

1 **Supplementary Information**

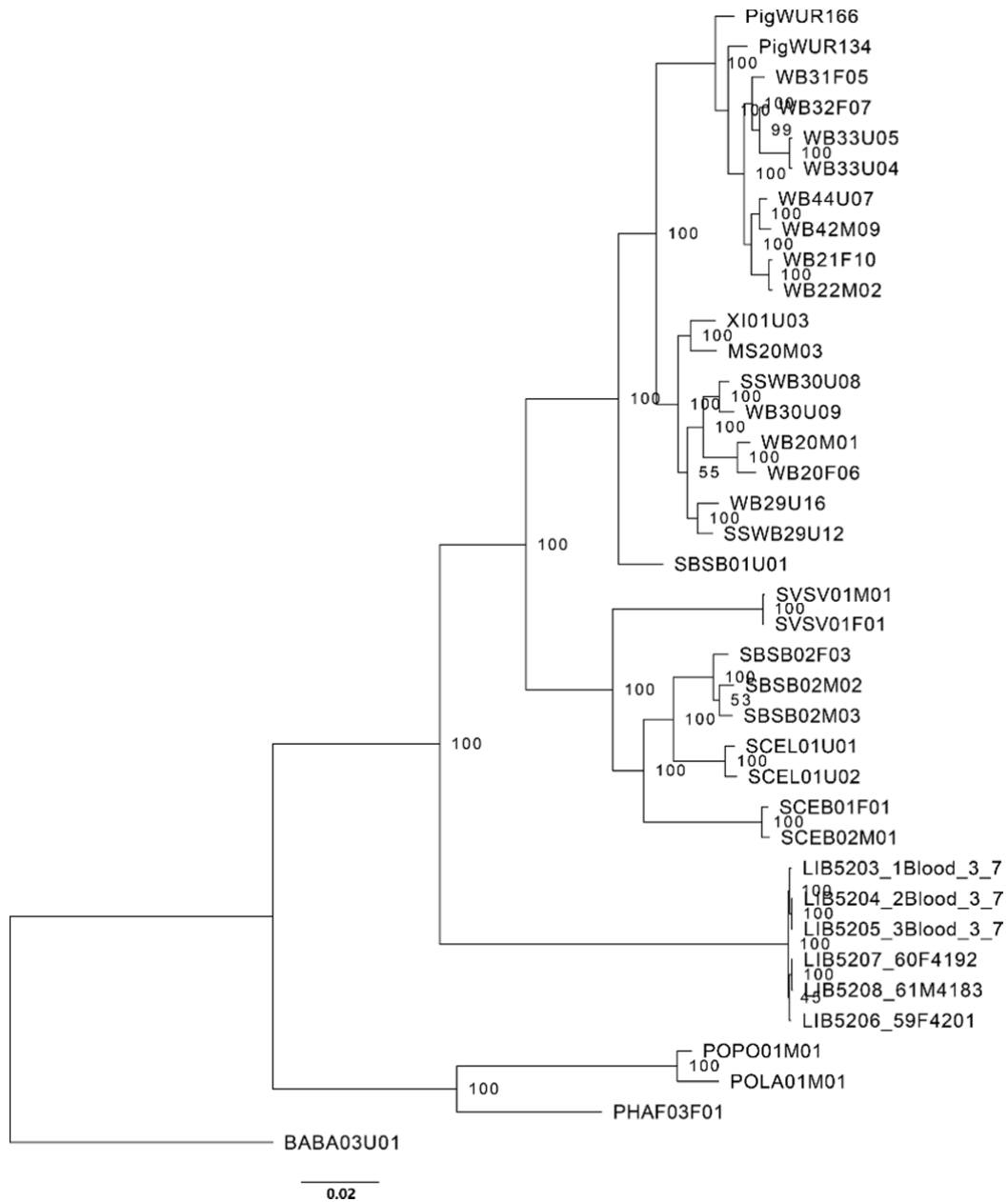
2

3 **Genomic analysis on pygmy hog (*Porcula salvania*) reveals**  
4 **multiple interbreeding during wild boar expansion**

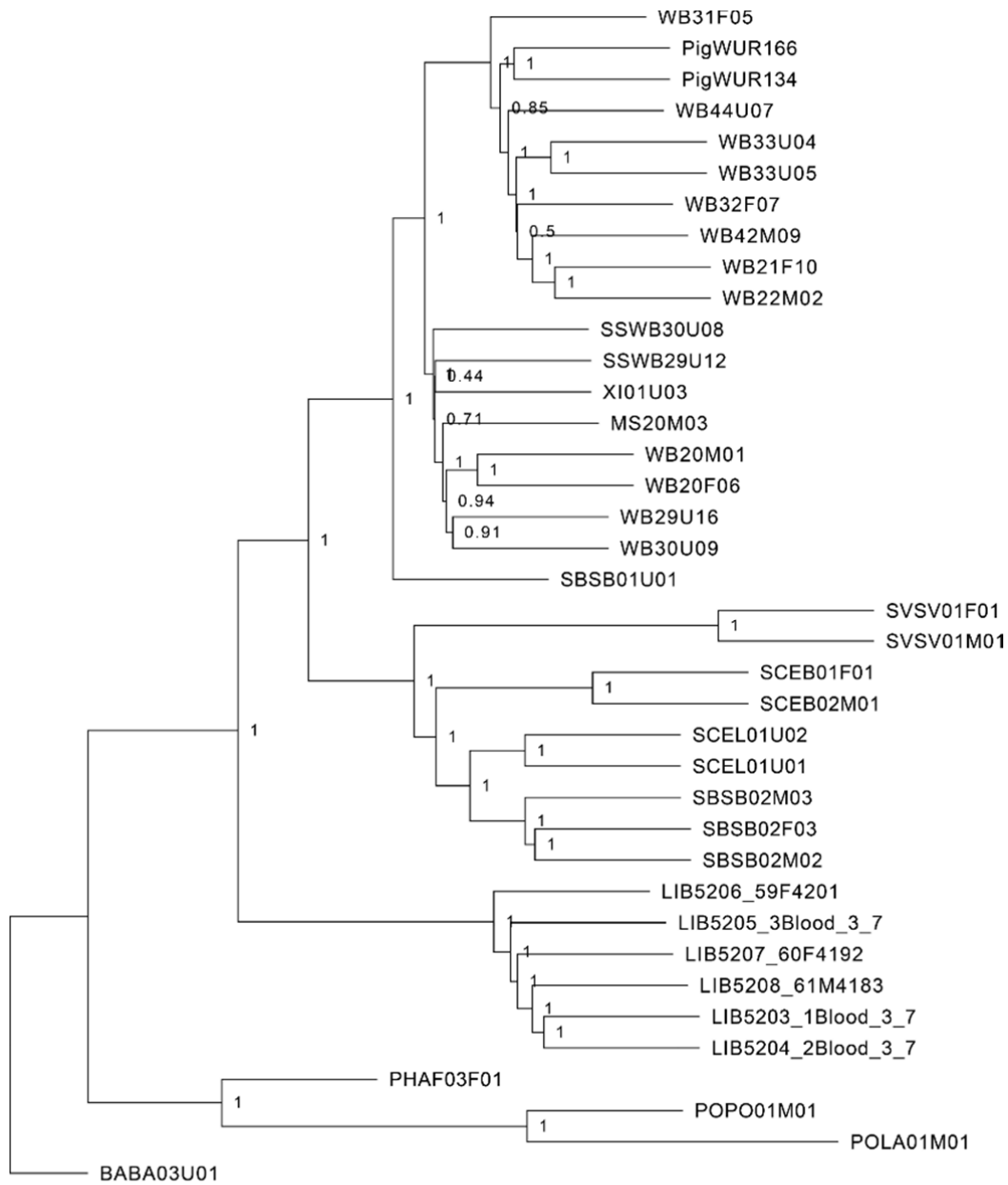
5

6 Liu et al.

7



8  
 9 Supplementary Fig. 1  
 10 Maximum likelihood phylogenetic tree constructed on concatenated coding  
 11 sequences using the supermatrix method (Optimization Likelihood: -42825135.3).  
 12 The numbers on the nodes represent bootstrap values based on 100 non-parametric  
 13 bootstrap replications. *Babyrousa babyrussa* was used as out group. See  
 14 Supplementary data 1 for name abbreviation.  
 15



16

0.02

17

Supplementary Fig. 2

18

Maximum likelihood phylogenetic tree based on supertree method.

19

Among the 1,0142 coding sequences that are used for generating gene trees, 2,659

20

gene trees with highest average bootstrap values (minimum 40) were used to build a

21

consensus species tree. The numbers on nodes represent concordance factors

22

(ranges from 0 to 1), which represent the proportion of the genome supporting the

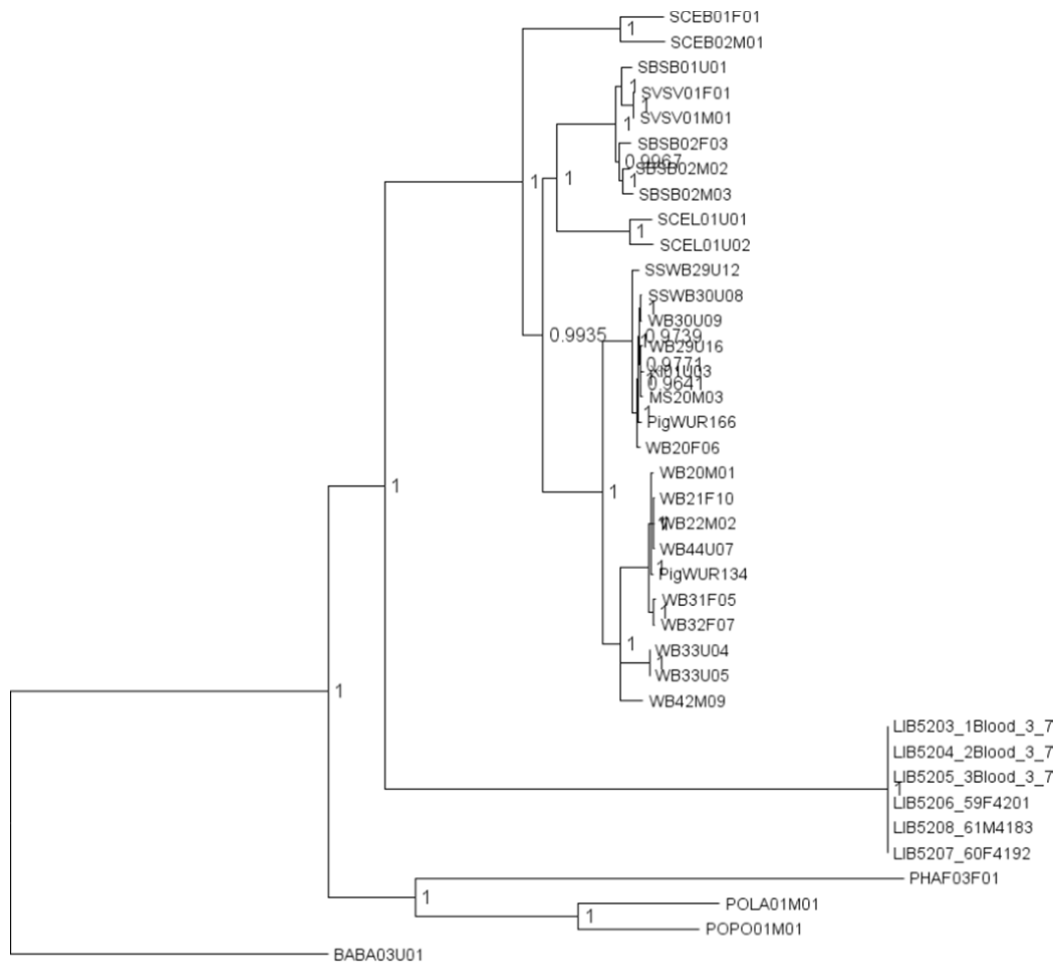
23

branch. *Babyrousa babyrussa* was used as out group. See Supplementary data 1 for

24

name abbreviation.

25

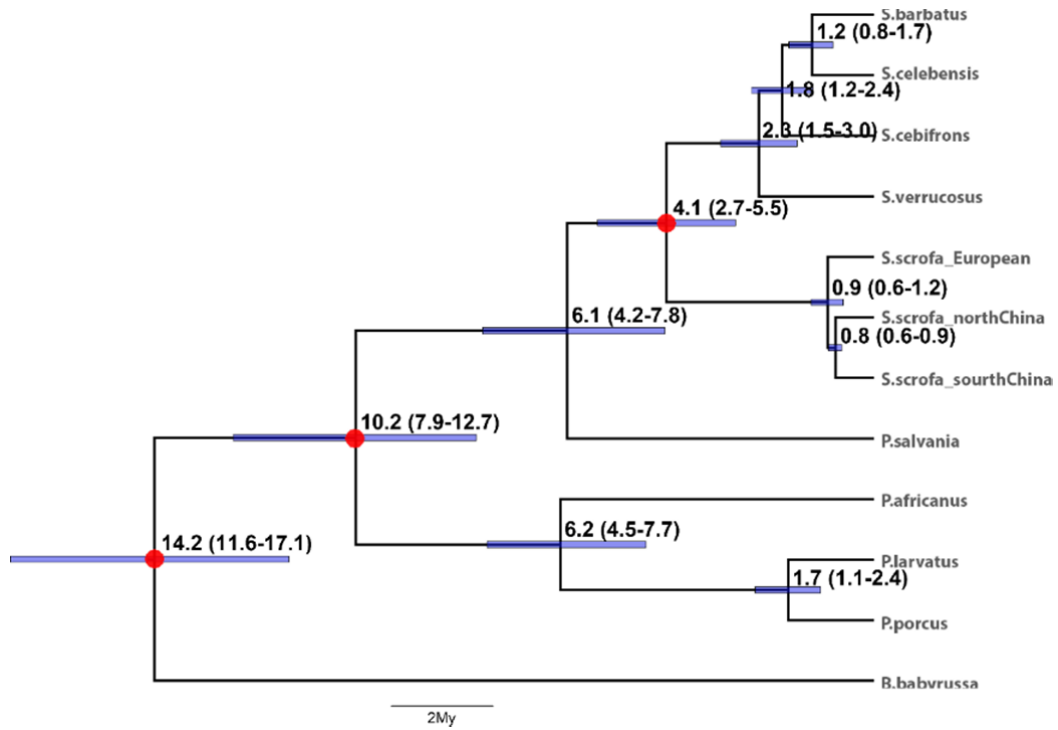


26  
27  
28  
29  
30  
31  
32  
33  
34

Supplementary Fig. 3

Mitochondrial phylogeny

A bayesian phylogeny of the 38 samples using complete mitochondrial genomes (16,613 bp). Pygmy hogs are the sister taxon of *Sus* species, with sub-Saharan suids as outgroup and the mitochondria replacement found by L.A.F Frantz et al 2014 and S.M. Funk et al 2007 is also supported<sup>1,2</sup>. See Supplementary data 1 for name abbreviation.



35

36 Supplementary Figure 4

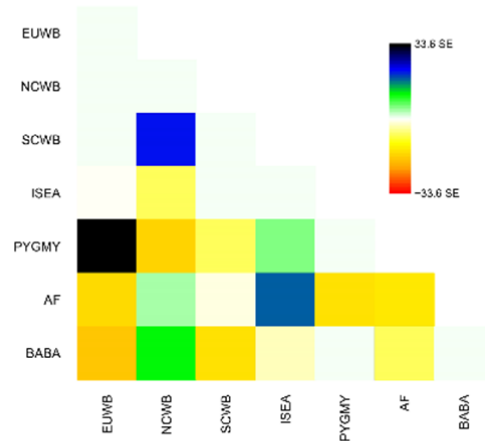
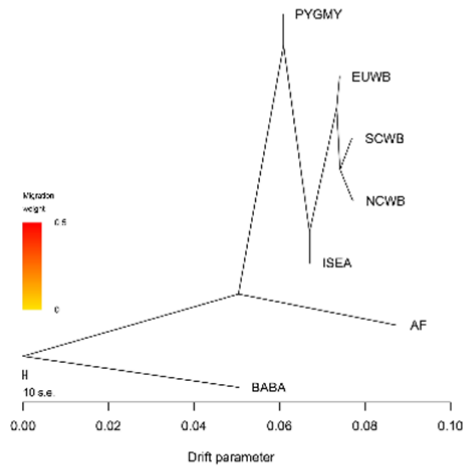
37 Phylogenetic time tree

38 Species tree with mean posterior age (in millions of years). Blue bars indicated 95%  
 39 confidence intervals. Red dots indicated the calibration points. We used a float prior  
 40 and a maximum bound age, with a scale parameter of  $c=2$ . For the root divergence,  
 41 we set the prior to  $(tU=2 [20 \text{ Mya}], p=0.1, c=2)$ . For MRCA of *Suinae* and *Sus*,  
 42 we used the same fossil calibration as in Frantz et al. 2013 ( $tL=0.55 [5.5 \text{ Mya}], p=0.9,$   
 43  $c=0.5$  and  $tL=0.2 [2 \text{ Mya}], p=0.1, c=0.5$ , respectively).

