Pervasive selection pressure in wild and domestic pigs

J. Leno-Colorado[1], S. Guirao-Rico[12], M. Pérez-Enciso[134], S. E. Ramos-Onsins[1*]

[1] Centre for Research in Agricultural Genomics (CRAG) Consortium CSIC-IRTA-UAB-UB. 08193 Bellaterra, Spain.

[2] Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, Barcelona, Spain.

[3] Universitat Autònoma de Barcelona, Department of Animal Science, Bellaterra, Spain.

[4] Institut Català de Recerca I Estudis Avançats (ICREA), Carrer de Lluís Companys 23, Barcelona, Spain.

*Corresponding author: sebastian.ramos@cragenomica.es

ABSTRACT

Animal domestication typically affected numerous polygenic quantitative traits, such as behavior, development and reproduction. However, uncovering the genetic basis of quantitative trait variation is challenging, since they are caused by small allele-frequency changes. To date, only a few causative mutations related to domestication processes have been reported, strengthening the hypothesis that small effect variants have a prominent role. So far, approaches on domestication have been limited to the detection of the global effect of domestication on deleterious mutations and on strong beneficial variants, ignoring the importance of variants with small selective effects. To overcome these difficulties, here we propose to estimate the proportion of beneficial variants based on the asymptotic MacDonald Kreitman (MK) method, according to estimates of variability based on frequency spectrum. We applied this approach to the pig species, analyzing 46 complete genome sequences from 20 European wild boars, 6 Iberian and 20 Large White pigs at different molecular scales: gene, metabolic pathway and whole-genome.

Descriptive variability analyses on pig populations indicate that domestic and wild pig populations do not differ in nonsynonymous fixed mutations. Instead, most variants are shared among them, despite that the phenotypes of wild and domestic individuals are clearly divergent. Additionally, asymptotic MK plots based on summary statistics show that small effects variants may affect the final calculation of α, the proportion of beneficial mutations. The distribution of fitness effects inferred with Approximate Bayesian Computation analysis indicates that both wild and domestic pigs display an important quantity of deleterious mutations at low frequency (~83% of total mutations) and a high number of nearly-neutral mutations (~17%) that may have a significant effect on the evolution of domestic and wild populations. Exclusive mutations show that recent demographic changes have severely affected the fitness of populations, especially of the local Iberian breed. Finally, the median proportion of the strong favorable mutations are very scarce in all cases (≤ 0.2%). The median estimated alpha values (weak and strong favorable) are 0.9% for wild and domestic pigs.

Keywords: Domestication, Distribution of fitness effects, Proportion of beneficial mutations, Approximate Bayesian calculation, Polygenic selection, Population genomics

INTRODUCTION

Domestic animal histories are evolutionary experiments that have often lasted for millennia, with the result of dramatic phenotypic changes to suit human needs. In addition, domestic species can be structured into subpopulations (breeds) that are partly or completely genetically isolated and can display a wide catalog of specific phenotypes. Therefore, they offer a material of utmost interest to study the interplay of demography and accelerated adaptation. However, as their demographic history can be quite complex, many events remain unknown or poorly documented nowadays.

The pig (*Sus scrofa*) is a particularly interesting species because of its domestication history and its relatively well-annotated genome. *S. scrofa* originated in Southeast Asia ~ 4 MYA and spread throughout Eurasia ~1.2 MYA, colonizing all climates except the driest (Frantz et al. 2013). Subsequently, the pig was domesticated from local wild boars independently in both Asia and Europe ~9,000 years ago. To complicate the story, modern European domestic pig breeds were crossed with Asian domestic pigs during the late 17th century and onwards. In breeds such as Large White (LW), approximately 30% of the genome is estimated to be of Asian origin (Bosse, Megens, Madsen, et al. 2014). Nevertheless, some local European breeds, such as the Iberian breed (IB), were spared genetic contact with Asian pigs and no evidence of genetic introgression has been found in this breed (Alves et al. 2003, Esteve-Codina et al. 2013). Moreover, domestic breeds have different recent demographic histories. For instance, the IB breed suffered a dramatic reduction of its effective population size during the last century (Alves et al. 2006), whereas commercial breeds such as Duroc or LW have been introgressed with Asian pigs (Bosse, Megens, Frantz, et al. 2014).

Differences in the effective population size, demographic histories and artificial selective pressures between pig populations could result in differences among their evolutionary rates. In addition to possible differences in the rate of evolution between populations, there may be differences in the rate of evolution between genes within genomes. For instance, it is known that the strength of the selection is affected by the position of the genes in the networks in which they participate. Genes that are more central in a network and are more connected with other genes are more evolutionarily constrained, while peripheral genes are more prone to be under adaptive selection (Fraser et al. 2002; Hahn and Kern 2005; Montanucci et al. 2011; Alvarez-Ponce and Fares 2012). Furthermore,

70 it has been observed that the evolutionary rate, within a metabolic pathway, increases as we move
71 downstream, possibly because upstream genes are more pleiotropic, since they are involved in
72 more functions and hence these genes are probably more conserved (Rausher, Miller, and Tiffin
73 1999; Riley, Jin, and Gibson 2003; Livingstone and Anderson 2009; Ramsay, Rieseberg, and
74 Ritland 2009).

75 So far, the nature of the underlying genetic changes caused by domestication and ensuing artificial
76 breeding is still under debate. While the most prevalent view is that regulatory changes have been
77 targeted (Anderson 2013), several other studies underline the influence of protein coding changes
78 (Rubin et al. 2012). Other authors have reported an increase in the rate of deleterious mutations in
79 domestic pigs compared to their wild counterparts (Cruz, Vilà, and Webster 2008; Renaut and
80 Rieseberg 2015; Pérez-Enciso et al. 2017; Leno-Colorado et al. 2017). Others, as in Makino et al.
81 (2018) detected a general pattern of reduction of variability in domestic populations in relation to
82 their wild counterpart, and a higher nonsynonymous/synonymous ratio across the frequency
83 spectrum. These patterns were compatible with the effect of strong bottlenecks in domestic
84 populations and the higher accumulation of deleterious mutations. Interestingly, the same authors
85 observed that the opposite trend has been observed in pigs. Moreover, most of the previous studies
86 have focused on genes of major effect with clear signals of selective sweeps. In those studies, the
87 hallmarks of positive selection were detected by a valley of reduced variation and/or population
88 differentiation that spans a relatively large region (e.g., Amaral et al. 2011, Rubin et al. 2012,
89 Frantz et al. 2013, Wilkinsonet al. 2013) but also by haplotype structure and homozygosity blocks
90 (e.g., Fang et al. 2011, Bosse et al. 2012, Li et al. 2013). Some of these studies have detected recent
91 breed specific signals of selection attributed to the domestication process (Li et al. 2014, Kim et
92 al. 2015). Nevertheless, the signals were too scarce to explain the domestication process. Other
93 studies have tried to elucidate the effect that domestication has at a whole-genome scale and on
94 the fitness of individuals of domestic populations (e.g., Cruz et al. 2008, MacEachern et al. 2009,
95 Kono et al. 2016, Perez-Enciso et al. 2016, Makino et al. 2018, Chen et al. 2018, Orlando and
96 Librado 2019). For instance, an excess of deleterious variants has been observed in a number of
97 domestic animal and plants (e.g., contrasting nonsynonymous versus synonymous polymorphism
98 ratios, Chen et al. 2018, using the MacDonald framework, MacEachern et al. 2009, contrasting
99 ancestors with ancient DNA, Orlando and Librado 2019, combining the frequency of
100 polymorphisms with functional effects and divergence, Kono et al. 2016, Makino et al. 2018).

101 Kono et al. (2016) and Perez-Enciso et al. (2016) found an excess of detrimental variants affecting
102 phenotypes of interest, suggesting, as we previously mention above, that protein sequence may
103 have a stronger influence than regulatory changes in the domestication process. Kono et al. (2016)
104 also showed that null alleles are uncommon in domestic animal species (also reviewed by
105 Anderson 2013), suggesting that phenotypic changes involved in domestication are produced by
106 the accumulation of consecutive mutations that modify these functions under selection. Finally,
107 the possible presence of beneficial mutations during the domestication process has also been
108 reported (Perez-Enciso 2016).

109

110 Here, we are interested in determining the proportion and the selective effects of protein-coding
111 variants in wild and domestic pig genomes to understand their role in the domestication process.
112 Particularly, we aimed to test the role of both new and extant mutations in the domestication
113 process and whether the phenotypes associated with domestic breeds are the product of a large
114 number of variants with weak selective effects, as suggested by previous results. To achieve this,
115 we have investigated the differential effects of selection on coding sequences at the different
116 molecular scales (gene, metabolic pathway and whole-genome) in two domestic and one wild pig
117 population using the McDonald-Kreitman framework (McDonald and Kreitman 1991, Eyre-
118 Walker 2006, Fay 2011) and have inferred the distribution of fitness effects (DFE) while taking
119 into account the effect of different demographic scenarios. Interestingly, the analysis was
120 performed using variability estimators that allow including positions with missing data (Ferretti,
121 Raineri, and Ramos-Onsins 2012).

122 Our results support the hypothesis that changes in allele frequencies in coding variants with weak
123 positive selective effect have been relevant for pig domestication, as evidenced by a relatively high
124 number of nonsynonymous variants segregating at medium and high frequencies and by the
125 obtained estimates of the Distribution of Fitness Effects in domestic pig populations.

126

127

128 **MATERIALS AND METHODS**

129 **Biological samples**

130 We analyzed a sample of 46 pig (*Sus scrofa*) genomes (Table S1). These pigs correspond to
131 European wild boars (WB, n = 20) and domestic pigs, which are represented by the Iberian
132 Guadyerbas (IB, n = 6) and LargeWhite (LW, n = 20) breeds. These two domestic breeds were
133 selected because they have very different interesting features: IB is a local breed that has been
134 under weak artificial selection intensity and with no documented evidence of Asian introgression.
135 LW, in contrast, is a commercial breed undergoing strong artificial selection with a deliberate
136 admixture with Asian pigs (Bosse, Megens, Madsen, et al. 2014; Groenen 2016). To analyze the
137 divergence between the different breeds, we used the consensus ancestral reference sequence
138 obtained from combining the information from several *Sus* species (*S. barbatus*, *S. cebifrons*, *S.*
139 *verrucosus*, *S. celebensi*, approximately 4.2 MYA of divergence) and the African warthog
140 (*Phacochoerus africanus*, around ~10 MYA of divergence) as an outgroup, as detailed in Bianco
141 et al. (2015). The sequences are available in public databases (Rubin et al. 2012; Ramírez *et al.*
142 2014; Bianco *et al.* 2015; Frantz et al. 2015; Moon *et al.* 2015, Esteve-Codina *et al.* 2013, Leno *et*
143 *al.* 2017) and were downloaded from the short read archive (SRA,
144 http://www.ncbi.nlm.nih.gov/sra).

