1

2 **Genomic analyses provide insights into genetic**
 3 **architecture of three-way crossbred pigs**

4

5 Yu Lin^{1,†}, Qianzi Tang^{1,†}, Yan Li¹, Mengnan He¹, Long Jin¹, Jideng Ma¹, Xun
 6 Wang¹, Keren Long¹, Zhiqing Huang², Xuwei Li¹, Yiren Gu³, Mingzhou Li^{1,*}

7

8 ¹Institute of Animal Genetics and Breeding, College of Animal Science and
 9 Technology, Sichuan Agricultural University, Chengdu 611130, China ²Institute
 10 of Animal Nutrition, Sichuan Agricultural University, Chengdu 611130, China
 11 ³Sichuan Animal Science Academy, Chengdu 610066, China

12

13 *Corresponding author: mingzhou.li@sicau.edu.cn

14 [†]These authors contributed equally to this work.

15

16

17

18

19

20 **Abstract**

21 **Background** : Crossbreeding is effective for improving performance in poultry
 22 and livestock, which is mainly attributed to heterosis. For pork production, a
 23 classic three-way crossbreeding system of Duroc × (Landrace × Yorkshire)
 24 (DLY) is widely used to produce terminal crossbred pigs with stable and
 25 prominent performance. Nonetheless, studies on the transmission of genetic
 26 information and gene expression pattern of DLY have been limited.

27 **Findings** : We analyzed population-scale SNPs based on 30 individuals from
 28 these three purebreds and identified 529.93 K SNPs of breed-of-origin of
 29 alleles. We also applied whole-genome sequencing of ten individuals from a
 30 DLY pig family as well as transcriptome of four representative tissues (adipose,
 31 skeletal muscle, heart, and liver) for six DLY individuals. Based on above, we
 32 identified a large number of high-confidence ASE genes, among which four
 33 ASE genes (*KMO*, *PLIN4*, *POPDC3* and *UGT1A6*) were found to be shared
 34 over all DLY individuals.

35 **Conclusion** : We suggest DLY is a more effective strategy of three-way
 36 crossbreeding among these three purebreds from genetic aspect. We suppose
 37 the numerous breed-of-origin of alleles have close association with improved
 38 performance of crossbred individuals. ASE may also play important roles on
 39 DLY three-way crossbreeding system. Our findings are valuable for
 40 understanding the transmission of genetic information and the gene
 41 expression in DLY three-way crossbreeding and may be used to guide
 42 breeding and production of pigs in the future.

43

44

45

46 **Background**

47 Crossbreeding for animals (especially agricultural livestock and poultry) is a
 48 classic and effective method to utilize heterosis, a phenomenon that progeny
 49 of diverse purebreds exhibits better performance than their parents [1,2,3].
 50 Using crossbreeding, alleles that are specific to different purebreds will be
 51 inherited by the offspring and generate a large number of heterozygous
 52 single-nucleotide polymorphisms (SNPs). These would be suggested to
 53 improve crossbred performance, especially in terms of growth rate,
 54 reproductive performance, production performance (meat, egg, and milk), and
 55 disease resistance [3,4].

56 Pigs (*Sus scrofa*) were domesticated at least ~9,000 years ago and have
 57 been used as a major source of animal protein in the human diet; there are
 58 over 730 distinct pig breeds worldwide [5,6]. The worldwide distribution of pigs
 59 is dominated by six international transboundary commercial pig breeds
 60 originating in Europe, namely, Berkshire, Duroc, Hampshire, Landrace,
 61 Piétrain, and Yorkshire, of which Duroc and Hampshire pigs were developed
 62 mainly in North America [7]. After the long-term practice of tests for the
 63 presence of combinations of abilities, a terminal crossbreeding system with
 64 three pig breeds, namely, Duroc × (Landrace × Yorkshire) (DLY), is generally
 65 used for commercial pork production [8]. Landrace and Yorkshire pigs share
 66 prominent traits for pork production, typically, long carcass length, thin
 67 subcutaneous fat layer, large hams, good mothering ability, and high
 68 muscularity in the carcass [9,10]. Duroc pig is mainly used to enhance the
 69 growth rate and intramuscular fat in this three-way crossbreeding system [9].

The pigs generated by this system exhibit a collection of excellent traits, such as high productivity, rapid growth, and desirable pork quality and pork production [9].

Despite the above background, studies on heterosis in pigs performed to date have been limited compared with those on plants (*Arabidopsis thaliana* [11,12], maize [13], and rice [14,15]), fish [16], birds [17,18], and other mammals [19,20]. To explore the transmission of genetic information in DLY, we analyzed the whole-genome data of 30 individuals from the three pig breeds and identified 529.93 K SNPs of breed-of-origin of alleles that probably contribute to the improved crossbred performance. To explore the gene expression pattern of DLY, we carried out whole-genome sequencing of ten individuals of a DLY pig family, as well as transcriptome data of four representative tissues (including adipose, skeletal muscle, heart and liver). From these, we identified a large number of high-confidence allele-specific expression (ASE) genes, among which four ASE genes (*KMO*, *PLIN4*, *POPDC3* and *UGT1A6*) were found to be shared over all DLY pigs, suggesting that ASEs may play important roles on DLY three-way crossbreeding. Our findings are valuable for understanding the genetic architecture of DLY three-way crossbreeding, which should be beneficial for pig breeding and production.