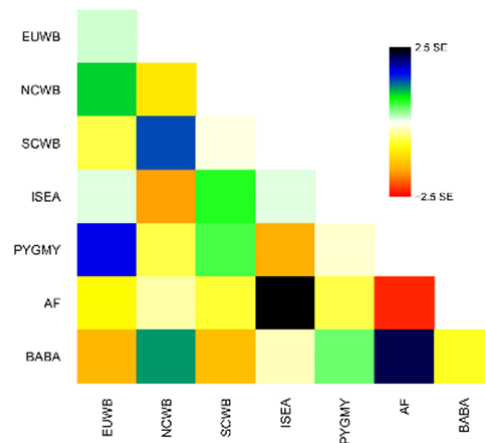
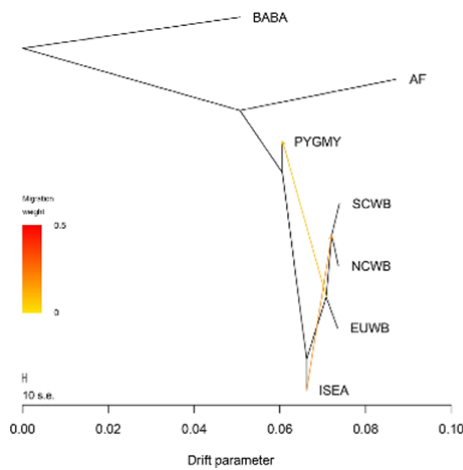
44

45

a) rejected tree model



b) graph with 2 migration edge



46

47 Supplementary Fig. 5

48 Admixture graph inferred using Treemix

49 a) A simple tree-like model without admixture fits the data poorly, as can be seen from

50 the matrix of residuals between empirical and modelled allele frequency covariance

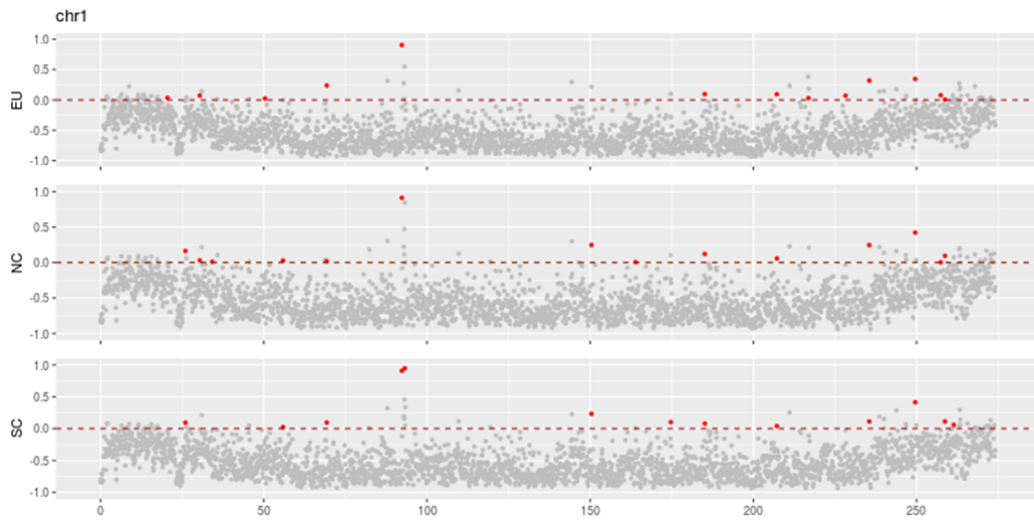
51 on the right. b) The optimal placement of two admixture event are from the common

52 ancestor of Eurasian wild boar to pygmy hog, as well as from ISEA Sus to common

53 ancestor of Asian wild boar.

