1 **The Y-chromosome clarifies the evolutionary history of *Sus scrofa* by**

2 **large-scale deep genome sequencing**

3

4 Huashui Ai[1,2], Jun Ren[1], Junwu Ma[1], Zhiyan Zhang[1], Wanbo Li[1], Bin Yang[1] and

5 Lusheng Huang[1,2].

6 [1] State Key Laboratory for Swine Genetic Improvement and Production Technology,

7 Jiangxi Agricultural University, 330045, Nanchang, China

8 [2] Corresponding authors: Lusheng Huang (lushenghuang@hotmail.com) or Huashui

9 Ai (aihsh@hotmail.com)

10

13

14

15

1

16 **Abstract**

17 The genetics and evolution of sex chromosomes are largely distinct from autosomes

18 and mitochondrial DNA (mtDNA). The Y chromosome offers unique genetic

19 perspective on male-line inheritance. Here, we uncover novel evolutionary history of

20 *Sus scrofa* based on 205 high-quality genomes from worldwide-distributed different

21 wild boars and domestic pig breeds. We find that only two haplotypes exist in the

22 distal and proximal blocks of at least 7.7 Mb on chromosome Y in pigs across

23 European and Asian continents. And the times of most recent common ancestors

24 (T_{MRCA}) within both haplotypes, approximately 0.14 and 0.10 million years, are far

25 smaller than their divergence time of around 1.07 million years. What's more, the

26 relationship between Sumatran and Eurasian continent *Sus scrofa* is much closer than

27 that we knew before. And surprisingly, European pigs share the same haplotype with

28 many Chinese pigs, which is not consistent with their deep splitting status on

29 autosome and mtDNA. Further analyses show that the haplotype in Chinese pigs was

30 likely introduced from European wild boars via ancient gene flow before pig

31 domestication about 24k years ago. Low mutation rates and no recombination in the

32 distal and proximal blocks on chromosome Y help us detect this male-driven ancient

33 gene flow. Taken together, our results update the knowledge of pig demography and

34 evolution, and might shed insight into the genetics and evolution studies on

35 chromosome Y in other mammals.

36

Introduction

38   The pig (*Sus scrofa*) provides human a main source of animal protein and serves as a

39   biomedical model for human disease (Groenen et al. 2012). Since pigs are usually

40   raised to yield meat and feed a majority of the world population, lots of studies have

41   been conducted to reveal the genetic mechanisms for the complex economic traits in

42   pigs, such as growth and fatness (Andersson et al. 1994), meat quality (Milan et al.

43   2000; Ma et al. 2014) and reproduction (Uimari et al. 2011). On the other hand, pigs

44   have experienced a long period of breeding in yards or fields adjunct to human

45   agricultural societies; they evolved similar eating pattern and dietary structure to

46   human beings (Fang et al. 2012). And pigs share high resemblances with humans in

47   terms of anatomy and physiology. It is generally believed that pigs can be treated as

48   the ideal animal model for studying human microbial infectious diseases (Meurens et

49   al. 2012), or as the promising candidate for development of tissue engineering

50   techniques and xenotransplantation (Eventov-Friedman et al. 2006; Ekser et al. 2012).

51   Therefore pig, as an important domestic animal, is very close related to human.

52   Elucidating pig evolutionary history increases our cognition to pigs, provides insights

53   to the development of two foundational functions of pigs, and helps us better

54   understand the demographic history of humans.

55   Currently, pig demography and evolutionary history have been adequately revealed

56   by lots of great works, which mainly included three aspects of zooarchaeological

57   analyses, mtDNA or ancient mtDNA studies, and autosomal genomic researches. It is

58   widely accepted that *Sus scrofa* emerged in Island South East Asia (ISEA) during the

59   climatic fluctuations of the early Pliocene about 3 to 4 million years ago (Mya) and

60   over the past one million year colonized almost the entire Eurasian continent (Frantz

61   et al. 2013; Groenen 2016). On the northern parts of Sumatran Island, one of

62 Southeast Asian islands, a wild boar population is also found, which split from

63 Eurasian wild boar around 1.6 to 2.4 Mya (Frantz et al. 2013; Groenen 2016). The

64 European and Asian wild boar populations diverged around 1 Mya, and North and

65 South Chinese *Sus scrofa* populations separated from each other during the Ionian

66 stage approximately 0.6 Mya (Groenen et al. 2012; Frantz et al. 2013). Pigs were

67 domesticated at least at two locations (Anatonia and China) ~10k years ago (Larson et

68 al. 2005; Larson et al. 2007; Frantz et al. 2013), and gradually formed a variety of

69 breeds in Europe and Asia (Kijas and Andersson 2001; Wang et al. 2011; Ottoni et al.

70 2013). While during and after domestication of pigs, long-term gene flows or

71 hybridization between wild boars and domestic pigs have been evidenced (Giuffra et

72 al. 2000; Frantz et al. 2015). Nowadays Eurasia has the most rich pig resources, and

73 about one third of worldwide pig breeds have adapted to divergent environment of

74 China (Wang et al. 2011). Recently, a complex pattern of admixture and introgression

75 between European and Asian domestic pigs, and African and American feral pigs has

76 been well documented, but most of pigs colonizing the American, African and

77 Australian continents originate from two highly distinct source populations of

78 European and Asian local pigs (Ramirez et al. 2009; White 2011; Noce et al. 2015).

79 These above studies about pig demography and evolution are mainly focused on

80 zooarchaeological evidences and genetic variations on autosomes and mtDNA. Few

81 works were conducted from the perspectives of whole sex chromosomes, especially

82 for the Y-chromosome, possibly due to lacking of enough genomic information. On

83 the X-chromosome, we have previously found an interesting event of possible ancient

84 interspecies introgression by sequencing 69 Chinese local pigs (Ai et al. 2015), which

85 made an important complement to pig evolutionary history. With the development of

86 sequencing technology, 13 de novo assembled pig genomes (Fang et al. 2012;

4

87   Groenen et al. 2012; Li et al. 2013; Vamathevan et al. 2013; Li et al. 2016) were

88   available for public access. But the version of Build 10.2 reference genome was still

89   widely used due to comprehensive gene annotation and rich variants information.

90   More recently, an improved assembly and gene annotation of the pig X Chromosome

91   and a draft assembly of the pig Y Chromosome (VEGA62) were also generated by

92   sequencing BAC and fosmid clones from Duroc animals and incorporating

93   information from optical mapping and fiber-FISH (Skinner et al. 2016). These

94   improved assemblies provide us make a profound survey to the evolutionary history

95   of pigs from the perspectives of sex chromosomes, especially from the unique genetic

96   perspective on male-line inheritance.

97   In the current work, we obtained high-quality whole-genome sequence data of 202

98   pigs from divergent populations and three outgroups including *Phacochoerus*

99   *africanus*, *Sus verrucosus* and *Sus celebensis*. Based on variants of the Y-

100   chromosome, we presented a deep investigation for male-line evolutionary history in

101   pigs from the global, with assist of autosomal and mtDNA information.