95

96

97

98 **Results**

99 **Simulation of three-way crossbreeding**

100 Genomic prediction for purebred pigs regarding their crossbred performance
 101 can be based on estimation of the effects of SNPs from purebreds on the
 102 performance of their crossbred offspring [21,22]. For crossbred performance,
 103 SNP effects might be breed-specific due to differences between purebreds in
 104 allele frequency (AF) and linkage disequilibrium (LD) patterns between SNPs
 105 and quantitative trait loci (QTL) [21,22].

106 To estimate the breed-specific effects in terminal crossbred pigs of DLY, we
 107 calculated the probability of heterozygous SNP for each SNP locus of
 108 simulated offspring generated by three kinds of three-way crossbreeding
 109 systems (**Figure 1a**). This was based on the ~12.56 million (M) SNPs of the
 110 publicly available genome sequencing data of 30 pig individuals from the three
 111 purebreds (including 11 Duroc, 9 Landrace, and 10 Yorkshire) with mean
 112 genome coverage of ~13.2x for each individual (**Table S1**). Compared with the
 113 other two kinds of three-way crossbreeding system (YDL and LYD), DLY
 114 exhibited a significant distinct distribution of SNP number (**Figure 1b**, mean
 115 Pearson's $r = -0.99$, $P < 0.01$), which was mainly due to more high-probability
 116 heterozygous SNPs in DLY (~949.05 K SNPs in DLY versus ~439.13 K SNPs
 117 in YDL and ~455.98 K SNPs in LYD when the probability of heterozygous SNP
 118 was higher than 60%; ~14.85 K SNPs in DLY versus 199 SNPs in YDL and
 119 549 SNPs in LYD when the probability of heterozygous SNP was higher than

120 90%).

121 To confirm the accuracy of the probability of heterozygous SNP simulated
 122 by our method, we sequenced a DLY pig family [including two grandparents (a
 123 male Landrace and a female Yorkshire), two parents (a male Duroc and a
 124 female LY crossbred pig), and six offspring (three male and three female DLY
 125 crossbred pigs)] (**Figure S2a**) and identified a total of ~14.12 M SNPs. This
 126 comprehensive collection of variations covered ~74.11% of the SNPs identified
 127 from the three downloaded pig breeds, and consistent with ~98.22% of the
 128 homozygous SNPs and ~98.36% of the heterozygous SNPs identified from
 129 Illumina's Porcine 60K Genotyping Bead-Chip (v.2) for each individual (**Tables**
 130 **S2** and **S3**). The percentage of heterozygous SNPs of each pig individual was
 131 calculated for 20 equal intervals (based on the probability of heterozygous
 132 SNP of simulated DLY offspring, from 0% to 100%, at intervals of 5%) to make
 133 a comparison with the expected percentage of heterozygous SNPs of
 134 simulated DLY offspring (**Figure 1c**). As expected, the distribution was
 135 significantly similar between the six DLY pig individuals and simulated DLY
 136 offspring (mean Pearson's $r = 0.98$, $P < 0.01$), but significantly different
 137 between the other four pig individuals and simulated DLY offspring (mean
 138 Pearson's $r = -0.39$, $P < 0.01$) (**Figure 1c**). This strongly confirmed the
 139 accuracy of the probability of heterozygous SNP simulated by our method.

140 The result presented above was most likely contributed to the smaller
 141 genetic divergence between L and Y (**Figure S1a** and **S1b**), which suggests
 142 that they may share a more recent ancestor. As a result, the simulated
 143 offspring of LY exhibited a significant distinct distribution of SNP number
 144 (**Figure S1d**, mean Pearson's $r = -0.93$, $P < 0.01$) compared with the other
 145 two kinds of two-way crossbreeding (DL and DY, **Figure S1c**), which was
 146 mainly due to fewer high-probability heterozygous SNPs in LY (~641.27 K
 147 SNPs in LY versus ~1.37 M SNPs in DL and ~1.27 M SNPs in YD when the

148 probability of heterozygous SNPs was higher than 60%; 1,211 SNPs in LY
149 versus ~43.62 K SNPs in DL and ~36.83 K SNPs in YD when the probability of
150 heterozygous SNPs was higher than 90%). Taking these findings together, we
151 suggest that DLY is a more effective strategy for three-way crossbreeding
152 among Duroc, Landrace, and Yorkshire breeds as it could generate a large
153 number of stably inherited heterozygous SNPs.

154 Identification of breed-of-origin of alleles

155 To identify the breed-of-origin of alleles that most likely contribute to the
156 prominent performance of crossbred individuals, we separately quantified the
157 genetic differentiation (F_{st}) for each SNP locus between D and L (D-L) as well
158 as D and Y (D-Y). We used an empirical procedure and selected SNP loci
159 simultaneously exhibiting a significant high value of F_{st} (the top 5% right tail,
160 where F_{st} was larger than 0.61) and a significant high value of probability of
161 heterozygous SNP of simulated DLY offspring (the top 5% right tail, where the
162 probability of heterozygous SNP was higher than 0.64) of the empirical
163 distribution (**Figure 2a** and **2b**). Using this approach, we identified ~399.48 K
164 and ~379.21 K SNPs of candidate breed-of-origin of alleles for D-L and D-Y,
165 respectively (**Figure 2a** and **2b**), which constituted a final combined collection
166 of 529.93 K SNPs of breed-of-origin of alleles (**Figure 2c**). As expected, these
167 identified SNPs of breed-of-origin of alleles exhibited a significant higher
168 percentage of heterozygous SNPs (~69.52%) than the genomic background
169 (~47.54%) in the six DLY pig individuals (**Figure 2d**, $P < 1.38 \times 10^{-6}$, paired
170 t -test).