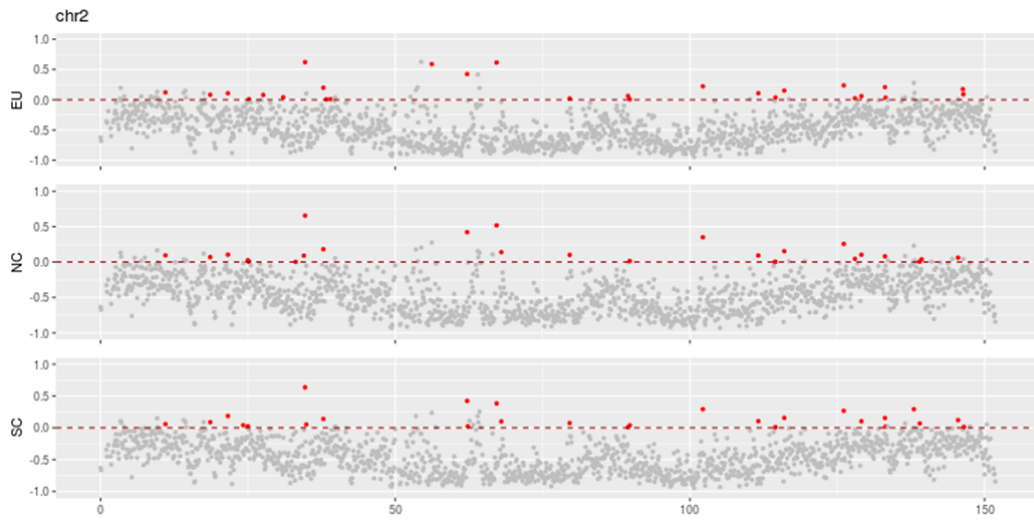
54

55



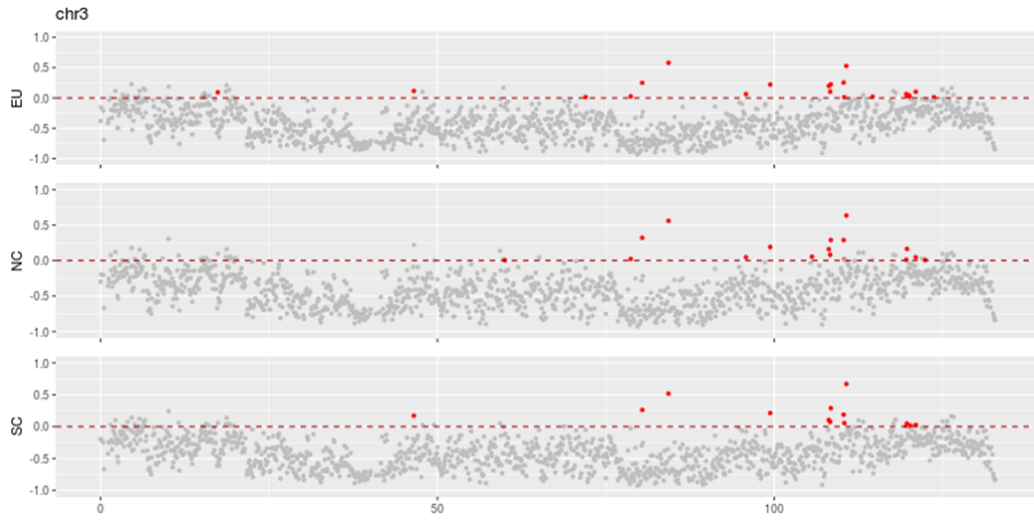
56

57



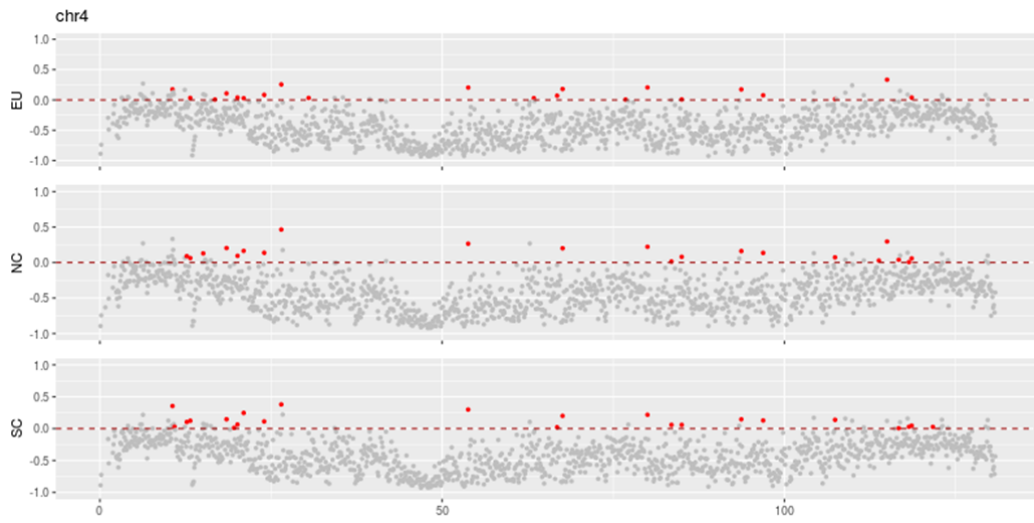
58

59



60

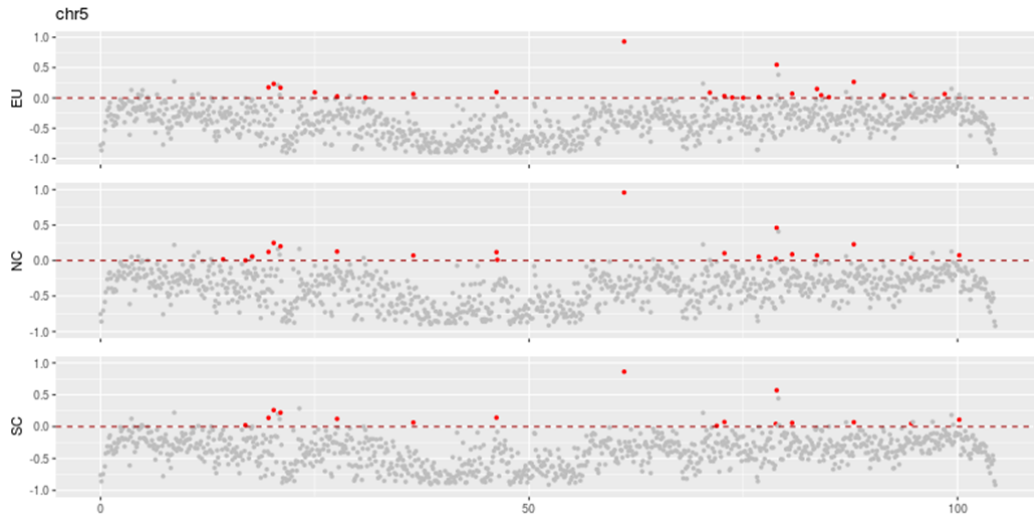
61



62

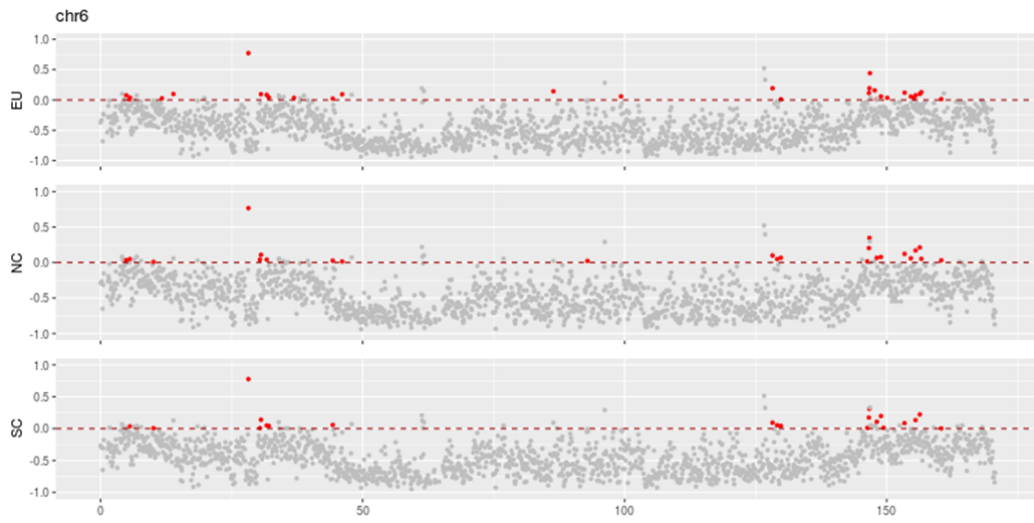
63





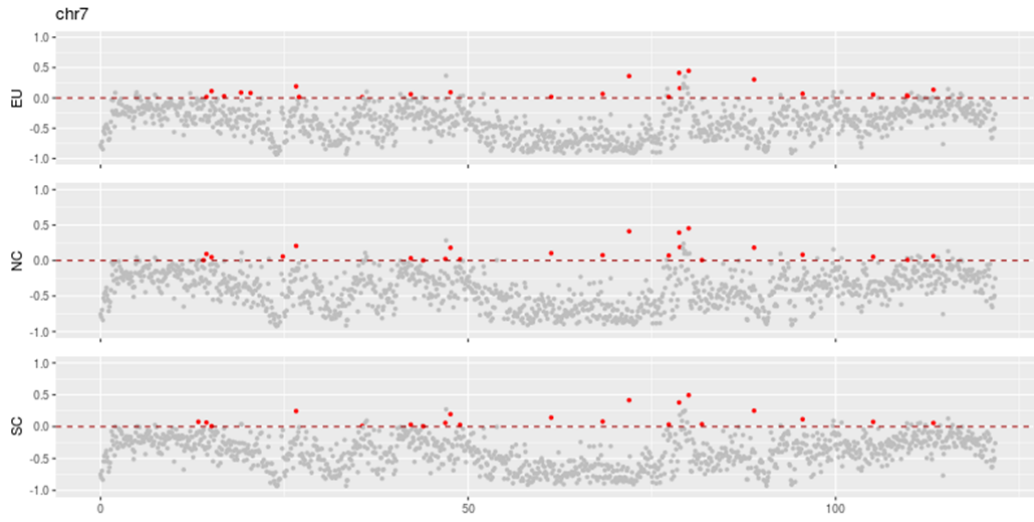
64

65



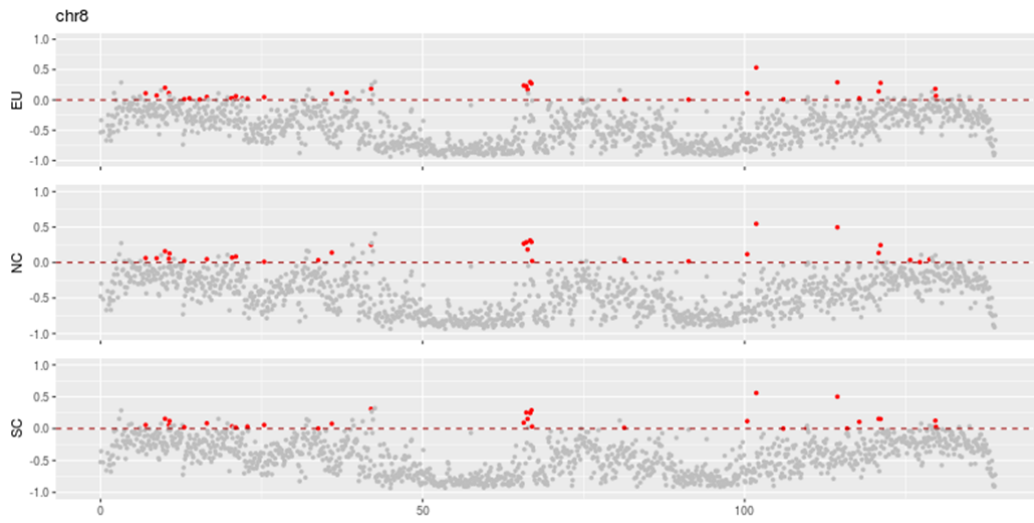
66

67



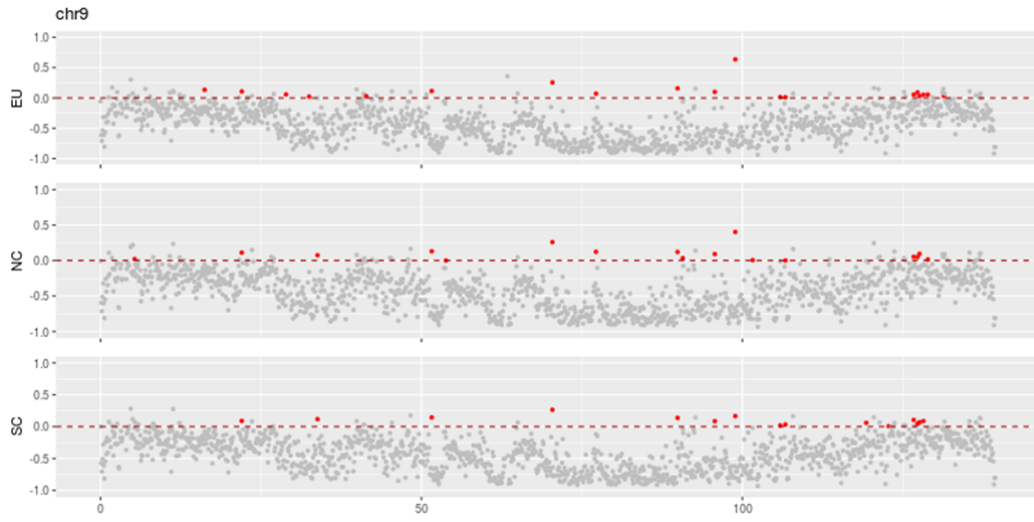
68

69



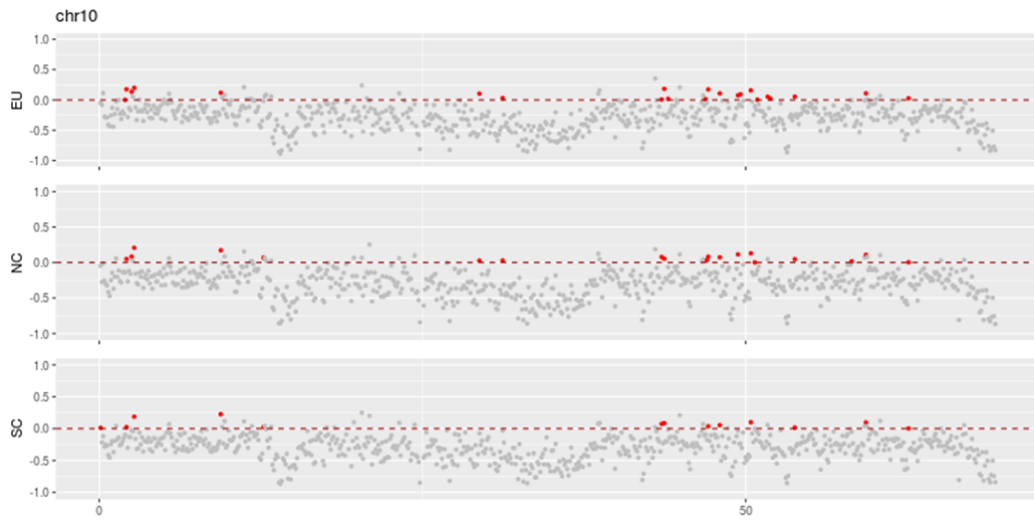
70

71



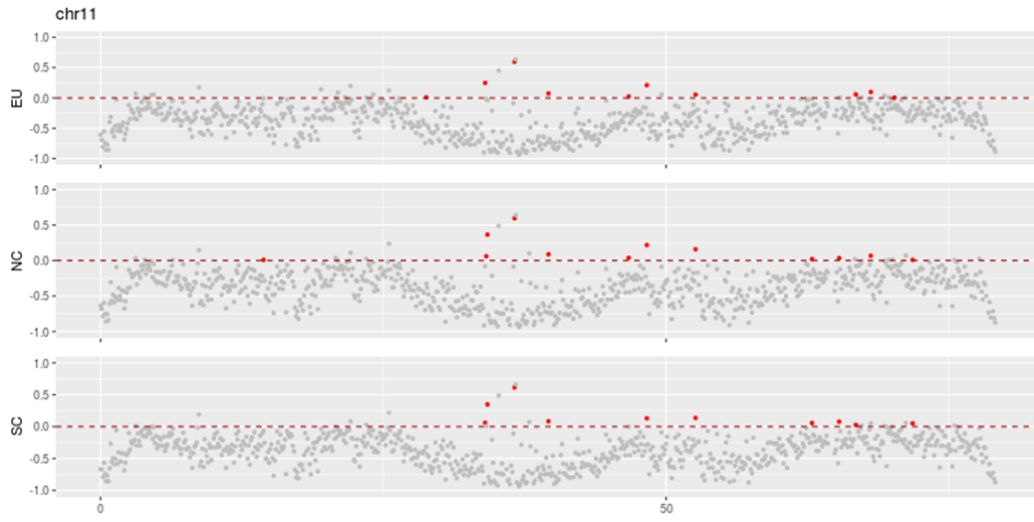
72

73



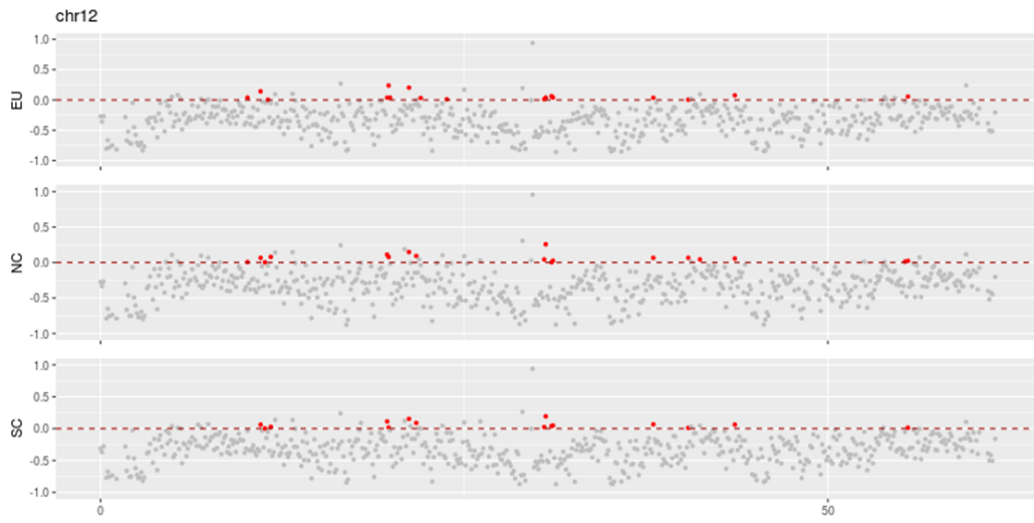
74

75



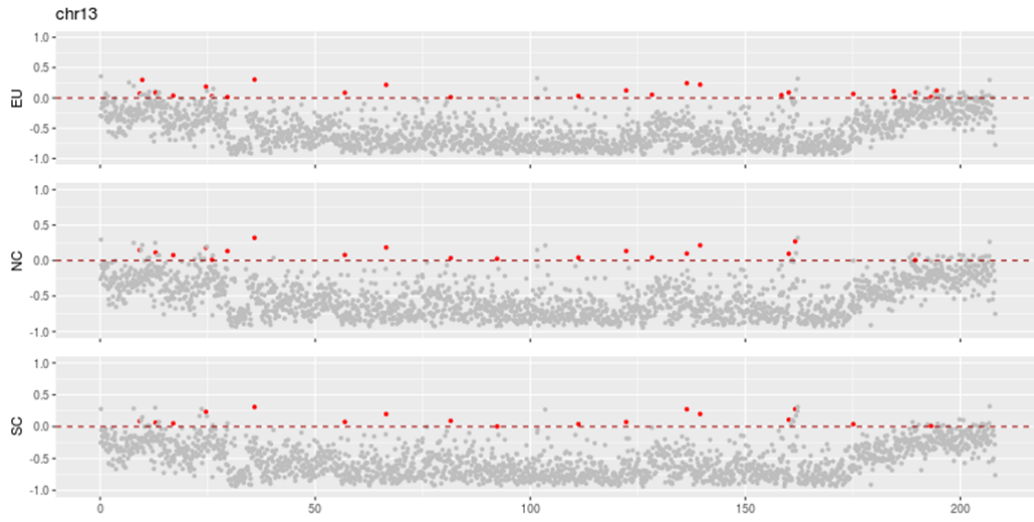
76

77



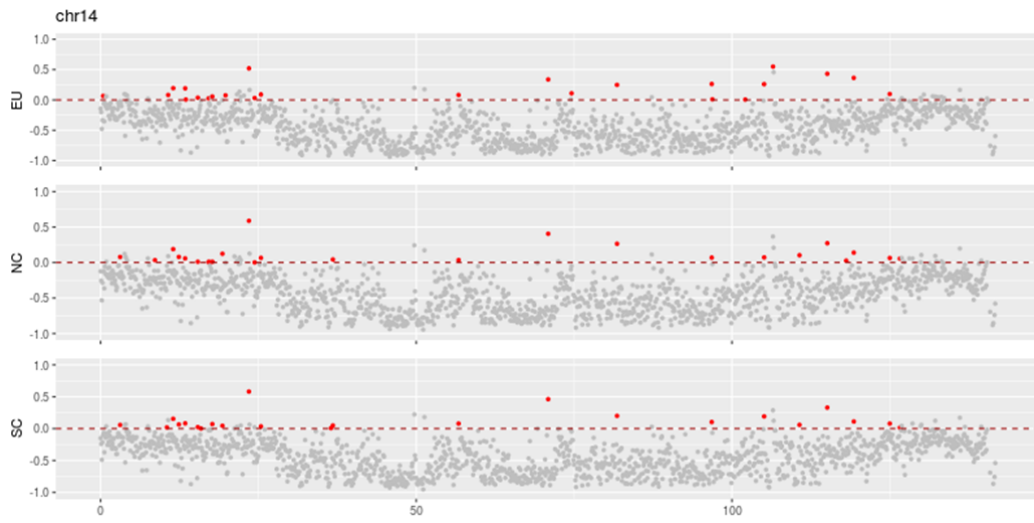
78

79



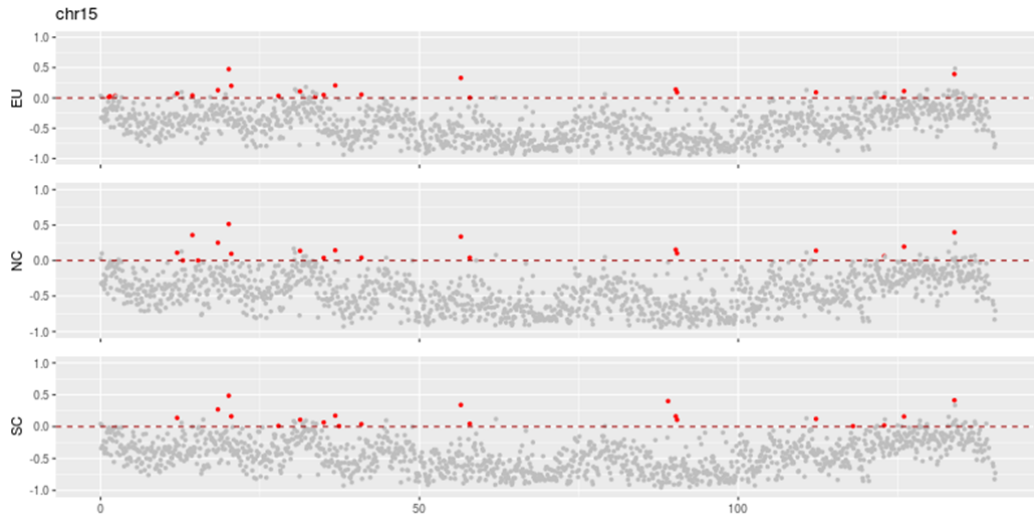
80

81



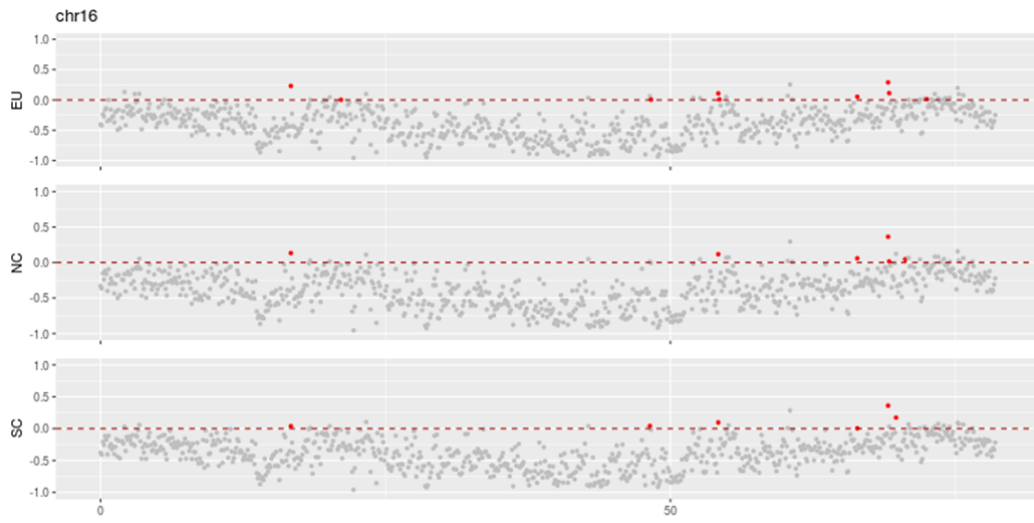
82

83



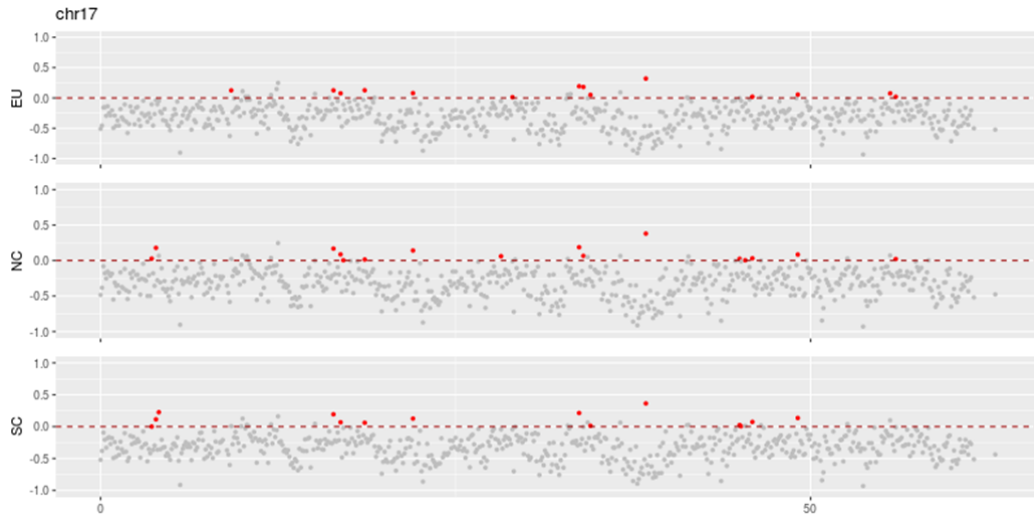
84

85



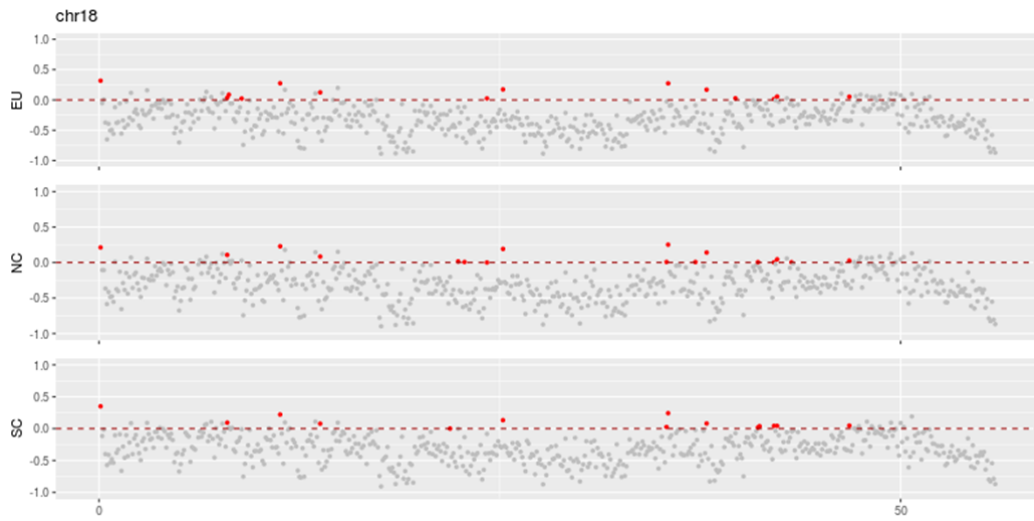
86

87



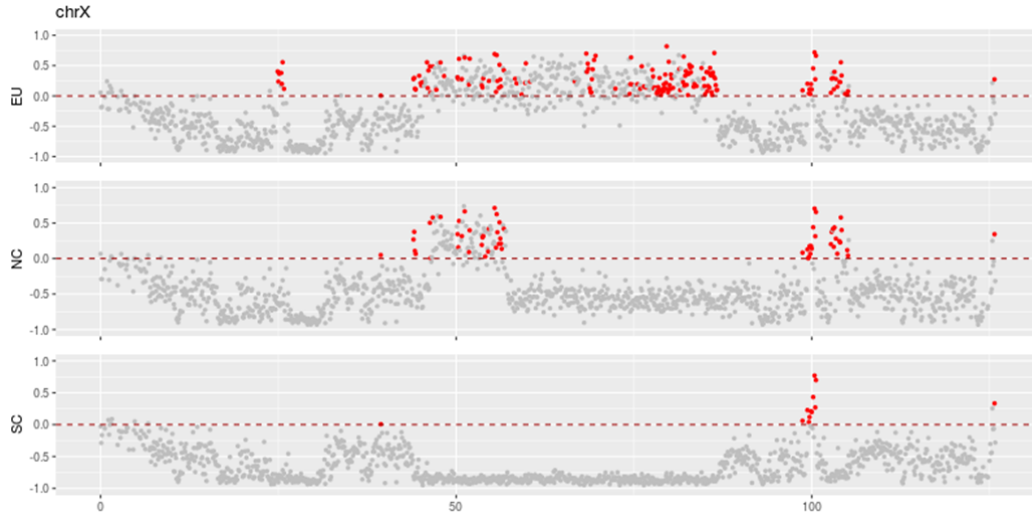
88

89



90

91



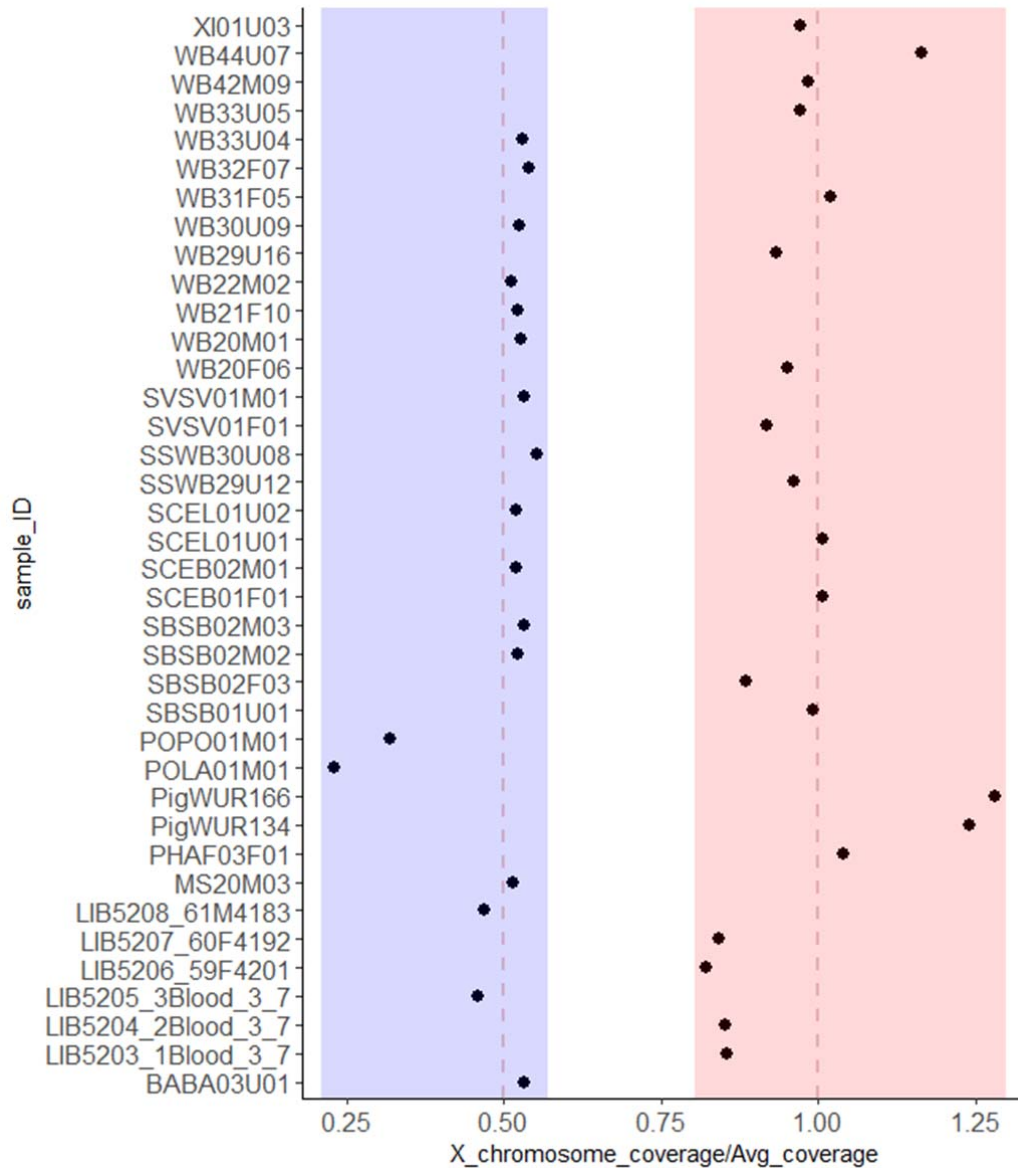
92

93 Supplementary Fig. 6

94 The D-statistic and  $f_d$  for testing introgression for 100-kb windows on each  
 95 chromosome with the tree topology (((ISEA, Pygmy hog), X), warthog), X=European  
 96 wildboar (EU), Northern China wildboar (NC) or Southern China wildboar (SC). We  
 97 also calculated DNA sequence divergence ( $d_{xy}$ ) between X and pygmy hog for each  
 98 window and color the windows red whose  $d_{xy}$  is in the lower 25% distribution of all the  
 99 windows and D-value > 0. Chr = chromosome, ISEA = Island of South east Asia pigs

