schematic diagram of primers design breakpoint for DUP1-4. (C) Amplification results of duplication breakpoint sequences DUP1-4 in 10 pigs, Marker 2000 bp. PDW: Pudong white pigs. LW: Large White pigs. JXH: Jiaxing black pigs. YSH: Yushan black pigs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• Supplementarymaterials.pdf