102

103

104

105

106

107

108

Results

110 We have previously sequenced 69 Chinese typical indigenous pigs to reveal the

111 genetic basis for porcine local climate adaptation and possible ancient interspecies

112 introgression on the X-chromosome (Ai et al. 2015). Here we increase 104 pigs for

113 next-generation sequencing, download public data of 39 pigs and 3 outgroups, and

114 combine our previous samples; we make a high-quality dataset of 205 individuals

115 (**Fig. 1A, Supplemental Table S1**). All data was mapped into the *Sus scrofa*

116 reference genome (build 10.2) using BWA (Li and Durbin 2009). A total of

117 48,745,075 SNPs were identified in the 205 genomes (**Supplemental Table S2**) using

118 Platypus (Rimmer et al. 2014) under the criterion of minor allele frequencies (MAF)

119 greater than 0.001 and call rates greater than 80%. All male data was remapped into

120 the Y-chromosome (VEGA62) using Bowtie2 (Langmead and Salzberg 2012). On the

121 new version of chromosome Y, a total of 49,103 SNPs were detected in the males

122 using Platypus under the above criterion of $MAF > 0.001$ and call rates $> 80\%$. These

123 SNPs were used for subsequent demographic and evolutionary analyses.

124 **Demographic parameters inferred by autosomal data**

125 We first constructed a neighbor-joining (NJ) tree for the above 205 animals using

126 47,009,938 SNPs on autosomes. All individuals from the same population gathered

127 together, and European pigs defined a branch clearly separated from Chinese pigs in

128 the NJ tree (**Fig. 1B**). Chinese wild boars clustered into one group, and Chinese

129 domestic pigs were roughly categorized into four groups, corresponding to their

130 geographical distributions: South China, West China, East China and North China

131 (**Fig. 1B**). We then conducted principal components (Price et al. 2006),

132 ADMIXTURE (Alexander et al. 2009) and Treemix (Pickrell and Pritchard 2012)

133 analyses to assess population structure of these animals. European pigs formed an

6

134   independent lineage that was genetically distinct from Chinese pigs in both the

135   ADMIXTURE results (**Fig. 1C**) and the principal components plot (**Fig. 1D**). The

136   Treemix analysis inferred the deep divergence not only between Sumatran pigs and

137   Eurasian pigs, but also between European and Chinese pigs, and North Chinese pigs

138   showed closer relationship with European pigs than other Chinese pigs (**Fig. 1E**). We

139   further inferred the history of population size of Sumatran, Chinese and European

140   pigs using the multiple sequentially Markovian coalescent method (Schiffels and

141   Durbin 2014). Obviously, the demographic profile of Sumatran pigs largely differed

142   from those of Chinese and European pigs. But from the ancient time of the Last

143   Glacial Maximum to the recent 1000 years, the curve of Sumatran pigs almost

144   overlapped with the one of European wild boars. Interestingly, the demographic

145   profiles of Chinese pigs began to diverge from those of European pigs about 0.3 Mya

146   (**Fig. 1F**), and had experienced less severe decline in population size during the Last

147   Glacial Maximum, which is consistent with the previous reports (Groenen et al. 2012;

148   Frantz et al. 2013). Notably, all the test pigs had experienced a bottleneck effect from

149   ~ 5000 to 2000 years ago, which was most severe than the past time. We

150   hypothesized that the activities of human domestication or human hunting might

151   contribute to the decease of pig or wild boar across the global.

152   Altogether, these results based on the autosomal data support the previous

153   conclusion that Chinese and European pigs represent two genetically divergent

154   ancestral populations and Sumatran wild boars are largely different with Eurasian pigs

155   (Larson et al. 2005; Groenen et al. 2012; Frantz et al. 2013).

156   **Odd haplotype patterns of both sex chromosomes in Eurasian pigs**

157   We then made a close examination on sex chromosomes using 1,730,532 and 49,103

158   SNPs on chromosomes X and Y, respectively. As expected, the low-recombinant low-

159 mutation-rate region of 52 Mb (from 49 to 101 Mb on Build 10.2 reference genome)

160 on the X chromosome exhibits three different major haplotypes: one found in

161 European pigs and North Chinese wild boars; one found in pigs from South and West

162 China; and the recombinant haplotype between the two haplotypes was observed in

163 North, East and West Chinese domestic pigs (**Supplemental Fig. S1**). This new view

164 of the odd haplotype pattern on the X-chromosome is well accordant with our

165 previous finding (Ai et al. 2015). Unexpectedly, the distal and proximal region on the

166 Y chromosome of at least 7.7 Mb (hereafter referred to as the SSCY region) displayed

167 two different haplotypes in all tested Euroasian pigs (European pigs, n=30; Chinese

168 pigs, n=71; **Fig. 2A**). The SSCY region was defined as the region from 8.9 to 10.6 Mb

169 and from 39.5 to 43.5 Mb on the VEGA62 version of Y chromosome, meanwhile plus

170 a 2-Mb unmapped contig of Y chromosome. The two haplotypes exhibit a deep

171 divergence as reflected by pairwise nucleotide diversity (**Fig. 2B**) and the phylogenic

172 tree (**Fig. 2C**).

173   However to our surprise, the distribution pattern of these two haplotypes on the Y

174 chromosome in Eurasian pigs was different from the haplotype pattern on the X

175 chromosome. One haplotype was exclusively found in pigs across China and hence

176 denoted as Asian haplotype; while the other one was found in European pigs and

177 some Chinese pigs from North, South and West China, therefore designated as

178 Eurasian haplotype (**Fig. 2A**). We further investigated the geographical distribution of

179 the two haplotypes in a larger panel of 426 male pigs from 82 diverse populations

180 across Eurasia and America using six tagging SNPs representing the SSCY region

181 (**Fig. 2D and Supplemental Table S3**). Again, the Asian haplotype was observed

182 only in East Asian pigs including Chinese pigs, Korean wild boars and Russian

183 Primorsky wild boars; and the Eurasian haplotype was found in European and

184   Chinese pigs, as well as American pigs, which were originated from European pigs

185   (**Fig. 2D**).

186   **Demography and Evolutionary history inferred by the Y-chromosome**

187   We estimated the divergence time of the haplotypes on the SSCY region using a strict

188   molecular clock implemented in BEAST (Drummond et al. 2012) under the GTR+Γ+I

189   model. In the Bayesian tree, there is a clear split between European pigs with the

190   Eurasian haplotype and Chinese pigs with the Asian haplotype (**Fig. 2C**). Their

191   divergence time was estimated to be 1.07 million years (**Table 1**). However, the time

192   of most recent common ancestors (T_{MRCA}) of European and Chinese pigs with the

193   Eurasian haplotype and of Chinese pigs with Asian haplotype were estimated to be

194   only about 0.10 and 0.14 million years, which are unusually smaller than their

195   splitting time (**Table 1**). And T_{MRCA} of Chinese pigs with the Eurasian haplotype was

196   inferred to be 24k years with 95% confidence intervals from 16k to 30k years (**Table**

197   **1**). Also surprisedly, Sumatran wild boar showed much closer relationship to Chinese

198   pigs with Asian haplotype with the splitting time of 0.90 million years (**Table 1**),

199   which is much smaller than the time of 2.1 million years as previously reported using

200   autosomal data (Frantz et al. 2013).