145

146 **Mapping and genotyping analysis**

147 Raw reads for each pig genome were mapped against the reference genome assembly (Sscrofa10.2,
148 Groenen et al. 2012) using *BWA mem* option (H. Li and Durbin 2009). PCR duplicates were
149 removed using *SAMtools rmdup* v 0.1.19 (H. Li et al. 2009) and mapped reads were realigned
150 around indels with the *GATK IndelRealigner* tool (McKenna et al. 2010). Genotype calling was
151 performed with *SAMtools mpileup* and *bcftools call* v 1.3.0 (H. Li et al. 2009) for each individual
152 separately. We set a minimum (5x) and a maximum depth (twice the average sample's depth plus
153 one) to call a SNP. Base quality was set to 20 (*P*-value=1e-2). Homozygous blocks (regions of
154 contiguous positions with the same nucleotide as the reference genome) were also called,
155 following the same criteria (i.e., minimum and maximum coverage and base quality) as with the
156 SNPs and using *samtools depth* utility, *BEDtools* (Quinlan 2014) and custom scripts (available at
157 https://github.com/miguelperezenciso/NGSpipeline, Pérez-Enciso *et al*., 2017). This resulted in a

158 *gVCF* file per individual with the combined information about variant calls and non-varying
159 positions. Next, each *gVCF* file was converted into a fasta file and all fasta files were subsequently
160 merged to obtain a multindividual *gVCF* file (Pérez-Enciso et al. 2017), which comprised the
161 whole set of the SNPs of the 46 pigs.

162

163 **Analysis of the population structure of the samples**

164 A principal component analysis (PCA) was performed using the total number of SNPs to analyze
165 the population structure. First, genotypes were converted to alternative allele frequency, being 0
166 for the homozygous reference genotype (0/0), 0.5 for the heterozygous genotype (0/1) and 1 for
167 the homozygous alternative genotype (1/1). For cases of missing genotype (./.), these were
168 replaced by the average SNP frequency across all individuals. We used the function *tcrossprod()*
169 from R version 3.3.1 (2016) to obtain the matrix of covariates from the frequencies matrix. Finally,
170 we obtained the principal components from the Eigen-value decomposition with the R function
171 *eigen()*.

172

173 **Estimation of levels and patterns of variability**

174 Genetic diversity and divergence per pig population were estimated using *mstatspop* software
175 (Nevado, Ramos-Onsins, and Perez-Enciso 2014; Bianco et al. 2015; Guirao-Rico et al. 2018,
176 available from the authors, https://github.com/cragenomica/mstatspop). The *multi-VCF* file was
177 converted into a *tfasta* (transposed *fasta*) file and *mstatspop* was run on either the whole genome,
178 using 5 Mb windows and at each functional coding region. We used four different estimators of
179 nucleotide variability that takes into account missing data (Ferretti, Raineri, and Ramos-Onsins
180 2012): Watterson (Watterson 1975), Tajima (Tajima 1983), Fu&Li (Fu and Li 1993) and
181 Fay&Wu's estimators (Fay and Wu 2000). Specifically, variability was estimated using the
182 Ferretti, Raineri, and Ramos-Onsins (2012) expression:

$$\hat{\theta} = \frac{1}{L}\sum_{x=1}^{L}\sum_{i=1}^{n_x-1} i\omega_{i,n_x}\xi_i(x), \quad \frac{1}{L}\sum_{x=1}^{L}\sum_{i=1}^{n_x-1}\omega_{i,n_x} = 1$$

183

184 *(Equation 1)*

185 where (ω_i) is the weight for the different variability estimators such as $\omega_i = n/(i(n-i)(1+\partial_{i,n-i}))$ for
186 the Watterson estimator, $\omega_i = n/(1+\partial_{i,n-i})$ for the Tajima estimator (both for folded spectrum), $\omega_i =$
187 i for the Fay&Wu estimator and $\omega_i = 1$, $\omega_{i>1} = 0$ for the Fu&Li estimator (Achaz 2009).

188

189 **Filtering for artifactual effects**

190 A preliminary analysis of the variability showed a moderate negative correlation (~ 0.3) between
191 the levels of variability and divergence and the proportion of missing data for each gene. To
192 eliminate this artifactual correlation, we plotted the estimators of variability and divergence versus
193 the ratio of missing data and eliminated those genes that showed a ratio of missing data greater
194 than 0.3. Since this filtering was not enough to completely remove the bias, we also removed genes
195 with extreme values of variability and divergence (higher than 99% quantile of the total genes).
196 The remaining ~13,500 genes (70% of the total annotated genes) showed a low or null correlation
197 with missing data and were used in the present analysis (Table S2).

198

199 **Estimation of the proportion of adaptive of variants**

200 Under the neutral model, the majority of polymorphisms segregating in a population are neutral
201 and only a small number of positively selected variants segregates for a short time on their way to
202 loss or fixation. Hence, most of the positive selected variants are only observed as fixed variants.
203 In addition, functional positions (nonsynonymous positions) are constrained compared to
204 nonfunctional positions (synonymous positions), and hence their evolutionary ratios are smaller.
205 In the neutral scenario, polymorphism and divergence (excluding the adaptive fixed variants) are
206 proportional to the mutation rate and to the constriction factor in the case of nonsynonymous
207 positions (McDonald and Kreitman 1991, Eyre-Walker 2006, Fay 2011). That is:

208 $$\frac{\theta_n}{\theta_s} = \frac{(1-\alpha)K_n}{K_s},$$

209 *(Equation 2)*

210 where θ_n the nonsynonymous variability, θ_s is the synonymous variability, K_n the nonsynonymous
211 divergence, K_s the synonymous divergence and α is the proportion of adaptive variants that have
212 been fixed. To estimate the proportion of nonsynonymous substitutions that are adaptive (α) we
213 reorder the previous expression:

214 $$\alpha = 1 - \frac{K_s}{K_n}\frac{\theta_n}{\theta_s}$$

215 *(Equation 3)*

216 A higher ratio of nonsynonymous to synonymous divergence versus polymorphisms suggests that
217 positive selection has fixed adaptive variants ($\alpha > 0$) and the opposite case ($\alpha < 0$) suggests the
218 presence of deleterious mutations segregating in the population.

219 If we consider that weak deleterious mutations are segregating in the population, we expect that
220 their relative proportion will be higher at lower frequency variants and low or null for fixed
221 deleterious mutations. Following the same notation as in equation 3:

222 $$\frac{\theta_{in}(1-\beta i)}{\theta_{is}} = \frac{(1-\alpha-\beta d)K_n}{K_s},$$

223 *(Equation 4)*

224 where i refers to the frequency at which the calculation of variability is estimated, βi is the
225 proportion of weakly deleterious polymorphic mutations at frequency i, βd is the proportion of
226 weakly deleterious fixed mutations. $\beta d < \beta i$ was assumed at any frequency. Then, solving for the
227 proportion of fixed adaptive variants (α):

228 $$\alpha = 1 - \beta d - (1-\beta i)\frac{K_s}{K_n}\frac{\theta_{in}}{\theta_{is}}$$

229 *(Equation 5)*

230 We see that in case calculating α without considering the effects of deleterious mutations, it would
231 be underestimated depending on the frequency at which the estimates of variability are calculated.
232 If we assume that the detrimental variants would never be fixed, a good estimator of α using
233 equation 3 would be the one that estimates variability based on high frequencies, as it would hardly
234 contain detrimental mutations. This is in agreement with the arguments used in Messer and Petrov
235 2013.

236

237 Similarly, if we also consider that weak positively selected variants are segregating in the
238 population, we expect that their relative proportion, compared to neutral ones, is higher at higher
239 frequencies:

240 $$\frac{\theta_{in}(1-\beta i-\gamma i)}{\theta_{is}}=\frac{(1-\alpha-\beta\,\mathrm{d}-\gamma d)K_n}{K_s},$$

241 *(Equation 6)*

242 where γi is the proportion of weakly advantageous polymorphic mutations at frequency i, and
243 γd is the proportion of weakly advantageous fixed mutations. Again, solving for the proportion of
244 fixed adaptive variants (α):

245 $$\alpha+\gamma d=1-\beta d-(1-\beta i-\gamma i)\frac{K_s}{K_n}\frac{\theta_{in}}{\theta_{is}}$$

246 *(Equation 7)*

247 In this case, the presence of adaptive variants segregating in the population would affect the
248 estimates of variability based on high frequency variants when using equation 3, which would
249 result in an underestimation of the proportion of fixed adaptive variants (α). Note that adaptive
250 variants stabilized at intermediate frequencies are not considered in this approach, which can be
251 an important source of adaptation considering the infinitesimal model.

252

253 If we focus on the effects on the polymorphisms, equation 6 suggests that the ratio of
254 nonsynonymous to synonymous polymorphisms would increase due to mutations having both
255 positive and negative effects. It is expected that the number of mutations with negative selection
256 coefficients would rapidly decrease as we move to intermediate and high frequencies, while the
257 opposite trend is expected for mutations with positive selection coefficients. Hence, higher ratios
258 of nonsynonymous to synonymous polymorphisms at higher frequencies may be explained by the
259 presence of advantageous mutations segregating in the population.

260

261 Furthermore, in cases where two populations are from the same species and no fixed mutations
262 between them we can estimate the possible differential effect of the selection (positive and
263 negative) at any frequency among populations from the ratios of synonymous to nonsynonymous
264 polymorphisms of the two populations ($R_\beta\gamma i$):

265 $$\frac{\theta_{in1}(1-\beta i1-\gamma i1)}{\theta_{is1}}=\frac{\theta_{in2}(1-\beta i2-\gamma i2)}{\theta_{is2}}$$

266 and

267 $$\frac{(1-\beta i1-\gamma i1)}{(1-\beta i2-\gamma i2)}=\frac{\theta_{is1}\theta_{in2}}{\theta_{in1}\theta_{is2}}=R_\beta\gamma i$$

268 *(Equation 8)*

269 In addition, a comparison of the $R_\beta\gamma i$ values calculated using different variability estimators
270 (hereafter $R_\beta\gamma i$ pattern) can be used to inform about the effects of the different types of selection.
271 Importantly, different demographic effects (e.g., bottlenecks) together with the presence of
272 mutations with small selective effects may also disturb the ratios of variability and hence must be
273 taken into account when interpreting the results.

274 The effect of linkage disequilibrium between selective (detrimental or adaptive) and neutral
275 variants should not affect the expected estimate of the proportion of adaptive variants, as it would
276 affect both synonymous and nonsynonymous positions in the same proportion. On the contrary,
277 the interaction of variants with opposite selective effects would possibly reduce the effect of
278 selection and would have a significant consequence on the estimation of adaptive fixed variants
279 (Hill and Robertson 1966; Booker and Keightley 2018).

280

281 **Bootstrap analysis**

282 Nonparametric bootstrap analysis was performed to estimate the null distribution of the α statistic
283 for each variability estimator and pig population. In each case, coding positions for synonymous
284 and nonsynonymous (separately) were randomly chosen with replacement and the α statistic was
285 calculated as in equation 2. This process was repeated 100 times.

286

287 **Simulations**

288 We carried out forward simulations using the software *SLiM* (Haller and Messer 2017) in order to
289 assess the interaction of the different selective effects and demographic factors affecting the
290 evolution of pig populations during domestication. We explored the expected values of nucleotide
291 diversity, divergence, α and $R_\beta\gamma$ under 63 different scenarios. For each scenario, we simulated
292 three populations corresponding to wild, domestic and an outgroup species. We first simulated
293 nine different scenarios that were classified into three main groups: i) standard neutral model
294 (SNM); ii) a model with negative selection (NS) and iii) a model with positive selection (PS). For
295 the models with selection, we let that selection operate from the ancestral species to the present
296 time. Each group of scenarios (SNM, NS and PS) was simulated with a constant effective
297 population size for the three populations or with a reduction or an expansion of the effective
298 population size in the branch leading to domestic pigs. A second group of simulations was
299 performed under more complex scenarios. In those simulations, we incorporated the combined
300 effect of negative and positive selective effects (using gamma and exponential distributions for the
301 selective coefficients, respectively) plus demographic effects such as expansion and reduction of
302 the effective population size in the domestic simulated populations and with or without migration
303 from the wild into the domestic populations (in total 54 complex simulated scenarios). Figure S1
304 shows a general scheme for the simulated populations and Table S3A-B shows the parameter
305 values used in these simulations. The obtained results were analyzed using the *mstatspop* software.

