ORIGINAL ARTICLE

Evolutionary Applications
Open Access WILEY

Whole-genome analysis reveals the hybrid formation of Chinese indigenous DHB pig following human migration

Yuzhan Wang[1] | Chunyuan Zhang[1] | Yebo Peng[1] | Xinyu Cai[1] | Xiaoxiang Hu[1] | Mirte Bosse[2] | Yiqiang Zhao[1]

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China

[2]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

Correspondence
Yiqiang Zhao, State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, 100193, China.
Email: yiqiangz@cau.edu.cn

Mirte Bosse, Animal Breeding and Genomics Centre, Wageningen University, Wageningen 6708 WD, The Netherlands.
Email: mirte.bosse@wur.nl

Abstract

Hybridization is widespread in nature and is a valuable tool in domestic breeding. The DHB (DaHuaBai) pig in South China is the product of such a breeding strategy, resulting in increased body weight compared with other pigs in the surrounding area. We analyzed genomic data from 20 Chinese pig breeds and investigated the genomic architecture after breed formation of DHB. The breed showed inconsistency in genotype and body weight phenotype, in line with selection after hybridization. By quantifying introgression with a haplotype-based approach, we proposed a two-step introgression from large-sized pigs into small-sized pigs to produce DHB, consistent with the human migration events in Chinese history. Combining with gene prioritization and allele frequency analysis, we identify candidate genes that showed selection after introgression and that may affect body weight, such as *IGF1R*, *SRC*, and *PCM1*. Our research provides an example of a hybrid formation of domestic breeds along with human migration patterns.

KEYWORDS
body weight, DaHuaBai pig, human migration, hybridization, *PCM1*, two-step introgression

1 | INTRODUCTION

Genetic introgression may result in the transmission of beneficial alleles. A growing body of evidence suggests a substantial role of introgression in improving domesticated breeds (Ackermann et al., 2019; Burgarella et al., 2019; Eizirik & Trindade, 2021). The introgression from yak enabled Tibetan taurine cattle to adapt to the extreme environment of the Qinghai–Tibet Plateau (Chen, Cai, et al., 2018; Qiu et al., 2012). Adaptive introgression can also occur between wild animals and domesticated breeds (Scheu et al., 2015; vonHoldt et al., 2010). Cao et al. (2021) found that introgression of wild sheep into domesticated sheep contributed to improving the ability of domestic sheep to resist pneumonia. Although many adaptive introgressions are related to environmental challenges, human-mediated genetic

introgression among pig populations is well documented in commercial breeding. Chinese indigenous pigs, such as Meishan, have contributed to European pig breeds since the nineteenth century (Rothschild et al., 1996; Zhao et al., 2018). Bosse et al. (2015) found that beneficial haplotype allele within the *AHR* locus from Asian pigs was favored via artificial selection after introgression and has played an important role in increasing sow fertility in European pigs.

China is one of the pig domestication centers (Larson et al., 2005). Recent studies suggest that there are possibly multiple domestication locations in China, and it is likely that there has been genetic exchange between different breeds since initial domestication and breed formation (Cai et al., 2019; Jin et al., 2012; Larson et al., 2010; Xiang et al., 2017; Yang et al., 2011). DHB (DaHuaBai) is a local breed in Guangdong province with a

---

distinct high body weight compared with other breeds in South China. The exact origin of the DHB breed has not yet been molecularly resolved. According to the book *Animal Genetics Resources in China*: *Pigs* (Wang et al., 2011), the establishment of the DHB breed may have involved human migration events from the north to the south about 1000 years ago and the immigrants (AD 907–967) to the HeQu river area of Lianzhou, Guangdong province, from Hunan Province, and the immigrants (AD 1127–1279) to the northern Guangdong Province from Jiangxi Province. DHB pig are suspected to be crosses of pigs from the north brought by the immigrants to hybridize with the local pigs in the south. This hypothesis is consistent with that the body weight of DHB is significantly higher than that of other pigs in Southern China, but is comparable with that of pigs from Northern China.

Due to the distinct body weight phenotype of the DHB breed and relevant historical documentation of human migration, it is hypothesized that the genome of DHB not only retains the signals of genetic introgression from pigs in the north to pigs in the south, but also it preferentially retains gene variants transmitted from the north that are involved in increased body weight. From this perspective, DHB constitutes an excellent model for studying human-mediated migration and introgression of Chinese indigenous pigs. Unfortunately, although there are as many as 100 more pig breeds recorded in the book of Wang et al. (2011), there is currently no systematic research on the origins and histories of Chinese indigenous pigs at the molecular level, including for DHB.

In this study, we used large-scale genotype and resequencing data to infer the origin and hybrid history of the DHB breed. We reported a possible two-step introgression to produce DHB along with human migration. Genes and variants that potentially affect phenotypes in DHB were identified via introgression and functional analysis, thus contributing to the large body weight of the breed. Together, the DHB breed demonstrates how genomic patterns reflect the history of human breeding efforts in creating a novel composite breed.

2 | MATERIALS AND METHODS

2.1 | SNP filtering and phasing

Chip data used in this study were retrieved from Yang et al. (2017). The data contain genotypes for 61,772 SNPs in 3482 individuals from 146 different pig breeds. We used probeID (https://support.illumina.com/array/array_kits/porcinesnp60_dna_analysis_kit/downloads.html) to convert the original Sscrofa10.2 genome coordinates to coordinates of Sscrofa11.1. We used plink1.9 (Chang et al., 2015) to filter the data (--mind 0.1: filters out all samples with missing call rates exceeding 0.1; --geno 0.1: filters out all variants with missing call rates exceeding 0.1), and bcftools v1.10 (Danecek et al., 2021) to remove duplicate sites. After filtering, the genotyping rate of chip data was 0.9956. BEAGLE5 was used

to phase (Browning & Browning, 2007) and impute (Browning et al., 2018) the entirety of SNP data with default parameters. After phasing and imputing, the complete chip dataset contains 54,323 SNPs and 2,102 individuals, but we only used 53,066 autosomal SNPs and a subset of domestic pigs from China. Semi-wild breeds, including pigs from Tibetan or Yunnan, as well as others without clear body weight records, were also excluded. Finally, we extracted 20 Chinese pig breeds and SBSB (*Sus barbatus*) from the processed data for further analysis.

2.2 | Population structure and phylogenetic analysis

The body weight information and illustrations of Chinese local pig breeds were retrieved from the book by Wang et al. (2011). The IBS distance matrix between individuals was calculated using plink1.9 (--genome; --distance-matrix). The distance matrix was used to build an unrooted neighbor-joining tree by FastME (Lefort et al., 2015); iTol (Letunic & Bork, 2019) was used for tree display. We used plink1.9 to prune the chip data with the parameter of --indep-pairwise 50 10 0.2. Principal component analysis (PCA) for the 20 Chinese breeds was performed using SMARTPCA (Patterson et al., 2006) based on the LD-pruned data. Admixture (Alexander et al., 2009) analysis was performed on the LD-pruned data using the program ADMIXTURE (v1.23), by varying $K = 2$–20. The admixture results were plotted using pophelper R package (Francis, 2017). The relationship between global ancestry estimated from admixture and body weight of pig breeds was assessed by linear regression. Cook's distance was then used to pinpoint the influential data points.

2.3 | *D*-statistic test of genetic introgression between DHB and large-sized breeds

D-statistic analysis was performed using qpDstat (AdmixTools) (Patterson et al., 2012) on DHB, small-sized breeds, large-sized breeds, and *Sus barbatus* as an outgroup using the combination *D* (DHB, small-sized breeds, large-sized breeds, *Sus barbatus*). A positive *D*-statistic with a *Z*-score of >3 indicated gene flow between DHB and large-sized breeds. All eligible combinations and corresponding *D*-statistic values were plotted.

2.4 | Genome-wide introgression detection based on haplotype similarity

Considering that the choice of block sizes may affect results, we determined the haplotype block size by measuring the linkage disequilibrium (LD) of the DHB population. We found the highest fraction of blocks showing average LD of greater than 0.2 in

blocks of five SNPs (Table S2). The genome (53,066 SNPs) was thus divided into 10,619 nonoverlapping blocks with a fixed number of five adjacent SNPs. To detect introgression signals, we adopted the method proposed by Zhang et al. (2019). If extensive introgression occurred from large-sized breeds into DHB, we would observe genomic loci of DHB exhibiting higher haplotype similarity to large-sized breeds than to small-sized breeds. Haplotype similarity was evaluated by cumulative Chi-squared statistics between DHB versus large-sized breeds and DHB versus small-sized breeds:

$$\chi^2 = \sum_{i=1}^{n} \frac{(A_i - T_i)^2}{T_i}$$

where A_i and T_i are observed, and expected blocks count for cell i. Overall, the higher the similarity between the two populations, the lower the Chi-square values. We then constructed $\Delta\chi^2$ as a Chi-square between DHB and large-sized breeds minus a Chi-square between DHB and small-sized breeds:

$$\Delta x^2 = \chi^2_{\text{DHB vs. Small}} - \chi^2_{\text{DHB vs. Large}}$$

To assess the significance, we performed permutation tests for each block and reported the blocks with a p value of ≤0.001 (Zhang et al., 2019). Each large-sized breed was tested separately with each of the four small-sized breeds.