100



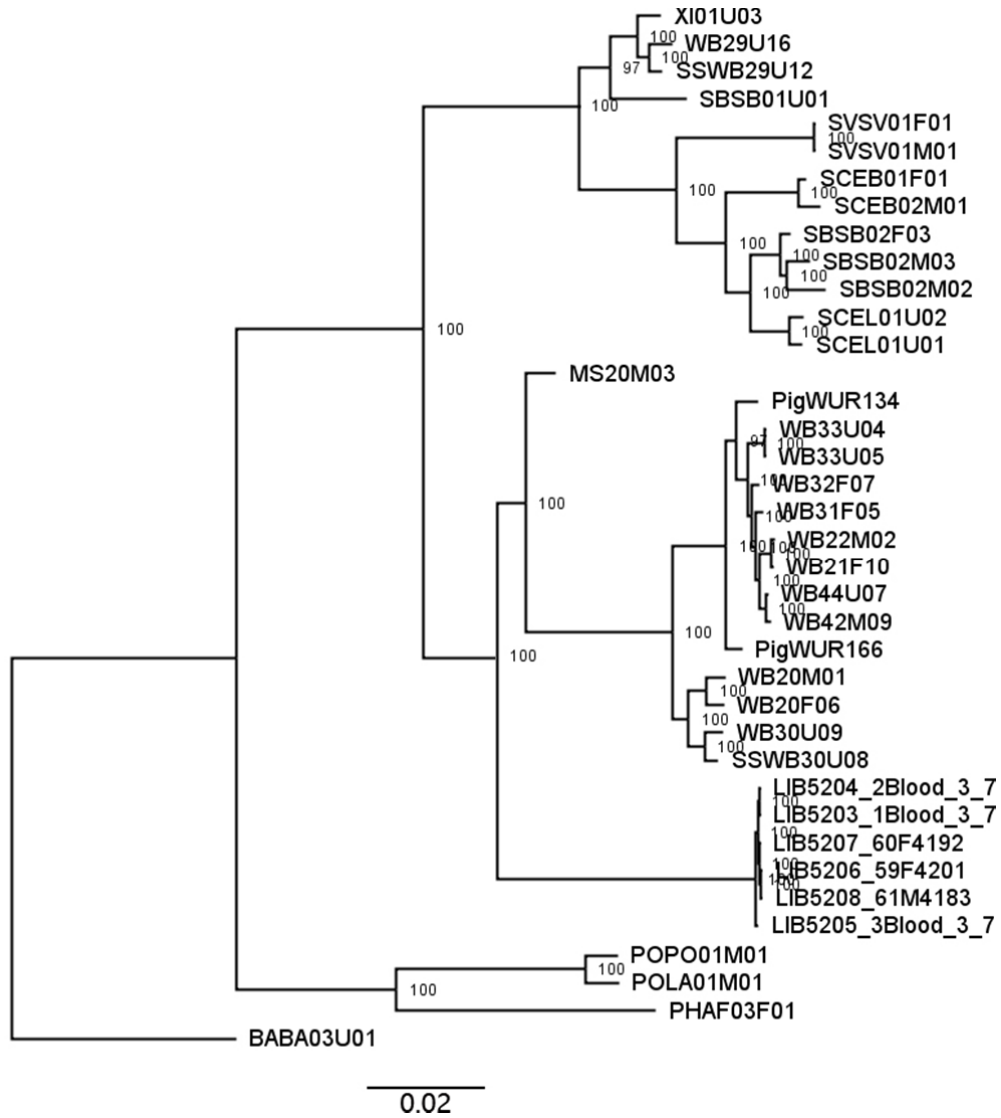


101

102 Supplementary Fig. 7

103 Observed fraction of X chromosome read depth compared to the average read depth  
 104 of autosomes in all individuals in this study.

105



107

108 Supplementary Fig. 8

109 X chromosome topology

110 The maximum likelihood tree of the whole X chromosome (Optimization Likelihood:

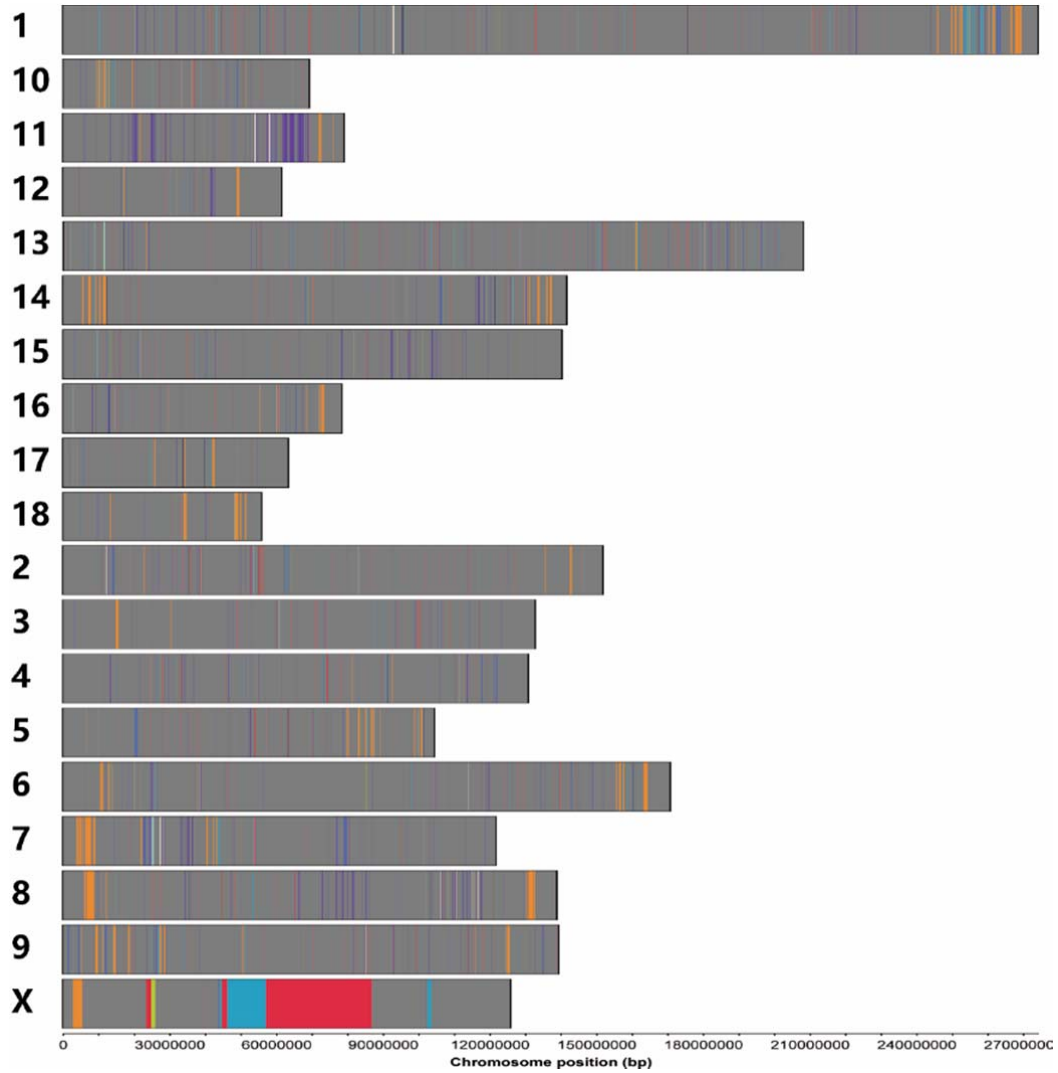
111 -223197633.0). The topology of the whole X chromosome tree was inconsistent with

112 the topology of autosomal phylogenetic trees (Supplementary Fig. 1&2). The numbers

113 represent node support inferred from 100 non-parametric bootstrap repetitions.

114 B.babyrussa is the outgroup. See Supplementary data 1 for name abbreviation.

115

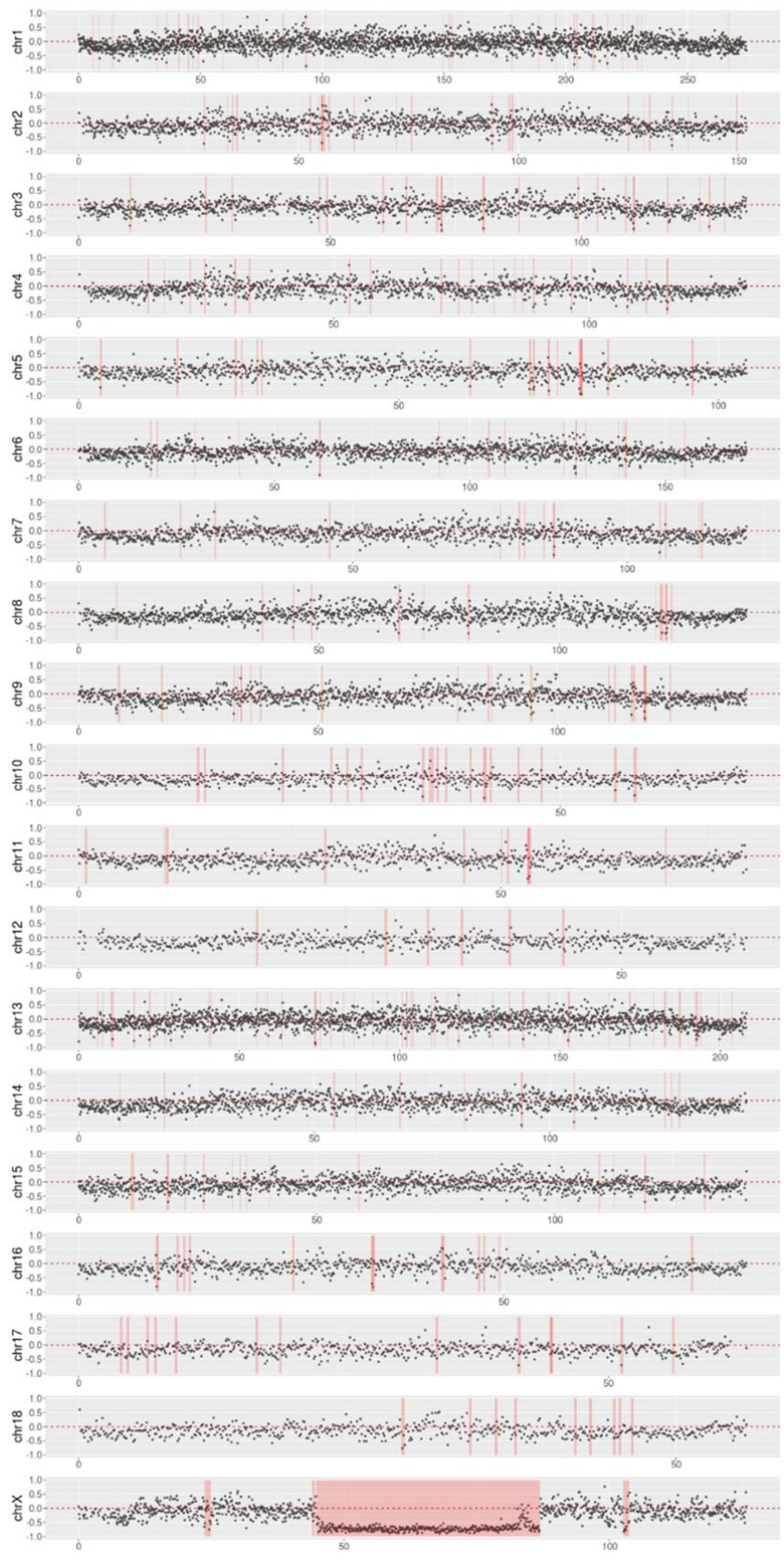


116

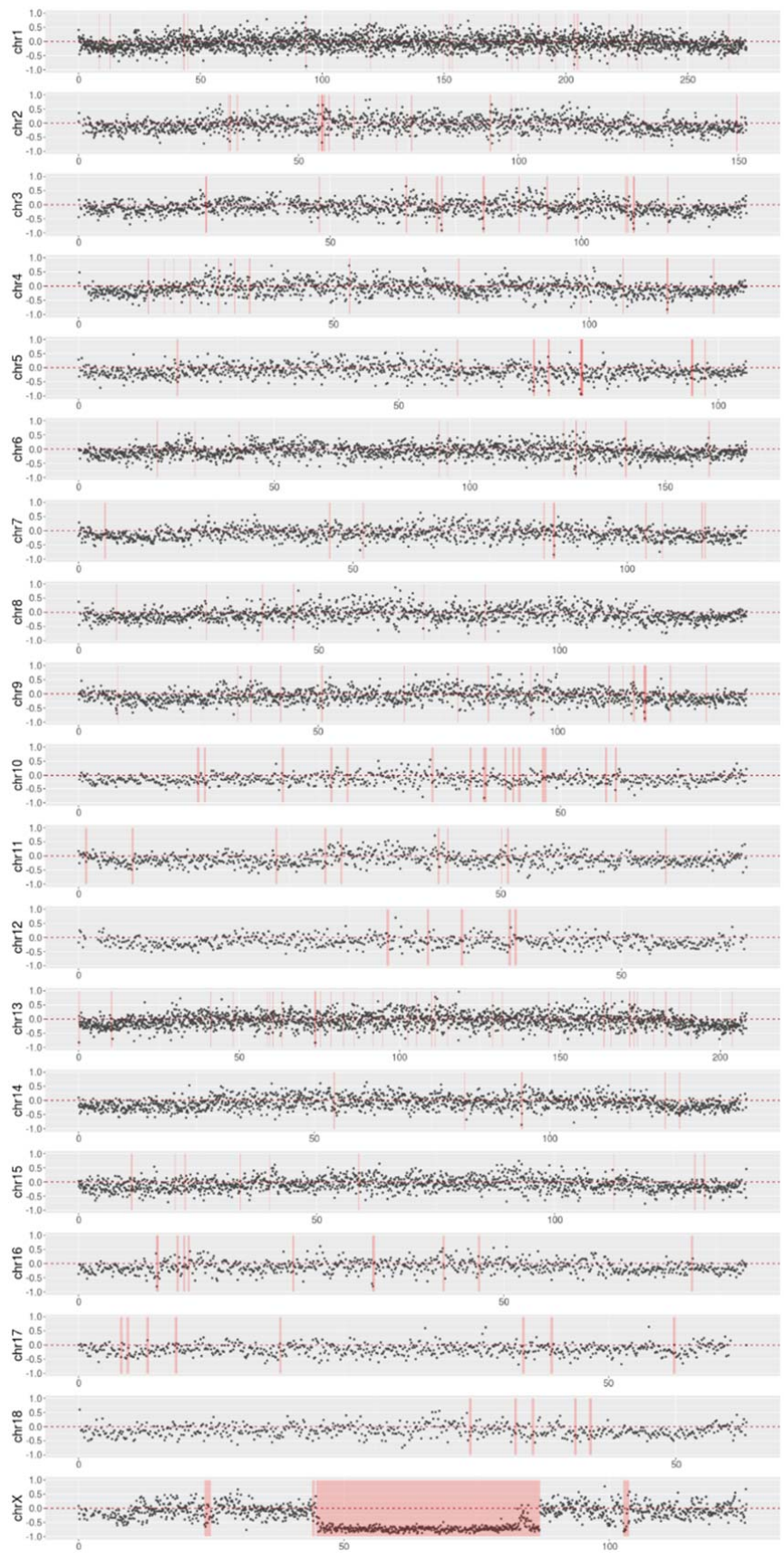
117 Supplementary Fig. 9

118 Distribution of chromosomal segments supporting the 12 most frequent tree  
 119 topologies assigned by SAGUARO to segments along the autosomes and the X  
 120 chromosome. The numbers given next to phylogenies indicate chromosome IDs. See  
 121 Supplementary data 3 for further details.