306

307 **Approximate Bayesian computation (ABC) analysis**

308 We used the ratio of the estimates of nucleotide variability ($\theta n/\theta s$) per nucleotide for
309 nonsynonymous versus synonymous positions (Fu&Li, Watterson, Tajima and Fay&Wu) and of
310 divergence (*Kn/Ks*) to infer the distribution of fitness effects (DFE) in coding regions. We
311 compared three evolutionary models that differ in the shape of the DFE using the algorithm
312 proposed by Tataru et al. (2017), which are the following: (i) model A: a model with a deleterious
313 gamma DFE with the mean and the shape of the gamma distribution as model parameters, (ii)
314 model C: a model with a gamma distribution of deleterious variants with two parameters (shape
315 and mean) and an exponential distribution of beneficial variants with one parameter (mean), and
316 the additional parameter of the proportion of beneficial versus deleterious variants, and (iii) model
317 D: a model with a discrete distribution of a priori values of possible selective coefficients (positive

318 and negative) and considering as parameters the proportion of each (positively and negatively
319 selected mutations). Some of the additional parameters, such as demographic or linkage effects,
320 were considered as nuisances, while others, such as errors in the polarity of unfolded mutations,
321 were fixed. Table S4 shows the parameters and the prior distributions used in the analysis. We
322 used *polyDFEv2* (Tataru et al 2019) to obtain the expected unfolded site-frequency spectrum
323 (SFS). With the aim of performing the ABC using summary statistics, the code of *polyDFEv2* was
324 slightly modified in order to print the SFS and the parameters for a large number of conditions.
325 For each model, one million iterations were run using different parameter conditions and the
326 resulting SFS for each condition were kept to later calculate the ratios of variability, divergence
327 and the α statistic. ABC analysis was performed using the R library *abc* (Csillery et al. 2012). We
328 performed a cross validation analysis to evaluate the ability of the approach to distinguish between
329 models using the *cv4postpr()* function, as suggested in the *abc* library documentation. The
330 confusion matrix indicated that these three models were quite distinguishable with a probability of
331 true classification from model A versus C/D of 0.70, from model C versus A/D of 0.60 and from
332 model D versus A/C of 0.67, with a tolerance of 0.05 (Table S5). Posterior probabilities of each
333 model given the observed data were obtained using the *postpr()* function and considering a
334 multinomial logistic and a rejection approach. Additionally, a goodness of fit analysis, which
335 compares the median of the distance between the accepted summary statistics and the observed
336 ones, was also performed to select the best model. Once the best model was chosen, the ability to
337 infer the parameters of the model was assessed using the *cv4abc()* function. Prediction errors for
338 the parameter inference of each model are shown in Table S6 and Figure S2. The parameters of
339 the best model were inferred with the *abc()* function using a local linear regression and a rejection
340 approach. Posterior predictive simulations were performed with the α statistic (instead of with the
341 ratios of variability and of divergence, to avoid circularity in the analysis) to determine whether
342 the simulated data generated from the estimated parameter of our best model resembled the
343 observed data (1000 replicates). Finally, the α values can be simply estimated using equation 10
344 from Tataru et al. (2017), which is a simple proportion of positive selective coefficients (s) values
345 in the case of the discrete distribution.

346

347 **Pathway analysis**

348 We downloaded the complete list of pathways and genes of *Sus scrofa* from KEGG v.20170213
349 (http://www.genome.jp/kegg/, Kanehisa et al. 2008). The list contained 471 pathways and 5480
350 genes. The median and mean number of genes per pathway was 26 and 43, respectively, and ranged
351 from 1 to 949. We filtered the pathways according to their size, removing pathways with less than
352 10 and more than 150 genes in order to discard pathways that were not informative or too generic
353 and complex. The final list contained 171 pathways and 3449 genes.

354 To analyze the selection pressure of each gene according to its position in the pathway, we obtained
355 different topological parameters. For that, we first downloaded the XML file of each pathway from
356 KEGG v.20170213. These files were analyzed with the *iGraph* R package (Csardi G. and Nepusz
357 T. 2006) to obtain the topological descriptors of each gene in each pathway. For each gene, three
358 different measures were computed: *betweenness* (number of shortest paths going through a vertex),
359 *in-degree* (number of in-going edges) and *out-degree* (number of out-going edges). These
360 parameters are measures of the importance of a gene within a pathway: *betweenness* is a centrality
361 feature, *in-degree* suggests the facility of a protein to be regulated and *out-degree* reflects the
362 regulatory role of a protein. We tested whether negatively and positively selected genes differed
363 in any of these statistics using a nonparametric Wilcoxon rank test, due to the extreme leptokurtic
364 distributions involved.

365

366 **RESULTS**

367 **Predominance of shared variants and similar selective effects on pig populations**

368 We found a total of 6,684,142 SNPs in autosomes, with 149,440 SNPs of these located in coding
369 regions. We found that 12.5% of the SNPs in the coding regions are shared among the three
370 populations, 32.2% are shared between at least two populations, 31.2% are exclusive to LW, 2.2%
371 are exclusive to IB and 34.4% are exclusive to WB (Table 1). Based on the PCA analysis and using
372 the total number of SNPs, we found that the individuals of each breed cluster together and are well
373 separated from other breeds (Figure S3). The proportion of private SNPs in each population is in
374 accordance with its specific demographic history (Esteve-Codina et al. 2013, Bosse, Megens,
375 Madsen, et al. 2014; Bosse, Megens, Frantz, et al. 2014).

376

377 For each breed, each coding position was classified as polymorphic, fixed (i.e., different alleles
378 from the outgroup) or ancestral allele (i.e., same allele as in the outgroup), with the aim of
379 identifying those variants that appeared previously or posteriorly to the domestication process
380 (Table 2). Surprisingly, we found very few fixed mutations between populations, indicating that
381 the phenotypic traits of each population are not associated with fixed coding variants. Similarly,
382 we found very few fixed coding variants in domestic (IB or LW) versus wild (WB). There are few
383 variants fixed in the domestic that are polymorphic in the wild population, suggesting that these
384 variants were previously present in wild breeds or, alternatively, were transferred into WB by gene
385 flow from introgressed domestic breeds. Most of the variants that are exclusive of a single breed
386 are polymorphic, which is in agreement with the recent origin of these variants. We found a large
387 number of fixed variants in the IB that are polymorphic in LW and WB, likely due to a reduction
388 of the effective population size of the IB breed. The ratio of nonsynonymous to synonymous
389 polymorphism was always lower than one and with similar values in the three populations
390 regardless of the variability estimator used. This result suggests that apparently, there are no
391 differential effects of selection between domestic and wild populations.

392

393 **Limited influence of genomic context and the network topology on selective patterns**

394 The heterogeneity in the recombination rate, the gene density, the %GC and the distribution of
395 CpG islands across the genome can affect the local levels of variability. A previous study on the
396 IB breed detected a strong correlation between recombination and variability, although no
397 correlation was observed between variability and gene density or GC content (Esteve-Codina et
398 al. 2013). However, the effect of these factors on the estimation of the proportion of adaptive
399 nonsynonymous mutations (α) has not been previously studied. We observed no correlation
400 between α and recombination, gene density, missing rate, %GC and CpG in any of the three breeds
401 (P-values > 0.01).

402

403 Next, we investigated the effect of gene network topology on the selective patterns. It has been
404 claimed that topology limits the 'evolvability' of genes and that highly connected genes are more
405 constrained and, consequently, less likely to be targets of positive selection. We compared the
406 network topology features (*betweenness*, *out-degree* and *in-degree*) of genes within pathways
407 regarding the value of α, grouping genes with positive versus negative α values. We found that
408 genes with negative α values show significant large values of the *betweenness* statistic in the three
409 pig breeds compared to genes with positive α values (P-value < 0.01; Figure S4). LW and WB
410 (but not IB, possibly because the low sample size) show significant values (P-values < 0.01) of the
411 *in-degree* statistic for genes with negative α values compared to genes with positive a values.
412 However, we did not observe significant differences in the *out-degree* values between genes with
413 negative and positive a values in any of the three breeds (Figure S4). These results suggest that, in
414 the three breeds, genes that are more central in a pathway are more evolutionary constrained
415 compared to peripheral genes. In addition, in LW and WB, the genes that are more constrained
416 tended to have a higher number of upstream genes that regulated them, which is also in agreement
417 with the central position of these genes in the pathway. We did not observe significant differences
418 in *in-degree* statistic in the IB breed between genes with negative and positive a values, perhaps
419 because of a relaxation of functional constraints as a consequence of the reduction of its effective
420 population size.

421

422 **Levels of nucleotide variation at protein coding regions is compatible with the history of the**
423 **surveyed pig populations and with positive selection**

424 To assess the selective effect of domestication, we first studied the pattern of variation at
425 synonymous and nonsynonymous positions using four estimators of variability that differentially
426 weight the SNP frequencies (See Material and Methods). Estimates of the levels of variability per
427 nucleotide at the genome level using different estimators are shown in Figure 1 and detailed in
428 Table S7. We expect that, under the Standard Neutral Model (SNM), all estimates of variability
429 should be similar while differences among them may indicate demographic and/or selective
430 effects. We observed that the levels of variability are different for each estimator within breeds
431 and also for the same estimator for different breeds. However, we observed a similar ratio of
432 nonsynonymous to synonymous polymorphisms for all estimators of variability, suggesting that
433 demographic effects are responsible for the differences in the levels of variability (Figure 1). The
434 less variable population is the IB breed, which shows far fewer singletons compared to WB and
435 LW, probably as a consequence of the reduction of its population size. Note than in all three
436 populations, high-frequency variants are proportionally more abundant than those at intermediate
437 frequencies, which would be compatible with the accepted demographic history of the surveyed
438 populations (i.e., introgression in the LW) but also with the presence of pervasive positive selection
439 in all three populations.

440

441 **α's and $R_\beta\gamma$ ratios might show a differential effect of selection due to domestication**

442 The differential effect of selection in the domestic and wild populations can be studied by
443 comparing the α values. Figure 2 and Table S8 show the genome-wide α values calculated using
444 the four variability estimators for each population. As expected, the α values are negative when α
445 is calculated using the estimate of variability based on low-frequency variants ($\alpha_{Fu\&Li}$), probably
446 reflecting the relatively high proportion of deleterious versus neutral mutations that are segregating
447 at low frequencies. We observed a similar value of $\alpha_{Fu\&Li}$ in all populations, suggesting a similar
448 proportion of segregating deleterious mutations, irrespective of the domestication process or other
449 demographic events (Figure 2A). Moreover, we observed less negative values of α, or even
450 positive (for LW in Figure 2A), when α is calculated based on variants at high frequencies,
451 according to expectations that point to a progressive elimination of deleterious mutations at higher
452 frequencies. Nevertheless, the pattern of α (i.e., the comparative a values calculated using the four
453 different variability estimators within each population) is very different in each population. WB

454 and IB show very low negative values of α for most of the estimators of variability, except for
455 $\alpha_{Fay\&Wu}$ in WB, which is zero. LW is the only breed that shows a linear increase of the α negative
456 values across the SFS, being even positive when calculated based on high frequencies ($\alpha_{Fay\&Wu}$).