To test whether the blocks of introgression from different large-sized breeds to DHB were randomly distributed across the genome, we simulated the data by randomly selecting an equal number of introgression blocks identified for each large-sized breed from the 10,619 blocks and then calculated overlapping blocks for each paired large-sized breed. The process was repeated 1000 times to construct a null distribution. The significance was estimated by comparing the real number against the 97.5% quantile of the null distribution.

2.5 | Selection signal analysis of the introgression regions

An in-house script was used to calculate the $H12$ statistic (Garud & Rosenberg, 2015) for each introgressed block in small-sized pigs and DHB. The formula of the $H12$ statistic is as follows:

$$H12 = (p_1 + p_2)^2 + \sum_{i=3}^{\infty} p_i^2$$

where p_i represents different types of haplotypes, with $\sum_{i=1}^{\infty} p_i = 1$ and $p_1 \geq p_2 \geq p_3 \geq p_4 \geq .. > 0$

Vcftools (v0.1.13) (Danecek et al., 2011) was used to calculate the genome-wide F_{st} at each SNP between DHB and small-sized breeds; the top 5% differentiation sites were considered as under selection. The Fisher's exact test was used to determine whether or not there are significantly more blocks under selection in the introgression regions compared with other regions in the DHB genome.

2.6 | GO enrichment and gene prioritization analysis

The human orthologs of pig genes were retrieved from BioMart (Ensembl) (Howe et al., 2020) for GO enrichment analysis, which was performed using metascape (Zhou et al., 2019) based on the human gene annotation database. We collected a set of 489 ortholog genes associated with human body weight from GWAS studies (https://www.ebi.ac.uk/gwas/efotraits/EFO_0004338) as training genes. Endeavour (Tranchevent et al., 2016) was used to prioritize the introgression genes, given the collected training genes.

2.7 | Construction of the demography model

We used the approximate Bayesian computation approach (ABCToolbox) (Wegmann et al., 2010) to fit a suitable model that can explain the demographic history of the DHB. Before ABC inference, some SNPs were removed by three criteria: first, SNPs located in CDS or within 100 kb distance to CDS; second, SNPs located in CpG islands; third, SNPs subjected to LD pruning by plink using "--indep-pairwise 50 10 0.2" (Frantz et al., 2015). Finally, we obtained 8,004 independent and putatively neutral SNPs for ABC inference. We included population statistics of ALL_H (mean heterozygosity for all populations), ALLSD_H (standard deviation of heterozygosity for all populations), MEAN_H (mean heterozygosity for each population), SD_H (standard deviation of heterozygosity for each population), and PAIRWISE_FST (pairwise F_{st}) in the model. ABCsampler and fastsimcoal26 were used to construct the model and to perform coalescent simulations. The observed and simulated summary statistics were calculated using arlsumstat. For each model, half a million replicates were simulated. ABCestimator was then used to calculate the marginal posterior distribution from 5000 (1%) retained simulations by the general linear model. Observed summary statistics located within the 95% interval of the retained simulated statistics indicate a good model fit. Bayes factor, which is usually used for model selection, was computed by comparing the marginal density estimated from ABCestimator for two models.

2.8 | Whole-genome sequencing analysis

Whole-genome sequencing data of 80 samples from 12 Chinese pig breeds (Table S13) were downloaded from the NCBI SRA database. We sequenced an additional six DHB pig and six SZL pigs (coverage depth: 7X-26X). For each sample, we used the fastp tool (Chen et al., 2018) with default parameters for quality filtering.

We generated gVCF for each sample against the reference genome of Sscrofa11.1 (GenBank accession: GCA_000003025.6) and performed joint calling for all individuals using the GTX-One platform, which is an FPGA-based hardware accelerator for GATK best practices workflow (Bu et al., 2020). We finally filtered out SNPs by setting the parameters of QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, and QUAL < 30 using GATK (4.1.9.0) VariantFiltration (DePristo et al., 2011).

For introgression regions in the DHB genome identified by the chip data, we recorded the genomic coordinates. Vcftools was again used to calculate the genome-wide F_{st} at each SNP between DHB and small-sized breeds. SNPs that had a missing rate of less than 10% and showed the top 1% of differentiation by F_{st} analysis were extracted and annotated by SnpEff (Cingolani et al., 2012).

3 | RESULTS

### 3.1 | Inconsistency of genotype and body weight phenotype of DHB

In this study, we included 338 individuals from 15 Chinese large-sized breeds, four Chinese small-sized breeds, the DHB (DaHuaBai) breed, and SBSB (*Sus barbatus*) as the outgroup (Figure 1, Table S1). All individuals were genotyped using the Illumina Porcine 60k Bead Chip. In total, 53,066 SNPs that passed our quality control were kept for further analyses (see Materials and Methods). Geographically,

LC (LuChuan), BMX (BaMaXiang), CJX (CongJiangXiaoEr), WZS (WuZhiShan), and DHB are from South China. SZL (ShaZiLing), GX (GanXiLiangTouWu), TC (TongCheng), LP (LePing) are from Central China. KL (KeLe), NJ (NeiJiang), RC (RongChang), GU (GuanLing) are from Northwest China. BM (BaMei), HT (HeTao), LWH (LaiWuHei), MZ (MinZhu) are from North China. JQH (JiangQuHai), JH (JinHua) and MS (MeiShan) are from East China (Figure 1). Except for DHB, the four breeds from the South are all small-sized, with an average body weight of no more than 80 kg (WZS: 27.05 kg, CJX: 39.54 kg, LC: 78.92 kg, and BMX: 34.4 kg). In contrast, pig breeds from the other regions, except for South China, are large-sized with an average body weight almost above 100 kg (e.g., GX: 131.49 kg, JH: 138.35 kg, and SZL: 117.01 kg) (Figure 2a). The body weight of DHB falls in the large-sized group, with an average of 110.73kg. Since the large-sized breeds were geographically located to the north of the small-sized breeds, in this study, we called the large-sized breeds northern breeds and the small-sized breeds southern breeds, unless otherwise specified.

Using the chip SNP data, we conducted PCA (principal component analysis) on the 20 Chinese pig populations (see Materials and Methods; Figure 2b; Figure S1). The PCA clustering result is similar according to the geographical locations. The group on the left contained 15 large-sized breeds (or pig breeds from North China), while the group on the right contained four small-sized breeds (or pig breeds from South China) plus DHB. The neighbor-joining tree based on the IBS distance matrix (Figure 2c) corroborated that, at the genetic level, the small-sized breeds plus the DHB form a monophyletic cluster. In contrast, the large-sized breeds were clustered into another branch.
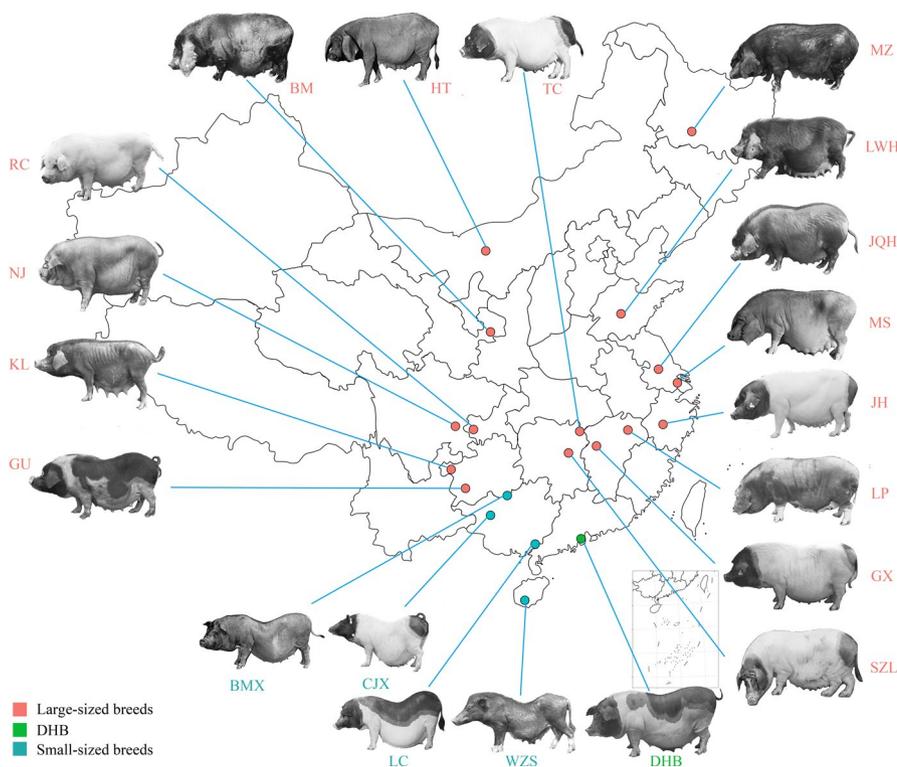
FIGURE 1 Distribution of Chinese local pig breeds included in this study. Large-sized (northern) breeds are abbreviated in red, small-sized (southern) breeds are abbreviated in dark green, and DHB is abbreviated in light green. The geographical locations of pig breeds were retrieved from Yang et al. (2017)