201   To learn more about the evolutionary history of chromosome Y, we further

202   calculated nucleotide variability at the genomic level and within the SSCY region in

203   these sequenced male animals. The SSCY region had a significantly lower level of

204   nucleotide diversity in comparison with autosomes (**Fig. 3**), which is not likely caused

205   by sex-biased genetic drift (**Supplemental Table S4 and S5**) and could be due to the

206   absence of recombination and reduced mutation rates within this region. Moreover,

207   nucleotide variability parameters including segregation sites, *theta* and *Pi* values were

208   all lower in European pigs with the Eurasian haplotype than Chinese pigs with the

209   Asian haplotype (**Fig. 3, Supplemental Table S4 and S5**). This could be explained

210   by the fact that European wild boars had suffered a more dramatically decrease of

211   population size than Asian wild boars during the Last Glacial Maximum period (**Fig.**

212   **1F**). Of note, Chinese pigs having the Eurasian haplotype had ~ six-fold lower values

213   of nucleotide variability than European pigs with this haplotype (**Supplemental**

214   **Table S4 and S5**), and selection signal was not observed in these Chinese pigs as

215   indicated by Tajima's D values (**Fig. 3, Supplemental Table S4 and S5**). A

216   reasonable explanation for this observation is that the Eurasian haplotype in Chinese

217   pigs is a European-originated genetic component and has been introgressed into Asian

218   wild boars via an ancient gene flow before domestication approximately 24k years

219   ago. This genetic component contributes to the current gene pool of a proportion of

220   Chinese domestic pigs, possibly via a complicated human-mediated dispersal after pig

221   domestication. Similar to our previous finding of the interspecies introgression on

222   chromosome X (Ai et al. 2015), our ability to detect this ancient male-driven gene

223   flow on chromosome Y is facilitated by the fact that the introgression fragment falls

224   in a region without recombination and low mutation rates and thus can be maintained

225   for a prolonged period. If the introgressed segment had not fallen in such a region, we

226   would likely never have detected the unusual haplotype pattern as recombination and

227   normal mutation rates may quickly degenerate the integrity of introgression fragments.

228   Of note, this evolutionary pattern has not been observed on the Y-chromosome of

229   other mammals like human (Poznik et al. 2016), dog (Shannon et al. 2015) and horse

230   (Wallner et al. 2013).

231

10

232 ## Discussion

233 **Chromosome Y clarifies the evolutionary history of *Sus scrofa***

234 In the present study, we show a previous unknown evolutionary history of the porcine

235 Y chromosome. In our findings, three key points surprise us. First, only two huge

236 different haplotypes were observed in the distal and proximal blocks of Chromosome

237 Y in a diverse panel of Eurasian pigs (**Fig. 2A**). The genomic and genetic analyses

238 reveal that these regions have two special evolutionary features: no recombination and

239 lower mutation rates, which have not been found in sex chromosomes of other

240 common mammals but only in pigs. Previously we found large region on

241 Chromosome X (48 Mb on the Wuzhishan reference X chromosome; 52 Mb on the

242 Build 10.2 version of Duroc reference X chromosome) harbors the characteristics of

243 low recombination and low mutation rates. We speculated that the enrichment of a 6-

244 kb poly(T) core sequence in the region might contribute to low recombination (Ai et

245 al. 2015). But what exactly contribute to low mutation rates on these regions with

246 large size? Still we don't know. This unknown biological mechanism, depressing the

247 nucleotide mutation on these regions, merits for further exploration. Second, from the

248 perspective of chromosome Y, we found that the relationship between Sumatran wild

249 boar and pigs from Eurasian continents is much closer than that inferred by autosomes.

250 Previously, Frantz et al. (2013) have observed the discordant phenomenon of

251 conflicting phylogenetic signal between mtDNA and autosomal chromosomes in

252 Sumatran wild boar. They explain this phenomenon by sunda-shelf admixture. Our

253 results of Chromosome Y confirm the hypothesis and make the whole event of inter-

254 specific gene flow more reasonable and much clearer. Third, many Chinese pigs share

255 the same haplotype of chromosome Y with European pigs, which is not consistent

256 with their deep splitting status on autosome and mtDNA. It is well recognized that

11

257   European and Asian pigs are two geographically and genetically distinct identities

258   (Groenen et al. 2012; Frantz et al. 2013). However, Asian germplasm had been

259   deliberately imported to Europe to improve commercial trait in European breeds

260   during the Industrial Revolution (Bosse et al. 2014), resulting in the fact that ~30% of

261   the genomes of European commercial pigs are derived from Chinese breeds (Groenen

262   et al. 2012). Here we show another interesting feature of *Sus scrofa* demography, i.e.

263   ancient hybridization between European and Asian wild boars provided an importance

264   source of male genetic components for Asian (Chinese) domestic pigs. These results

265   collectively depict a complex demographic pattern of *Sus scrofa*, significantly

266   advance our knowledge of pig evolutionary history.

267   **Male-driven gene flow on Chromosome Y is not likely a recent but an ancient**

268   **event**

269   Given the geographic distance between Europe and Asia, introgression and

270   hybridization between European and Asian domestic pigs were certainly extremely

271   rare before Ferdinand Magellan circumnavigated the globe. More recently, two waves

272   of gene flow between European and Asian domestic pigs have been well documented.

273   One occurred at the onset of Industrial Revolution and early nineteenth centuries.

274   During this period, Asian pigs mainly from South China were introduced into British

275   to improve local breeds, resulting in mosaic modern genomes of British-derived

276   commercial breeds (such as Large White, Landrace and Duroc) that possess ~30% of

277   Asian haplotypes (White 2011; Groenen et al. 2012). Interestingly, the gene flow

278   orientation shifted afterwards. Since the early twentieth centuries, European improved

279   breeds having desirable performance of lean pork production were imported into

280   China via the Russian, British and German colonies. These breeds including Large

281   White, Berkshire, Duroc and Landrace were introgressed into some Chinese

12

282   indigenous breeds like Kele and Licha, leaving hybridization signals in the genomes

283   of these breeds (Ai et al. 2014). This raises one possibility that the gene flow observed

284   on chromosome Y could take place after the twentieth centuries via the introgression

285   of European germplasm into Chinese local pigs. This possibility seems to be

286   supported by the observation of very few segregating sites and small Theta and Pi

287   values in Chinese pigs with Eurasian Y haplotype (**Supplemental Table S4**) and of a

288   short branch of these pigs in the Bayesian tree (**Fig. 2C**). However, we can definitely

289   exclude this possibility based on the following four arguments, therefore supporting

290   the conclusion that the male-driven gene flow on chromosome Y is an ancient event

291   before domestication around 24k years ago.

292   First, nearly all Chinese pig samples used in this study were collected from nucleus

293   herds raised in national conservation farms. These herds are geographically isolated

294   populations. Except for North Chinese domestic pigs, there is a lack of any historical

295   records describing the importation of European breeds into these populations. Indeed,

296   we did not find European ancient lineages in autosomes of most of Chinese domestic

297   pigs in the ADMIXTURE analysis (**Fig. 1C**), which could be detected if there was a

298   recent gene flow between these European and Chinese pigs.

299   Second, we did not observe any evidence of admixture between European and

300   Chinese domestic pigs on chromosome X. All Chinese pigs with the Eurasian SSCY

301   haplotype do not have the European haplotype of 52 Mb on chromosome X

302   (**Supplemental Fig. S1**). If European germplasm were recently introgressed into

303   Chinese domestic pigs during the last century, it is impossible to retain only Eurasian

304   SSCY segments while completely loss the European chromosome X haplotype in all

305   introgressed individuals via genetic drift or artificial selection within a short time

306   interval of ~100 years. Unlike European commercial breeds, no intensive selection for

13

307   specific traits has been documented for Chinese indigenous breeds during the past

308   century.