457

458 We did not observe similar patterns of α between domestic breeds compared to WB (Figure 2A).
459 The differences in the ratio of synonymous to nonsynonymous variability between the two
460 different breeds is summarized by the $R_\beta\gamma$ ratio (Figure 3A). We observed deviations from $R_\beta\gamma$
461 = 1 when the ratio was calculated based on high-frequency variants ($\alpha_{Fay\&Wu}$). Although the ratio
462 of the two populations is difficult to interpret because of their different underlying demographics,
463 some trends can be observed. WB shows an excess of nonsynonymous variants segregating at
464 intermediate frequencies (WB-IB, WB-LW), which might be explained by a past bottleneck that
465 increased deleterious mutations at intermediate frequencies. In addition, the $R_\beta\gamma$ ratio in LW-IB
466 shows an incremental pattern from low to high frequencies, which is compatible with an increase
467 of nonsynonymous beneficial variants on their way to fixation in LW.

468

469 **α's and $R_\beta\gamma$ ratios based on from exclusive and shared polymorphisms might reflect**
470 **changes in selection patterns before and after domestication**

471 We observed a high ratio of nonsynonymous to synonymous exclusive singletons ($\alpha_{Fu\&Li}$, Figure
472 2B), suggesting that they have deleterious effects in all populations. Nevertheless, the values of α
473 calculated based on intermediate frequency variants (α_{Tajima}) in the WB and IB populations are
474 lower than to those based on low-frequency variants, which point to the action of positive selection
475 maintaining nonsynonymous variants at higher frequencies in the case of WB and to an attenuated
476 effect of deleterious mutations in IB due to a population size decline. In the case of LW, the
477 observed pattern of α is similar to that calculated with all SNPs (Figure 2B). Likewise, the R_βγ
478 statistic shows the same pattern as that calculated using all SNPs but with a higher magnitude of
479 its value (Figure 3B).

480

481 On the other hand, the α values based on shared variants are in general more moderate (closer to
482 zero) than those based on exclusive or all SNPs (Figure 2C), likely because shared nonsynonymous

483 polymorphisms are older and hence, expected to be more functionally constrained than exclusive
484 ones. Additionally, the values of α based on singletons ($\alpha_{Fu\&Li}$) are less negative than those based
485 on intermediate-frequency variants. Again, this pattern might indicate that selection is involve in
486 the increase of the ratio nonsynonymous to synonymous polymorphisms up to intermediate
487 frequencies. The $R_\beta\gamma$ statistic shows similar patterns than those observed for all variants but with
488 values much closer to 1, indicating a small or moderate selective effect on the shared variants
489 compared to all variants (Figure 3C).

490

491 When we calculated the α values from shared variants only between the two domestic breeds, we
492 found an inverse pattern regarding that calculated from all SNPs in each population, with high
493 positive values of α based on low frequencies and very negative values when α is calculated based
494 on high-frequency variants (Figure 2D). Some possible explanations might be the active
495 elimination of new nonsynonymous variants to preserve differences among breeds ($\alpha_{Fu\&Li}$) and
496 either the effect of the ancestral population structure (wild versus domestic), or the presence of
497 nonsynonymous variants targeted by the process of domestication that shifts them toward high
498 frequencies ($\alpha_{Fay\&wu}$).

499

500 **Absolute values of α are dependent of the molecular scale but patterns remain similar**

501 In addition to the genome-wide analysis, α was estimated using three additional molecular scale
502 levels: i) gene level, ii) genes within windows of 5 Mb, and iii) genes within the same pathway.
503 Figure 4 shows the median of the distributions of the α values for each scale level. In general, the
504 distribution of α values estimated at the genome-wide level are concordant with those estimated at
505 the gene level, genes within windows and pathway level in each breed. However, differences in
506 the value of α within each breed are notorious depending on the scale level examined. The median
507 estimates of α are generally lower at the gene scale level and most of them are very negative, while
508 at the genome-wide scale, the α values are closer to zero. However, the distribution of α values
509 can have a large variance at the gene scale since few variants are used for its estimation. We
510 identified the regions and pathways that showed extreme α values (Table S9 and S10). We found
511 a large number of genes having $\alpha = 1$ (highest value) because the number of polymorphic
512 nonsynonymous variants per gene was zero. We also found a moderately high correlation of α

513 values between breeds (rho ~ 0.7, Pearson correlation using pathways) suggesting that in general,
514 these breeds are under similar selective effects. For shared and exclusive variants, we generally
515 observed the same pattern, from genes to whole-genome, that is, larger α values at the gene level
516 and closer to zero α values at the larger scale. Only shared variants of the IB breed exhibited similar
517 α values from genes to the whole-genome scale. The differences in absolute α values could be
518 explained because of the nature of this ratio statistic, in which the mean can be more displaced to
519 more negative values due to a reduced number of functional variants (*i.e.,* few or null segregating
520 nonsynonymous variants).

521

522 **Simulated data under scenarios that include the joint effect of demography and selective**
523 **events were more similar to the observed data**

524 We used computer simulations to study how the different demographic and selective events
525 occurred during domestication affected nucleotide variation. We simulated populations mimicking
526 the process of domestication using *SLiM* software (Haller and Messer 2017) coupled with several
527 demographic events, including changes of the population size and/or migration. We analyzed the
528 genome-wide patterns of α and the $R_\beta\gamma$ statistic produced by 63 simulated scenarios that included
529 different demographic events and selective forces acting separately (simple scenarios) or jointly
530 (complex scenarios). The results of the simulation study are summarized in Figures S5-S46. The
531 observed patterns of α calculated from all variants in the surveyed populations are not compatible
532 with simple scenarios that only consider demographic or positive selection forces (Figure S5).
533 Rather, simulated α trends (irrespective of the magnitude of α) fit a scenario with a predominant
534 effect of negative selection (Figure S5). However, the $R_\beta\gamma$ statistic do not fit any of the simulated
535 simple scenarios (Figure S6). When more complex scenarios were considered, the general trends
536 of the patterns of α generated by scenarios with both negative and positive selection resembled
537 those observed in WB and LW (Figures S7-S13). Notice that this is true only for those scenarios
538 that include some demographic or migration events (Figures S10-S12). On the other hand, the IB
539 population would fit a scenario without positive selection and with a recent population size
540 reduction (Figure S11). The trends in the $R_\beta\gamma$ statistic are, in broad strokes, concordant with the
541 conclusions extracted from the comparison between the observed and simulated patterns of α
542 (Figures S13-S18).

543

544 The observed patterns of α values from exclusive variants are different from those based on the
545 total variants in WB and IB populations (Figure S19). These patterns cannot be fully explained by
546 any of the complex simulated scenarios proposed when considering all variants, except for the IB
547 population, which might be compatible with a scenario with negative selection and population size
548 reduction (Figures S22-S26). The observed $R_\beta\gamma$ are also compatible with the scenarios combining
549 both types of selection and population size reduction (WB-IB) or with scenarios that include
550 migration (WB-LW; Figures S27-S32). Finally, the observed patterns of α and $R_\beta\gamma$ statistics
551 calculated from shared variants are also compatible with scenarios having deleterious plus
552 beneficial mutations (Figures S33-S46).

553

554 **A model that assumes a discrete distribution of beneficial and deleterious mutations would**
555 **fit better the observed data in the ABC analysis**

556 We used an approximate Bayesian computation (ABC) analysis to infer the DFE separately for
557 each population using the ratios of nonsynonymous to synonymous variants obtained from the
558 whole-genome analysis (see Materials and Methods). Three different models implemented in
559 *polyDFE2* software (Tataru et al. 2019) were tested. Model A, which assumes a gamma
560 distribution of deleterious mutations; model C, which assumes a gamma distribution for
561 deleterious mutations and an exponential distribution for beneficial mutations; and model D, that
562 assumes a discrete distribution of beneficial and deleterious mutations. In these models, we
563 included demographic and linkage effects as nuisance parameters (Tataru et al. 2017). Goodness
564 of fit (GoF) analysis, which is a measure of the adjustment of the prior chosen models to the data,
565 revealed that the simulated data under the different models do not always fit well to the observed
566 data (Table S11A), suggesting that the real data fit only to a very restrictive parameter range of
567 each model. This is especially pronounced for exclusive variants, where the simulations under the
568 different models fit only marginally to the observed data. Indeed, the model with a wider parameter
569 range versus the real data is usually the discrete model D. However, if we also take into account
570 the values of the posterior probabilities, which is the probability assigned to each model relative
571 to the other models of the analysis, we found that the best fit model differed among populations
572 (model C to WB and LW and model D to the IB breed; Table S11A). Finally, posterior predictive

573 analysis indicated that the observed α values (Figure 5) and variability
574 nonsynonymous/synonymous ratios (Figure S47) cannot be obtained using the estimated
575 parameters of any model, although they were closer to model D. The parameters of the DFE
576 inferred for each population are shown in Table S12. The obtained results indicate that the DFE is
577 quite similar among all three populations, which is not entirely surprising because they share a
578 long-term history. Despite there is a lot of uncertainty in the inferred estimates, the obtained results
579 show that the DFE contains a large fraction of very deleterious variants, with approximately 83%
580 of the variants being strongly deleterious ($S \leq -200$; considering the posterior distribution and using
581 the rejection method), and with approximately 17% of the variants being neutral, weak beneficial
582 and weak deleterious mutations ($-2 \leq S \leq +2$). The median proportion of adaptive mutations (α)
583 estimated from the discrete distribution (by summing weak and strong beneficial proportion of
584 mutations in Model C) is approximately 0.9% (Table S12A).

585

586 **The inferred DFE is different when based on exclusive and shared variants**

587 Although the inference of the DFE is going to be distorted by choosing only a subsection of the
588 variants (e.g., exclusive variants are mostly very recent), we considered that it can give some clues
589 about past events that could be more related to the domestication processes. Regarding to exclusive
590 variants, the simulations under the three different models show a low fit to the observed data.
591 Indeed, in some cases, the GoF is less than 1% (Table S11B). Although the posterior probability
592 is higher for model C, the posterior predictive simulations show that none of the models can
593 reproduce well the observed data. In general, the posterior predictive simulations under model D
594 yield more similar values to the observed data. However, only the posterior predictive simulations
595 under this model are reasonably similar to the observed data for the LW breed (Figure 6, Figure
596 S48). The results obtained for shared variants are very similar to those considering the total
597 variants, supporting the hypothesis that shared mutations have a predominant effect (compared to
598 exclusive variants) at the whole genome. Finally, the median proportion of adaptive variants (α)
599 estimated from the discrete distribution (Model C) gave similar results to the total mutations, but
600 surprisingly estimated a slightly higher proportion for the Iberian breed (Table S12C; Figure S49-
601 S50).

602

603
604 **DISCUSSION**

605 The polygenic nature of the domestication traits in animals often precludes identifying their
606 underlying genes since the domestic phenotypes are caused by subtle allele-frequency changes of
607 variants distributed throughout the genome that are very difficult to detect. In addition, the study
608 of the effects of selection using genome sequences that contain a nonnegligible fraction of missing
609 data is challenging and needs the use of appropriate methods to account for these positions.
610 Statistics that exploit the frequency of the variants while accounting for missing data are
611 particularly appropriate for such analyses.