171 Although we observed similar occurrences between breed-of-origin of
172 alleles and the genomic background (**Table S4**, Pearson's $r = 0.99$, $P < 0.01$)
173 for five different genomic elements (i.e., upstream, downstream, intergenic,
174 intronic, and exonic), we also identified a large number of functional categories

(GO and KEGG) for SNPs of breed-of-origin of alleles located in upstream (Table S5), downstream (Table S6), exonic (Table S7), and intronic regions (Table S8), some of which were strongly associated with growth rate (epidermal growth factor receptor signaling pathway, $P = 0.034$; cartilage development, $P = 4.1 \times 10^{-3}$; skeletal system morphogenesis, $P = 4.9 \times 10^{-3}$; regulation of cell proliferation, $P = 0.029$; post-embryonic development, $P = 0.048$), disease resistance (MHC class II receptor activity, $P = 1.1 \times 10^{-3}$; response to antibiotics, $P = 0.029$; antigen processing and presentation of exogenous peptide antigen via MHC class II, $P = 3.8 \times 10^{-3}$; inflammatory mediator regulation of TRP channels, $P = 4.9 \times 10^{-4}$; T-cell costimulation, $P = 0.048$), pork production (lipid catabolic process, $P = 6.9 \times 10^{-3}$; acyl-CoA hydrolase activity, $P = 0.047$; regulation of fat cell differentiation, $P = 0.041$; long-chain fatty-acyl-CoA biosynthetic process, $P = 0.016$; regulation of lipolysis in adipocytes, $P = 0.025$), and reproductive performance (oocyte meiosis, $P = 0.026$; ovarian steroidogenesis, $P = 0.013$; oxytocin signaling pathway, $P = 5.1 \times 10^{-4}$).

To test the influence of the breed-of-origin of alleles on protein-coding genes, we selected 50 nonsynonymous SNPs of breed-of-origin of alleles simultaneously exhibiting the highest values of probability of heterozygous SNP and high-confidence heterozygous SNPs in DLY individuals (five or more heterozygous SNPs for six DLY pig individuals) (Figure 3a). These selected nonsynonymous SNPs correspond to 50 genes that were supposed to have close associations with the improved performance of crossbred individuals (Figure 3a). A representative example of these is *AKAP9* (A-kinase anchoring protein 9), which is a member of structurally diverse proteins that have the common function of binding to the regulatory subunit of protein kinase A (PKA) [23, 24]. The detail information of variant in *AKAP9* was showed in Figure 3b. Mice with *AKAP9* knockout displayed decreased body fat and body weight, hematopoietic abnormalities, and an atypical plasma chemistry profile,

204 suggesting that this gene may be important for fat and body weight
205 development [25,26]. Detailed information for the other genes is presented in
206 **Table S9**. Taken together, our findings provide a valuable resource to further
207 explore the mechanism behind heterosis in DLY three-way crossbreeding.

208 **Allele-specific expression (ASE)**

209 To explore the gene expression pattern during the progress of DLY three-way
210 crossbreeding, we sequenced the transcriptome of four representative tissues,
211 including adipose (energy metabolism), skeletal muscle (body movement and
212 supporter), heart (hematopoiesis and circulatory), and liver (digestion and
213 detoxification) for six DLY individuals (**Table S10**). We detected 2,071–2,261
214 (65.35%–71.35%) heterozygous SNPs in these six individuals overlapping
215 with the 3,169 exonic SNPs of breed-of-origin of alleles; the majority of the
216 overlapping heterozygous SNPs (~96.03%) could be distinguished between
217 the paternal and maternal alleles (**Figure S3a**).

218 We used a strict criteria (P -value < 0.05 and fold change > 2) to identify
219 high-confidence ASEs in adipose (**Figure S4**), skeletal muscle (**Figure S5**),
220 heart (**Figure S6**), and liver (**Figure S7**) and found that ASEs were significantly
221 enriched in breed-of-origin of alleles (**Figure 4e**, mean ASE percentage of
222 ~3.79% for breed-of-origin of alleles versus ~2.04% for the genomic
223 background, $P < 4.82 \times 10^{-3}$, paired t -test). We compared the distribution of
224 ASEs in six DLY pigs and found that four ASEs were shared among all DLY
225 pigs, while the others were individual-specific (**Figure 5a**). We then focused on
226 these four ASEs to explore the influence of ASEs on DLY three-way
227 crossbreeding system.