309   Third, our mtDNA data show that 20 European pigs (males=9, females=11)

310   including White Duroc, Large White, Landrace and Pietrain have mtDNA of Chinese

311   origin (**Supplemental Table S2**). This is most likely attributable to the recent

312   importation of Chinese breeds to Europe during the 1900s. However, all 134 Chinese

313   pigs (males=71, females=63) lack Europe-originated mtDNA (**Supplemental Table**

314   **S2**). This also conflicts with the hypothesis of a recent gene flow between European

315   and Chinese domestic pigs.

316   Forth, we calculated a negative and positive value of Tajima's D for the SSCY

317   haplotype in European pigs (-1.631) and Chinese pigs (2.350) with the Eurasian

318   haplotype, respectively (**Fig. 3**). For a recently introgressed haplotype, a surplus of

319   rare alleles appearing after the introgression usually creates smaller (i.e. more

320   negative) value of Tajima's D. Therefore, it is theoretically expect to obtain a more

321   negative value of D in presumably introgressed Chinese pigs, i.e. those with the

322   Eurasian SSCY haplotype. However, the prediction is contradictory to our

323   observation, thus excluding the possibility of a recent gene flow on SSCY.

324   Together, we conclude that the male-driven gene flow on SSCY is an ancient event,

325   in which hybridization between European and Asian wild boars occurred before

326   domestication (~24k years ago) and provided an important source of male genetic

327   components for Asian (Chinese) domestic pigs. The very few segregating sites and

328   small values of Theta and Pi in Chinese pigs with the Eurasian SSCY haplotype are

329   most likely due to the unusually evolutionary feature of this region: low mutation

330   rates and no recombination.

331

332 **Methods**

333 **Samples and genome sequencing**

334 We sequenced the genomes of 163 Chinese and European pigs, including 24 South

335 Chinese domestic pigs, 33 North Chinese domestic pigs, 36 West Chinese domestic

336 pigs, 33 East Chinese domestic pigs, 6 South Chinese wild boars and 31 European

337 domestic pigs. Of these animals, Chinese pigs were from 16 geographically divergent

338 breeds, European pigs were from 4 commercial breeds (**Supplemental Table S1**), and

339 59 Chinese pigs have been sequenced in our previous study (Ai et al. 2015).

340   The genome sequencing was conducted as previously described (Ai et al. 2015).

341 Briefly, genomic DNA was extracted from ear tissues using a standard phenol-

342 chloroform method, and then sheared into fragments of 200-800 bp according to the

343 Illumina DNA sample preparation protocol. These treated fragments were end-

344 repaired, A-tailed, ligated to paired-end adaptors and PCR amplified with 500 bp (or

345 350 bp) inserts for library construction. Sequencing was performed to generate 100 bp

346 (or 150 bp) paired-end reads on a HiSeq 2000 (or 2500) platform (Illumina) according

347 to the manufacture's standard protocols.

348 **SNP calling**

349 We downloaded the genome sequence data of 39 pigs, one African warthog

350 (*Phacochoerus africanus*), one Java warty pig (*Sus verrucosus*) and one Celebes

351 warty pig (*Sus celebensis*) from the NCBI SRA database

352 (https://www.ncbi.nlm.nih.gov/sra). These data were integrated into the sequence data

353 obtained in this study, resulting in a 205-sample high-quality data set (**Supplemental**

354 **Table S1 and Supplemental Table S2**). Clean reads from all individuals were

355 aligned to the *Sus scrofa* reference genome (build 10.2) using BWA (Li and Durbin

356 2009). The mapped reads were subsequently processed by sorting, indel realignment,

15

357   duplicate marking, and low quality filtering using Picard (http://picard.sourceforge.net)

358   and GATK (McKenna et al. 2010). Sequencing coverage and depth of each sample

359   were calculated using genomecov implemented in Bedtools (Quinlan and Hall 2010).

360   A two-round procedure of SNP calling was performed using Platypus (Rimmer et al.

361   2014). In the first round, SNPs were individually called with default parameters, and

362   high-quality called SNPs were merged together. These merged SNPs were treated as

363   known variants to guide the second-round genotyping for all the samples via Platypus

364   with   parameters   of   "--source=KnownVariants.vcf.gz   --minPosterior=0   --

365   getVariantsFromBAMs=0". All SNPs except those on chromosome Y were filtered

366   with the criterion of MAF > 0.001 and SNP call rates > 80%. For SNPs on the Y-

367   chromosome, only male individuals were explored to call SNPs under the criterion of

368   MAF > 0.001 and call rates > 80%.

369   **Population genetic analysis using autosome data**

370   A total of 47,009,938 qualified SNPs on autosomes were used to calculate genetic

371   distance among all individuals using Plink as previously described (Ai et al. 2013). A

372   neighbor-joining tree was then constructed for all individuals using Neighbor in

373   PHYLIP   v3.69   (Felsenstein   2005)   and   visualized   by   FigTree   software

374   (http://beast.bio.ed.ac.uk/FigTree). Population genetic structure was inferred using the

375   Maximum Likelihood approach implemented in ADMIXTURE v1.20 (Alexander et

376   al. 2009). The ADMIXTURE program was run in an unsupervised manner with a

377   variable number of clusters (K = 2 to 5). Principal component (PC) analysis was

378   conducted using Smartpca in EIGENSOFT v6.0 (Price et al. 2006). To avoid artifacts

379   caused by linkage disequilibrium (LD), we excluded SNPs with $r^2 \geq 0.2$ in the PC

380   analysis.

381   TreeMix (Pickrell and Pritchard 2012) was employed to infer the patterns of

382   historical splits and mixtures among Eurasian pig populations in the context of *Suidae*,

383   with no migration events and 5000 SNPs grouping together in a LD block. We

384   inferred demographic history of Eurasian pigs using the Multiple Sequentially

385   Markovian Coalescent (MSMC) method (Schiffels and Durbin 2014). To ensure the

386   quality of consensus sequence, we only used representative samples (n=40) of high

387   sequencing depth for each geographic population (n=20) with parameters set as

388   follow: "-p 2*2+50*1+1*4+1*6 --fixedRecombination". The generation time (g) was

389   set as 5 years, and a standard mutation rate (μ) of $2.5×10^{-8}$ was used as previously

390   described (Groenen et al. 2012).

391   **Evolutionary history analysis using chromosome Y data**

392   The bowtie2 software (Langmead and Salzberg 2012) was employed to align filtered

393   clean reads from all male individuals to the chromosome Y reference sequence

394   (VEGA62 version) (Skinner et al. 2016) with the parameter of "--no-mixed --no-

395   discordant --no-unal". Then the two-round SNP calling was performed using Platypus

396   (Rimmer et al. 2014) same as for autosomal SNPs. A total of 49,103 SNPs on

397   chromosome Y passed the criterion of MAF > 0.001 and call rates > 80%. Shapeit2

398   (Delaneau et al. 2013) was used to phase these SNPs with the parameters of "-X --

399   burn 14 --prune 16 --main 40". Phased alleles on the distal and proximal blocks were

400   linked together to form sequences, which were then used to reconstruct a phylogenetic

401   tree using BEAST (Tamura et al. 2013) under the GTR+Γ+I model of evolution.