612

613 Here, we provide a novel approach that combines the use of different estimators of variability that
614 account for missing data with the asymptotic approach proposed by Messer and Petrov (2013) and
615 Uricchio et al. 2019). This approach allowed us to study and to interpret the effects of the
616 domestication process on genomic variation as an alternative to the use of the full SFS. Although
617 it can be less precise (we used only four statistics to capture the entire trend of α across the SFS),
618 it helps to reduce the variance and facilitates the visual interpretation. The analyses performed here
619 include an exhaustive comparative study of the observed patterns of functional versus neutral
620 diversity in domestic pigs and wild boars, a forward simulation study for several diverse
621 evolutionary scenarios and the inference of DFE parameters given different selective models.

622 Note that the DFE was inferred using Bayesian calculations (ABC) instead of exact Bayesian or
623 Likelihood methods. Despite ABC requires additional steps and validation analysis and is in
624 general less precise, it allows contrasting models and inferring parameters from complex datasets
625 or data containing missing information (Beaumont et al. 2002).

626

627 **Selection pressure on pigs and the process of domestication**

628 A number of analyses with the aim to explain the process of domestication have been performed
629 already, using the MacDonald and Kreitman extension methods, or using other estimates such as
630 variability or divergence at functional or synonymous positions (MacEachern et al. 2009, Kono et
631 al. 2016, Makino et al. 2018). For instance, Kono et al. (2016) analyzed derived frequencies in

632 domesticated barley populations at different functional classes (deleterious, tolerated) and
633 observed a higher quantity of deleterious variants at low frequencies compared to tolerated or to
634 synonymous mutations. Makino et al. (2018) investigated the ratio of functional to neutral variants
635 at different frequencies in domestic and in wild populations of several animal and plant species,
636 including Asian and European pigs. They observed generally lower levels of synonymous and
637 nonsynonymous variation and a higher ratio of functional to neutral variants in the domestic
638 species compared with their wild counterparts. This ratio was negatively correlated with the
639 frequency of the variants, consistent with a higher number of detrimental variants in domestic
640 populations. The authors claimed that these patterns were compatible with the expected effect of
641 a bottleneck as a consequence of domestication, and with an increase in nonsynonymous variants
642 produced by the lesser efficacy of purifying selection at smaller population sizes, although the
643 presence of positive selection (hitchhiking) was not discarded. However, the opposite pattern was
644 observed in European wild boars and domestic pigs (a higher variability and a lower ratio of
645 nonsynonymous to synonymous in domestic pig populations compared to wild boars). The authors
646 argue that this can be explained by the highly variable patterns produced by bottlenecks, the strong
647 population contraction of European wild boars during the last glaciation and the presence of gene
648 flow between wild and domestic populations.

649

650 As in Makino et al. (2018), we do not observe an increase in functional diversity in domestic versus
651 wild populations. This result may also be explained by several recent events occurring in these
652 populations: (i) differences in the recent history of local and commercial domesticated populations,
653 that is, high inbreeding in Iberian local pigs but recent gene flow of the commercial pigs with
654 Asian pigs; (ii) demographic effects in the wild boar population that may have reduced their
655 diversity (Groenen et al 2016) or increase the variance of the patterns (see simulations, Figures
656 S5-S46); (iii) differential adaptive forces in local versus commercial pigs, with a recent high
657 selective pressure in this last population.

658

659 The two domestic breeds analyzed here have very different recent histories: the IB is a local
660 Spanish breed (Guadyerbas) that suffered a strong bottleneck during the 1970s (Esteve-Codina et
661 al. 2013) and with no evidence of introgression whereas the LW breed was admixed with pigs of

662 Asian origin (Bosse, Megens, Madsen, et al. 2014). Currently, approximately 20–35% of the LW
663 genome has been estimated to be of Asian origin (Groenen et al. 2012, Bosse et al. 2012, Frantz et
664 al. 2015, Bianco et al. 2015, Ai et al. 2015). Accordingly, the IB breed shows the lowest levels of
665 synonymous and nonsynonymous variation among the breeds studied, probably because of its
666 small effective population size and because the individuals from the IB sample come from a very
667 closed population of pigs. However, we expected to find a higher variability in LW compared to
668 WB due to the process of Asian introgression that this breed has undergone. Surprisingly, we
669 detected very similar levels of variability between them. However, the high levels of variability
670 were observed for variants that belonged to different frequency ranges in these two populations:
671 singletons in WB and in high-frequency derived alleles in LW. This difference may be due mainly
672 to the effects of gene flow in LW but also in some extend to the selective programs applied to the
673 commercial breed.

674

675 **Domestication hallmarks at pig coding regions**

676 The paucity of fixed variants at coding positions in the three breeds indicates that the observed
677 heritable phenotypic differences among the breeds are either due to: i) few selective sweeps, ii)
678 positive selection at noncoding functional regions which were not analyzed in this work, iii)
679 changes in the frequencies of nonsynonymous variants without being fixed. In the first hypothesis,
680 we expect that domestication process should fix the adaptive variants for those genes underlaying
681 the phenotypes of interest. However, we found no fixed variants between domestic breeds and
682 wild pigs, although they show the right breed phenotype. We checked the α values for those genes
683 that were previously reported to show signals of positive selection using other approaches
684 (Groenen 2016). We found that these genes show little or no nonsynonymous polymorphisms or
685 fixed variants (Table S13). This absence of variability is typical from regions under selective
686 sweeps, although not necessarily implicating that these genes are the targets of domestication since
687 there are no variants fixed or close to fixation at their coding regions. We only found significant
688 values of α over zero at gene KIT in the IB breed, genes IGF2R and JMJD1C in the LW breed and
689 gene LRRTM3 in the WB population. The second hypothesis implies that the functional regions
690 implicated in domestication would be out of coding regions (promotors, enhancers, etc. (e.g., Li et
691 al 2018, Rubin et al 2012, Anderson 2012). However, although being a promising hypothesis, we

692 did not analyze those regions because it requires a very accurate analysis of homology and their
693 associated functionality, which is very complicated at the genome level, especially for non-model
694 species. The third hypothesis suggests that the domesticated phenotype is caused by a moderate
695 change in the frequency of a relatively large number of variants with small selective effects. In this
696 case, depending on the size of the selective effect, the number of fixed shared variants may be
697 significant at the genome level, or alternatively, there would be only changes in the frequencies of
698 the variants without reaching fixation. In the last case, the functional variants involved in
699 domestication should be segregating in the analyzed populations. These positively selected
700 variants segregating at high frequencies, together with the presence of deleterious mutations also
701 segregating at low frequencies, would be reflected as an excess of non-neutral compared to
702 divergence. Hence, negative α values calculated using statistics based on high frequencies variants
703 should be observed in cases where there is a significant proportion of positive selection variants
704 that have not yet being fixed. We have observed this pattern in WB and the IB breed, but not in
705 LW. Furthermore, the estimation of DFE from ABC analysis showed that the global proportion of
706 beneficial mutations (weak and strong) is relatively small (~0.9%) and similar in all wild and
707 domestic populations. Nevertheless, this proportion of mutations is substantial in absolute
708 numbers. Although speculative, these mutations may change the fate of these populations that are
709 affected by natural or artificial selection.

710

711 We expected that shared variants between populations would be enriched by selective pressures
712 that predate domestication. Although they may reflect biological constraints at the species level,
713 these shared polymorphisms can also be the source of phenotypic variation in a polygenic selective
714 scenario such that a change in their frequencies (in an infinitesimal scenario) would result in the
715 observed phenotypic differences among the breeds. Furthermore, private variants (those
716 segregating only in one breed) may reflect recent and breed-specific selective hallmarks and hence,
717 would be responsible for the observed differences between domesticated and wild breeds. In both
718 cases, shared and exclusive polymorphisms are contributing to the differences in the SFS between
719 functional and nonfunctional positions. We expect that adaptive changes would increase the ratio
720 of nonsynonymous to synonymous polymorphisms and that this should be reflected as an increase
721 in the negative value of the α statistic. Our simulated domestication process indicates that the effect
722 of positive selection irrespective of being either strong and affecting a small percentage of variants

723 or weak and affecting a large percentage of variants is not reflected as marked changes in the
724 estimated patterns of α. This could be due to the short time since the change in the DFE occurred
725 but also by the interaction of positive and negative selection and demographic processes. In fact,
726 the observed α patterns are compatible with demographic effects (population size reduction in WB
727 and IB and gene flow in LW) but also with the effect of positive selection in LW. Finally, the
728 obtained results in the ABC analysis based on total variants show a clear genome-wide effect of
729 the action of purifying selection. We also observed a minor effect of purifying selection in IB and
730 WB when the analysis was performed based on exclusive variants, which suggest a reduction of
731 the population size of these two populations. Nevertheless, we had some difficulties in adjusting
732 the observed data to pertinent DFE models, especially when the analysis was performed based on
733 exclusive variants. Although the models used here are very simple and contain few parameters,
734 the real observations contained high heterogeneity that could not be fitted to these models. The
735 reasons may be technical, conceptual (undetected correlations that distort model assumptions) or
736 biological (too simplistic models to explain the real data). In any case, a model that assumes a
737 discrete distribution of beneficial and deleterious mutations (model D) seems to generally explain
738 the observed data better. The change of the DFE when the analysis was performed based on shared
739 variants is undistinguishable from that based on all variants, indicating that exclusive variants
740 should be more useful to detect the effects of the change of selective effects.

741

742 **Final remarks**

743 The observed patterns of variability are compatible with the presence of deleterious mutations
744 segregating in all three breeds and with weak signals of positive selection. Nevertheless, when the
745 variants are split into shared and exclusive, we observed patterns that are in line with the simulated
746 data under different demographic scenarios joint with the action of positive and negative selection.
747 We found a clear effect of deleterious mutations at low-frequency variants and a mild effect of
748 positive selection at high frequencies. Additional analyses contrasting evolutionary models that
749 consider the effects of standing variation whose effect change under domestication may shed more
750 light and will help to understand the patterns of variation due to the domestication process.

751

752 **ACKNOWLEDGMENTS**

753 We acknowledge L. Silió and M. C. Rodríguez for helpful comments on the manuscript. This work was
754 supported by Ministerio de Economia y Competitividad grants AGL2013-41834-R (MEC, Spain),
755 AGL2016-78709-R (MEC, Spain) and by the CERCA Programme/Generalitat de Catalunya. We
756 acknowledge financial support from the Spanish Ministry of Economy and Competitiveness,
757 through the "Severo Ochoa Programme for Centres of Excellence in R&D" 2016-2019 (SEV-
758 2015-0533). S.G-R. was supported by a Beautriu de Pinós postdoctoral fellowship (AGAUR; 2014
759 BP-B 00027).

760

761 **LITERATURE CITED**

762 Achaz G. 2009. "Frequency Spectrum Neutrality Tests: One for All and All for One." *Genetics*
763 183: 249–58. https://doi.org/10.1534/genetics.109.104042.

764 Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. 215 "Adaptation and possible ancient
765 interspecies introgression in pigs identified by whole- genome sequencing." *Nat Genet.*
766 2015;47:217–25.