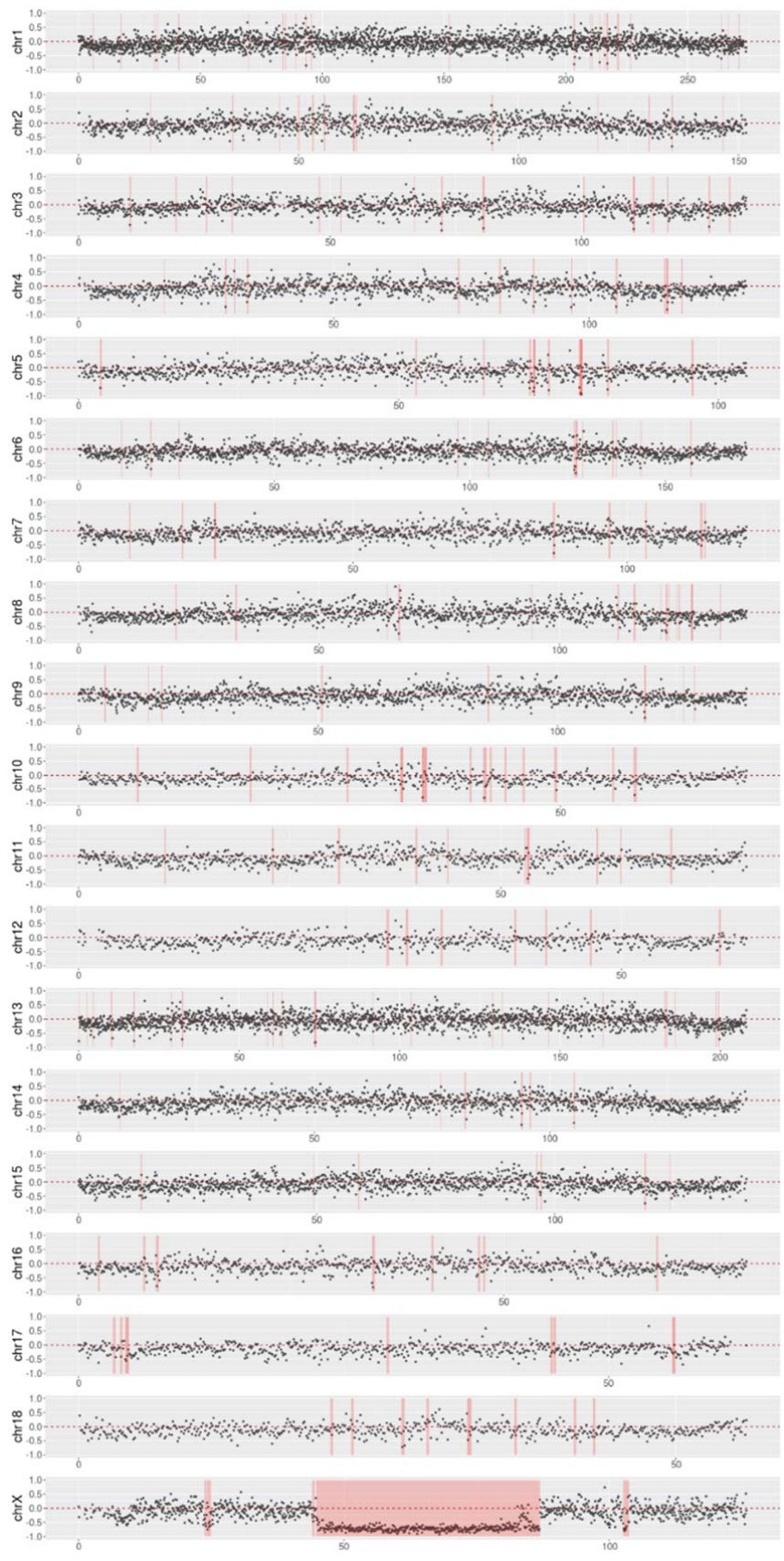
122



124 Supplementary Fig. 10  
125 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
126 X. For the tree topology (((ISEA, EUWB), pigmy hog), warthog). The ABBA-BABA  
127 method and its related statistics compute the excess of shared derived mutation  
128 between two taxa compared to a control not connected by gene flow to the others.  
129 The outgroup allows mutations to be polarized. A mean D value of 0 is expected if the  
130 two taxa are not connected by gene flow. Red shaded areas show the introgression  
131 regions inferred by Saguaro using the same quadruplet. ISEA=Island of South Asia  
132 pigs, EUWB = European wild boar.  
133

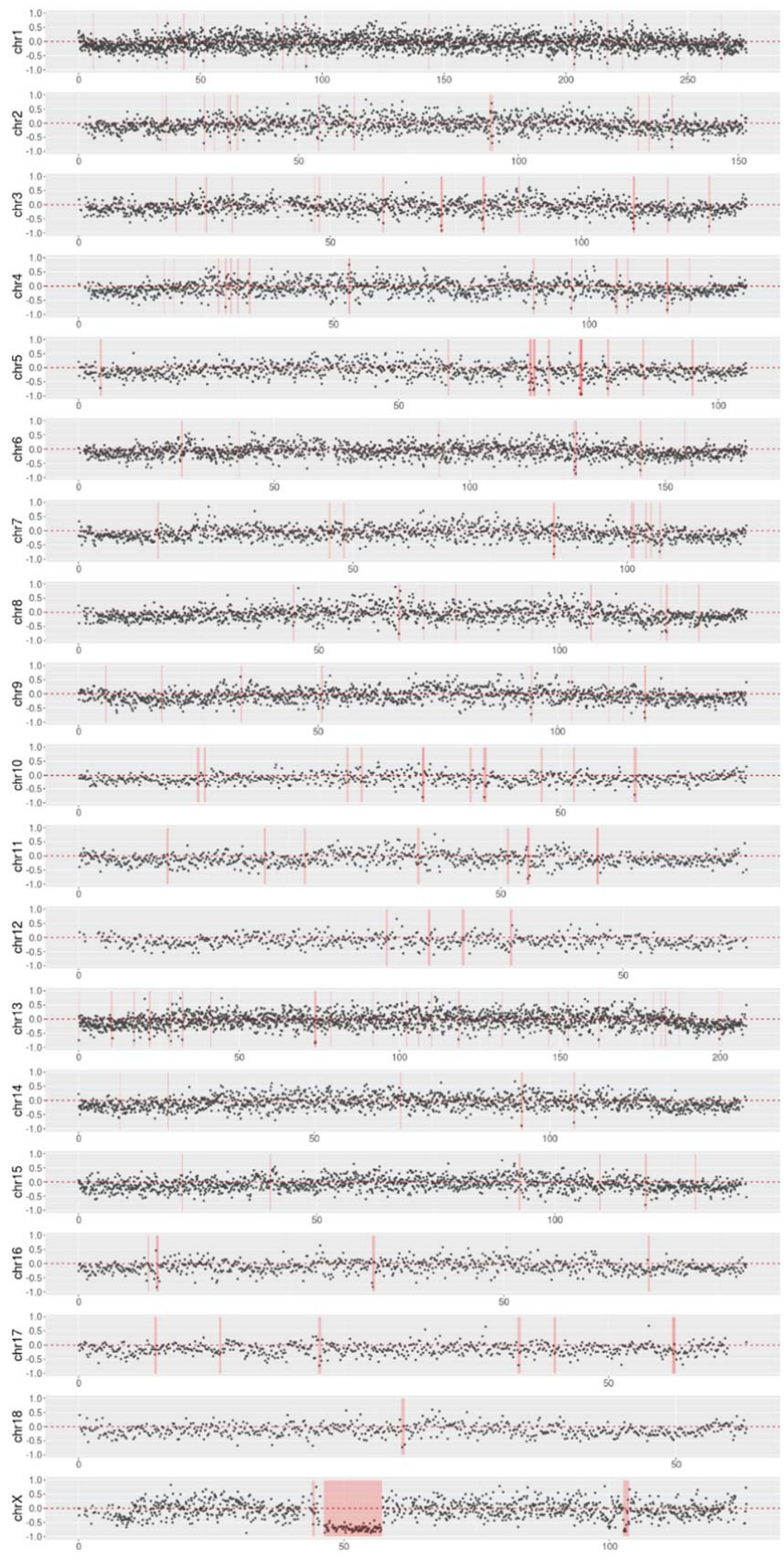


135 Supplementary Fig. 11  
136 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
137 X. For the tree topology (((ISEA, EUD), pygmy hog), warthog). Red shaded areas  
138 show the introgression regions inferred by Saguaro using the same quadruplet.  
139 ISEA=Island of South Asia pigs, EUD = European domesticated pig.  
140  
141

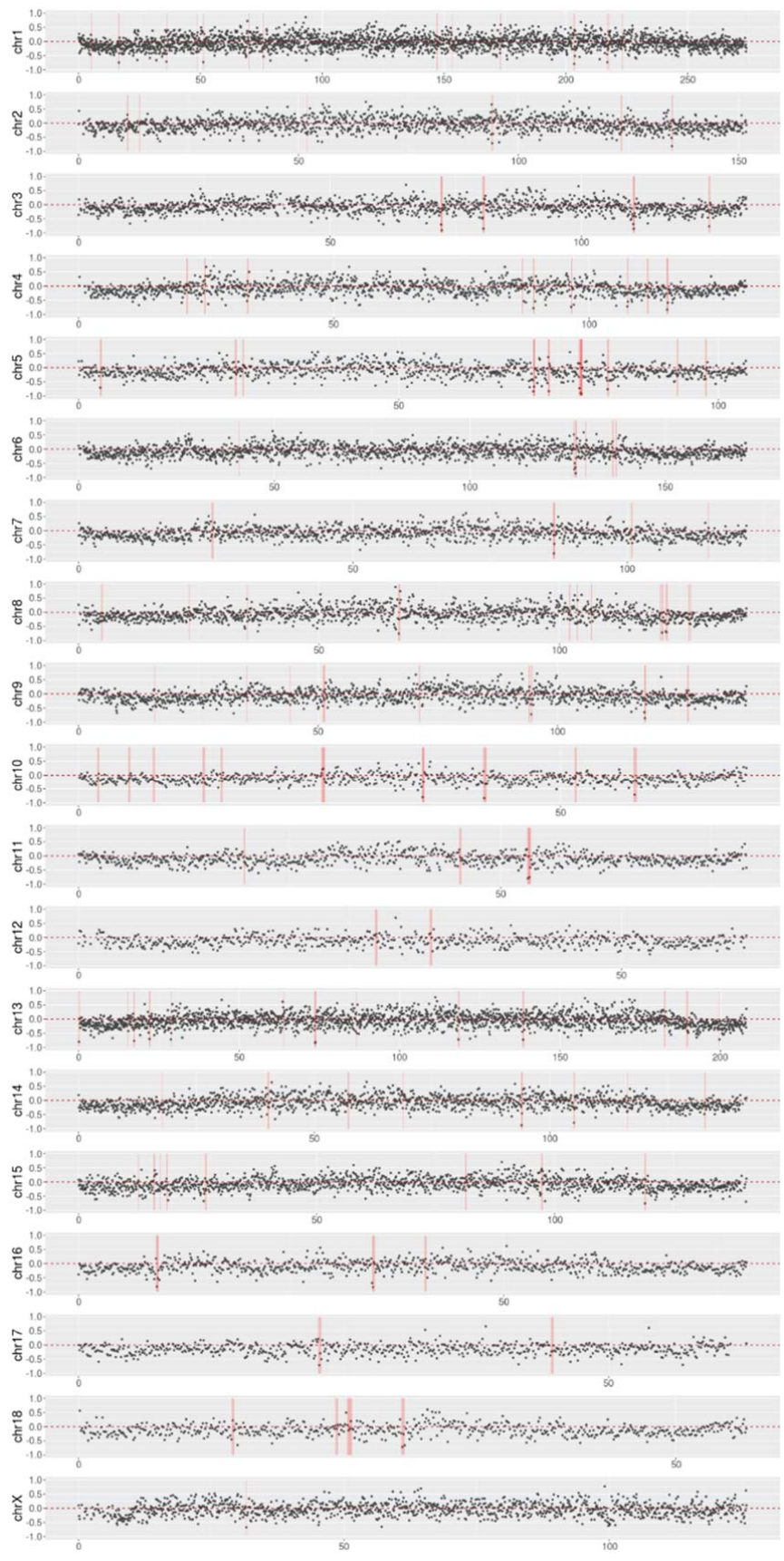




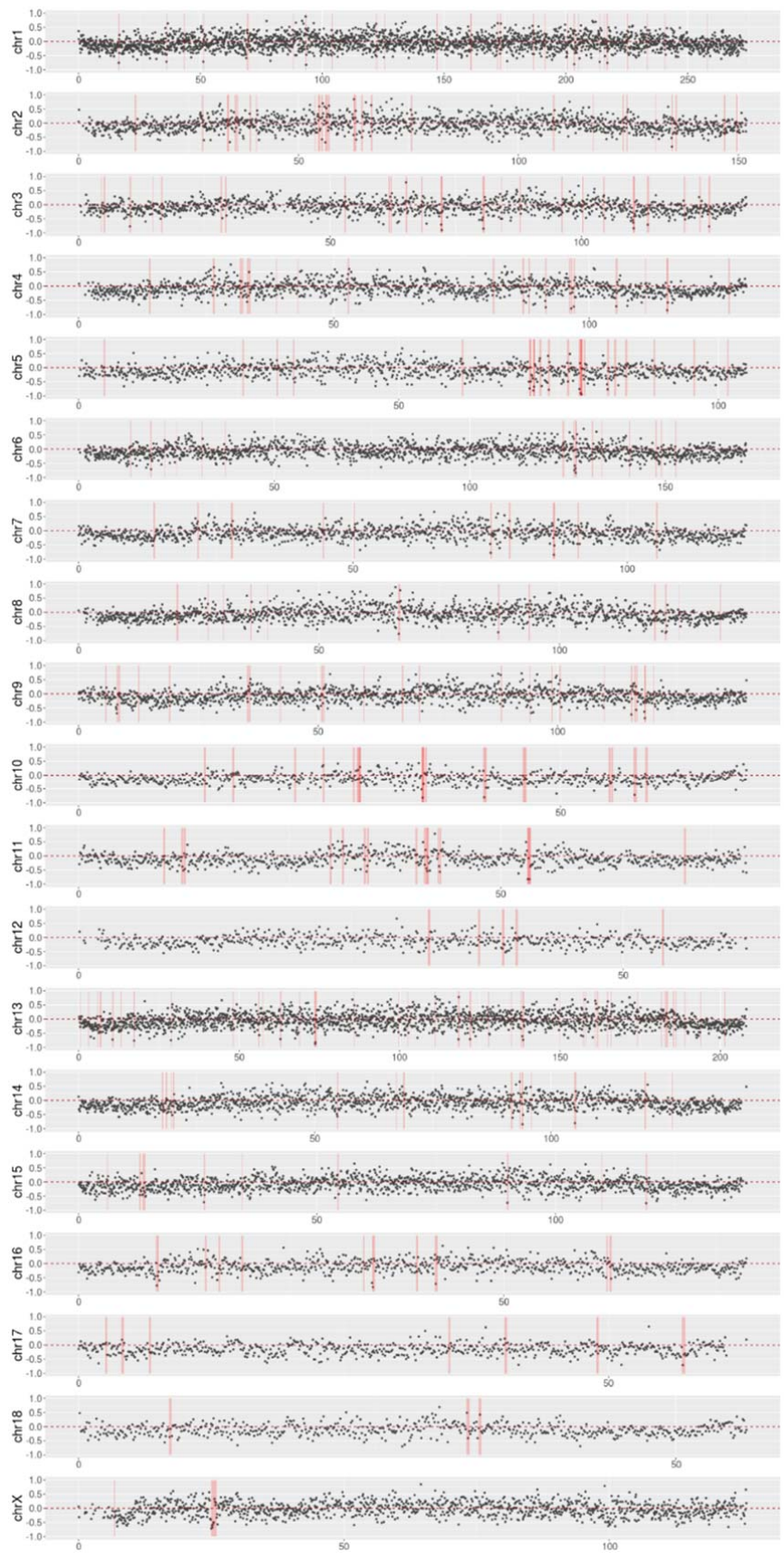
143 Supplementary Fig. 12  
144 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
145 X. For the tree topology (((ISEA, NCWB), pygmy hog), warthog). Red shaded areas  
146 show the introgression regions inferred by Saguaro using the same quadruplet.  
147 ISEA=Island of South Asia pigs, NCWB = North Chinese wild boar.  
148  
149