(a)

(b)

(c)
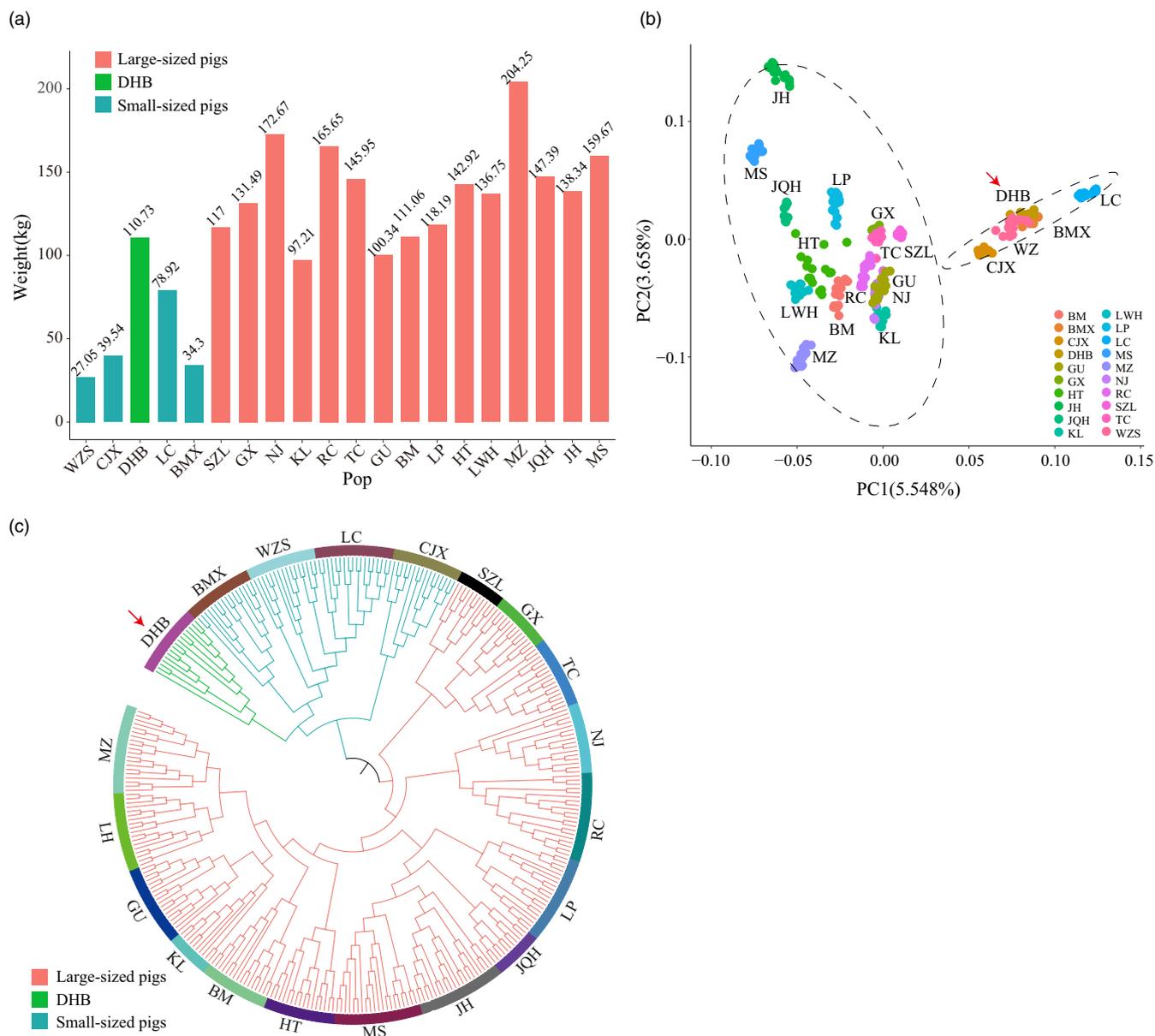
FIGURE 2 DHB was clustered differently in terms of body weight and genetic composition. (a) Body weight of the 20 Chinese local pig breeds. Body weight is the sow weight of the breed collected from the book of Wang et al. (2011). (b) PCA projection of pig breeds using whole-genome SNPs. PC1 and PC2 explained 5.55% and 3.65% genetic variance, respectively. 95% confidence ellipses were drawn for the two groups of large-sized and small-sized breeds. (c) NJ tree based on IBS distance matrix of all pig individuals

In summary, although DHB belongs to the genetic cluster of small-sized breeds from South China, their larger body weight, together with the historical documentation of human migration, suggested potential crossbreeding with the northern pigs.

3.2 | Genomic evidence of hybridization between large-sized breeds from the North and small-sized breeds from the South to produce DHB

Considering the contradictory classification of DHB with respect to body weight and genotype, one possibility is that the genetic makeup of DHB is a mixture derived from large-sized and small-sized breeds,

with the majority of their genetic composition coming from small-sized breeds. We conducted admixture analysis for all 20 Chinese breeds using $K = 2$–20 (Figure S2). When $K = 2$, the genetic composition was mostly reflected by geographical location (Figure 3a). As shown in Figure 3a, the genetic composition of LC (in green) was 99% pure; thus, we used LC as a proxy for unadmixed small-sized breeds, or southern breeds. Similarly, we used MS and JH as proxies for unadmixed large-sized breeds, or northern breeds, of which the genetic composition of MS and JH (in red) is 99% pure. Small-sized breeds such as BMX, WZS, and CJX share 89%, 80%, and 77% ancestral composition with LC, respectively. Other large-sized breeds share ancestral composition with MS and JH from 55% to 93%. When plotting the genetic ancestry composition onto the map, as

(a)

(b)

(c)
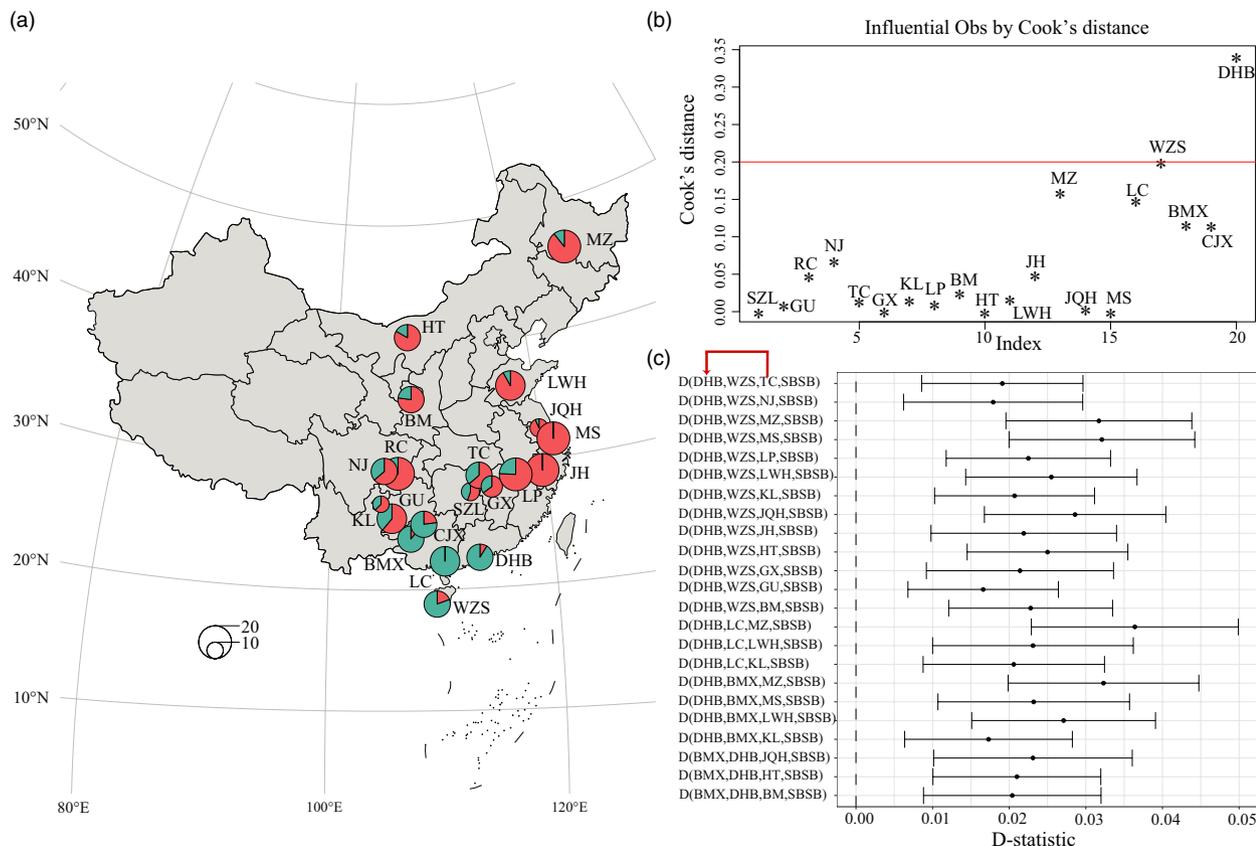
FIGURE 3 Genetic evidence of hybridization of DHB. (a) Genetic structure of pig breeds for $K = 2$. The average ancestral composition of each breed was averaged for all individuals of the breed. The sample size of the breed determined the size of the pie chart. (b) Cook's distance identified DHB as the most influential outliers in the regression model. The red line indicates the threshold of 4/n, where n is the number of breeds. (c) D-statistic analysis with D (DHB, small-sized breeds, large-sized breeds, SBSB) (y-axis) for detection introgression from large-sized breeds into DHB. Four small-sized breeds and 15 large-sized breeds formed 60 D-statistic test combinations