402   Splitting times and 95% highest posterior density intervals in the tree were estimated

403   using a Bayesian Markov chain Monte Carlo method implemented in BEAST

404   (Tamura et al. 2013) with 1,000,000 MCMC samples. The node age of *Sus verucosus*

405   was set to be 4.2 million years (Frantz et al. 2013) as the calibration constraint.

17

406   To investigate global distribution of the haplotypes within the distal and proximal

407   region on chromosome Y, we employed six tagging SNPs representing this region

408   from porcine 60K chip in 426 Eurasian pigs from 82 geographically divergent

409   populations. Pairwise nucleotide differences per site within (d_x) and between (d_{xy})

410   populations were calculated as previously described (Ai et al. 2015). Segregating sites,

411   Theta, Pi and Tajima's D values were calculated for autosomes and the SSCY region

412   in 50 kb windows at a step size of 25 kb using c++ library of libsequence (Thornton

413   2003). A coalescence simulation for the SSCY region was performed in Chinese pigs

414   with the Asian haplotype, Chinese pigs with the Eurasian haplotype and European

415   pigs with the Eurasian haplotype using the ms software (Hudson 2002) under

416   demographic model inferred from the MSMC (Schiffels and Durbin 2014) results

417   without recombination.

Data access

419   The raw sequence reads from this study have been submitted to the NCBI Sequence

420   Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number

421   SUB2302970.

422

427

Author's contributions

429   L.H., J.R. and H.A. designed the study. H.A. and J.R. designed the bioinformatics

430   analysis process. H.A. performed the analyses of bioinformatics, population genetics

431   and evolutionary history. B.Y., J.M., and Z.Z. performed sample collection. W.L. and

432   H.A. performed mapping of raw sequencing reads. H.A., J.R. and L.H. wrote and

433   revised the paper.

Competing interests

435   The authors declare that they have no competing interest.

436

Figure legends

Figure 1. Demographic history of Eurasian pigs inferred using autosomal SNP data.

440   (A) Geographical locations of Eurasian pigs analyzed in this study. EP, European pigs

441   including Creole, Duroc, Iberian, Large White, Mangalica, Landrace, Pietrain and

442   wild boars; ECDP, East Chinese domestic pigs including Erhualian and Jinhua;

443   WCDP, West Chinese domestic pigs including Baoshan, Neijiang and Tibetan pigs in

444   Yunnan and Sichuan provinces; SCDP, South Chinese domestic pigs including

445   Bamaxiang, Luchuan and Wuzhishan; NCDP, North Chinese domestic pigs including

446   Bamei, Hetao, Laiwu and Min; CWB, Chinese wild boars. (B) Neighbor joining

447   phylogenic tree of all sequenced pigs. SWB, Sumatran wild boars. *S. celebensis*

448   (Celebes wild boar), *S. verrucosus* (Java warty pig) and *P. africanus* (African

449   warthlog) were used as outgroups. (C) ADMIXTURE analysis with $K = 2$-5. Colors in

450   each column represent ancestry proportion. (D) Principal component analysis plots

451   based on the first two principal components. (E) Relationships among Eurasian pigs

452   inferred using Treemix. (F) Effective population sizes of Eurasian pigs inferred using

453   MSMC. The period of the Last Glacial Maximum (LGM; ~20,000 years ago) is

454   shaded in grey.

455 **Figure 2. Demographic history of Eurasian pigs based on the data of**

456 **chromosome Y.**

457 (A) The pattern of haplotype on chromosome Y sharing in Eurasian pigs. The

458 haplotypes were reconstructed for each individual using all qualified SNPs on the Y

459 chromosome. Alleles that are identical or different from the ones on the VEGA62

460 reference genome are indicated by orange or blue, respectively. (B) The plots of dx

461 and dxy (the number of pairwise differences per site) statistics in a window size of

462 200 kb on chromosome Y. These statistics were calculated for pigs with the Eurasian

463 or Asian haplotypes. U_Y indicates the unmapped contig on chromosome Y. (C)

464 Bayesian trees of Eurasian pigs constructed using SNP data of the SSCY region and

465 mtDNA, and neighbor-joining tree of these pigs based on autosomal data. Inferred

466 divergence time is shown in Y-axis of the Bayesian trees. The abbreviations of EP,

467 ECDP, NCDP, WCDP, CWB, SCDP and SWB are as in **Figure 1**. *S. verrucosus*

468 (Java warty pig) was set as outgroup. (D) The geographical distribution of the

469 Eurasian (red) and Asian (blue) haplotypes within the proximal and distal region on

470 the Y chromosome (the SSCY region) in Eurasian pigs. The two haplotypes were

471 phased using six tagging SNPs on the porcine 60K Chip (Illumina) within this region.

472 ISEA haplotype (black), the haplotype found in a Sumatran wild boar.

473 **Figure 3. Comparison of nucleotide variability within the proximal and distal**

474 **region on the Y chromosome and on autosomes**.

475 Statistics of segregation sites, theta, Pi and Tajima's D values were calculated in a

476 window size of 50 kb for European pigs with the Eurasian haplotype, Chinese pigs

477 with the Eurasian haplotype and Chinese pigs with the Asian haplotype on the Y

478 chromosome, respectively.

479

20

480 **Table**

481 **Table 1. Time estimates for the T_{MRCA} of phylogenetic nodes of particular**

482 **interest on the SSCY region**

Node	Time estimate of T_{MRCA} (Million years)	95% highest posterior density (HPD) interval
All the *Sus scrofa*	1.070	0.945-1.156
Sumatran wild boar and Chinese pigs with Asian haplotype	0.903	0.778-0.977
All the pigs with Asian haplotype	0.136	0.111-0.152
All the pigs with Eurasian haplotype	0.102	0.087-0.125
Chinese pigs with Eurasian haplotype	0.024	0.016-0.030

483

484 **Supplemental materials**

485 **Supplemental Table S1.**

486 Samples used in this study. See excel file "Supplemental Table S1.xlsx".

487 **Supplemental Table S2.**

488 Sequencing statistics of 205 samples used in this study. See excel file "Supplemental

489 Table S2.xlsx".

490 **Supplemental Table S3.**

491 The geographical distribution of the two haplotypes in a large panel of 426 male pigs

492 from the global. See excel file "Supplemental Table S3.xlsx".

493 **Supplemental Table S4.**

494 Z-test showing that gene drift is not likely a cause of the two haplotypes on

495 chromosome Y in Eurasian pigs. See excel file "Supplemental Table S4.xlsx".

496 **Supplemental Table S5.**

497 Coalescent simulations showing that gene drift is not likely a cause of the two

498 haplotypes on chromosome Y in Eurasian pigs. See excel file "Supplemental Table

499 S5.xlsx".

500 **Supplemental Fig. S1**

501 **The pattern of haplotype on the X chromosome sharing in Eurasian pigs.** The

502 haplotypes were reconstructed for each individual using all qualified SNPs on this

503 chromosome. Alleles that are identical or different from the ones on the Duroc

504 reference genome are indicated by orange or blue, respectively. The abbreviations of

505 EP, NCDP, ECDP, WCDP, SCDP and CWB are as in **Figure 1**.

506

507

508 **References**

509 Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W et

510 al. 2015. Adaptation and possible ancient interspecies introgression in pigs

511 identified by whole-genome sequencing. *Nature genetics* **47**: 217-225.