767 Alvarez-Ponce, David, and Mario A. Fares. 2012. "Evolutionary Rate and Duplicability in the
768 Arabidopsis Thaliana Protein-Protein Interaction Network." *Genome Biology and Evolution*
769 4 (12): 1263–74. https://doi.org/10.1093/gbe/evs101.

770 Alves, E,, C Ovilo, C Rodríguez, and L Silió. 2003. "Mitochondrial DNA sequence variation and
771 phylogenetic relationships among Iberian pigs and other domestic and wild pig
772 populations." *Animal Genetics* 34:319-324.

773 Alves, Estefania, A I Fernández, Carmen Barragán, C Ovilo, C Rodríguez, and L Silió. 2006.
774 "Inference of Hidden Population Substructure of the Iberian Pig Breed Using Multilocus
775 Microsatellite Data." *Spanish Journal of Agricultural Research* 4 (1): 37–46.
776 http://www.inia.es/gcontrec/pub/alves-fernandez-barragan-..._1141287881015.pdf.

777 Amaral AJ, Ferretti L, Megens HJ, Crooijmans RPMA, Ni H, Ramos-Onsins SE, Perez-Enciso
778 M, Schook LB, Groenen MAM. 2011 "Genome-wide footprints of pig domestication and
779 selection revealed through massive parallel sequencing of pooled DNA." *PLoS One.*
780 2011;6(4):e14782. doi:10.1371/journal.pone.0014782

Andersson L. 2012 "How selective sweeps in domestic animals provide new insight into biological mechanisms." J Intern Med. 2012;271:1–14.

Andersson L. 2013 "Molecular consequences of animal breeding." *Curr Opin Genet Dev*. 23(3):295-301. doi: 10.1016/j.gde.2013.02.014.

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002;162(4):2025-2035.

Bianco, Erica, Bruno Nevado, Sebastian E. Ramos-Onsins, and Miguel Pérez-Enciso. 2015. "A Deep Catalog of Autosomal Single Nucleotide Variation in the Pig." *PLoS ONE* 10 (3): 1–21. https://doi.org/10.1371/journal.pone.0118867.

Booker, Tom R, and Peter D Keightley. 2018. "Understanding the Factors That Shape Patterns of Nucleotide Diversity in the House Mouse Genome." *Molecular Biology and Evolution* 35 (12): 2971–88. https://doi.org/10.1093/molbev/msy188.

Bosse, Mirte, Hendrik-Jan Megens, Laurent A. F. Frantz, Ole Madsen, Greger Larson, Yogesh Paudel, Naomi Duijvesteijn, et al. 2014. "Genomic Analysis Reveals Selection for Asian Genes in European Pigs Following Human-Mediated Introgression." *Nature Communications* 5: 4392. https://doi.org/10.1038/ncomms5392.

Bosse, Mirte, Hendrik-Jan Megens, Ole Madsen, Laurent A. F. Frantz, Yogesh Paudel, Richard P. M. A. Crooijmans, and Martien A. M. Groenen. 2014. "Untangling the Hybrid Nature of Modern Pig Genomes: A Mosaic Derived from Biogeographically Distinct and Highly Divergent Sus Scrofa Populations." *Molecular Ecology* 23 (16): 4089–4102. https://doi.org/10.1111/mec.12807.

Bosse, Mirte, Hendrik Jan Megens, Ole Madsen, Yogesh Paudel, Laurent A. F. Frantz, Lawrence B. Schook, Richard P.M.A. Crooijmans, and Martien A.M. Groenen. 2012. "Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape." *PLoS Genetics* 8 (11). https://doi.org/10.1371/journal.pgen.1003100.

Chen J, Ni P, Li X, Han J, Jakovlić I, Zhang C, Zhao S. 2018 "Population size may shape the accumulation of functional mutations following domestication." *BMC Evolutionary Biology* 18:4 DOI 10.1186/s12862-018-1120-6

Cruz F, Vilà C, and Webster MT. 2008. “The Legacy of Domestication: Accumulation of Deleterious Mutations in the Dog Genome.” *Molecular Biology and Evolution* 25 (11): 2331–36. https://doi.org/10.1093/molbev/msn177.

Csardi G., and Nepusz T. 2006. “The Igraph Software Package for Complex Network Research.” *InterJournal, Complex Systems* 1695. http://igraph.org.

Csilléry K, Olivier François O. and Blum MGB. 2012 " abc: an R package for approximate Bayesian computation (ABC) " Methods in Ecology and Evolution 2012, 3, 475–479.

Esteve-Codina, Anna, Yogesh Paudel, Luca Ferretti, Emanuele Raineri, Hendrik-Jan Megens, Luis Silió, María C Rodríguez, Martien A. M. Groenen, Sebastian E. Ramos-Onsins, and Miguel Pérez-Enciso. 2013. “Dissecting Structural and Nucleotide Genome- Wide Variation in Inbred Iberian Pigs.” *BMC Genomics* 14 (148): 1. https://doi.org/10.1186/1471-2164-14-148.

Eyre-Walker, Adam. 2002. “Changing Effective Population Size and the McDonald-Kreitman Test.” *Genetics* 162: 2017–24. http://www.genetics.org/content/genetics/162/4/2017.full.pdf.

Eyre-Walker A. 2006 "The genomic rate of adaptive evolution." *Trends Ecol Evol* 21(10): 569–575.

Fang M, Larson G, Ribeiro HS, Li N, Andersson L. 2011 "Contrasting mode of evolution at a coat color locus in wild and domestic pigs." *PLoS Genet.* 5:e1000341.

Fay, Justin C, and Chung-I Wu. 2000. “Hitchhiking Under Positive Darwinian Selection.” *Genetics* 155: 1405–13. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1461156/pdf/10880498.pdf.

Fay JC, Wyckoff GJ, Wu CI. 2001 "Positive and negative selection on the human genome." *Genetics* 158(3):1227–1234.

Ferretti, Luca, Emanuele Raineri, and Sebastian E. Ramos-Onsins. 2012. “Neutrality Tests for Sequences with Missing Data.” *Genetics* 191 (4): 1397–1401. https://doi.org/10.1534/genetics.112.139949.

Fisher, R. A. 1919. “XV.—The Correlation between Relatives on the Supposition of Mendelian

838 Inheritance." *Transactions of the Royal Society of Edinburgh* 52 (02): 399–433.
839 https://doi.org/10.1017/S0080456800012163.

840 Frantz, Laurent A. F., Joshua G. Schraiber, Ole Madsen, Hendrik-Jan Megens, Mirte Bosse,
841 Yogesh Paudel, Gono Semiadi, et al. 2013. "Genome Sequencing Reveals Fine Scale
842 Diversification and Reticulation History during Speciation in Sus." *Genome Biology* 14 (9):
843 R107. https://doi.org/10.1186/gb-2013-14-9-r107.

844 Frantz, Laurent A. F., Joshua G. Schraiber, Ole Madsen, Hendrik-Jan Megens, Alex Cagan,
845 Mirte Bosse, Yogesh Paudel, Richard P. M. A. Crooijmans, Greger Larson, and Martien A.
846 M. Groenen. 2015. "Evidence of Long-Term Gene Flow and Selection during
847 Domestication from Analyses of Eurasian Wild and Domestic Pig Genomes." *Nature*
848 *Genetics* 47 (10): 1141–48. https://doi.org/10.1038/ng.3394.

849 Fraser, Hunter B, Aaron E Hirsh, Lars M Steinmetz, Curt Scharfe, and Marcus W Feldman.
850 2002. "Evolutionary Rate in the Protein Interaction Network." *Science* 296 (5568): 750–52.
851 https://doi.org/10.1126/science.1068696.

852 Fu, Yun-Xin, and Wen-Hsiung Li. 1993. "Maximum Likelihood Estimation of Population
853 Parameters." *Genetics* 134: 1261–70.
854 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1205593/pdf/ge13441261.pdf.

855 Groenen, Martien A. M. 2016. "A Decade of Pig Genome Sequencing: A Window on Pig
856 Domestication and Evolution." *Genetics, Selection, Evolution : GSE Sel Evol* 48 (23): 1–9.
857 https://doi.org/10.1186/s12711-016-0204-2.

858 Groenen, Martien A. M., Alan L. Archibald, Hirohide Uenishi, Cristopher K. Tuggle, Yasuhiro
859 Takeuchi, Max F. Rothschild, Claire Rogel-Gaillard, et al. 2012. "Analyses of Pig Genomes
860 Provide Insight into Porcine Demography and Evolution." *Nature* 491 (7424): 393–98.
861 https://doi.org/10.1038/nature11622.

862 Guirao-Rico, Sara, Oscar Ramirez, Ana Ojeda, Marcel Amills, and Sebastian E. Ramos-Onsins.
863 2018. "Porcine Y-Chromosome Variation Is Consistent with the Occurrence of Paternal
864 Gene Flow from Non-Asian to Asian Populations." *Heredity* 120 (1): 63–76.
865 https://doi.org/10.1038/s41437-017-0002-9.

866 Hahn, Matthew W., and Andrew D. Kern. 2005. "Comparative Genomics of Centrality and

867 Essentiality in Three Eukaryotic Protein-Interaction Networks." *Molecular Biology and*
868 *Evolution* 22 (4): 803–6. https://doi.org/10.1093/molbev/msi072.

869 Haller, Benjamin C., and Philipp W. Messer. 2017. "SLiM 2: Flexible, Interactive Forward
870 Genetic Simulations." *Molecular Biology and Evolution* 34 (1): 230–40.
871 https://doi.org/10.1093/molbev/msw211.

872 Hill, W G, and A Robertson. 1966. "The Effect of Linkage on Limits to Artificial Selection."
873 *Genetical Research* 8 (3): 269–94. http://www.ncbi.nlm.nih.gov/pubmed/5980116.

874 Kanehisa, Minoru, Michihiro Araki, Susumu Goto, Masahiro Hattori, Mika Hirakawa, Masumi
875 Itoh, Toshiaki Katayama, et al. 2008. "KEGG for Linking Genomes to Life and the
876 Environment." *Nucleic Acids Research* 36 (SUPPL. 1): 480–84.
877 https://doi.org/10.1093/nar/gkm882.

878 Kim H, Song KD, Kim HJ, Park W, Kim J, Lee T, et al. 2015 "Exploring the genetic signature of
879 body size in Yucatan miniature pig." *PLoS One*. 10:e0121732.

880 Kono TJY, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, Smith KP, Tiffin P, Fay JC,
881 Morrell PL. 2016 "The Role of Deleterious Substitutions in Crop Genomes." *Mol. Biol.*
882 *Evol.* 33(9):2307–2317. https://doi:10.1093/molbev/msw102

883 Leno-Colorado, Jordi, Nick J Hudson, Antonio Reverter, and Miguel Pérez-Enciso. 2017. "A
884 Pathway-Centered Analysis of Pig Domestication and Breeding in Eurasia." *G3* 7 (7):
885 2171–84. https://doi.org/10.1534/g3.117.042671.

886 Li C, Zou C, Cui Y, Fu Y, Fang C, Li Y, Li J, Wang W, Xiang H, Li C. 2018 "Genome-wide
887 epigenetic landscape of pig lincRNAs and their evolution during porcine domestication."
888 Epigenomics. 2018 Dec;10(12):1603-1618. doi: 10.2217/epi-2017-0117.

889 Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-
890 Wheeler Transform." Journal Article. *Bioinformatics* 25 (14): 1754–60.
891 https://doi.org/10.1093/bioinformatics/btp324.