shown in Figure 3a, it was evident that two distinct ancestral groups existed. Pig breeds with mixed genetic composition were geographically located in-between.

Although the body weight of DHB approaches that of large-sized breeds, admixture results showed that the ancestral composition of DHB was highly shared with LC (91%) but less shared with MS and JH (9%). We next regressed the body weight of all breeds against their ancestry ratio shared with LC (admixture analysis $K = 2$). A significant negative correlation was observed, with a Pearson correlation coefficient of −0.748 (95% CI: −0.894 to (−0.457)) (Figure S3). Cook's distance analysis identified the DHB as the most influential outlier affecting the regression model (Cook's distance: 0.3245; Figure 3b), suggesting that the specific genetic material in DHB was highly associated with increased body weight.

To perform a formal test of genetic introgression from large-sized breeds into DHB, D-statistic was used to detect possible introgression in 60 combinations (15 large-sized breeds versus 4 small-sized breeds), using SBSB as an outgroup, and small-sized breeds (either for WZS, BMX, LC, or CJX) and DHB as sibling groups. We found that 23 out of 60 tests supported introgression to DHB with confidence (Figure 3c).

To gain detailed insights into the introgression of large-sized breeds into the DHB, we sought genomic regions in the DHB genome that were genetically closer to large-sized breeds and more distant to small-sized breeds, based on a haplotype similarity approach developed by Zhang et al. (2019) (see Materials and Methods), and we constructed a genome-wide introgression map of DHB (Figure 4; Table S3). Consistent with the D-statistics results, among the introgression regions, GX showed the most extensive introgression into DHB, with summed introgression lengths of 58 MB in total. SZL (48 MB), JH (40 MB), JQH (53 MB), and MS (42 MB) also showed a high degree of introgression to DHB (Figure S4).

3.3 | Model of two-step introgression to DHB

The $H12$ statistic is widely used for measuring haplotype diversity. It calculates the sum of the squares of haplotype frequencies, combining the two most common haplotypes into a single frequency (see Materials and Methods). We found higher values of $H12$ in DHB than in most small-sized breeds in the introgression regions (Table S4). We also calculated genome-wide F_{st} between DHB and small-sized breeds and found significantly more blocks under selection in the
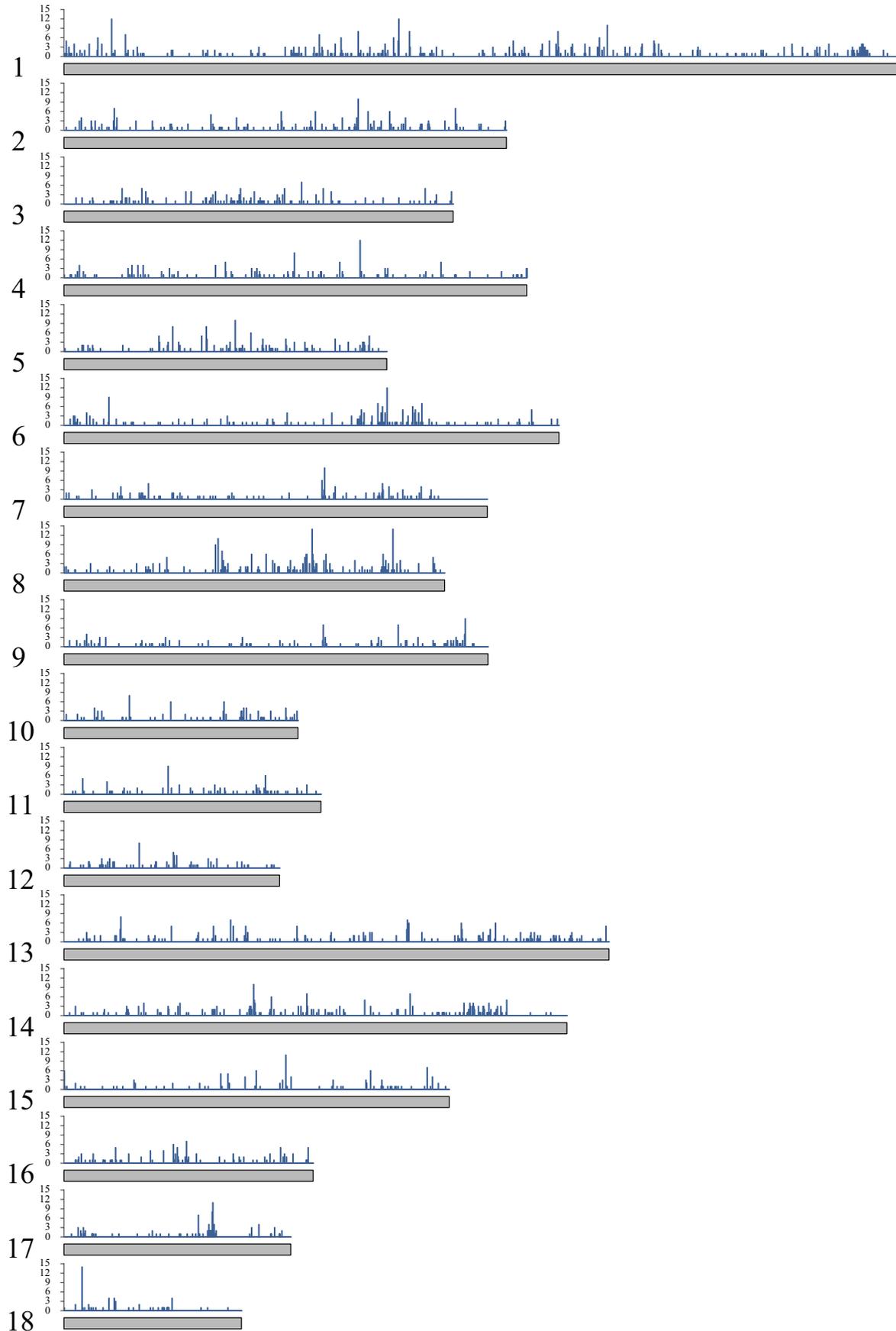
FIGURE 4 Genome-wide introgression map of DHB from large-sized breeds. Only data from chromosomes 1–18 are plotted. The *y*-axis indicates the number of large-sized breeds that had introgressed into DHB

introgression regions, compared with blocks in the nonintrogression regions (Fisher's exact test, $p < 0.01$; Table S5). Together, this evidence indicated that the introgression regions we identified in the DHB genome were under positive selection.

The introgression blocks shared between paired large-sized breeds were highly overlapped compared with random expectation (see Materials and Methods; Figure S5). By investigating the number and length of blocks of introgression from different large-sized breeds, we noticed that the large-sized breeds geographically closer to DHB displayed gradually more introgression length (Figure S4). This pattern of gradual introgression also excluded the possibility that the regions shared by multiple large-sized breeds were caused by incomplete lineage sorting due to random chance. However, we found that pig breeds with a higher degree of introgression tended to be less constrained when checking the ratio of introgressed blocks under selection (Figure 5a; Table S6). We hypothesized that this was due to a chain-manner introgression from the north to the south, with the transmission of beneficial alleles originally from the north because more extensive and noisy introgressions are expected to happen at the end of the introgression chain. As the GX seemed to be the last step in the introgression route to DHB, we also assessed the overlap of introgression regions between GX and other large-sized breeds that were possibly in the introgression chain (Figure 5b). We observed significantly more overlapped introgressions with GX and these breeds than random expectation (Figure S5, GX, and SZL: 63.97 more blocks, GX and JQH: 60.19 more blocks, GX and MS: 50.93 more blocks, GX and JH: 47.68 more blocks). From a geographical point of view, GX is located in Jiangxi Province and SZL

is located in Hunan Province, and both are geographically closer to DHB. JH, JHQ, and MS are located in Jiangsu and Zhejiang province, which are geographically distant from DHB but are next to SZL and GX. Our observation agreed with the "chain introgression" hypothesis, in which pig breeds from Jiangsu or Zhejiang in the north initially introgressed into pig breeds in Hunan or Jiangxi in the middle, followed by a second wave of introgression to the South that produced DHB.