228 *KMO* (kynurenine 3-monooxygenase) contained a nonsynonymous SNP
229 (position 12,489,146 on chromosome 10), which was detected to express
230 significant more maternal allele (~85.46%) than paternal allele (~14.54%) in

liver of all DLY individuals (**Figure 5b**, $P < 10^{-16}$, paired t -test). *KMO* encodes a mitochondrial outer-membrane protein that catalyzes the hydroxylation of an L-tryptophan metabolite, L-kynurenine, to form L-3-hydroxykynurenine; recent studies have also suggested that kynurenine-3-monooxygenase deficiency is associated with disorders of the brain (e.g., schizophrenia and tic disorders) and of the liver [27,28,29]. *PLIN4* (perilipin 4) contained a synonymous SNP (position 74,324,610 on chromosome 2), which was detected to express significant more paternal allele (~79.27%) than maternal allele (~20.73%) in adipose ($P < 3.5 \times 10^{-10}$, paired t -test) as well as significant more paternal allele (~75.87%) than maternal allele (~24.13%) in skeletal muscle ($P < 2.3 \times 10^{-5}$, paired t -test) of all DLY individuals (**Figure 5b**). *PLIN4* is one of the perilipin members of the PATS family, which share high amino acid sequence similarity and associate with lipid droplets [30]. It may have a close association with insulin resistance and obesity risk due to its protection of lipid droplets from lipases [31]. Studies have revealed close relationships between perilipin genes and body weight and body composition during weight maintenance [32], suggesting that *PLIN4* plays an important role in adipose storage and development for DLY crossbred pigs. *POPDC3* (popeye domain containing 3) contained a synonymous SNP (position 71,380,966 on chromosome 1), which was detected to express significant more paternal allele (~78.01%) than maternal allele (~21.99%) in skeletal muscle ($P < 4.5 \times 10^{-8}$, paired t -test) as well as significant more paternal allele (~76.55%) than maternal allele (~23.45%) in heart ($P < 1.7 \times 10^{-8}$, paired t -test) of all DLY individuals (**Figure 5b**). *POPDC3* encodes a member of the POP family of proteins containing three putative transmembrane domains [33]. Although the function of this gene is not very clear, it is predominantly expressed in skeletal and cardiac muscle, suggesting that this gene has an important function in these tissues [34]. *UGT1A6* (UDP glucuronosyltransferase family 1 member A6) contained a nonsynonymous SNP (position 133,998,238 on chromosome 15), which was detected to specifically express paternal allele in liver of all DLY individuals

(**Figure 5b**). *UGT1A6* is related to the function of the liver due to its function of transforming small lipophilic molecules (steroids, bilirubin, hormones, and drugs) into water-soluble, excretable metabolites [35,36], suggesting this an important role associated with liver. Taking these findings together, we suggest that ASEs may also play important roles in improving the performance of DLY crossbred pigs.

Discussion

In this study, we investigated the DLY three-way crossbreeding system via genomic analyses. Based on ~12.56 M population-scale SNPs (30 individuals from three pig breeds), ~14.12 M family-scale SNPs (including ten pig individuals), and transcriptome data (including adipose, skeletal muscle, heart, and liver), we identified 529,930 breed-of-origin of alleles and numerous high-confidence ASEs, including four shared ASE genes (*KMO*, *PLIN4*, *POPDC3*, and *UGT1A6*). These findings should be valuable for pig breeding and production in the future and provide an excellent model of crossbreeding analysis for many other agricultural animals. We expect that the application of more resources and technologies will reveal more comprehensive and significant discoveries about genetic mechanism of heterosis for pigs in the future.

Materials and methods

Genome sequencing data of pigs

In this study, we downloaded genome sequencing data of 30 pig individuals (including 11 Duroc pigs, 9 Landrace pigs and 10 Yorkshire pigs) from NCBI (Supplementary URLs) (**Table S1**). To produce a tidy dataset for population variants calling, we randomly selected 33 Gb (~13.2x coverage) data for each

286 pig individual (**Table S1**). We also sequenced a DLY pig family including 10 pig
 287 individuals (**Figure S2a**) using Illumina Hiseq 4000 platform. In total, we
 288 generated ~962 Gb paired-end 100-bp (PE100) sequencing data for 10 pig
 289 individuals, with an average genome coverage of ~38.48x (**Table S2**).

290

291 Mapping and SNP calling

292 The genome sequencing data of pigs in this study were first mapped to
 293 reference pig genome (version Sus_scrofa.11.1, Supplementary URLs) using
 294 BWA (version 0.7.8) [37] with default options. We used “MarkDuplicates”
 295 module in package Picard (version 1.48, Supplemental URLs) to remove
 296 duplicated reads. The module “HaplotypeCaller” in Genome Analysis Toolkit
 297 (GATK; version 3.7) [38] was used to identify SNPs, which were next filtered by
 298 following criteria: QUAL < 30.0, QD < 2.0, MQ < 40.0, FS > 60.0. Additionally,
 299 we required genome coverage equal or larger than 4 and SNP integrity equal
 300 or larger than 0.7, which were analyzed by VCFtools (version 1.15) [39]. Finally,
 301 we detected a total of ~12.56 M SNPs of the three downloaded pig breeds
 302 (**Table S1**) and ~14.12 M SNPs of the DLY pig family (**Table S2**). We used
 303 Illumina’s porcine 60K Genotyping Bead- Chip (v.2) to validate the accuracy of
 304 SNP identification for each individual of DLY pig family (**Table S3**). To test
 305 whether the 10 individuals of DLY pig family had a correct sibship, we used
 306 PLINK (version 1.07) [40] to calculate PI_HAT value between two different
 307 individuals using the ~14.12 M family-scale SNPs (**Figure S2b**). The expected
 308 value of PI_HAT is 0.5 between offspring and parents and decreased to 0.25
 309 between grandparents and offspring. If two pig individuals shared the same
 310 parents, the value of PI_HAT is close to 0.5. The extreme value of 0 means
 311 two individuals have no sibship. Based on this method, we verified the correct
 312 sibship of the sequenced DLY pig family (**Figure S2b**).