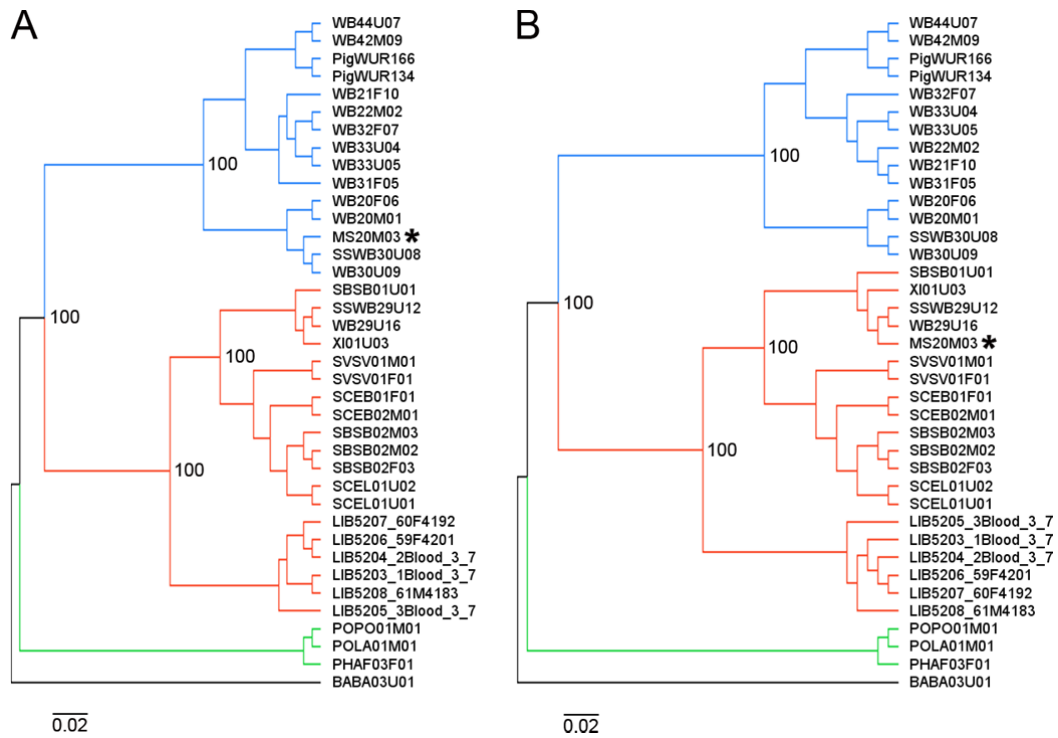
151 Supplementary Fig. 13  
152 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
153 X. For the tree topology (((ISEA, NCD), pygmy hog), warthog). Red shaded areas  
154 show the introgression regions inferred by Saguaro using the same quadruplet.  
155 ISEA=Island of South Asia pigs, NCD = North Chinese domesticated pig.  
156



158 Supplementary Fig. 14  
159 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
160 X. For the tree topology (((ISEA, SCWB), pygmy hog), warthog). Red shaded areas  
161 show the introgression regions inferred by Saguaro using the same quadruplet.  
162 ISEA=Island of South Asia pigs, SCWB = South Chinese wild boar.  
163

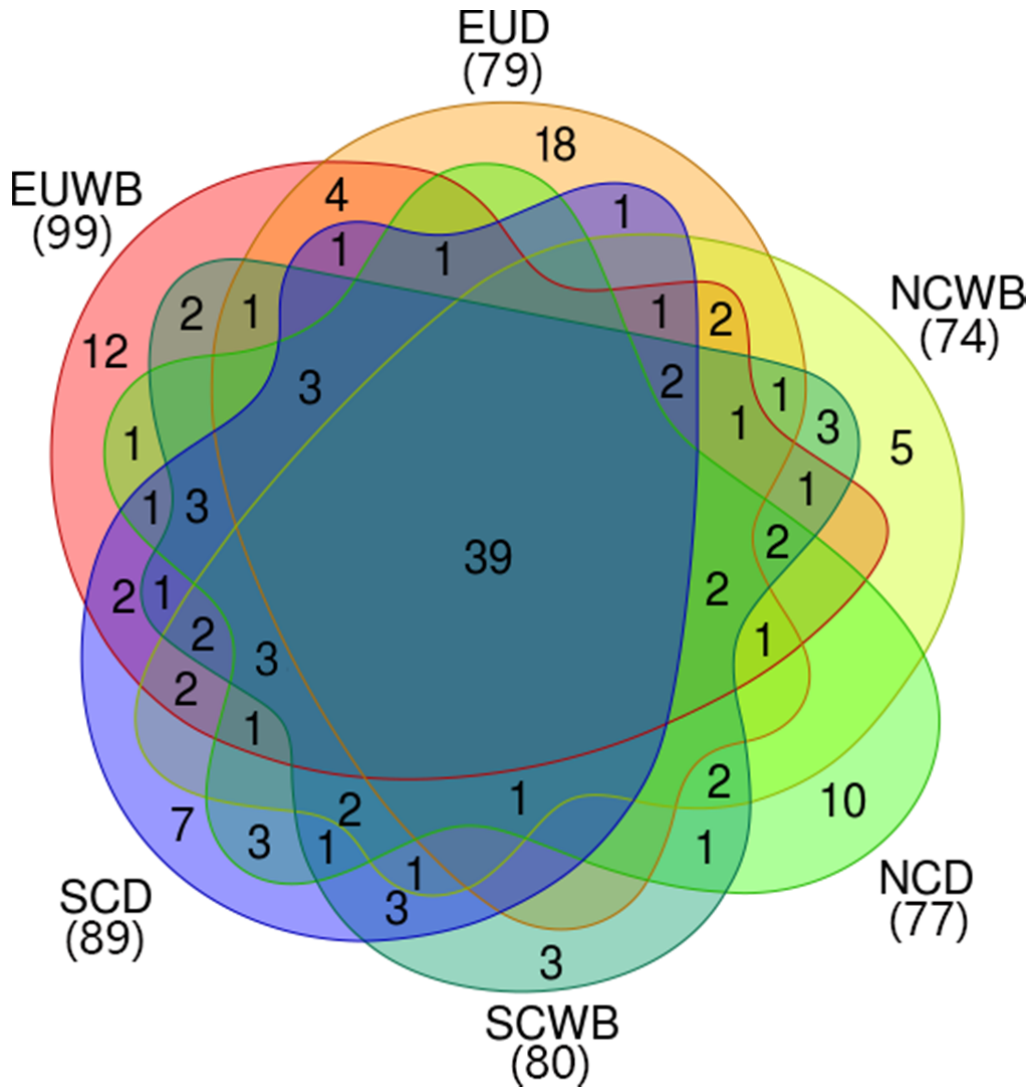


165 Supplementary Fig. 15  
166 The D-statistic for testing introgression for 100-kb windows on chromosome 1-18 and  
167 X. For the tree topology (((ISEA, SCD), pygmy hog), warthog). Red shaded areas  
168 show the introgression regions inferred by Saguaro using the same quadruplet.  
169 ISEA=Island of South Asia pigs, SCD = South Chinese domesticated pig.  
170



171  
172 Supplementary Fig. 16  
173 X chromosome topology for the two introgression regions.  
174 Maximum likelihood tree for the two genealogical discordant regions on chromosome  
175 X. A. Maximum likelihood tree using the first discordant region spanning 52.0~57.8Mb  
176 on X chromosome (Optimization Likelihood: -18385084.9). B. Maximum likelihood  
177 tree built using the second discordant region spanning 57.8~91.5Mb on X  
178 chromosome (Optimization Likelihood: -50166590.1). Meishan (North China domestic  
179 breed) was marked with asterisk (\*) to highlight that this breed has different  
180 haplotypes in the two regions. See Supplementary data 1 for name abbreviation. The  
181 numbers at the nodes represent bootstrap values based on 100 non-parametric  
182 bootstrap replications  
183



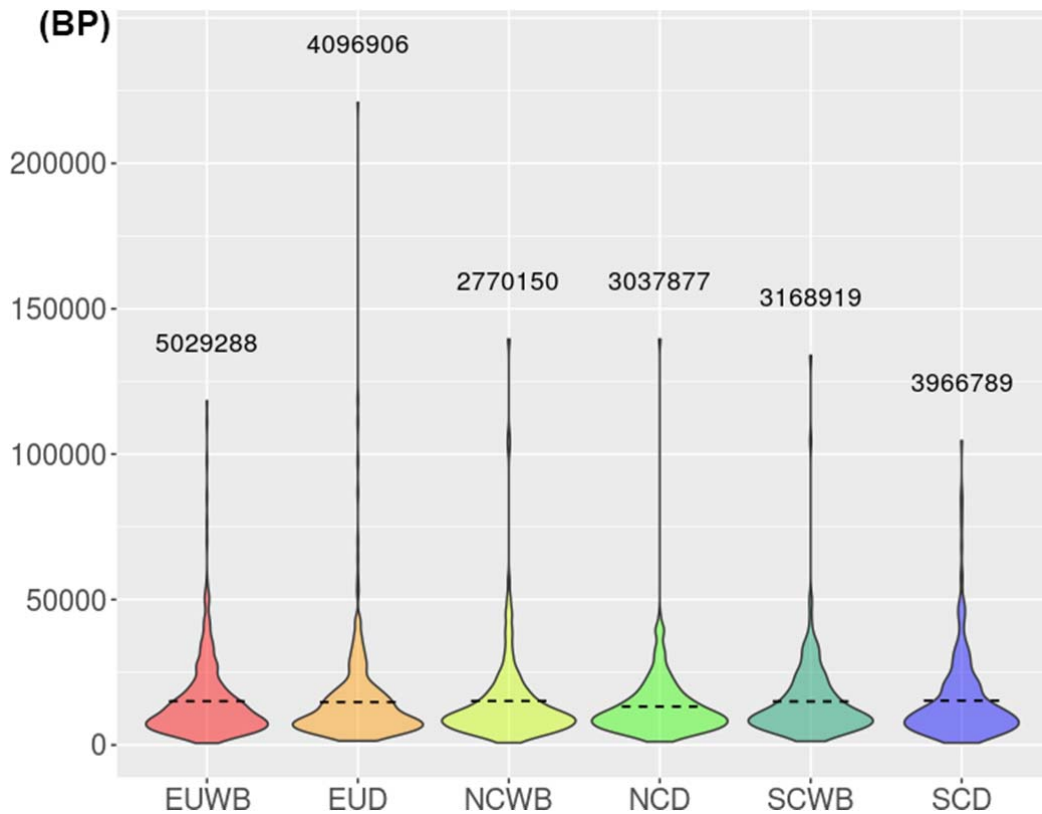


184

185 Supplementary Fig. 17

186 The Venn diagram shows shared autosomal segments with 'ghost' introgression  
 187 signal in sliding-window ABBA-BABA analysis among different swine populations. For  
 188 European pigs, we calculated the average D-value for the clear 'ghost' introgression  
 189 region on X chromosome and any autosomal windows with D-value lower than this  
 190 will be regarded as introgression region. Numbers in parentheses refer to the total  
 191 amount of introgression windows in each population. SCWB = South Chinese wild  
 192 boar, SCD = South Chinese domesticated pig, NCWB = North Chinese wild boar,  
 193 NCD = North Chinese domesticated pig, EUWB = European wild boar, EUD =  
 194 European domesticated pig.

195



196

197 Supplementary Fig. 18

198 Violin plot representing the length of autosomal 'ghost' introgression region in different

199 population inferred by Saguaro. Numbers on top of each violin are the total length of

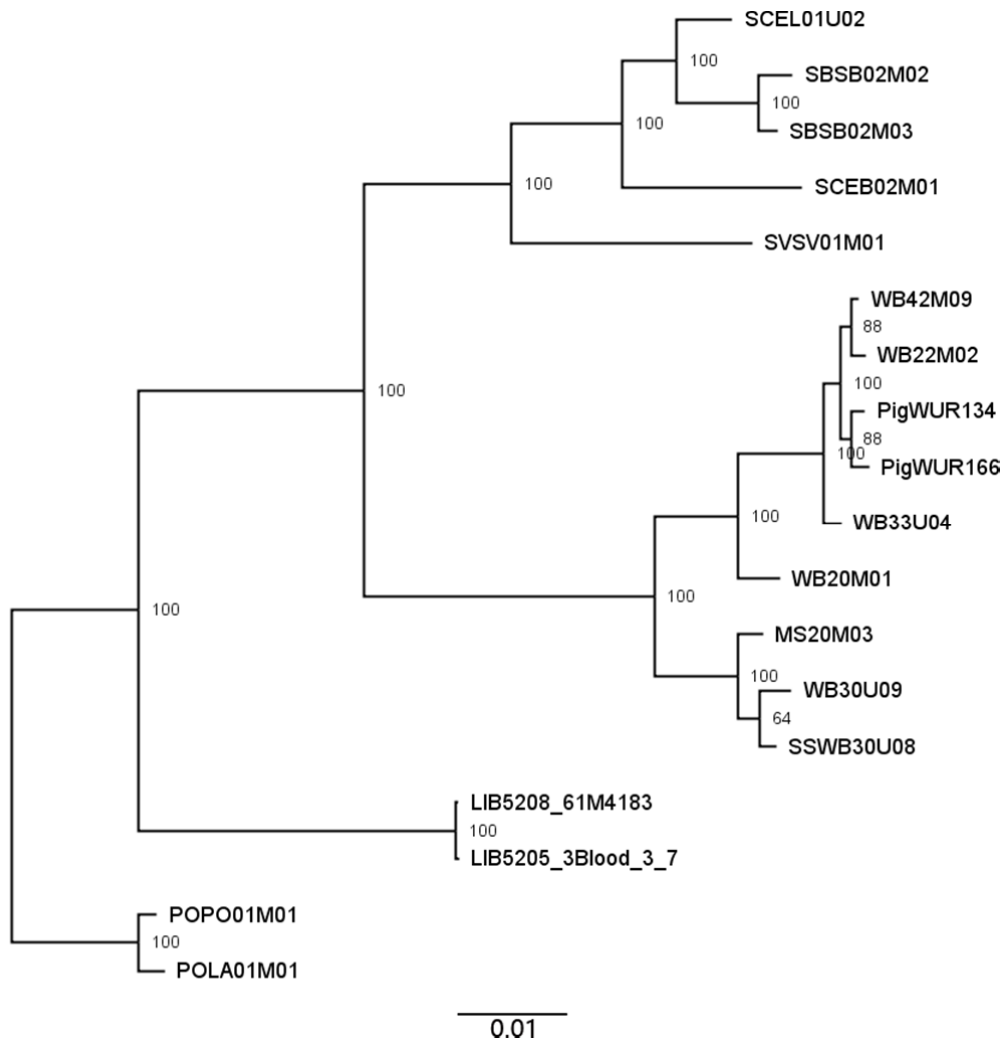
200 introgression region. Dash lines indicated the average length. SCWB = South

201 Chinese wild boar, SCD = South Chinese domesticated pig, NCWB = North Chinese

202 wild boar, NCD = North Chinese domesticated pig, EUWB = European wild boar, EUD

203 = European domesticated pig.

204



205

206 Supplementary Fig. 19

207 Y chromosome topology

208 The maximum likelihood tree of the non-recombining Y chromosome (Optimization

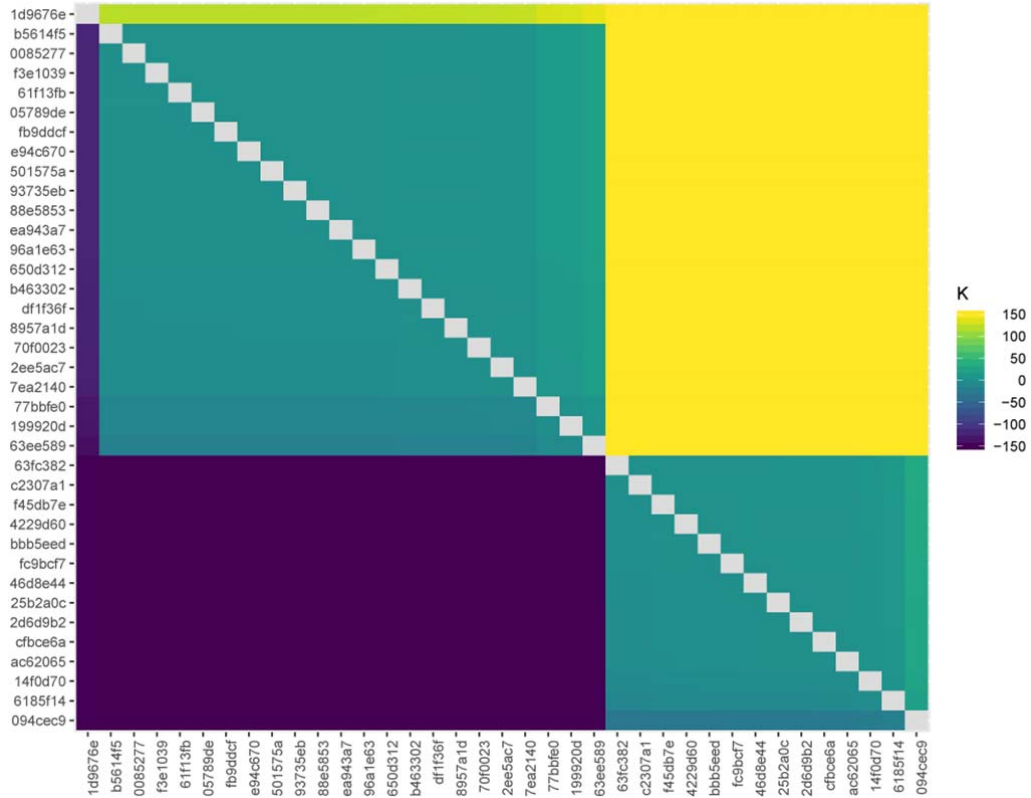
209 Likelihood: -21216029.3). The topology of Y chromosome tree was consistent with the

210 topology of autosomal phylogenetic trees (Supplementary Fig. 1&2). The numbers

211 represent node support inferred from 100 non-parametric bootstrap repetitions.

212 *B.babyrusa* is the outgroup. See Supplementary data 1 for name abbreviation.