We next used an approximate Bayesian approach and coalescent simulation to build differentiation models for DHB. With our prior knowledge of ancestral composition and the introgression pattern of DHB, we could design the model more efficiently. In our first model, the Chinese domestic pigs were divided into north and south groups. All four small-sized breeds were grouped into one group, and the top five introgression large-sized breeds (GX, SZL, JH, JQH, and MS) were grouped into the other group. The large-sized breeds from the north and small-sized breeds from the south were differentiated from a common ancestor. The result of the D-statistic showed that there was genetic introgression between large-sized breeds and DHB, but the direction was not determined. Therefore, we designed two models with opposite introgression directions. Model 1 (Figure 6a) was designed based on our findings that DHB is genetically close to the small-sized breeds, while possibly having been introgressed from large-sized breeds. In Model 2 (Figure S6), the group of breeds was consistent with Model 1 (Figure 6a); the difference was in the direction of introgression, from DHB into large-sized breeds. The Bayes factor obtained by ABCestimator was 1.9, which indicated that the model shown in Figure 6a could better fit the observations compared with Model 2 (Figure S6). As shown
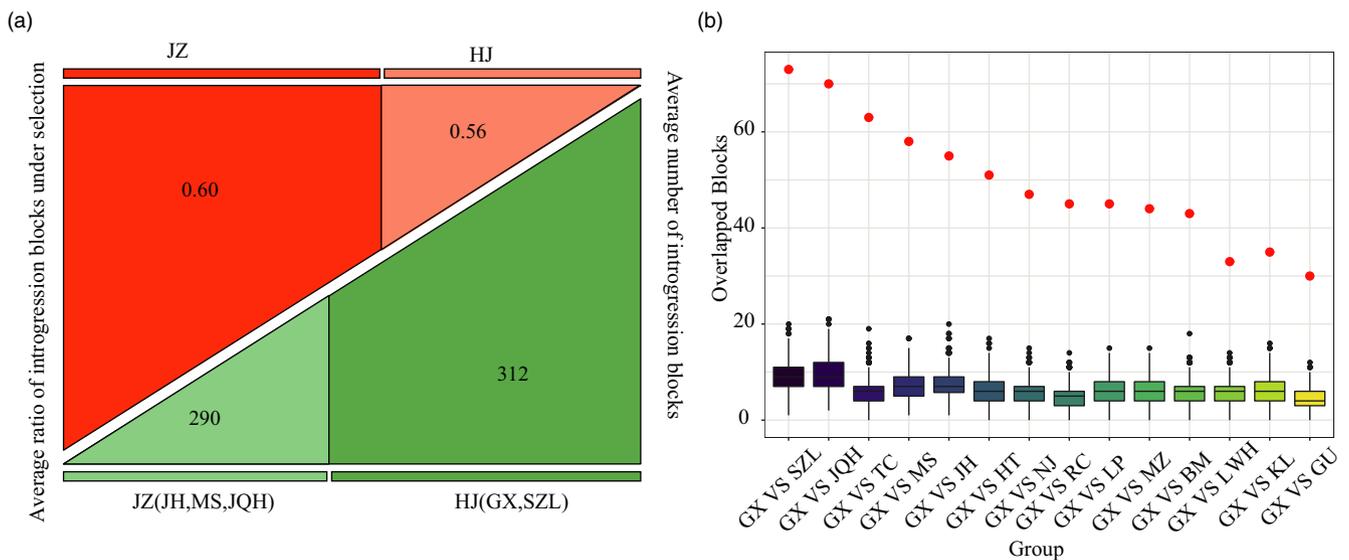
FIGURE 5 Genetic evidence of two-step introgression to DHB. (a) Pig breeds with a higher degree of introgression were less constrained in the introgressed region. We divided the top five pig breeds with highest introgression into two groups according to their geographic location, namely JZ (JH, MS, and JQH) and HJ (GX and SZL), and calculated the average number of introgression blocks (in green) and the ratio of the introgression blocks under selection, as detected by top 5% F_{st} value (in red) for the two groups. (b) Observed and expected overlapped introgression blocks between GX and other large-sized breeds. GX served as an intermediate introgression in the two-step introgression chain. The box diagram represents the random sampling distribution from 1000 simulations; red dots indicate the number of overlapped blocks observed
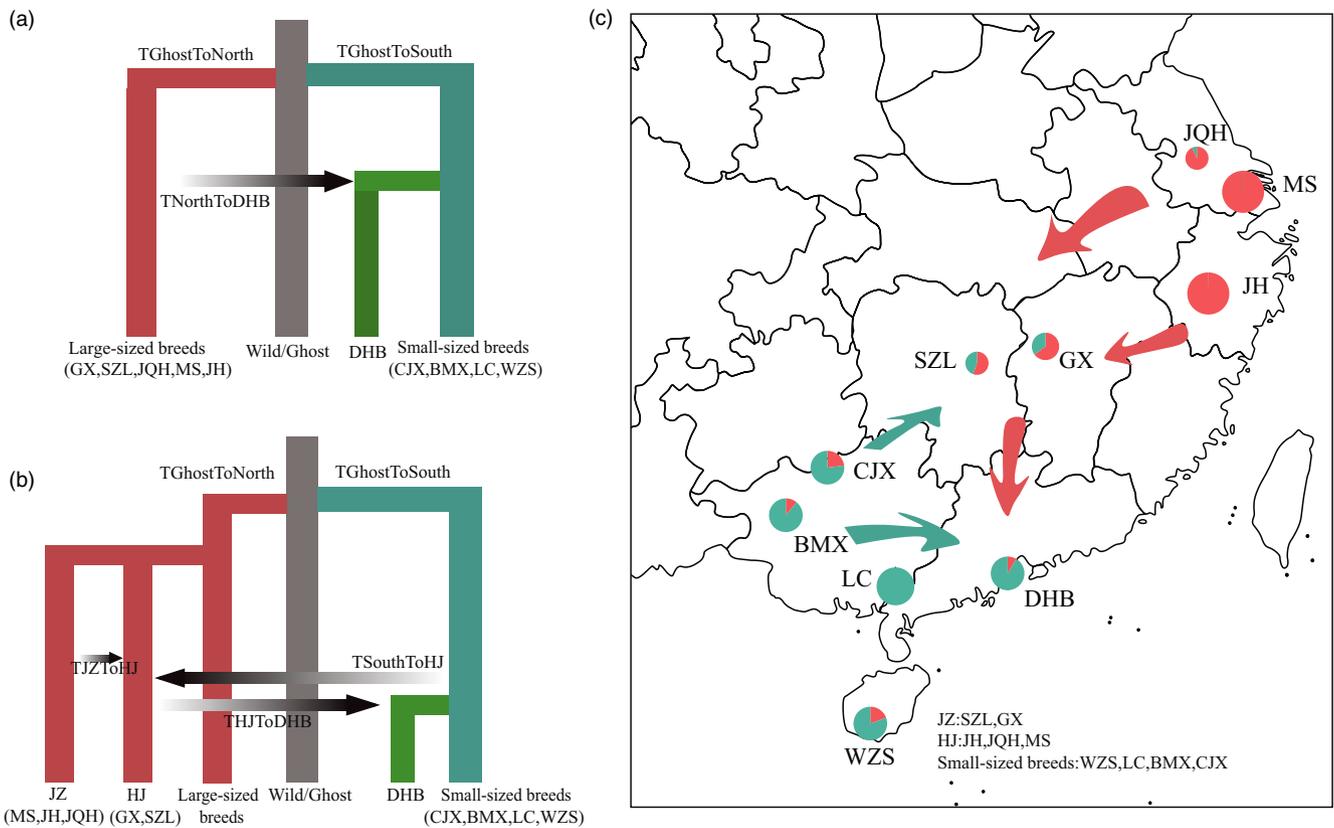
FIGURE 6 Demographic model of DHB breed formation. (a) Simple demographic model of DHB breed formation. In this model, DHB was derived from small-sized breeds from the south with introgression of pigs from the north. (b) A two-step introgression demographic model of DHB breed formation. In this model, GX and SZL served as agents for initial introgression, and a second wave of introgression from GX and SZL migrated to the south and produced DHB. (c) Breeds migration route map. The color and size of the pie chart are the same as in Figure 2a

in Figure 6a, DHB differentiated from the pigs from the south, along with the migration of pigs from the north to the south. Model 1 suggested that the establishment of the DHB breed started from 1558.76 years ago (50% highest posterior density intervals (HPD): 1165.06–1828.83) with 33.86% (50% HPD: 18.80–40.33%) genetic composition initially introgressed from the pigs from the north to the pigs in the south (Table S7; Figure S7a).