313 **Calculation of the probability of heterozygous SNP**

314 To calculate the probability of heterozygous SNP of simulated offspring
 315 generated by different crossbreeding systems (including two-way and
 316 three-way crossbreeding), we calculated frequency of three genotypes
 317 (homozygous identical with reference, homozygous distinct from reference,
 318 heterozygous) of the ~12.56 M population-scale SNPs for Duroc (11
 319 individuals), Landrace (9 individuals) and Yorkshire (10 individuals)
 320 respectively. Based on the above results, we calculated the frequency of the
 321 same three genotypes for the simulated offspring generated by three kinds of
 322 two-way crossbreeding systems (DL, LY and YD, **Figure S2c**). Similarly, the
 323 frequency of the same three genotypes of simulated offspring generated by
 324 three kinds of three-way crossbreeding systems (**Figure 1a**) were inferred
 325 using same method.

326 **Population genetic analyses**

327 The phylogenetic tree was inferred using the package treebest (version 1.9.2,
 328 Supplementary URLs) under the p-distances model using the ~12.56 M
 329 population-scale SNPs with a bootstrap of 100. We performed PCA (principal
 330 component analysis) with software GCTA (version 1.91.3beta, Supplementary
 331 URLs) using the ~12.56 M population-scale SNPs. F_{st} value for each SNP
 332 locus was calculated by VCFtools (version 1.15) [39] between D and L as well
 333 as D and Y respectively.

334 **Annotation of breed-of-origin of alleles**

335 We performed annotation for SNPs of breed-of-origin of alleles with
 336 ANNOVAR (version 3.5c) [41]. The current pig gene annotation file (version
 337 Sus_scrofa.11.1.93, Supplementary URLs) was used to build ANNOVAR

338 database of pig and SNPs of breed-of-origin of alleles were classified into 5
339 different genomic elements, including upstream, downstream, exonic, intronic
340 and intergenic regions (**Table S4**).

341 **Function enrichment analyses**

342 We performed function enrichment analyses with DAVID (Database for
343 Annotation, Visualization and Integrated Discovery, version 6.8) [42]. We
344 prefer to use human annotation database rather than pig so we downloaded
345 ortholog gene information between pig and human from Ensembl genome
346 browser website (Supplementary URLs) and translated pig genes to human
347 ortholog genes.

348 **Allele-specific expression (ASE) analysis**

349 To identify allele-specific expression (ASE) during the progress of DLY
350 three-way crossbreeding system, we sequenced 4 representative tissues,
351 including adipose (subcutaneous adipose), skeletal muscle (longissimus dorsi
352 muscle), heart and liver, generating a total of ~111.5 Gb transcriptome data
353 (**Table S8**). We used software vcf2diploid (version 0.2.6a, Supplementary
354 URLs) to create paternal genome and maternal genome based on the ~14.12
355 M family-scale SNPs for the 6 DLY pig individuals respectively. Transcriptome
356 data were mapped to paternal and maternal genome with bowtie (version 1.1.2)
357 [43] and AlleleSeq software (version 1.2a) [44] was used to calculate the
358 *P*-value for imbalanced expression level. We identified a high-confidence ASE
359 simultaneously with significant *P*-value smaller than 0.05 and significant
360 percentage of paternal allele coverage smaller than 0.33 or larger than 0.66
361 (also means fold-change > 2).

362 **Data access**

363 The whole-genome sequencing data of ten pig individuals and transcriptome
364 data the six DLY pig individuals of the DLY family are accessible at NCBI
365 BioProject under accession number of PRJNA507853. The genotyping data of
366 the Illumina's porcine 60K Genotyping Bead- Chip (v.2) have been submitted
367 to the NCBI Gene Expression Omnibus (GEO) under accession number of
368 GSE123327.

369 FUNDING

370 This work was supported by grants from the National Key R & D Program of
371 China (2018YFD0500403), the National Natural Science Foundation of China
372 (31772576, 31522055 and 31802044), the Sichuan Province & Chinese
373 Academy of Science of Science & Technology Cooperation Project
374 (2017JZ0025), the Science & Technology Support Program of Sichuan
375 (2016NYZ0042 and 2017NZDZX0002), the Earmarked Fund for China
376 Agriculture Research System (CARS-35-01A) .