512 Ai H, Huang L, Ren J. 2013. Genetic diversity, linkage disequilibrium and selection

513 signatures in chinese and Western pigs revealed by genome-wide SNP markers.

514 *PLoS One* **8**: e56001.

515 Ai H, Yang B, Li J, Xie X, Chen H, Ren J. 2014. Population history and genomic

516 signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics* **15**: 834.

517 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry

518 in unrelated individuals. *Genome Res* **19**: 1655-1664.

519 Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K,

520 Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I et al. 1994.

521 Genetic mapping of quantitative trait loci for growth and fatness in pigs.

522   *Science* **263**: 1771-1774.

523 Bosse M, Megens HJ, Frantz LA, Madsen O, Larson G, Paudel Y, Duijvesteijn N,

524   Harlizius B, Hagemeijer Y, Crooijmans RP et al. 2014. Genomic analysis

525   reveals selection for Asian genes in European pigs following human-mediated

526   introgression. *Nat Commun* **5**: 4392.

527 Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for

528   disease and population genetic studies. *Nat Methods* **10**: 5-6.

529 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with

530   BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**: 1969-1973.

531 Ekser B, Ezzelarab M, Hara H, van der Windt DJ, Wijkstrom M, Bottino R, Trucco M,

532   Cooper DK. 2012. Clinical xenotransplantation: the next medical revolution?

533   *Lancet* **379**: 672-683.

534 Eventov-Friedman S, Tchorsh D, Katchman H, Shezen E, Aronovich A, Hecht G,

535   Dekel B, Rechavi G, Blazar BR, Feine I et al. 2006. Embryonic pig pancreatic

536   tissue transplantation for the treatment of diabetes. *PLoS Med* **3**: e215.

537 Fang X, Mou Y, Huang Z, Li Y, Han L, Zhang Y, Feng Y, Chen Y, Jiang X, Zhao W et

538   al. 2012. The sequence and analysis of a Chinese pig genome. *Gigascience* **1**:

539   16.

540 Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) Version 3.6. Distributed

541   by the author, Department of Genome Sciences, University of Washington,

542   Seattle.

543 Frantz LA, Schraiber JG, Madsen O, Megens HJ, Bosse M, Paudel Y, Semiadi G,

544   Meijaard E, Li N, Crooijmans RP. 2013. Genome sequencing reveals fine

545   scale diversification and reticulation history during speciation in Sus. *Genome*

546   *Biology* **14**: 1719-1728.

547   Frantz LA, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, Paudel Y,
548   Crooijmans RP, Larson G, Groenen MA. 2015. Evidence of long-term gene
549   flow and selection during domestication from analyses of Eurasian wild and
550   domestic pig genomes. *Nature genetics* **47**: 1141-1148.

551   Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. 2000. The origin
552   of the domestic pig: independent domestication and subsequent introgression.
553   *Genetics* **154**: 1785-1791.

554   Groenen MA. 2016. A decade of pig genome sequencing: a window on pig
555   domestication and evolution. *Genet Sel Evol* **48**: 23.

556   Groenen MA Archibald AL Uenishi H Tuggle CK Takeuchi Y Rothschild MF Rogel-
557   Gaillard C Park C Milan D Megens HJ et al. 2012. Analyses of pig genomes
558   provide insight into porcine demography and evolution. *Nature* **491**: 393-398.

559   Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of
560   genetic variation. *Bioinformatics* **18**: 337-338.

561   Kijas JM, Andersson L. 2001. A phylogenetic study of the origin of the domestic pig
562   estimated from the near-complete mtDNA genome. *J Mol Evol* **52**: 302-308.

563   Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*
564   *Methods* **9**: 357-359.

565   Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
566   alignment of short DNA sequences to the human genome. *Genome Biol* **10**:
567   R25.

568   Larson G, Cucchi T, Fujita M, Matisoosmith E, Robins J, Anderson A, Rolett B,
569   Spriggs M, Dolman G, Kim TH. 2007. Phylogeny and ancient DNA of Sus
570   provides insights into neolithic expansion in Island Southeast Asia and
571   Oceania. *Proceedings of the National Academy of Sciences* **104**: 4834-4839.

572   Larson G, Dobney K, Albarella U, Fang M, Matisoosmith E, Robins J, Lowden S,

573   Finlayson H, Brand T, Willerslev E. 2005. Worldwide phylogeography of wild

574   boar reveals multiple centers of pig domestication. *Science* **307**: 1618-1621.

575   Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler

576   transform. *Bioinformatics* **25**: 1754-1760.

577   Li M, Chen L, Tian S, Lin Y, Tang Q, Zhou X, Li D, Yeung CK, Che T, Jin L et al.

578   2016. Comprehensive variation discovery and recovery of missing sequence in

579   the   pig   genome   using   multiple   de   novo   assemblies. *Genome   Res*

580   doi:10.1101/gr.207456.116.

581   Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J et al.

582   2013. Genomic analyses identify distinct patterns of selection in domesticated

583   pigs and Tibetan wild boars. *Nature genetics* **45**: 1431-1438.

584   Ma J, Yang J, Zhou L, Ren J, Liu X, Zhang H, Yang B, Zhang Z, Ma H, Xie X et al.

585   2014. A splice mutation in the PHKG1 gene causes high glycogen content and

586   low meat quality in pig skeletal muscle. *PLoS Genet* **10**: e1004710.

587   McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella

588   K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit:

589   a MapReduce framework for analyzing next-generation DNA sequencing data.

590   *Genome Res* **20**: 1297-1303.

591   Meurens F, Summerfield A, Nauwynck H, Saif L, Gerdts V. 2012. The pig: a model

592   for human infectious diseases. *Trends Microbiol* **20**: 50-57.

593   Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, Rogel-Gaillard C, Paul

594   S, Iannuccelli N, Rask L et al. 2000. A mutation in PRKAG3 associated with

595   excess glycogen content in pig skeletal muscle. *Science* **288**: 1248-1251.

596   Noce A, Amills M, Manunza A, Muwanika V, Muhangi D, Aliro T, Mayega J,

597   Ademun R, Sanchez A, Egbhalsaied S et al. 2015. East African pigs have a

598   complex Indian, Far Eastern and Western ancestry. *Anim Genet* **46**: 433-436.

599 Ottoni C, Flink LG, Evin A, Georg C, De Cupere B, Van Neer W, Bartosiewicz L,

600   Linderholm A, Barnett R, Peters J et al. 2013. Pig domestication and human-

601   mediated dispersal in western Eurasia revealed through ancient DNA and

602   geometric morphometrics. *Mol Biol Evol* **30**: 824-832.

603 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from

604   genome-wide allele frequency data. *PLoS Genet* **8**: e1002967.

605 Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q,

606   McCarthy SA, Narechania A, Kashin S et al. 2016. Punctuated bursts in

607   human male demography inferred from 1,244 worldwide Y-chromosome

608   sequences. *Nature genetics* **48**: 593-599.

609 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006.