Since we suggested a chain-manner introgression as above, and GX and SZL that are geographically closer to the place where the small-sized breeds live appeared to be half admixed. In our third model (Figure 6b), we incorporated the scenario of two-step introgression in which GX and SZL served as agents for genetic introgressions. In this model, we estimated that pigs in Zhejiang migrated to Hunan or Jiangxi about 1200 years ago (50%HPD: 866.05–1449.98), and a second wave of introgression of pigs in Hunan or Jiangxi migrated to the south and produced DHB about 800 years ago (50% HPD: 487.42–1044.56) (Figure 6c; Table S8; Figure S7b).

3.4 | Functional analysis of introgression genes

Due to the evidence from Cook's distance, selection signals along the introgression route, and the increased body weight of DHB, we

expected that the introgressed regions in DHB would harbor important genes controlling body weight. Among the 388 blocks that showed evidence of introgressions from more than three large-sized breeds, 855 genes were retrieved. GO enrichment analysis revealed that the introgression genes were significantly enriched in muscle contraction (GO: 0006936), blood circulation (GO: 0008015), and cardiac muscle contraction (GO: 0060048) (Table S9).

To obtain a thinned set of best candidates, we prioritized the 855 genes using the Endeavour method, given the known genes associated with human body weight as the training set. In total, 503 genes were successfully prioritized (Table S10). Table 1 lists the top 20 genes based on the prioritized *p* value. Among the best candidates, *IGF1R* encodes for an insulin receptor. Previous studies showed that this gene affected body weight in many species, such as mice, Japanese quail, and cattle (Moe et al., 2007; Szewczuk et al., 2013; Todd et al., 2007). Zhao et al. (2018) also found that the *IGF1R* was selected in Meishan pigs, which may be related to the high fertility of this breed. *SRC* is related to food intake in mice (Yang et al., 2019), and *ADCY2* is reported to affect human subcutaneous fat content, both of which are functionally relevant (Daily et al., 2019).

Since the chip data only offer a compromised resolution, to investigate candidate mutations in more detail, we collected whole-genome resequencing data from a public database of 80 Chinese

Gene	Chromosome	Block start	Block end	Number of introgressed populations
IGF1R	1	137008027	137430730	5
SRC	17	40326252	40517532	4
ADCY2	16	74541662	74616199	3
MAP3K7	1	58446382	58544570	3
NRIP1	13	179884140	180051342	3
FOS	7	98401765	98508334	3
RBPJ	8	20011888	20111114	3
IL6ST	16	35149902	35219808	5
CAMK4	2	115928254	116128462	10
DCC	1	102981316	103118451	3
PCM1	17	5557099	5831348	3
PRKG1	14	98108806	98237722	7
PDE4D	16	37970230	38177214	7
ZFHX4	4	59481109	59596863	3
CAMK2A	2	151206576	151328030	3
RABGAP1	1	263806904	264148357	3
PRKCQ	10	64441239	64539431	4
GRIA1	16	69441441	69459646	3
ID2	3	127438077	127529210	3
EDNRB	11	49735789	50089826	3

TABLE 2 SNP sites with high impact

Chromosome	Gene	Position	Ref/Alt	Effect
chr2	LOC100525562	63348710	T/G	stop_gained
chr3	IMMT	58600395	A/G	splice_acceptor_variant&intron_variant
chr13	MRAP	195965524	A/G	splice_acceptor_variant&intron_variant
chr17	PCM1	5704650	A/T	stop_lost

pigs and sequenced six DHB pig and six SZL pigs to scan for functional SNPs in the introgression regions (Materials and Methods). We annotated SNPs in the introgression region with high F_{st} signals between DHB and small-sized breeds using SnpEff and identified four SNPs with high-impact effect (Table 2; Table S11) and 161 SNPs with moderate-impact effect (Table S11). The PCM1 gene, which was reported to associate with early-onset obesity in children, caught our attention (Pettersson et al., 2017). In the chip data, the dominant haplotype (ATACT, chr17:5557099–5831348) in PCM1 was at a high frequency in DHB and large-sized breeds, but at a low frequency in small-sized breeds (Figure 7a). Within the haplotype, the frequency of the allele A at site chr17:5719831 is close to fixed in DHB and most large-sized pigs, while the allele A frequency is relatively low in small-sized pigs (Figure 7a). Of the four SNPs that we identified as high-impact using resequencing data, the reference allele A in position chr17: 5704650, which was at a higher frequency in most large-sized pigs but at a lower frequency in small-sized pigs (Figure 7b), produces an early stop codon, causing the PCM1

protein to be truncated compared with alternative allele T (Materials and Methods). We used the STRING (https://www.string-db.org/) to extract proteins that are functionally associated with PCM1 (Figure 7c). Of the 25 proteins in the protein–protein interaction (PPI) network using a p value threshold of 10^{-16} (PPI enrichment), RSPH3, GABARAPL2, SETDB1, and CHAF1A were also identified as introgressed genes by our haplotype similarity approach (Table S3). Among the genes, GABARAPL2 is associated with total and high-density lipoprotein–cholesterol (Capel et al., 2008), and SETDB1 is an antiadipogenic factor playing a role in obesity (Okamura et al., 2010). These genes, together with other genes with high-impact variants, are worth further experimental validation.

4 | DISCUSSION

Most studies on genetic introgression focus on introgression between highly divergent lineages, such as interspecies hybridization
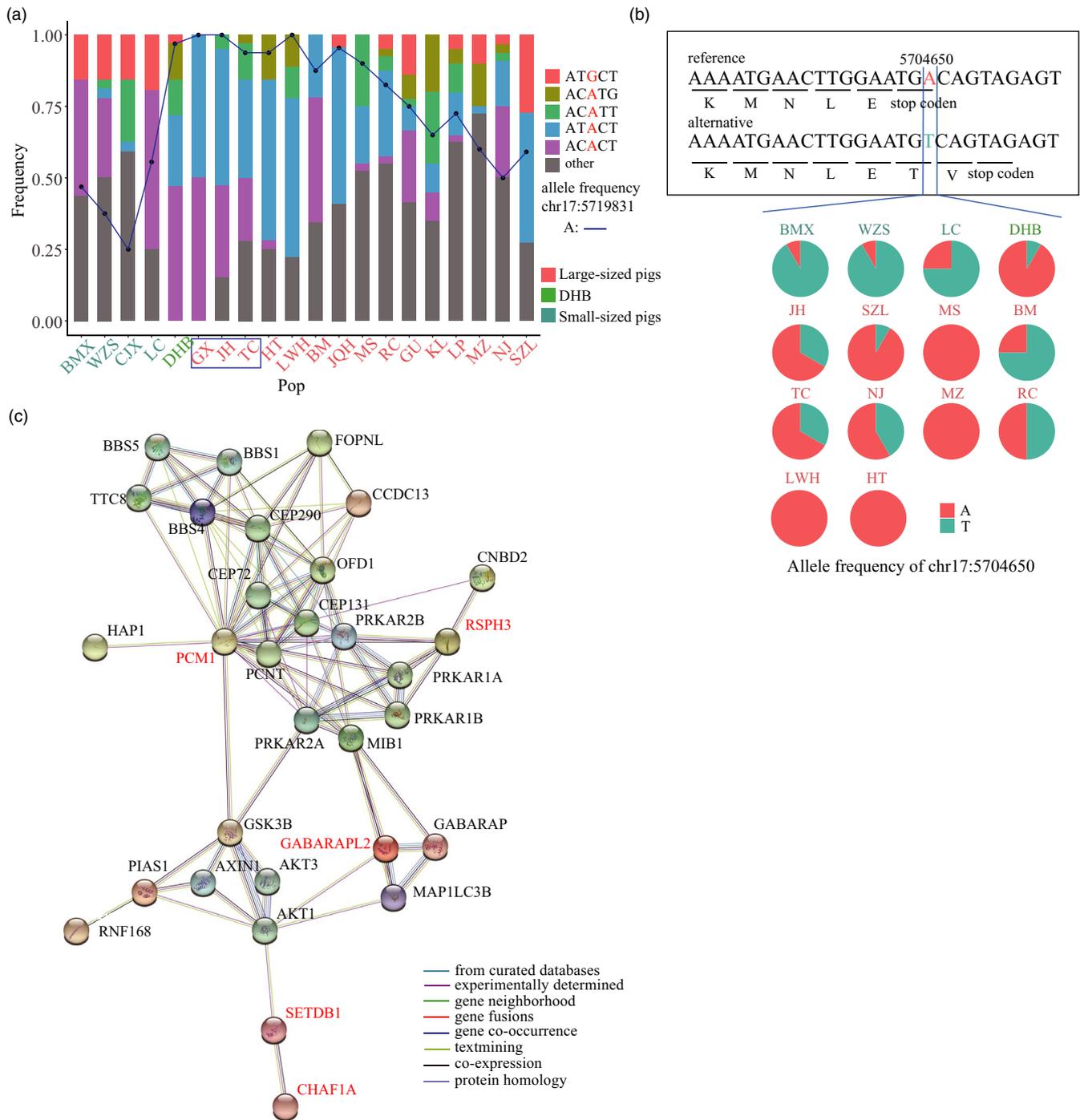
FIGURE 7 Frequency and functional analysis of introgression gene *PCM1*. (a) Multiple haplotypes co-existed in introgressed block chr17:555709-5831348. The frequency of haplotype ATACT was higher in large-sized pigs and DHB but lower in small-sized pigs. The blue box indicates that genetic introgression into DHB had been detected in these breeds. Two alleles (AG) were observed at site 5719831bp on chromosome 17, marked in red in the legend. The black dot and the blue line indicate the frequency of the A allele in all breeds. (b) The alternative allele T on chr17:5704650 introduces two extra amino acids in the protein sequence of gene *PCM1* compared to reference allele A. The frequency of this SNP in each population is represented by a pie chart, in which red refers to allele A and blue refers to alternative T. (c) Interaction network centered by *PCM1*; only the top 25 STRING interactants are shown in the figure. Genes in red indicate that they are identified as introgressed genes in this study