377

378 URLs. NCBI, <https://www.ncbi.nlm.nih.gov/>; Reference pig genome ,[ftp://ftp.en](ftp://ftp.ensembl.org/pub/release92/fasta/susscrofa/dna/Sus_scrofa.Sscrofa11.1.dna.top_level.fa.gz)
379 [sembl.org/pub/release92/fasta/susscrofa/dna/Sus_scrofa.Sscrofa11.1.dna.top](ftp://ftp.ensembl.org/pub/release92/fasta/susscrofa/dna/Sus_scrofa.Sscrofa11.1.dna.top_level.fa.gz)
380 [level.fa.gz](http://picard.sourceforge.net/); Picard, <http://picard.sourceforge.net/>; treebest, [http://treesoft.so](http://treesoft.sourceforge.net/treebest.shtml)
381 [urceforge.net/treebest.shtml](http://www.gcta-ga.org/); GCTA, <http://www.gcta-ga.org/>; Reference pig
382 annotation file, [ftp://ftp.ensembl.org/pub/release93/gtf/sus_scrofa/Sus_scrofa.](ftp://ftp.ensembl.org/pub/release93/gtf/sus_scrofa/Sus_scrofa.Sscrofa11.1.93.gtf.gz)
383 [Sscrofa11.1.93.gtf.gz](http://asia.ensembl.org/index.html); Ensembl genome browser, [http://asia.ensembl.org/i](http://asia.ensembl.org/index.html)
384 [ndex.html](http://alleleseq.gersteinlab.org/tools.html); vcf2diploid, <http://alleleseq.gersteinlab.org/tools.html>.

385

386 References

- 387 1. James A. Birchler, Donald L. Auger, Nicole C. Riddle. In Search of the
388 Molecular Basis of Heterosis. *Plant Cell*. 2003; 15(10) : 2236-9.

- 389 2. Zhang J, Chai J, Luo Z, et al. Meat and nutritional quality comparison of
390 purebred and crossbred pigs. *Anim Sci J*. 2018; 89(1) : 202-210.

- 391 3. Sevillano CA, Vandenplas J, Bastiaansen JW, et al. Empirical
392 determination of breed-of-origin of alleles in three-breed cross pigs. *Genet
393 Sel Evol*. 2016; 48(1) : 55.

- 394 4. Vandenplas J, Calus MP, Sevillano CA, et al. Assigning breed origin to
395 alleles in crossbred animals. *Genet Sel Evol*. 2016; 48(1) : 61.

- 396 5. Choi JS, Lee HJ, Jin SK, et al. Comparison of Carcass Characteristics and
397 Meat Quality between Duroc and Crossbred Pigs. *Korean J Food Sci Anim
398 Resour*. 2014; 34(2) : 238-44.

- 399 6. Chen K, Baxter T, Muir WM, et al. Genetic resources, genome mapping
400 and evolutionary genomics of the pig (*Sus scrofa*). *Int J Biol Sci*. 2007; 3(3) :
401 153-65.

- 402 7. O'Brien PJ, Shen H, Cory CR, et al. Use of a DNA-based test for the
403 mutation associated with porcine stress syndrome (malignant hyperthermia)
404 in 10,000 breeding swine. *J Am Vet Med Assoc*. 1993; 203(6) : 842-51.

- 405 8. Meinert L, Christiansen SC, Kristensen L, et al. Eating quality of pork from
406 pure breeds and DLY studied by focus group research and meat quality
407 analyses. *Meat Sci*. 2008; 80(2) : 304-14.

- 408 9. Choi YS, Lee JK, Jung JT, et al. Comparison of Meat Quality and Fatty Acid
409 Composition of Longissimus Muscles from Purebred Pigs and Three-way
410 Crossbred LYD Pigs. *Korean J Food Sci Anim Resour*. 2016; 36(5) :
411 689-696.

- 412 10. Ruusunen M, Puolanne E, Sevon-Aimonen ML, et al. Carcass and meat
413 quality traits of four different pig crosses. *Meat Sci.* 2012; 90(3) : 543-7.

- 414 11. Yang M, Wang X, Ren D, et al. Genomic architecture of biomass heterosis
415 in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2017; 114(30) : 8101-8106.

- 416 12. Oakley CG, Ågren J, Schemske DW. Heterosis and outbreeding
417 depression in crosses between natural populations of *Arabidopsis thaliana*.
418 *Heredity (Edinb).* 2015; 115(1) : 73-82.

- 419 13. Hu X, Wang H, Diao X, et al. Transcriptome profiling and comparison of
420 maize ear heterosis during the spikelet and floret differentiation stages.
421 *BMC Genomics.* 2016; 17(1) : 959.

- 422 14. Huang X, Yang S, Gong J, et al. Genomic architecture of heterosis for yield
423 traits in rice. *Nature.* 2016; 537(7622) : 629-633.

- 424 15. Huang X, Yang S, Gong J, et al. Genomic analysis of hybrid rice varieties
425 reveals numerous superior alleles that contribute to heterosis. *Nat*
426 *Commun.* 2015; 6 : 6258.

- 427 16. Gao Y, Zhang H, Gao Q, et al. Transcriptome analysis of artificial hybrid
428 pufferfish Jiyan-1 and its parental species: implications for pufferfish
429 heterosis. *PLoS One.* 2013; 8(3) : e58453.

- 430 17. Wang YH, Zhang HY, Zhang CL, et al. The fear gene stathmin alleles
431 generated heterosis on feed efficiency parameters in Peking ducks. *Anim*
432 *Biotechnol.* 2012; 23(4) : 233-40.