610   Principal components analysis corrects for stratification in genome-wide

611   association studies. *Nature genetics* **38**: 904-909.

612 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing

613   genomic features. *Bioinformatics* **26**: 841-842.

614 Ramirez O, Ojeda A, Tomas A, Gallardo D, Huang LS, Folch JM, Clop A, Sanchez A,

615   Badaoui B, Hanotte O et al. 2009. Integrating Y-chromosome, mitochondrial,

616   and autosomal data to analyze the origin of pig breeds. *Mol Biol Evol* **26**:

617   2061-2072.

618 Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Consortium WGS, Wilkie AO,

619   McVean G, Lunter G. 2014. Integrating mapping-, assembly- and haplotype-

620   based approaches for calling variants in clinical sequencing applications.

621   *Nature genetics* **46**: 912-918.

622   Schiffels S, Durbin R. 2014. Inferring human population size and separation history

623   from multiple genome sequences. *Nature genetics* **46**: 919-925.

624   Shannon LM, Boyko RH, Castelhano M, Corey E, Hayward JJ, McLean C, White ME,

625   Abi Said M, Anita BA, Bondjengo NI et al. 2015. Genetic structure in village

626   dogs reveals a Central Asian domestication origin. *Proc Natl Acad Sci U S A*

627   **112**: 13639-13644.

628   Skinner BM, Sargent CA, Churcher C, Hunt T, Herrero J, Loveland JE, Dunn M,

629   Louzada S, Fu B, Chow W et al. 2016. The pig X and Y Chromosomes:

630   structure, sequence, and evolution. *Genome Res* **26**: 130-139.

631   Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular

632   Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

633   Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis.

634   *Bioinformatics* **19**: 2325-2327.

635   Uimari P, Sironen A, Sevon-Aimonen ML. 2011. Whole-genome SNP association

636   analysis of reproduction traits in the Finnish Landrace pig breed. *Genet Sel*

637   *Evol* **43**: 42.

638   Vamathevan JJ, Hall MD, Hasan S, Woollard PM, Xu M, Yang Y, Li X, Wang X,

639   Kenny S, Brown JR et al. 2013. Minipig and beagle animal model genomes

640   aid species selection in pharmaceutical discovery and development. *Toxicol*

641   *Appl Pharmacol* **270**: 149-157.

642   Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, Brem G. 2013. Identification of

643   genetic variation on the horse y chromosome and the tracing of male founder

644   lineages in modern breeds. *PLoS One* **8**: e60015.

645   Wang L, Wang A, Wang L, Li K, Yang G, He R, Qian L, Xu N, Huang R, Peng Z et al.

646   2011. Animal Genetic Resourses in China: Pigs. *Beijing: China Agricultural*

647   *Press In Chinese*.