or introgression of wild counterparts into domesticated species. Due to incomplete population differentiation, it is more difficult to quantify introgression between evolutionarily closely related breeds within species. By utilizing the haplotype similarity approach, we

investigated hybridization between domesticated pig breeds on a relatively short timescale. For the first time, at the molecular level, we studied the origin and history of the DHB breed that seem to be the product of hybridization of large-sized and small-sized breeds.

Based on the quantification of introgression of multiple breeds, we proposed a two-step introgression from the north to the south, which is in line with historical records as well as our demographic modeling.

In our demographic model, we inferred that the pigs from Jiangsu and Zhejiang provinces migrated to Hunan and Jiangxi provinces about 1200 years ago (50% HPD: 866.05–1449.98) (AD 552–1135), and a second wave of migration happened about 800 years ago (50% HPD: 487.42–1044.56) (AD 957–1513) from Hunan and Jiangxi to South China (Guangdong and Guangxi province), finally producing the DHB breed. During the second half of the Tang dynasty (AD 755–907), rebellion and uprisings erupted all over the country. The country later broke down into warring states, and short-lived military regimes alternated frequently, referred to as the "Five Dynasties and Ten Kingdoms" (AD 907–967). During this period, the southern part of China was more stable than the northern area; the economic center of gravity shifted southward as well as the northern population. The years AD 960–1279 corresponded to the Song Dynasty in Chinese history. The Mongols conquered the Song Dynasty at the end of the thirteenth century. A large number of refugees migrated from the north to the south with their production materials, such as local domestic pigs. The inferred timing from ABC reconstruction is highly accordant with the migration history in China and also verifies the documentation in the book of *Animal Genetics Resources in China*: *Pigs* (Wang et al., 2011). Our results suggested that population mobility may be one of the reasons for the establishment of a new breed, which happened passively and accompanied major historical events.

Body weight is a polygenic trait controlled by many genes with minor additive effects. To identify genes associated with body weight in animals generally requires an advanced intercross population and a large number of descendants. Since the body weight of DHB is significantly higher than other breeds from the place to live, with a clear history of genetic introgression of DHB, we therefore sought out and narrowed down candidate genes controlling body weight from the identified introgressed regions. It constitutes a cost-efficient approach for gene mapping compared with traditional GWAS. A similar approach was previously employed by Bosse et al. (2015) to locate phenotypically important genes introgressed in European commercial pig breeds.

In most cases, the initial admixture proportion would not be high in the recipient population. Introgressed genetic materials that failed to contribute positively to the fitness of the recipient populations would likely be wiped out by random drifts or negative selection. The introgression signals captured by our haplotype similarity method are thus more likely to represent adaptive introgressions. This is supported by the overrepresentation of overlaps of introgression blocks from multiple large-sized breeds, as well as the observation that introgression regions in DHB are highly divergent to small-sized breeds. The analyses of the introgression route showed that it was more likely that the beneficial genes were transmitted originally from the north since the introgressed genes from the middle-upper area of the introgression route were shown under stronger selections.

By analyzing additional resequencing data, we pinpointed candidate haplotypes and mutations from biologically relevant genes that may have an effect on body weight. We also found that candidate genes were enriched in blood circulation (GO: 0008015) and cardiac muscle contraction (GO: 0060048), suggesting that besides muscle development, body weight may also be related to improved blood circulation. Besides body weight, spine shape also differs between large-sized and small-sized breeds, as illustrated in Figure 1. The spine of large-sized breeds and DHB tends to be straight, while the spine of small-sized breeds appears to be curved. The gene *ACP2*, which controls the thoracic spine shape in mice (Saftig et al., 1997), was found to be introgressed from the large-sized breeds to DHB (Table S3). However, it is yet unclear if the spine shape is associated with body weight or if it is an independent trait under selection in the DHB.

In conclusion, we reported human migration waves reflected in the hybrid genome composition of DHB pig. We also reported candidate genes and mutations that potentially affected body weight in DHB by introgression and functional analysis. Together we provide compelling evidence of the impact of human migration and selection on the formation of this unique Chinese pig breed.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ORCID

Yuzhan Wang https://orcid.org/0000-0003-4463-4075
Mirte Bosse https://orcid.org/0000-0003-2433-2483
Yiqiang Zhao https://orcid.org/0000-0002-0076-3476

REFERENCES

Ackermann, R. R., Arnold, M. L., Baiz, M. D., Cahill, J. A., Cortés-Ortiz, L., Evans, B. J., Grant, B. R., Grant, P. R., Hallgrimsson, B., Humphreys, R. A., Jolly, C. J., Malukiewicz, J., Percival, C. J., Ritzman, T. B., Roos, C., Roseman, C. C., Schroeder, L., Smith, F. H., Warren, K. A., ... Zinner, D. (2019). Hybridization in human evolution: Insights from other organisms. *Evolutionary Anthropology, 28*(4), 189–209. https://doi.org/10.1002/evan.21787

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Bosse, M., Lopes, M. S., Madsen, O., Megens, H.-J., Crooijmans, R. P. M. A., Frantz, L. A. F., Harlizius, B., Bastiaansen, J. W. M., & Groenen, M. A. M. (2015). Artificial selection on introduced Asian haplotypes

shaped the genetic architecture in European commercial pigs. *Proceedings of the Royal Society B-Biological Sciences*, *282*(1821), 20152019. https://doi.org/10.1098/rspb.2015.2019

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, *81*(5), 1084–1097. https://doi.org/10.1086/521987

Bu, L., Wang, Q. I., Gu, W., Yang, R., Zhu, D. I., Song, Z., Liu, X., & Zhao, Y. (2020). Improving read alignment through the generation of alternative reference via iterative strategy. *Scientific Reports*, *10*(1), 18712. https://doi.org/10.1038/s41598-020-74526-7

Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux, Y., & Berthouly-Salazar, C. (2019). Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. *Frontiers in Plant Science*, *10*, 4. https://doi.org/10.3389/fpls.2019.00004

Cai, Y., Quan, J., Gao, C., Ge, Q., Jiao, T., Guo, Y., Zheng, W., & Zhao, S. (2019). Multiple domestication centers revealed by the geographical distribution of Chinese native pigs. *Animals*, *9*(10), 709. https://doi.org/10.3390/ani9100709

Cao, Y. H., Xu, S. S., Shen, M., Chen, Z. H., Gao, L., Lv, F. H., & Li, M. H. (2021). Historical introgression from wild relatives enhanced climatic adaptation any resistance to pneumonia in sheep. *Molecular Biology and Evolution*, *38*(3), 838–855. https://doi.org/10.1093/molbev/msaa236

Capel, F., Viguerie, N., Vega, N., Dejean, S., Arner, P., Klimcakova, E., Martinez, J. A., Saris, W. H. M., Holst, C., Taylor, M., Oppert, J. M., Sørensen, T. I. A., Clément, K., Vidal, H., & Langin, D. (2008). Contribution of energy restriction and macronutrient composition to changes in adipose tissue gene expression during dietary weight-loss programs in obese women. *Journal of Clinical Endocrinology & Metabolism*, *93*(11), 4315–4322. https://doi.org/10.1210/jc.2008-0814

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., Hu, S., Huang, S., Zhang, H., Zheng, Z., Song, W., Ma, Z., Ma, Y., Dang, R., Zhang, Z., Xu, L., Jia, Y., Liu, S., Yue, X., … Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications*, *9*(1), 2337. https://doi.org/10.1038/s41467-018-04737-0