- 433 18. Siwendu NA, Norris D, Ngambi JW, et al. Heterosis and combining ability
434 for body weight in a diallel cross of three chicken genotypes. *Trop Anim*
435 *Health Prod.* 2013; 45(4) : 965-70.

- 436 19. Thirstrup JP, Larsen PF, Pertoldi C, et al. Heterosis and genetic variation in
437 the litter size of purebred and crossbred mink. *J Anim Sci.* 2014; 92(12) :
438 5406-16.
- 439 20. Clasen JB, Norberg E, Madsen P, et al. Estimation of genetic parameters
440 and heterosis for longevity in crossbred Danish dairy cattle. *J Dairy Sci.*
441 2017; 100(8) : 6337-6342.
- 442 21. Claudia A. Sevillano, Jeremie Vandenplas, John W. M. Bastiaansen, et al.
443 Genomic evaluation for a three-way crossbreeding system considering
444 breed-of-origin of alleles. *Genet Sel Evol.* 2017; 49(1) : 75.
- 445 22. Sevillano CA, Ten Napel J, Guimarães SEF, et al. Effects of alleles in
446 crossbred pigs estimated for genomic prediction depend on their
447 breed-of-origin. *BMC Genomics.* 2018; 19(1) : 740.
- 448 23. Jo YS, Kim MS, Yoo NJ, et al. Frameshift Mutations of AKAP9 Gene in
449 Gastric and Colorectal Cancers with High Microsatellite Instability. *Pathol*
450 *Oncol Res.* 2016; 22(3) : 587-92.
- 451 24. Deepak Venkatesh, Dolores Mruk, Jan M. Herter, et al. AKAP9, a Regulator
452 of Microtubule Dynamics, Contributes to Blood-Testis Barrier Function. *Am*
453 *J Pathol.* 2016; 186(2) : 270–284.
- 454 25. Skarnes WC, Rosen B, West AP, et al. A conditional knockout resource for
455 the genome-wide study of mouse gene function. *Nature.* 2011; 474(7351) :
456 337-42.
- 457 26. Van der Weyden L, White JK, Adams DJ, et al. The mouse genetics toolkit:
458 revealing function and mechanism. *Genome Biol.* 2011; 12(6) : 224.
- 459 27. Smith JR, Jamie JF, Guillemin GJ, et al. Kynurenine-3-monooxygenase: a

- 460 review of structure, mechanism, and inhibitors. *Drug Discov Today*. 2016;
461 21(2) : 315-24.
- 462 28.Campbell BM, Charych E, Lee AW, et al. Kynurenines in CNS disease:
463 regulation by inflammatory cytokines. *Front Neurosci*. 2014; 8 : 12.
- 464 29.Buness A, Roth A, Herrmann A, et al. Identification of metabolites, clinical
465 chemistry markers and transcripts associated with hepatotoxicity. *PLoS*
466 *One*. 2014; 9(5) : e97249.
- 467 30.Cusano NE, Kiel DP, Demissie S, et al. A Polymorphism in a gene encoding
468 Perilipin 4 is associated with height but not with bone measures in
469 individuals from the Framingham Osteoporosis Study. *Calcif Tissue Int*.
470 2012; 90(2) : 96-107.
- 471 31.Greenberg AS, Egan JJ, Wek SA, et al. Perilipin, a major hormonally
472 regulated adipocyte-specific phosphoprotein associated with the periphery
473 of lipid storage droplets. *J Biol Chem*. 1991; 266(17) : 11341-6.
- 474 32.Soenen S, Mariman EC, Vogels N, et al. Relationship between perilipin
475 gene polymorphisms and body weight and body composition during weight
476 loss and weight maintenance. *Physiol Behav*. 2009; 96(4-5) : 723-8.
- 477 33.Brand T, Schindler R. New kids on the block: The Popeye domain
478 containing (POPDC) protein family acting as a novel class of cAMP effector
479 proteins in striated muscle. *Cell Signal*. 2017; 40 : 156-165.
- 480 34.Andrée B, Hillemann T, Kessler-Icekson G, et al. Isolation and
481 characterization of the novel popeye gene family expressed in skeletal
482 muscle and heart. *Dev Biol*. 2000; 223(2) : 371-82.
- 483 35.Ritter JK, Chen F, Sheen YY, et al. A novel complex locus UGT1 encodes

- 484 human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes
485 with identical carboxyl termini. *J Biol Chem.* 1992; 267(5) : 3257-61.
- 486 36.Shrestha B, Reed JM, Starks PT, et al. Evolution of a major drug
487 metabolizing enzyme defect in the domestic cat and other felidae:
488 phylogenetic timing and the role of hypercarnivory. *PLoS One.* 2011; 6(3) :
489 e18046.
- 490 37.Li H, Durbin R. Fast and accurate short read alignment with
491 Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14) : 1754-60.
- 492 38.McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a
493 MapReduce framework for analyzing next-generation DNA sequencing
494 data. *Genome Res.* 2010; 20(9) :1297-303.
- 495 39.Danecek P, Auton A, Abecasis G, et al. The variant call format and
496 VCFtools. *Bioinformatics.* 2011; 27(15) : 2156-8.
- 497 40.Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for
498 whole-genome association and population-based linkage analyses. *Am J*
499 *Hum Genet.* 2007; 81(3) : 559-75.
- 500 41.Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic
501 variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;
502 38(16) : e164.
- 503 42.Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for
504 Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;
505 4(5) : P3.
- 506 43.Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient
507 alignment of short DNA sequences to the human genome. *Genome Biol.*

508 2009; 10(3) : R25.