648   White S. 2011. From globalized pig breeds to capitalist pigs: a study in animal

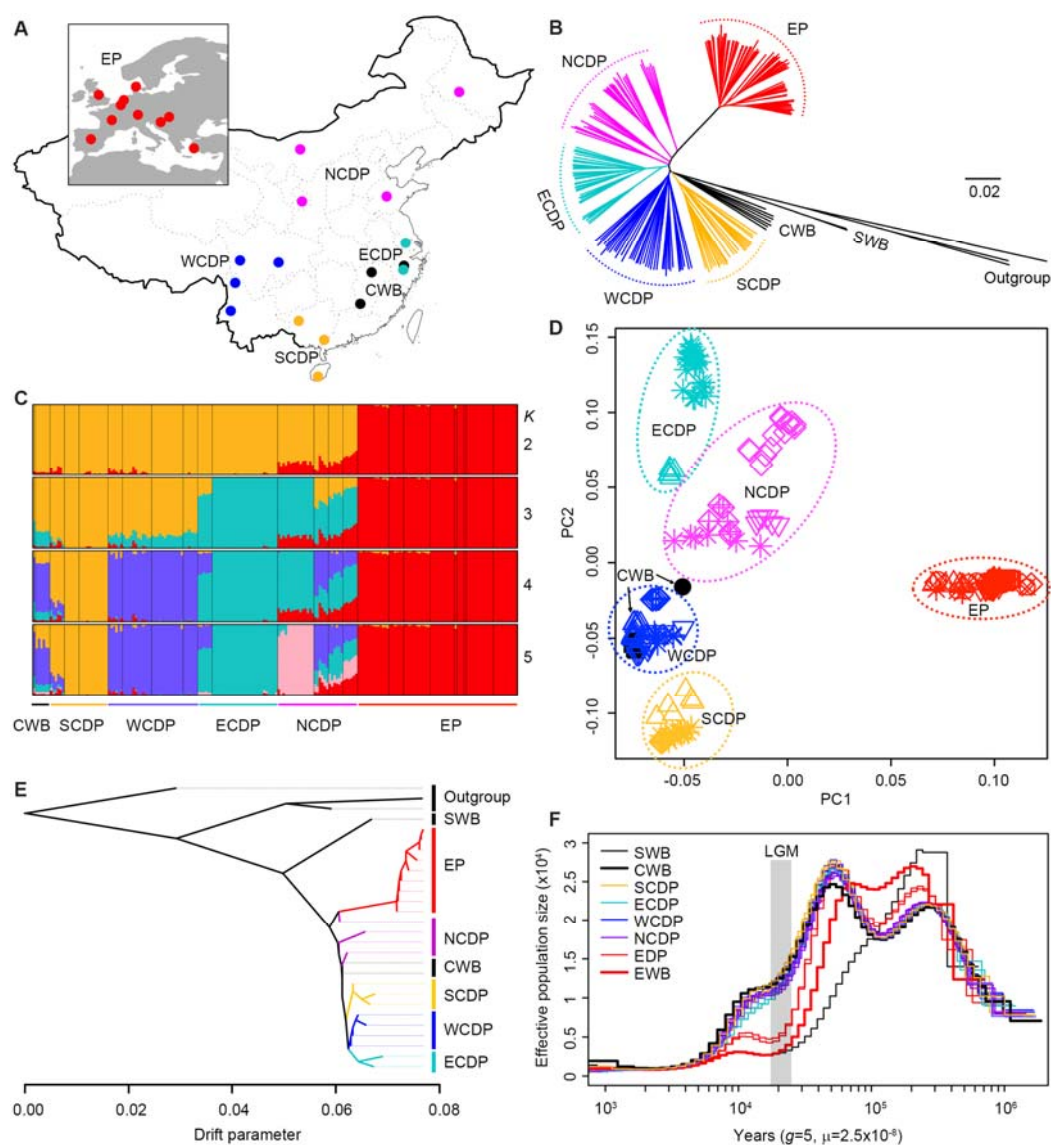649   cultures and evolutionary history. *Environmental History* **16**: 94-120.
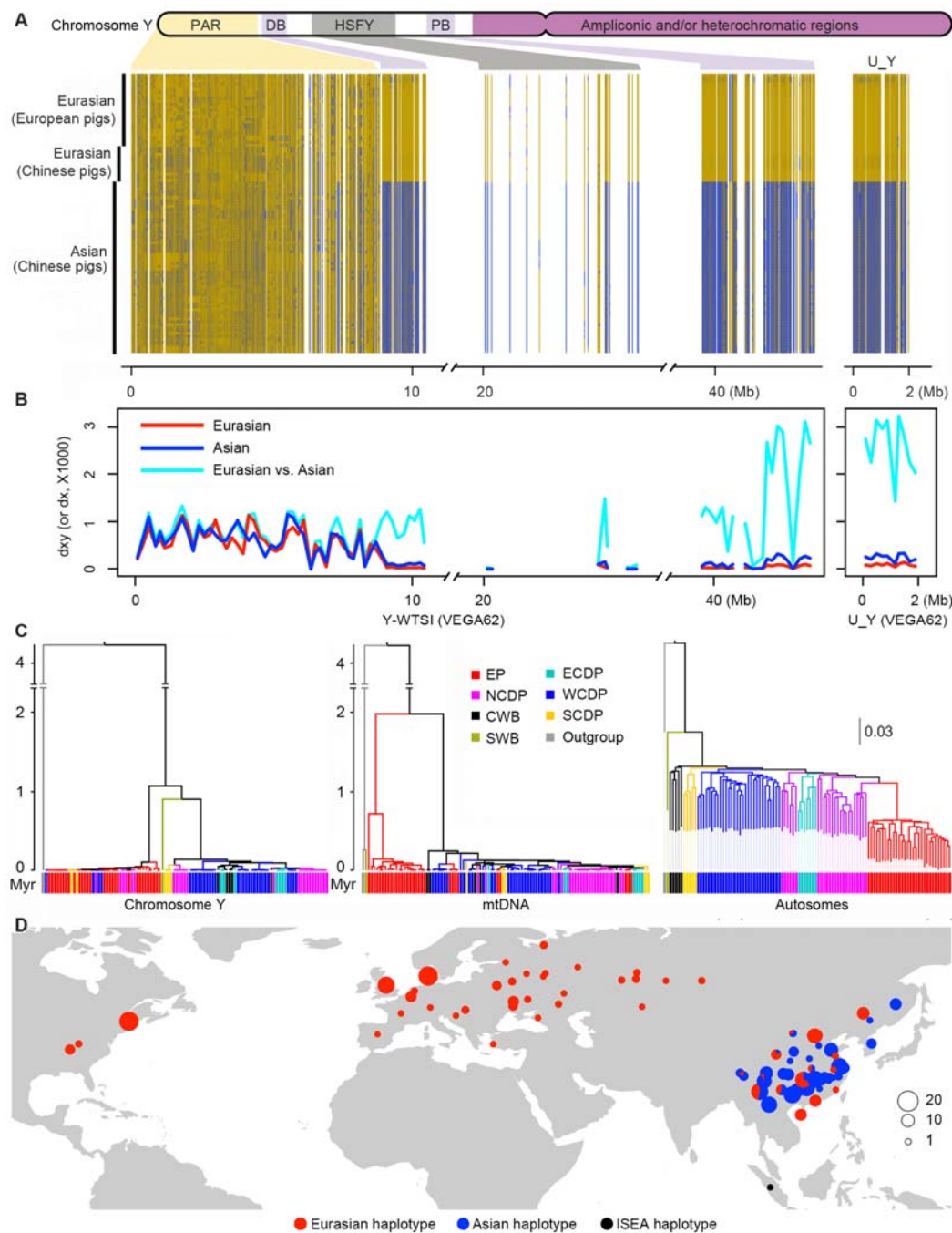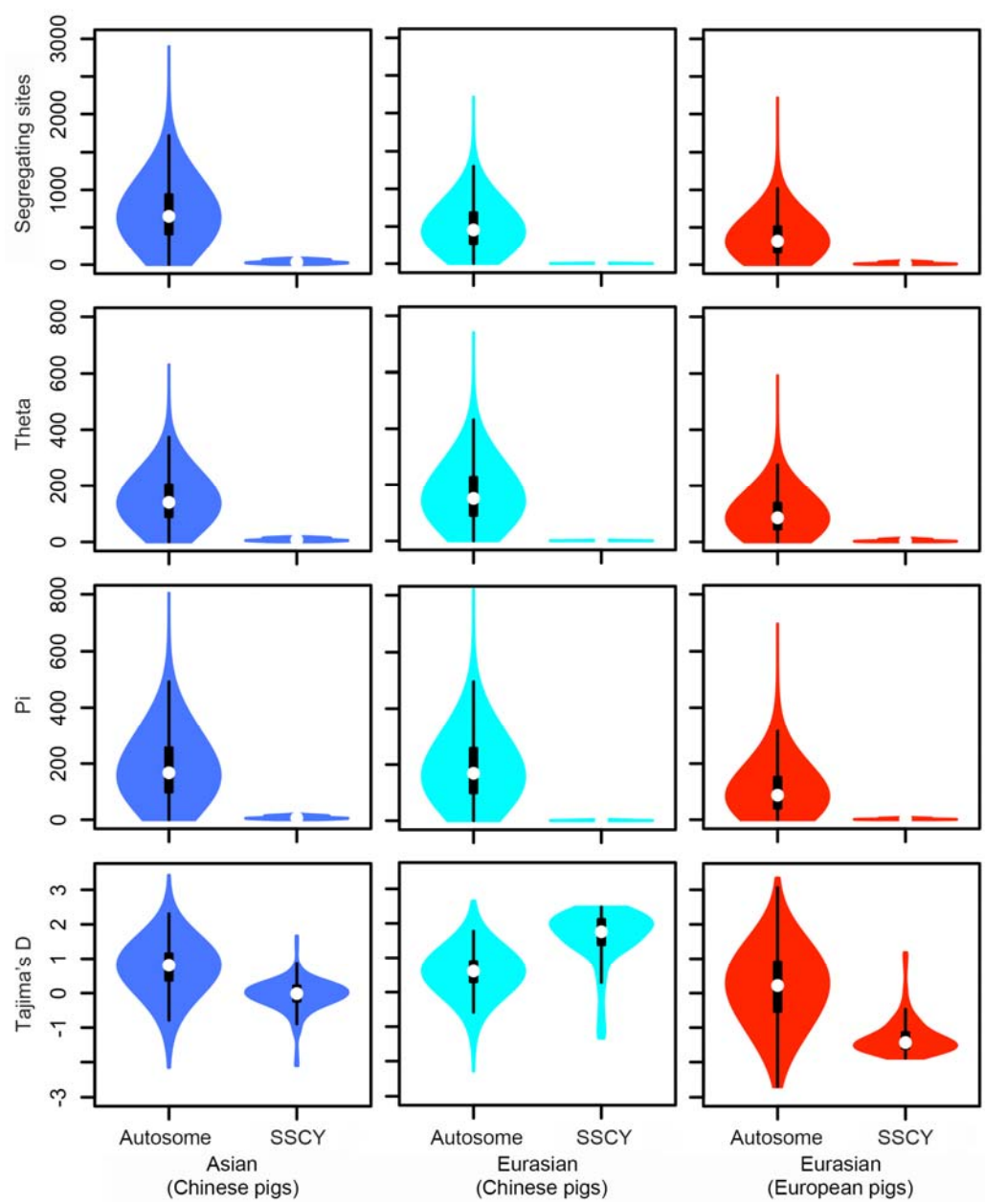
650

651

652

653

654   **Fig. 1**

655

656

657

658   **Fig. 2**

659

660

661

662

663   **Fig. 3**

664

665

666

667 **Supplemental Fig. S1**

668