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Daily, J. W., Yang, H. J., Liu, M. L., Kim, M. J., & Park, S. (2019). Subcutaneous fat mass is associated with genetic risk scores related to proinflammatory cytokine signaling and interact with physical activity in middle-aged obese adults. *Nutrition & Metabolism*, *16*(1), 75. https://doi.org/10.1186/s12986-019-0405-0

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Eizirik, E., & Trindade, F. J. (2021). Genetics and evolution of mammalian coat pigmentation. *Annual Review of Animal Biosciences*, *9*(1), 125–148. https://doi.org/10.1146/annurev-animal-022114-110847

Francis, R. M. (2017). POPHELPER: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H.-J., Cagan, A., Bosse, M., Paudel, Y., Crooijmans, R. P. M. A., Larson, G., & Groenen, M. A. M. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature Genetics*, *47*(10), 1141–1148. https://doi.org/10.1038/ng.3394

Garud, N. R., & Rosenberg, N. A. (2015). Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. *Theoretical Population Biology*, *102*, 94–101. https://doi.org/10.1016/j.tpb.2015.04.001

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., … Flicek, P. (2020). Ensembl 2021. *Nucleic Acids Research*, *49*(D1), D884–D891. https://doi.org/10.1093/nar/gkaa942

Jin, L., Zhang, M., Ma, J., Zhang, J., Zhou, C., Liu, Y., Wang, T., Jiang, A.-A., Chen, L., Wang, J., Jiang, Z., Zhu, L. I., Shuai, S., Li, R., Li, M., & Li, X. (2012). Mitochondrial DNA evidence indicates the local origin of domestic pigs in the upstream region of the Yangtze River. *PLoS One*, *7*(12), e51649. https://doi.org/10.1371/journal.pone.0051649

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., & Cooper, A. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, *307*(5715), 1618–1621. https://doi.org/10.1126/science.1106927

Larson, G., Liu, R., Zhao, X., Yuan, J., Fuller, D., Barton, L., Dobney, K., Fan, Q., Gu, Z., Liu, X.-H., Luo, Y., Lv, P., Andersson, L., & Li, N. (2010). Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(17), 7686–7691. https://doi.org/10.1073/pnas.0912264107

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*(10), 2798–2800. https://doi.org/10.1093/molbev/msv150

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Moe, H. H., Shimogiri, T., Kamihiraguma, W., Isobe, H., Kawabe, K., Okamoto, S., Minvielle, F., & Maeda, Y. (2007). Analysis of polymorphisms in the insulin-like growth factor 1 receptor (IGF1R) gene from Japanese quail selected for body weight. *Animal Genetics*, *38*(6), 659–661. https://doi.org/10.1111/j.1365-2052.2007.01653.x

Okamura, M., Inagaki, T., Tanaka, T., & Sakai, J. (2010). Role of histone methylation and demethylation in adipogenesis and obesity. *Organogenesis*, 6(1), 24–32. https://doi.org/10.4161/org.6.1.11121

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pettersson, M., Viljakainen, H., Loid, P., Mustila, T., Pekkinen, M., Armenio, M., Andersson-Assarsson, J. C., Mäkitie, O., & Lindstrand, A. (2017). Copy number variants are enriched in individuals with early-onset obesity and highlight novel pathogenic pathways. *The Journal of Clinical Endocrinology & Metabolism*, 102(8), 3029–3039. https://doi.org/10.1210/jc.2017-00565

Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D. M., Auvil, L., Capitanu, B., Ma, J., Lewin, H. A., Qian, X., Lang, Y., Zhou, R., Wang, L., Wang, K., ... Liu, J. (2012). The yak genome and adaptation to life at high altitude. *Nature Genetics*, 44(8), 946–949. https://doi.org/10.1038/ng.2343

Rothschild, M., Jacobson, C., Vaske, D., Tuggle, C., Wang, L., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren, D., Southwood, O., van der Steen, H., Mileham, A., & Plastow, G. (1996). The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proceedings of the National Academy of Sciences of the United States of America*, 93(1), 201–205. https://doi.org/10.1073/pnas.93.1.201

Saftig, P., Hartmann, D., Lüllmann-Rauch, R., Wolff, J., Evers, M., Köster, A., Hetman, M., von Figura, K., & Peters, C. (1997). Mice deficient in lysosomal acid phosphatase develop lysosomal storage in the kidney and central nervous system. *Journal of Biological Chemistry*, 272(30), 18628–18635. https://doi.org/10.1074/jbc.272.30.18628

Scheu, A., Powell, A., Bollongino, R., Vigne, J.-D., Tresset, A., Çakırlar, C., Benecke, N., & Burger, J. (2015). The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genetics*, 16(1), 54. https://doi.org/10.1186/s12863-015-0203-2

Szewczuk, M., Zych, S., Wójcik, J., & Czerniawska-Piatkowska, E. (2013). Association of two SNPs in the coding region of the insulin-like growth factor 1 receptor (IGF1R) gene with growth-related traits in Angus cattle. *Journal of Applied Genetics*, 54(3), 305–308. https://doi.org/10.1007/s13353-013-0155-z

Todd, B. J., Fraley, G. S., Peck, A. C., Schwartz, G. J., & Etgen, A. M. (2007). Central insulin-like growth factor 1 receptors play distinct roles in the control of reproduction, food intake, and body weight in female rats. *Biology of Reproduction*, 77(3), 492–503. https://doi.org/10.1095/biolreprod.107.060434

Tranchevent, L.-C., Ardeshirdavani, A., ElShal, S., Alcaide, D., Aerts, J., Auboeuf, D., & Moreau, Y. (2016). Candidate gene prioritization with Endeavour. *Nucleic Acids Research*, 44(W1), W117–W121. https://doi.org/10.1093/nar/gkw365

vonHoldt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahloun, A., Geffen, E., Pilot, M., ... Wayne, R. K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, 464(7290), 898–902. https://doi.org/10.1038/nature08837

Wang, L., Wang, A., Wang, L., Li, K., Yang, G., He, R., & Qian, L. (2011). *Animal genetics resources in China: Pigs*. China Agricultural Press.

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11(1), 116. https://doi.org/10.1186/1471-2105-11-116

Xiang, H., Gao, J., Cai, D., Luo, Y., Yu, B., Liu, L., Liu, R., Zhou, H., Chen, X., Dun, W., Wang, X. I., Hofreiter, M., & Zhao, X. (2017). Origin and dispersal of early domestic pigs in northern China. *Scientific Reports*, 7, 5602. https://doi.org/10.1038/s41598-017-06056-8

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R. P. M. A., Zinovieva, N., Schook, L. B., Archibald, A., Gatphayak, K., Knorr, C., Triantafyllidis, A., Alexandri, P., Semiadi, G., Hanotte, O., Dias, D., Dovč, P., Uimari, P., Iacolina, L., Scandura, M., ... Megens, H.-J. (2017). Genome-wide SNP data unveils the globalization of domesticated pigs. *Genetics Selection Evolution*, 49(1), 71. https://doi.org/10.1186/s12711-017-0345-y

Yang, S., Zhang, H., Mao, H., Yan, D., Lu, S., Lian, L., Zhao, G., Yan, Y., Deng, W., Shi, X., Han, S., Li, S., Wang, X., & Gou, X. (2011). The local origin of the Tibetan pig and additional insights into the origin of Asian pigs. *PLoS One*, 6(12), e28215. https://doi.org/10.1371/journal.pone.0028215

Yang, Y., van der Klaauw, A. A., Zhu, L., Cacciottolo, T. M., He, Y., Stadler, L. K. J., Wang, C., Xu, P., Saito, K., Hinton, A., Yan, X., Keogh, J. M., Henning, E., Banton, M. C., Hendricks, A. E., Bochukova, E. G., Mistry, V., Lawler, K. L., Liao, L., ... Xu, Y. (2019). Steroid receptor coactivator-1 modulates the function of Pomc neurons and energy homeostasis. *Nature Communications*, 10, 1718. https://doi.org/10.1038/s41467-019-08737-6

Zhang, C., Lin, D., Wang, Y., Peng, D., Li, H., Fei, J., Chen, K., Yang, N., Hu, X., Zhao, Y., & Li, N. (2019). Widespread introgression in Chinese indigenous chicken breeds from commercial broiler. *Evolutionary Applications*, 12(3), 610–621. https://doi.org/10.1111/eva.12742

Zhao, P., Yu, Y., Feng, W., Du, H., Yu, J., Kang, H., Zheng, X., Wang, Z., Liu, G. E., Ernst, C. W., Ran, X., Wang, J., & Liu, J.-F. (2018). Evidence of evolutionary history and selective sweeps in the genome of Meishan pig reveals its genetic and phenotypic characterization. *Gigascience*, 7(5). https://doi.org/10.1093/gigascience/giy058

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wang, Y., Zhang, C., Peng, Y., Cai, X., Hu, X., Bosse, M., & Zhao, Y. (2022). Whole-genome analysis reveals the hybrid formation of Chinese indigenous DHB pig following human migration. *Evolutionary Applications*, 15, 501–514. https://doi.org/10.1111/eva.13366