509 44. Rozowsky J, Abyzov A, Wang J, et al. AlleleSeq: analysis of allele-specific
510 expression and binding in a network framework. *Mol Syst Biol.* 2011; 7 :
511 522

512

513

514

516

517 **Figure 1. Simulation of three-way crossbreeding and validation of the**
 518 **accuracy of the probability of heterozygous SNP.** (a) Three kinds of
 519 simulated three-way crossbreeding system among Duroc, Landrace, and
 520 Yorkshire. We did not consider the influence of sex chromosomes (X and Y) on
 521 crossbreeding, so each simulated three-way crossbreeding system
 522 represented two forms [e.g., $D \times (L \times Y)$ is equivalent to $D \times (Y \times L)$]. (b) SNP
 523 number statistics of simulated offspring of three kinds of three-way
 524 crossbreeding for 20 equal intervals (based on the probability of a
 525 heterozygous SNP, from 0% to 100%, with intervals of 5%). Pearson's r was
 526 inferred between each of two kinds of three-way crossbreeding. (c)
 527 Percentages of heterozygous SNPs of the sequenced ten pig individuals for 20
 528 equal intervals (based on the probability of heterozygous SNP of simulated
 529 DLY offspring, from 0% to 100%, with intervals of 5%) were calculated.
 530 Pearson's r was inferred for each comparison between simulated DLY
 531 offspring (red line) and sequenced pig individual.

532

533

534

535

536

537

538

Figure 2. Identification of breed-of-origin of alleles. (a) Distribution of probability of heterozygous SNP of simulated DLY offspring and F_{st} value between D and L for each SNP locus. A total of ~399.48 K SNPs of candidate

548 breed-of-origin of alleles were selected that simultaneously showed a
549 significant high probability of heterozygous SNP (the top 5% right tail, where
550 the probability of heterozygous SNP was larger than 0.64) and a significant
551 high F_{st} value (the top 5% right tail, where F_{st} value was larger than 0.61). (b)
552 Distribution of probability of heterozygous SNP of simulated DLY offspring and
553 F_{st} value between D and Y for each SNP locus. A total of ~379.21 K SNPs of
554 candidate breed-of-origin of alleles were selected that simultaneously showed
555 a significant high probability of heterozygous SNP (the top 5% right tail, where
556 the probability of heterozygous SNP was larger than 0.64) and a significant
557 high F_{st} value (the top 5% right tail, where F_{st} value was larger than 0.61). (c)
558 The full collection of 529.93 K SNPs of breed-of-origin of alleles combining
559 candidate breed-of-origin of alleles of D-L and candidate breed-of-origin of
560 alleles of D-Y. (d) Paired t -test was used to compare the percentage of
561 heterozygous SNPs of breed-of-origin of alleles and the genomic background
562 in the six DLY pig individuals.

563

564

565 **Figure 3. Allele distribution and genes related to nonsynonymous SNPs**
 566 **of breed-of-origin of alleles.** (a) We showed the allele distribution among D,

567 L, and Y for 1000 exonic SNPs of breed-of-origin of alleles with the highest
 568 probability of heterozygous SNP and selected 50 nonsynonymous SNPs of
 569 breed-of-origin of alleles that simultaneously showed the highest probability of
 570 heterozygous SNP and high-confidence heterozygous SNPs in DLY pig
 571 individuals (five or more heterozygous SNPs for six DLY pig individuals).
 572 These 50 nonsynonymous SNPs of breed-of-origin alleles correspond to 50
 573 genes. (b) Multiple protein sequence alignment of *AKAP9* among pig and 15
 574 other mammals. The nucleotide on the reference pig genome chr9 :
 575 72,027,468 was identified as a breed-of-origin allele, caused an amino acid
 576 conversion from S to F in the Duroc breed.

577

578

579

580

581

582

583

584

585

586

587

588

589

Figure 4. Identification of allele-specific expression (ASE). (a) ASE identification of adipose for DLY individual 1. (b) ASE identification of skeletal muscle for DLY individual 1. (c) ASE identification of heart for DLY individual 1. (d) ASE identification of liver for DLY individual 1. (e) Paired *t*-test was used for comparison of ASE occurrence between breed-of-origin of alleles and the

596 genomic background for adipose, skeletal muscle, heart, and liver.

597

598

599

600

601 **Figure 5. ASE distribution over six DLY pig individuals.** (a) We analyzed
602 the ASE distribution over the six DLY pig individuals for adipose, skeletal
603 muscle, heart, and liver. All SNP loci were classified into three categories [not
604 heterozygous SNP (gray), not ASE (light blue), and ASE (red)]. (b) Distribution
605 of the percentages of paternal allele (green) and maternal allele (yellow) for
606 the four shared ASEs over six DLY pig individuals.

607