892 Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth,
893 Gonçalo R. Abecasis, Richard Durbin, and Subgroup Genome Project Data Processing.
894 2009. "The Sequence Alignment/Map Format and SAMtools." Journal Article.
895 *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

896 Li, Mingzhou, Shilin Tian, Long Jin, Guangyu Zhou, Ying Li, Yuan Zhang, Tao Wang, et al.
897 2013. “Genomic Analyses Identify Distinct Patterns of Selection in Domesticated Pigs and
898 Tibetan Wild Boars.” *Nature Genetics* 45 (12): 1431–38. https://doi.org/10.1038/ng.2811.

899 Li M, Tian S, Yeung CK, Meng X, Tang Q, Niu L, Wang X, Jin L, Ma J, Long K, Zhou C, Cao
900 Y, Zhu L, Bai L, Tang G, Gu Y, Jiang A, Li X, Li R. 2014 "Whole-genome sequencing of
901 Berkshire (European native pig) provides insights into its origin and domestication." *Sci*
902 *Rep.* 4:4678. doi: 10.1038/srep04678.

903 Livingstone, Kevin, and Stephanie Anderson. 2009. “Patterns of Variation in the Evolution of
904 Carotenoid Biosynthetic Pathway Enzymes of Higher Plants.” *Journal of Heredity* 100 (6):
905 754–61. https://doi.org/10.1093/jhered/esp026.

906 MacEachern S, McEwan J, McCulloch A, Mather A, Savin K, Goddard M. 2009 "Molecular
907 evolution of the Bovini tribe (Bovidae, Bovinae): is there evidenceof rapid evolution or
908 reduced selective constraint in Domestic cattle?" *BMC Genomics* (10) *179;* https:// doi:
909 10.1186/1471-2164-10-179.

910 Makino T, Rubin CJ, Carneiro M, Axelsson E, Andersson L, Webster MT. 2018 "Elevated
911 Proportions of Deleterious Genetic Variation in Domestic Animals and Plants." *Genome*
912 *Biol Evol.* 10(1):276-290. https://doi: 10.1093/gbe/evy004.

913 McDonald, J H, and M Kreitman. 1991. “Accelerated Protein Evolution at the Adh Locus in
914 Drosophila.” *Nature* 351: 652–54.

915 McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew
916 Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce
917 Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20
918 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

919 Messer, Philipp W., and Dmitri A. Petrov. 2013. “Frequent Adaptation and the McDonald-
920 Kreitman Test.” *Proceedings of the National Academy of Sciences of the United States of*
921 *America* 110 (21): 8615–20. https://doi.org/10.1073/pnas.1220835110.

922 Montanucci, Ludovica, Hafid Laayouni, Giovanni Marco Dall’Olio, and Jaume Bertranpetit.
923 2011. “Molecular Evolution and Network-Level Analysis of the N-Glycosylation Metabolic
924 Pathway across Primates.” *Molecular Biology and Evolution* 28 (1): 813–23.

https://doi.org/10.1093/molbev/msq259.

Moon, Sunjin, Tae-Hun Kim, Kyung-Tai Lee, Woori Kwak, Taeheon Lee, Si-Woo Lee, Myung-Jick Kim, et al. 2015. “A Genome-Wide Scan for Signatures of Directional Selection in Domesticated Pigs.” *BMC Genomics* 16 (1): 1–12. https://doi.org/10.1186/s12864-015-1330-x.

Nevado, Bruno, Sebastian E. Ramos-Onsins, and Miguel Perez-Enciso. 2014. “Resequencing Studies of Nonmodel Organisms Using Closely Related Reference Genomes: Optimal Experimental Designs and Bioinformatics Approaches for Population Genomics.” *Molecular Ecology* 23 (7): 1764–79. https://doi.org/10.1111/mec.12693.

Orlando L, Librado P. 2019 "Origin and Evolution of Deleterious Mutations in Horses." *Genes* 10, 649; doi:10.3390/genes10090649

Pérez-Enciso, M., G. de los Campos, N. Hudson, J. Kijas, and A. Reverter. 2017. “The ‘Heritability’ of Domestication and Its Functional Partitioning in the Pig.” *Heredity* 118: 160–68. https://doi.org/10.1038/hdy.2016.78.

Quinlan, Aaron R. 2014. “BEDTools: The Swiss-Army Tool for Genome Feature Analysis.” *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]* 47 (January): 11.12.1-11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.

Ramírez, Oscar, William Burgos-Paz, Encarna Casas, Maria Ballester, Erica Bianco, Iñigo Olalde, Gabriel Santpere, et al. 2014. “Genome Data from a Sixteenth Century Pig Illuminate Modern Breed Relationships.” *Heredity* 114 (2): 175–84. https://doi.org/10.1038/hdy.2014.81.

Ramsay, Heather, Loren H Rieseberg, and Kermit Ritland. 2009. “The Correlation of Evolutionary Rate with Pathway Position in Plant Terpenoid Biosynthesis.” *Molecular Biology and Evolution* 26 (5): 1045–53. https://doi.org/10.1093/molbev/msp021.

Rausher, Mark D, Richard E Miller, and Peter Tiffin. 1999. “Patterns of Evolutionary Rate Variation Among Genes of the Anthocyanin Biosynthetic Pathway.” *Molecular Biology and Evolution* 16 (2): 266–74. https://watermark.silverchair.com/mbev_16_02_0266.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAdwwggHYBgkqhkiG9w0BBwagggHJMIIB

954 xQIBADCCAb4GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMB_IVBQyX1n-
955 F9I4pAgEQgIIBj8r0BkU_bdeIWEuY_7bUxMFToA8Yo-26snF_yPY.

956 Renaut, Sebastien, and Loren H. Rieseberg. 2015. "The Accumulation of Deleterious Mutations
957 as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae
958 Crops." *Molecular Biology and Evolution* 32 (9): 2273–83.
959 https://doi.org/10.1093/molbev/msv106.

960 Riley, Rebecca M., WeI Jin, and Greg Gibson. 2003. "Contrasting Selection Pressures on
961 Components of the Ras-Mediated Signal Transduction Pathway in Drosophila." *Molecular*
962 *Ecology* 12 (5): 1315–23. https://doi.org/10.1046/j.1365-294X.2003.01741.x.

963 Rubin, Carl-Johan, Hendrik-Jan Megens, Alvaro Martinez Barrio, Khurram Maqbool, Shumaila
964 Sayyab, Doreen Schwochow, Chao Wang, et al. 2012. "Strong Signatures of Selection in
965 the Domestic Pig Genome." *Proceedings of the National Academy of Sciences of the United*
966 *States of America* 109 (48): 19529–36. https://doi.org/10.1073/pnas.1217149109.

967 Tajima, Fumio. 1983. "Evolutionary Relationship of DNA Sequences in Finite Populations."
968 *Genetics* 105: 437–60.
969 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1202167/pdf/437.pdf.

970 Uricchio LH, Petrov DA, Enard D. 2019 "Exploiting selection at linked sites to infer the rate and
971 strength of adaptation." *Nat Ecol Evol*. 3(6):977-984. doi: 10.1038/s41559-019-0890-6.

972 Watterson, G.A. 1975. "On the Number of Segregating Sites in Genetical Models without
973 Recombination." *Theoretical Population Biology* 7 (2): 256–76.
974 https://doi.org/10.1016/0040-5809(75)90020-9.

975 Wilkinson S, Lu ZH, Megens HJ, Archibald AL, Haley C, Jackson IJ, Groenen MA, Crooijmans
976 RP, Ogden R, Wiener P. 2013 "Signatures of diversifying selection in European pig
977 breeds." PLoS Genet. 2013 Apr;9(4):e1003453. doi: 10.1371/journal.pgen.1003453.

978 Zeder, Melinda A. 2012. "The Domestication of Animals." *Journal of Anthropological Research*
979 68 (2): 161–90. https://doi.org/10.3998/jar.0521004.0068.201.

980

981 **TABLES**

982 ***Table 1.*** *Number of SNPs (in whole genome, in genes and in coding regions), classified by its*
983 *presence in each population.*

984 ***Table 2.*** *Combinations of SNPs from coding regions according to its allelic status in each*
985 *population (A: Ancestral allele, F: Fixed allele, P: Polymorphic allele). SNPs that are missing in*
986 *any of the populations are not considered in this table.*

987

988 **FIGURES**

989 **Figure 1.** Levels of variation at synonymous (A) and nonsynonymous (B) sites for each pig
990 population and variability estimates and for shared and exclusive variants. WB; wild boar
991 population; IB, Iberian breed; LW, Large White breed.

992 **Figure 2.** Estimates of α for each pig population. Total variants (A), exclusive variants (B), shared
993 variants (C) and shared variants between IB and LW (D). Bootstrap intervals at 95% are indicated
994 by a line at each bar. WB; wild boar population; IB, Iberian breed; LW, Large White breed.

995 **Figure 3.** Estimates of R_βγ for each pig population and for all, exclusive and shared variants.
996 WB; wild boar population; IB, Iberian breed; LW, Large White breed.

997 **Figure 4.** Estimates of the median values of α for each pig population, different molecular scales
998 and for all, exclusive and shared variants. For each population, the order of different α's is: Fu&Li,
999 Watterson, Tajima and Fay&Wu. WB; wild boar population; IB, Iberian breed; LW, Large White
1000 breed.

1001 **Figure 5**. Posterior distribution of the α values for total variants. Four different estimators of alpha
1002 (Fu&Li, Watterson, Tajima and Fay&Wu, see Materials and Methods) are used. Box plots indicate
1003 simulated distributions of α values. Red line indicates observed α values.

1004 **Figure 6**. Posterior distribution of the α values for exclusive variants. Four different estimators of
1005 alpha (Fu&Li, Watterson, Tajima and Fay&Wu, see Materials and Methods) are used. Box plots
1006 indicate the simulated distributions of α values. Red line indicates observed α values.

1007

1008 **SUPPLEMENTARY MATERIAL**

1009 See Supplementary Material file added to see the Tables (S1-S13) and Figures (S1-S50).

Table 1. Number of SNPs present in the whole genome, genes and coding regions and classified according to their presence in pig populations.