213

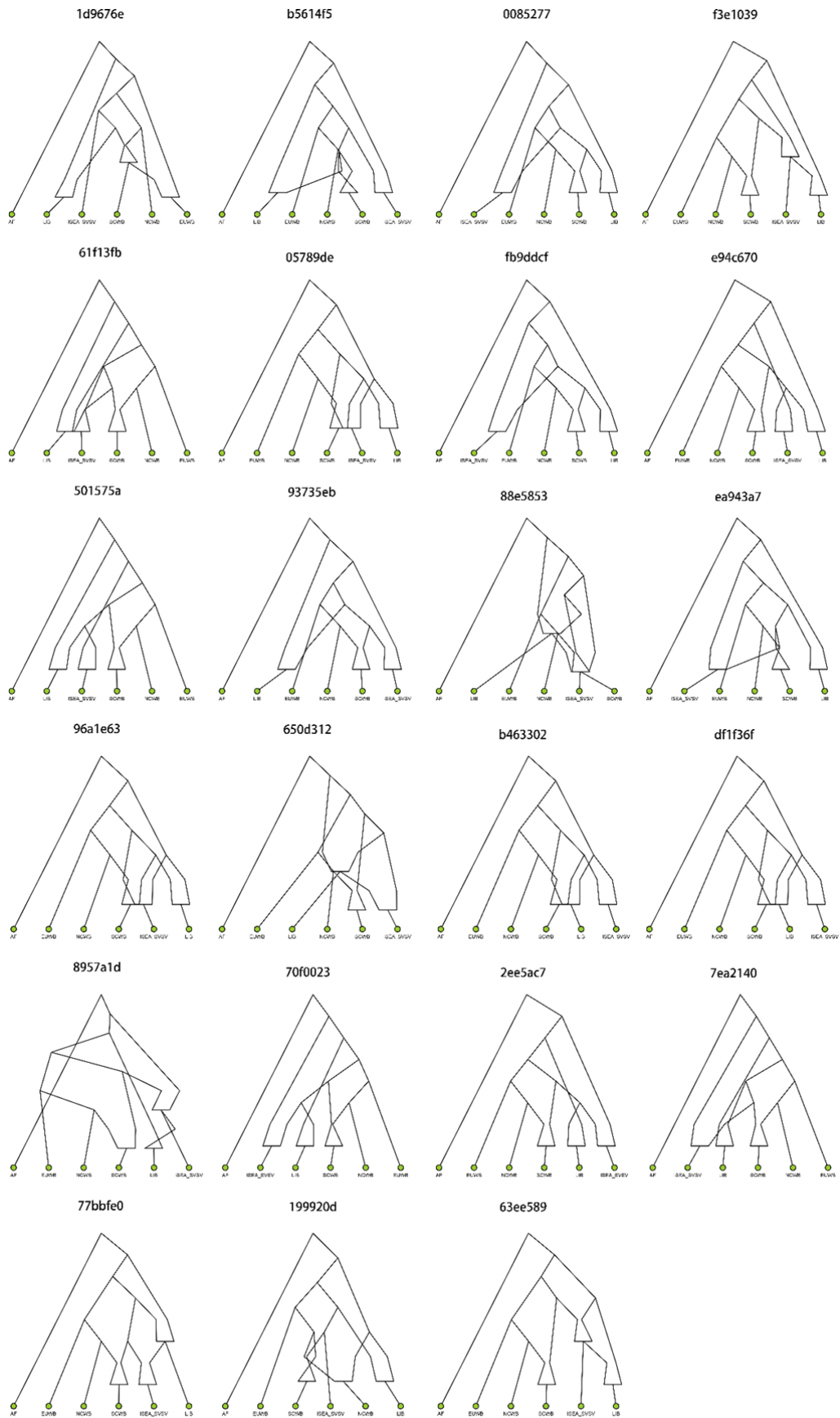


214

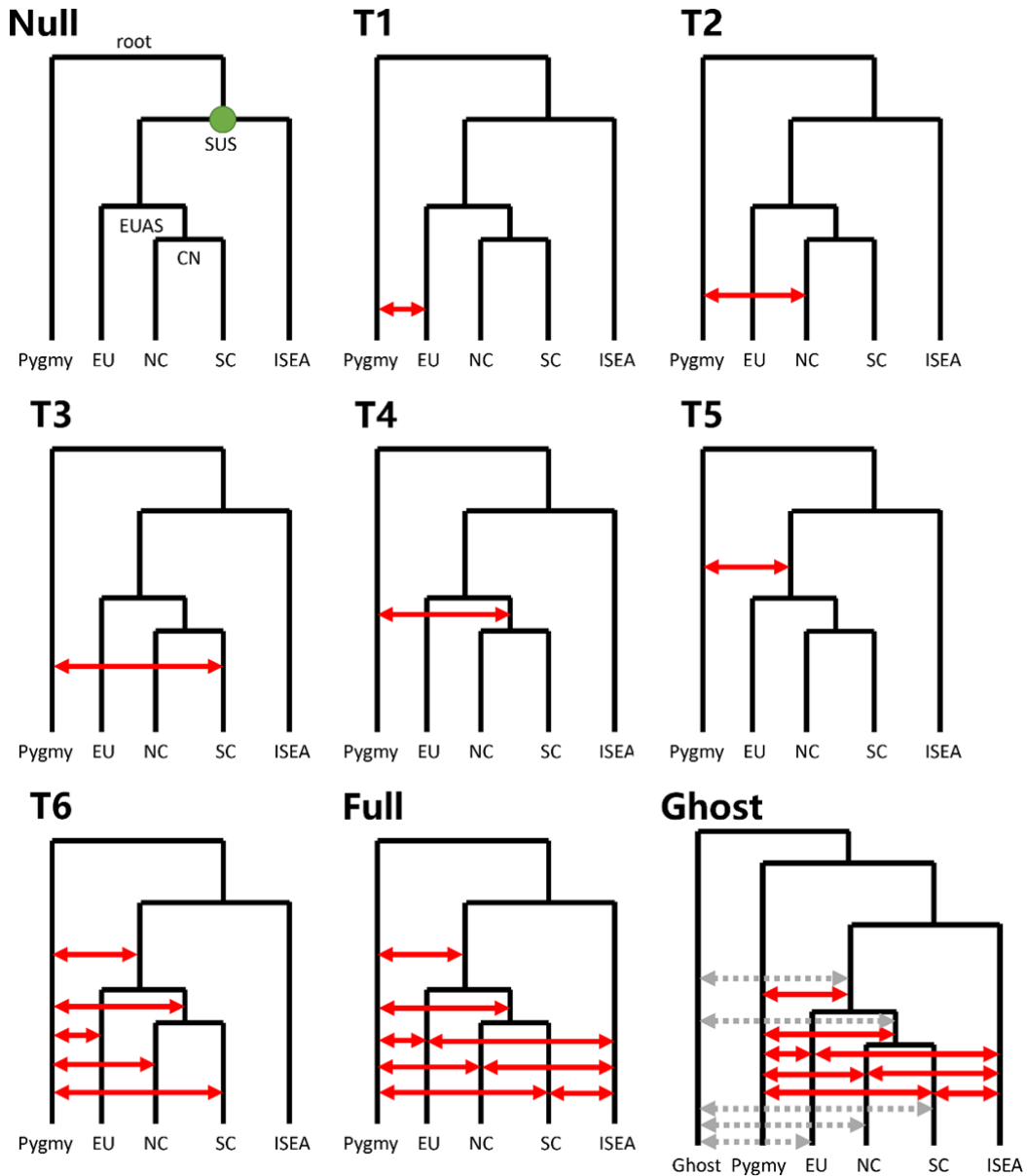
215 Supplementary Fig. 20

216 Heatmap showing the comparison of the Bayes factors for all model pairs.  $|K| > 3$  is  
 217 considered as the significant threshold. Coordinates of the scale indicated the internal  
 218 code for each model.

219



221 Supplementary Fig. 21  
222 23 most well supported models (based on K) used for ADMIXTUREGRAPH. Strings  
223 on top of each diagram indicated the internal code for each model. (AF=  
224 *Phacochoerus africanus*; LIB=pygmy hog; ISEA\_SVSV=*Sus verrucosus*;  
225 EUWB=European wildboar; NCWB=North Chinese wildboar; SCWB=South Chinese  
226 wildboar)  
227



228

229

Supplementary Fig. 22

230

Demographic Models used for G-Phocs analyses. Two-way arrows indicated the

231

migration band designated in each model. Gray dashed arrow indicated the migration

232

band which ghost population involved. Green dot in Null model shows the fossil

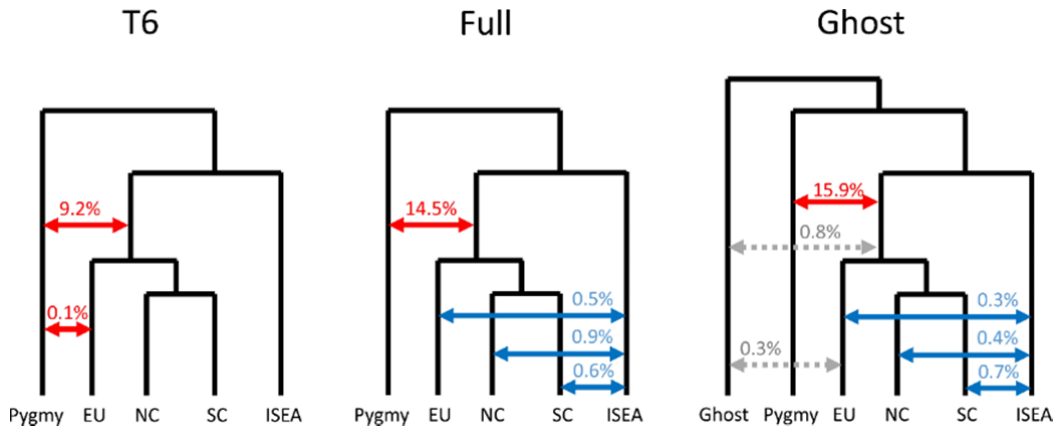
233

calibration. ISEA=Island of South Asia pigs, SC = South Chinese pig, NC = North

234

Chinese pig, EU = European pig.

235



236

237

Supplementary Fig. 23

238

Migration bands are shown in blue, red and gray (see Supplementary Fig. 23) with

239

associated values indicating estimates of total migration rates, which equal the

240

probability that a lineage will migrate through the band during the time period when

241

the two populations co-occur. This figure only shows a subset of models tested and

242

only significant migration signals were delineated (See Supplementary data 4 for

243

further details). G-Phocs result is consistent with our admixture analysis between

244

pygmy hog and *Sus scrofa*. In our model with ghost population, European pigs

245

appeared to have more introgression from ghost population than other pigs, which

246

supports the hypothesis that European pig used to hybrid with other species. Notably,

247

G-Phocs always estimated a migration signal from *Sus scrofa* to other species. A

248

possible reason is that *Sus scrofa* population remain the biggest effective population

249

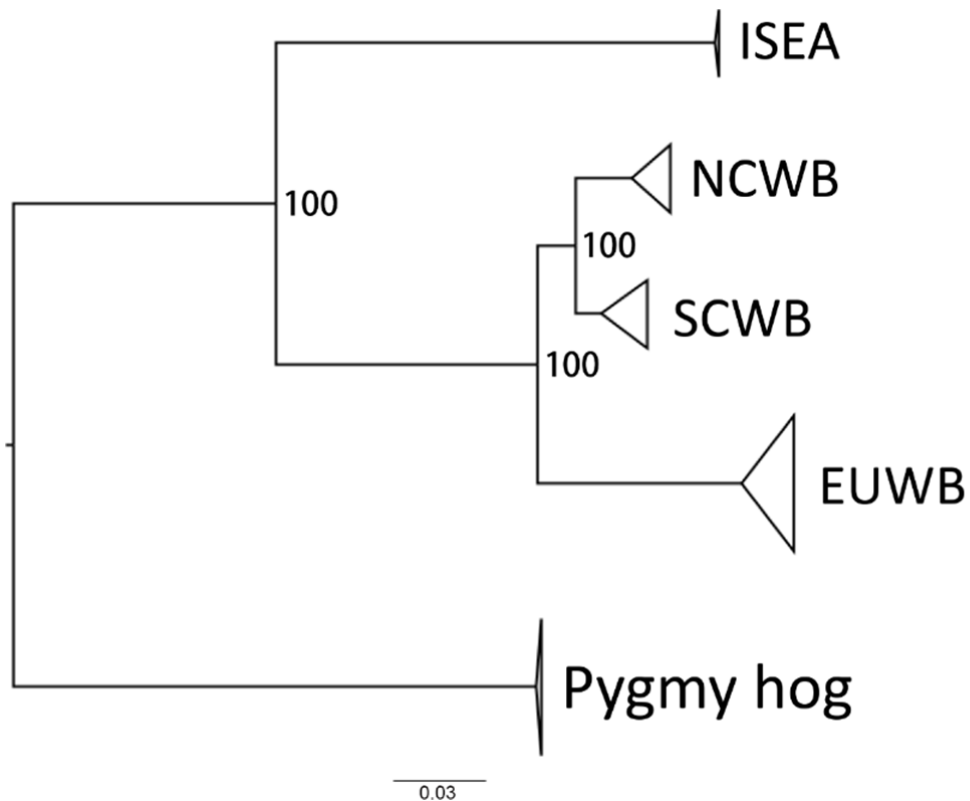
size and genetic diversity, which may overweight ILS. ISEA=Island of South Asia pigs,

250

SC = South Chinese pig, NC = North Chinese pig, EU = European pig.

251



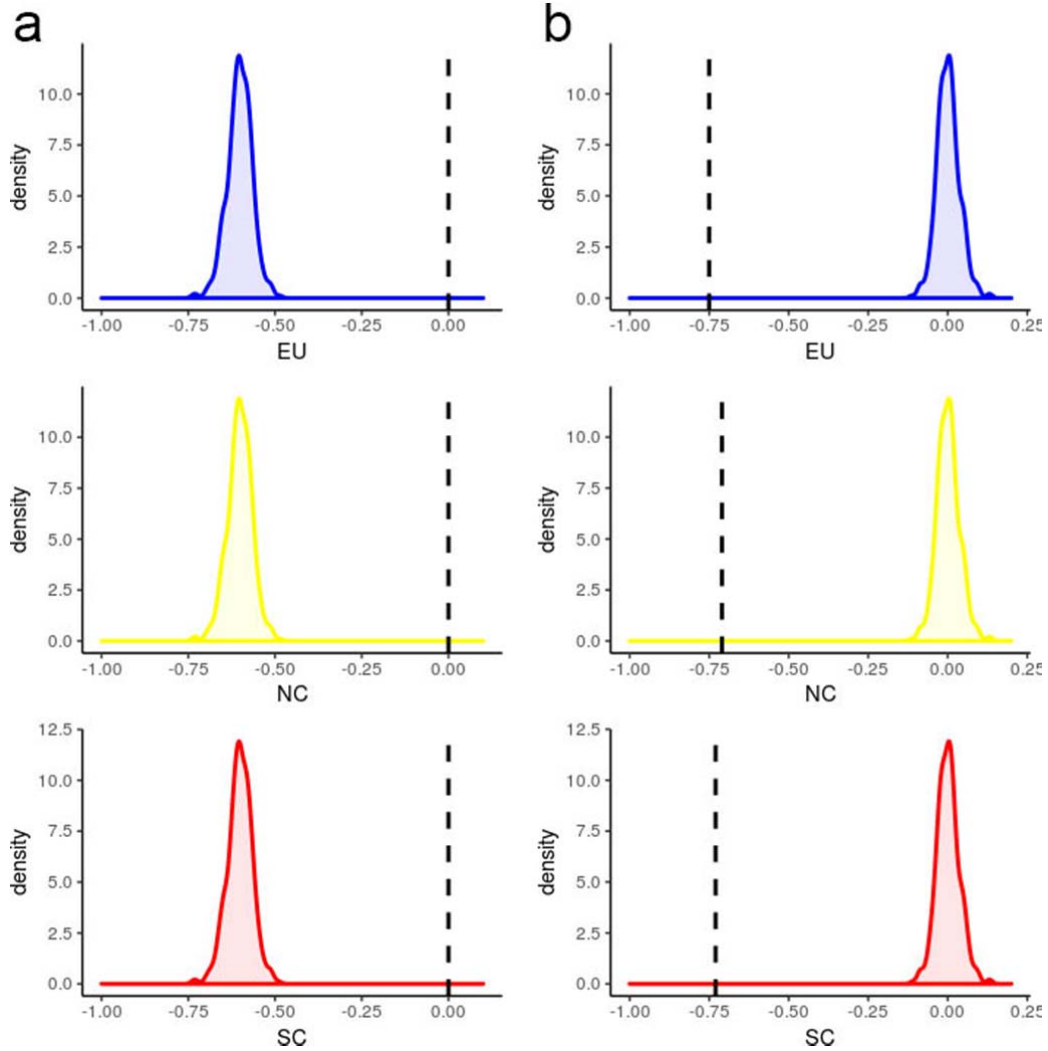


252

253 Supplementary Fig. 24

254 NJ tree based on selected 10,000 neutral loci used in G-PhoCS. Node labels show  
 255 bootstrap values. The result is accordance with our main topology in consensus and  
 256 concatenation tree. . ISEA=Island of South Asia pigs, SCWB = South Chinese wild  
 257 boar, NCWB = North Chinese wild boar, EUWB = European wild boar.

258

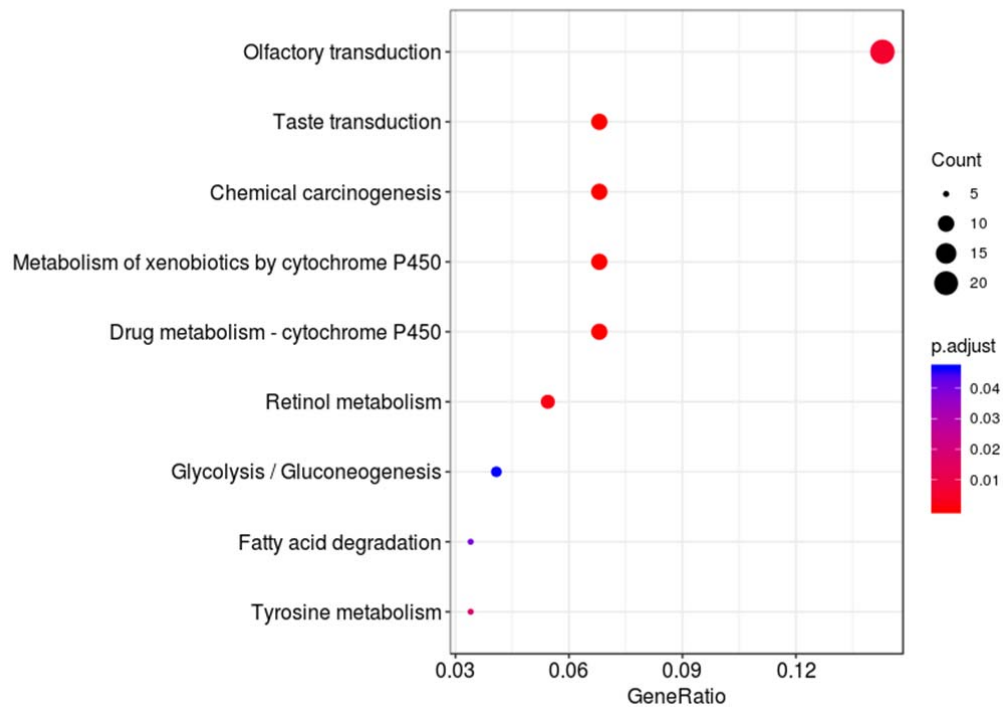


260

261 Supplementary Fig. 25

262 Null distributions of D. a) Each density curve corresponds to the D values (((ISEA,  
 263 Pygmy hog), X), warthog) obtained under null models without gene flow, and the dash  
 264 vertical line corresponds to the threshold value for the introgression fragments in the  
 265 real data. b) Each density curve corresponds to the D values (((ISEA, X), pygmy hog),  
 266 warthog) obtained under null models without gene flow, and the dash vertical line  
 267 corresponds to the D values observed in the read data. All observed D values are  
 268 significant ( $P < 0.001$ ).

269



270

271 Supplementary Fig. 26

272 Functional enrichment result for pigmy hog/*Sus scrofa* introgression gene set

273 Analysis by gene ontology terms and pathways. The number of horizontal axis is the

274 number of enriched genes. The bubble size indicates the ratio of genes in each term

275 or pathway, and different colours correspond to different adjusted p-values. The

276 p-values are adjusted by Benjamini-Hochberg method.

277

278 **Supplementary Note**

279

280 **Phylogenomic analysis**

281 We characterized the evolutionary history of pygmy hog by using one-to-one  
282 autosomal orthologous genes (1:1 - pig:cow). We applied both supermatrix and  
283 supertree phylogenomic methods based on maximum likelihood (ML) using  
284 *Babyrousa babyrussa* as an outgroup<sup>3,4</sup>. The concatenation and consensus methods  
285 resulted in same main topology (Fig. 2a, Supplementary Fig. 1, Supplementary Fig. 2).  
286 The phylogenetics analyses clearly show that *Suinae* consists of three separate  
287 clades, sub-Saharan suids, pygmy hog (*Porcula salvania*) and the genus *Sus*. The  
288 most basal split within the *Suinae* are sub-Saharan suids followed by a highly  
289 supported split of pygmy hogs (BS=100 in supermatrix and CF=1 in supertree) from  
290 all *Sus* species. There has been a longstanding taxonomic debate over the  
291 classification of pygmy hog as either basal to or nested inside the genus *Sus*<sup>5,6</sup>. Up to  
292 now, only one case of molecular classification has been reported for pygmy hog  
293 based on a short fragment of mitochondrial genome<sup>2</sup>. To compare our phylogenetics  
294 results to an earlier study using fragments of mitochondrial DNA, we carried out a  
295 Bayesian phylogenetic analysis using complete mitochondrial genomes  
296 (Supplementary Fig. 3). The resulting topology is consistent with previous studies  
297 confirming pygmy hog as basal to *Sus* and showing mitochondrial replacements  
298 between Sundaland *Sus* species<sup>7</sup>.

299

300 **Estimation of divergence time**

301 We selected autosomal genomic loci supporting the main topology to obtain the basal  
302 divergence between the studied taxa (Fig. 2a, Supplementary Fig. 4). The divergence  
303 between *Babyrousa* and the other *Suidae* (sub-Saharan suids, pygmy hog and *Sus*)  
304 took place during middle Miocene ~14.2 Mya (95% HPD = 17.1-11.6). The basal split  
305 of *Suinae*, between extant sub-Saharan African *Suidae* and the Pygmy-*Sus* clade,  
306 took place during the late Miocene ~10.2 Mya (95% HPD = 12.7-7.9). This split was  
307 followed by the initiated diversification of the ancestors of extant sub-Saharan suids<sup>8-</sup>  
308 <sup>11</sup>. Our results show that the divergence between *Phacochoerus* and *Potamochoerus*  
309 during ~6.2 Mya (95% HPD = 7.7-4.5). Our divergence time estimates suggest that  
310 the initial divergence of the pygmy hog and *Sus* during Miocene/Pliocene boundary  
311 ~6.1Mya (95% HPD = 7.8-4.2). Within the *Sus* genus, DNA analysis suggests the two  
312 major groups of extant species (*Sus scrofa* and ISEA *sus*) diverged from a common  
313 ancestor in early Pliocene ~4.1Mya (95% HPD = 5.5-2.7).

314

315 **Admixture between pygmy hogs and *Sus***

316 We used D-statistics (D) to test whether pygmy hogs had autosomal introgression  
317 from some present-day *Sus* species. We measured the difference in the percent  
318 matching by D (P1, P2, Pygmy hog, sub-Saharan suids) that does not differ  
319 significantly from zero when the derived alleles in the pygmy hog match alleles in the  
320 two *Sus* species/breeds equally often. If D is positive, pygmy hog alleles match alleles  
321 in the second *Sus* species (P2) more often, while if D is negative, pygmy hog alleles

322 match alleles in the first *Sus* species (P1) more often. We performed this test using 11  
323 *Sus* populations (Supplementary data 2). We found that pygmy hogs have a clear  
324 overrepresentation of derived autosomal alleles with *Sus scrofa* compared to the *Sus*  
325 species from Island of south east Asian (ISEA) (Fig. 2b), suggesting that some  
326 admixture between pygmy hog and *Sus scrofa* may have occurred.

327 To further examine this autosomal genome-wide patterns of admixture between  
328 Pygmy hog and *Sus scrofa*, we calculated the D-statistics and fd-statistics in (ISEA,  
329 Pygmy hog, *Sus scrofa*, warthog) for every 100 kb non-overlapping window across  
330 the autosomal genome separately and used D and fd both >0 as a minimum threshold  
331 for introgressed intervals. In order to distinguish between introgression and ancestral  
332 variation, we also calculated DNA sequence divergence (dxy) for each candidate  
333 introgression interval and compared this with the autosomal chromosomal mean dxy,  
334 because introgressed regions generally show lower genetic divergence<sup>12</sup>. We  
335 compared the result between *Sus scrofa* from Europe, Northern China and Southern  
336 China. We identified 636 intervals of which 427 (67.1%) were shared among *Sus*  
337 *scrofa* (Supplementary Fig. 6), which suggested an ancestral gene flow between  
338 pygmy hog and the common ancestor of *Sus scrofa*.

339

#### 340 **X chromosome analysis**

341 We reconstructed a phylogenetic tree based on the whole X chromosome and found  
342 remarkable inconsistency between the phylogenetic topology of the X chromosome  
343 and the autosomes, more specific in that pygmy hog significantly group with European  
344 and North Chinese *Sus scrofa* to the exclusion of Southern *Sus scrofa* and ISEA *Sus*  
345 species (Supplementary Fig. 8). To investigate this discrepancy, we used SAGUARO,  
346 which combines a hidden Markov model and a self-organizing map to infer both the  
347 trees and window boundaries. We mapped occurrences of the six most frequently  
348 observed unrooted ML topologies of six Suidae species along the 19 chromosomes  
349 (Fig. 3, Supplementary data 3, Supplementary Fig. 9). SAGUARO identified a  
350 ~40.6Mb (46.2-86.8 Mb) region in X chromosome, where pygmy hog clustered with  
351 ISEA and South China *Sus scrofa* breeds, while the European *Sus scrofa* clade  
352 appeared to be outgroup to this cluster. In the meantime, Northern Chinese domestic  
353 pig shows a splicing pattern, which clustered with European pig/North China pig in  
354 46.2-57.1 Mbp and clustered with South China pig in 57.1-86.8 Mbp. To further  
355 confirm this observation, we carried out D-statistic analysis in sliding windows,  
356 expecting symmetry ( $D = 0$ ) under the null hypothesis corresponding to no gene flow  
357 derived from the chromosomes (ISEA *Sus*, *Sus scrofa*, Pygmy hog, warthog)  
358 (Supplementary Fig. 10-15). Since the fixed frame-size sliding-window test may  
359 over-estimating the length of introgression region, for each quadruplet, we also ran  
360 individual SAGUARO analysis. The results suggest that within this genomic region,  
361 sequences of European/North China pigs have an ancient origin from a ghost lineage  
362 older than the split of pygmy hogs but younger than Sub-saharan suids. Due to the  
363 distinct haplotype between northern and southern Chinese *Sus scrofa* population in X  
364 chromosome, there are two prominent topologies drawn in the putative introgression  
365 region. For those two genetic regions, we reconstructed phylogenetic tree separately

366 and obtained the same phylogenetic relationship among *Suinae* as SAGUARO  
367 estimated (Supplementary Fig. 16). We also identified the autosomal regions which  
368 support the introgression event found in the X chromosome. The sliding-window  
369 D-statistic analyses suggest a nearly equal amount of introgressed windows among  
370 different population (Supplementary Fig. 17). Similar average length of introgression  
371 region was estimated by the quadruplet SAGUARO analyses present (Supplementary  
372 Fig. 18).

373

#### 374 **D statistic under models of no gene flow**

375 We used the divergence time estimated from MCMCtree. We assumed an effective  
376 population size of 100,000 individuals and a generation time of 5 years, a  
377 recombination rate 0.8 cM/Mb ( $0.8 \times 10^{-8}$  per bp per generation) and a mutation rate  
378  $1.25 \times 10^{-8}$ . All simulations were done in a msprime-based software momi2<sup>13</sup>. We  
379 simulated 10000 loci with length of 100kb:

380 #an simplify model for Suidae phylogeny without gene flow

381 model = momi.DemographicModel(N\_e=1e5, gen\_time=5, muts\_per\_gen=1.25e-8)

382 model.add\_leaf("PYGMY", N=1e5)

383 model.add\_leaf("EU", N=1e5)

384 model.add\_leaf("NC", N=1e5)

385 model.add\_leaf("SC", N=1e5)

386 model.add\_leaf("ISEA", N=1e5)

387 model.add\_leaf("AF", N=1e5)

388 model.move\_lineages("NC", "SC", t=800000)

389 model.move\_lineages("EU", "SC", t=900000)

390 model.move\_lineages("SC", "ISEA", t=4100000)

391 model.move\_lineages("ISEA", "PYGMY", t=6100000)

392 model.move\_lineages("PYGMY", "AF", t=10200000)

393 We compared these simulations with read data from our study. D-statistics were  
394 calculated using quadruplets (((ISEA, Pygmy hog), X), warthog) and (((ISEA, X),  
395 pygmy hog), warthog). Comparing the D values to the simulations under the null

396 model, we find that the observed D statistics are in the extreme tail, with  $p < 0.001$  for  
397 all comparisons. Thus, these simulations suggest that it is very unlikely that D values  
398 that low are due to incomplete lineage sorting in the ancestral population, consistent  
399 with our claim of introgression. However, we need to be aware that this method  
400 requires a very detailed and precise demographic model to get a better assessment.  
401 The historical demographic information of pygmy hog and the ancestral population of  
402 Suinae species are still deficient. Here, we can only fit in a simplified model.  
403 Inaccurate of effective ancestral population size and bottleneck event may led to  
404 over-/underestimation of ILS.

405  
406

#### 407 **Demographic model for Pygmy hog and *Sus***

408 To better assess the demographic implications of a separate species origin, rather  
409 than one due entirely to admixture, we performed demographic inference by applying  
410 G-PhoCS (Generalized Phylogenetic Coalescent Sampler) to simple branching  
411 models. G-PhoCS analyses are computational demanding, thus, due to  
412 computational constraints, we focused on 12 individuals, from the following different  
413 population or species: pygmy hog, ISEA (represented by *S. verrucosus* which has  
414 little detectable admixture with *Sus scrofa*), European wild boar, North and South  
415 China wild boar. Our objective was to infer rates of gene flow from between pygmy  
416 hog and *Sus scrofa* in the context of a complete demographic model that includes  
417 population divergence and changes in ancestral population sizes. We applied  
418 G-PhoCS to a multiple sequence alignment of neutral genome fragments of the 12  
419 genomes. From this neutral dataset, we first constructed a NJ tree which is  
420 concordance with the main topology of the used population (Supplementary Fig. 24).  
421 We therefore assumed a plausible topology based on the main phylogenetic tree for  
422 the population phylogeny and fitted a variety of different gene flow models  
423 (Supplementary Fig. 22). G-PhoCS models' migration bands allow a test of admixture  
424 from D-statistic. Under the assumed branching structure, we inferred high probability  
425 of gene flow between the common ancestor of *Sus scrofa* and pygmy hog. Using  
426 G-PhoCS, we were also able to examine signatures of admixture in the ISEA  
427 population, and found significant gene flow between the ISEA *Sus* and Asian wild boar,  
428 which is consisted with former research<sup>7</sup>. Furthermore, we also applied a model with a  
429 basal ghost population to test the results obtained from the X-chromosome analysis.  
430 This model confirmed a post-speciation gene flow between the common ancestor of  
431 wild boar and the ghost population (Supplementary data 4, Supplementary Fig. 23).  
432 Also, European wild boar shows a secondary gene flow from the ghost population.  
433 We interpret this pattern as isolation by distance, as mainland Asia is the start point of  
434 *Sus scrofa* expansion, which allow the ancestor of extant Asian pig have more chance  
435 to hybrid with non-introgression population. As we can't properly match a  
436 non-introgression population into our model, G-PhoCS may assign the exceeding  
437 introgression allele in European wild boar as a potential gene flow signal.

438  
439

#### 439 **Enrichment analysis**

440 We applied a functional annotation analysis using PANTHER v.11<sup>14</sup> on the candidate  
441 introgressed genes (Supplementary data 5). Genes from the pygmy hog/Sus scrofa  
442 introgression and the Sus scrofa/ghost lineage introgression were analyzed  
443 separately. Gene-enrichment analyses were performed using clusterProfiler<sup>15</sup>. False  
444 discovery rate (FDR) was performed to adjust P-values using the Benjamini and  
445 Hochberg method. A P-value of <0.05 was used as the cut-off criterion. For the pygmy  
446 hog/Sus scrofa introgression gene set, significant enriched KEGG pathway was  
447 showed in Supplementary Fig. 27. For the Sus scrofa/ghost lineage introgression, we  
448 only identified a limited number of candidate genes we identified, and no significant  
449 enrichment was found.  
450  
451



452 **References**

- 453 1. Bosse, M. *et al.* Genomic analysis reveals selection for Asian genes in  
454 European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392  
455 (2014).
- 456 2. Funk, S. M. *et al.* The pygmy hog is a unique genus: 19th century taxonomists  
457 got it right first time round. *Mol. Phylogenet. Evol.* **45**, 427–436 (2007).
- 458 3. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and  
459 post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- 460 4. Frantz, L. A. F. Speciation and domestication in Suiformes: a genomic  
461 perspective. (Wageningen University, 2015).
- 462 5. Hodgson, B. H. On a new form of the hog kind or Suidae. *J. Asiat. Soc. Bengal*  
463 *16*, **16**, 423–428 (1847).
- 464 6. Oliver, W. L. R. & Deb Roy, S. The pygmy hog (*Sus salvanius*). *Pigs, Peccaries*  
465 *Hippos status Surv. Conserv. action plan. IUCN, Gland* 121–129 (1993).
- 466 7. Frantz, L. A. F. *et al.* Genome sequencing reveals fine scale diversification and  
467 reticulation history during speciation in *Sus*. *Genome Biol.* **14**, R107 (2013).
- 468 8. Gongora, J. *et al.* Rethinking the evolution of extant sub-Saharan African suids  
469 (*Suidae*, *Artiodactyla*). *Zool. Scr.* **40**, 327–335 (2011).
- 470 9. Brunet, M. & White, T. D. Deux nouvelles espèces de Suini (Mammalia, Suidae)  
471 du continent Africain (Éthiopie; Tchad). *Comptes Rendus l'Academie Sci. - Ser.*  
472 *Ila Sci. la Terre des Planetes* **332**, 51–57 (2001).
- 473 10. Haile-Selassie, Y. & Simpson, S. W. A new species of *Kolpochoerus*  
474 (Mammalia: Suidae) from the Pliocene of Central Afar, Ethiopia: Its Taxonomy  
475 and Phylogenetic Relationships. *J. Mamm. Evol.* **20**, 115–127 (2013).
- 476 11. Pickford, M. Ancestors of broom's pigs. *Trans. R. Soc. South Africa* **67**, 17–35  
477 (2012).
- 478 12. Smith, J. & Kronforst, M. R. Do *Heliconius* butterfly species exchange mimicry  
479 alleles? *Biol. Lett.* **9**, 20130503–20130503 (2013).
- 480 13. Tortereau, F. *et al.* A high density recombination map of the pig reveals a  
481 correlation between sex-specific recombination and GC content. *BMC*  
482 *Genomics* **13**, 586 (2012).
- 483 14. Westbury, M. V *et al.* Extended and Continuous Decline in Effective Population  
484 Size Results in Low Genomic Diversity in the World's Rarest Hyena Species,  
485 the Brown Hyena. *Mol. Biol. Evol.* **35**, 1225–1237 (2018).
- 486 15. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for  
487 Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* **16**,  
488 284–287 (2012).
- 489