	Number of SNPs	**Shared between WB, IB and LW**	**Shared between WB and IB**	**Shared between WB and LW**	**Shared between IB and LW**	**Exclusive of WB**	**Exclusive of IB**	**Exclusive of LW**
Whole-genome	24,869,699	7,293,787	666,927	4,017,107	138,378	4,239,052	385,504	8,128,944
Genes	6,684,142	1,964,562	98,433	1,152,555	48,351	1,138,370	100,550	2,181,321
Coding regions	149,440	18,611	3,252	25,044	1,177	51,432	3,356	46,568

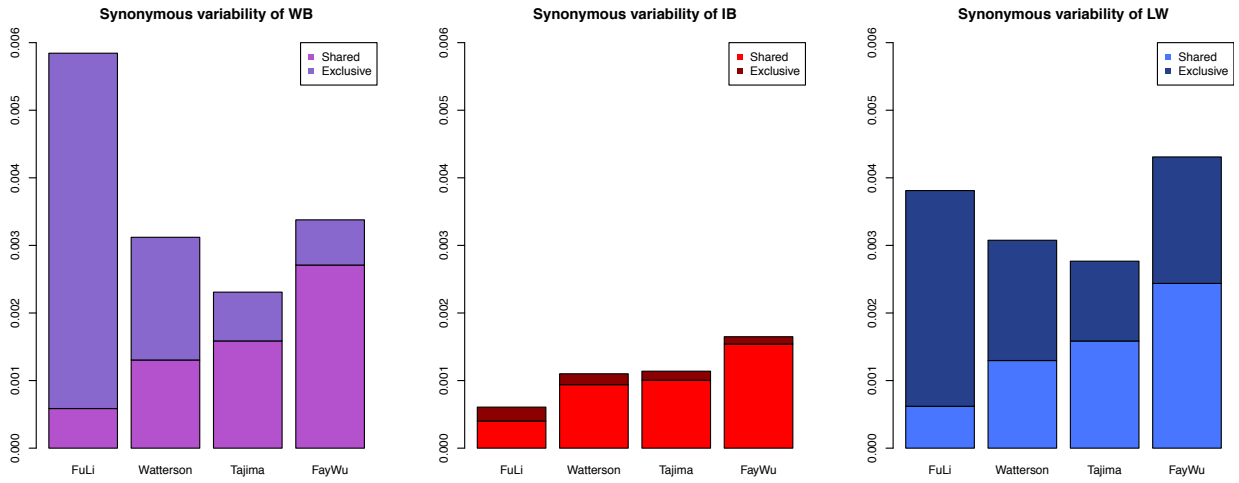
Table 2. Number of synonymous and nonsynonymous SNPs for different combinations of the allelic status. SNPs that are missing in any of the populations are not considered in this table.

IB	LW	WB	Synonymous	Non-synonymous
F	F	F	20297	9342
P	P	P	11712	7597
A	A	F	0	0
A	F	A	0	0
F	A	A	3	5
A	A	P	30314	20988
A	P	A	26027	15035
P	A	A	1833	1588
A	F	F	0	0
F	A	F	1	0
F	F	A	1	0
A	P	P	10128	7930
P	A	P	1676	1254
P	P	A	700	363
A	F	P	11	1
A	P	F	0	2
F	A	P	30	30
P	A	F	0	0
F	P	A	8	4
P	F	A	1	1
F	P	P	4924	2378
P	F	P	242	139
P	P	F	81	52
F	F	P	1140	489
F	P	F	4911	2073
P	F	F	38	22
			114078	69293

F: position with a fixed derived variant. **P**: Polymorphic position. **A**: position with the ancestral varian

Figure 1

A


Synonymous variability of WB
Synonymous variability of IB
Synonymous variability of LW
Shared
Exclusive
0.000
0.001
0.002
0.003
0.004
0.005
0.006
FuLi
Watterson
Tajima
FayWu

B

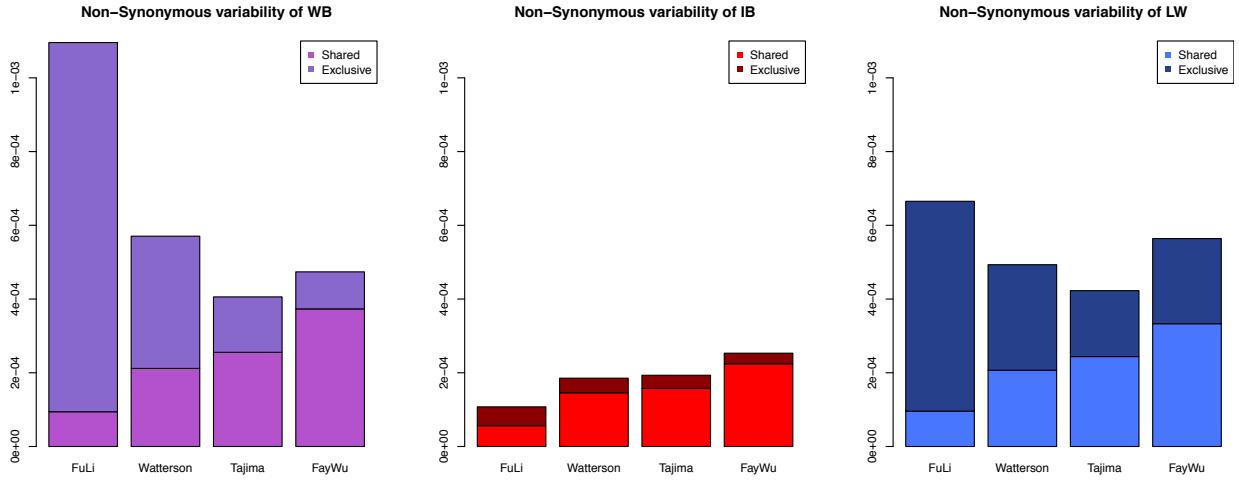

Non−Synonymous variability of WB
Non−Synonymous variability of IB
Non−Synonymous variability of LW
Shared
Exclusive
0e+00
2e−04
4e−04
6e−04
8e−04
1e−03
FuLi
Watterson
Tajima
FayWu

Figure 2

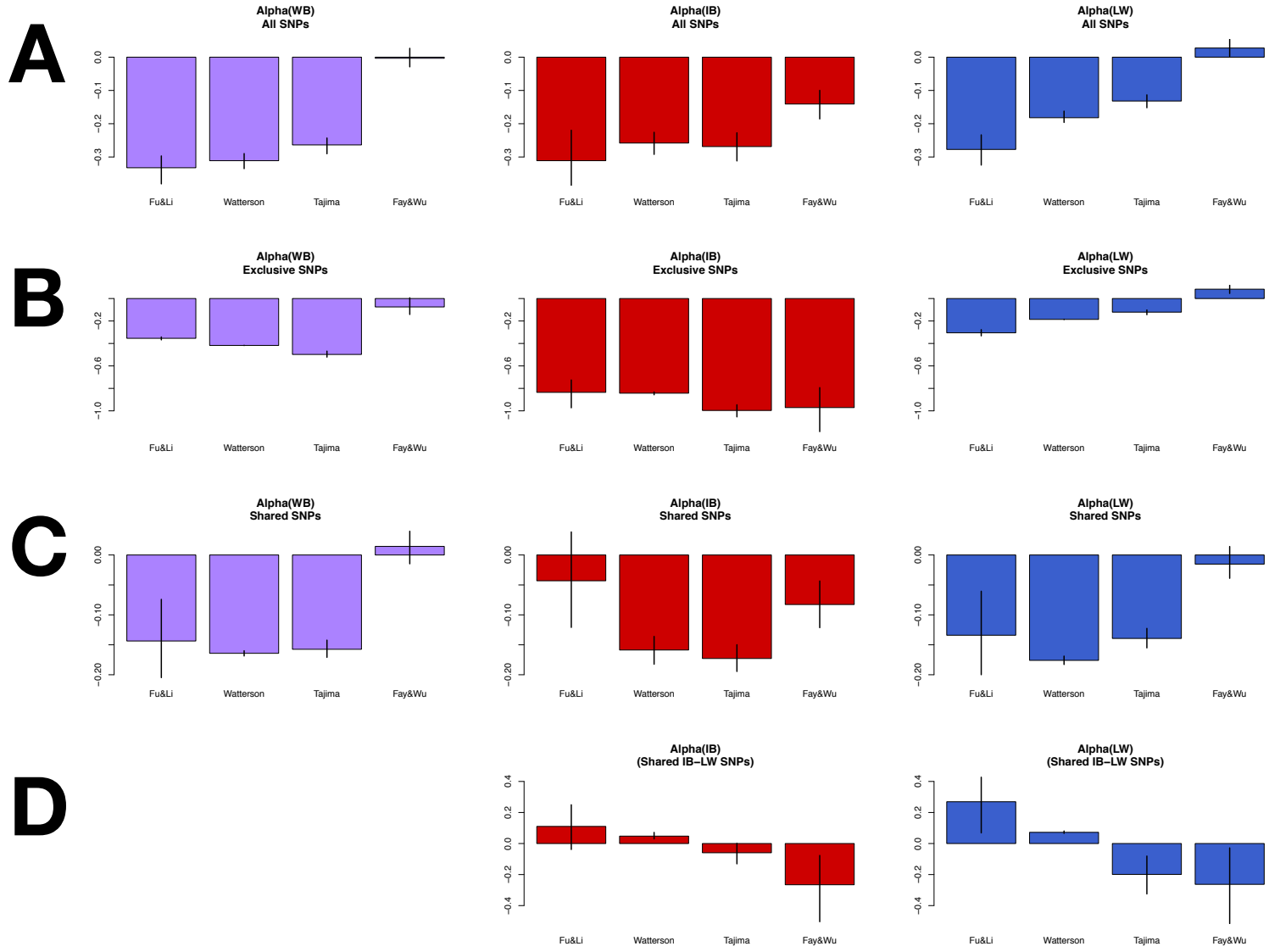

A
Alpha(WB)
All SNPs
Alpha(IB)
All SNPs
Alpha(LW)
All SNPs
B
Alpha(WB)
Exclusive SNPs
Alpha(IB)
Exclusive SNPs
Alpha(LW)
Exclusive SNPs
C
Alpha(WB)
Shared SNPs
Alpha(IB)
Shared SNPs
Alpha(LW)
Shared SNPs
D
Alpha(IB)
(Shared IB−LW SNPs)
Alpha(LW)
(Shared IB−LW SNPs)
Fu&Li
Watterson
Tajima
Fay&Wu

Figure 3

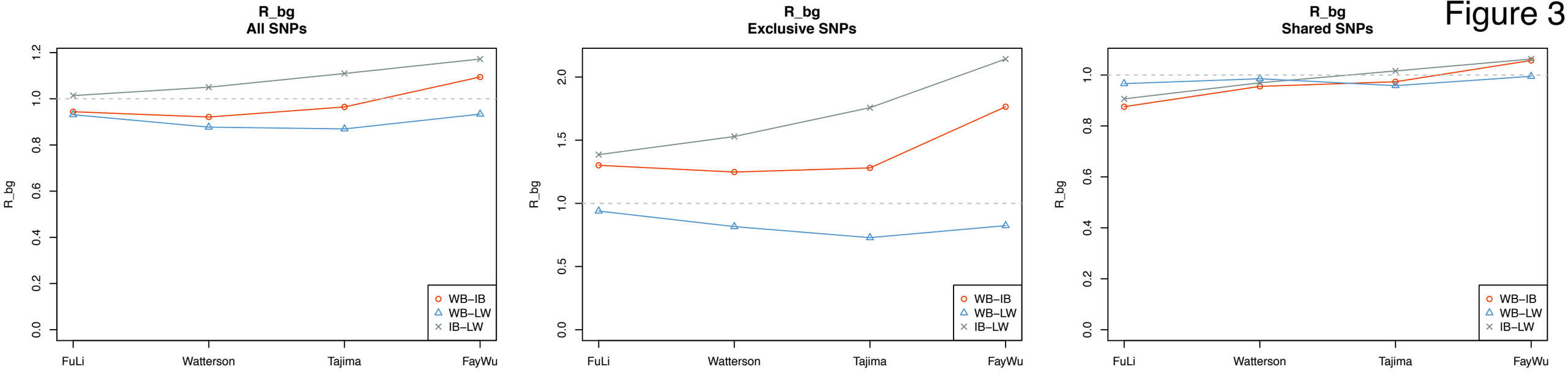

R_bg
All SNPs
R_bg
Exclusive SNPs
R_bg
Shared SNPs
R_bg
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.5
1.5
2.0
FuLi
Watterson
Tajima
FayWu
WB–IB
WB–LW
IB–LW

Figure 4

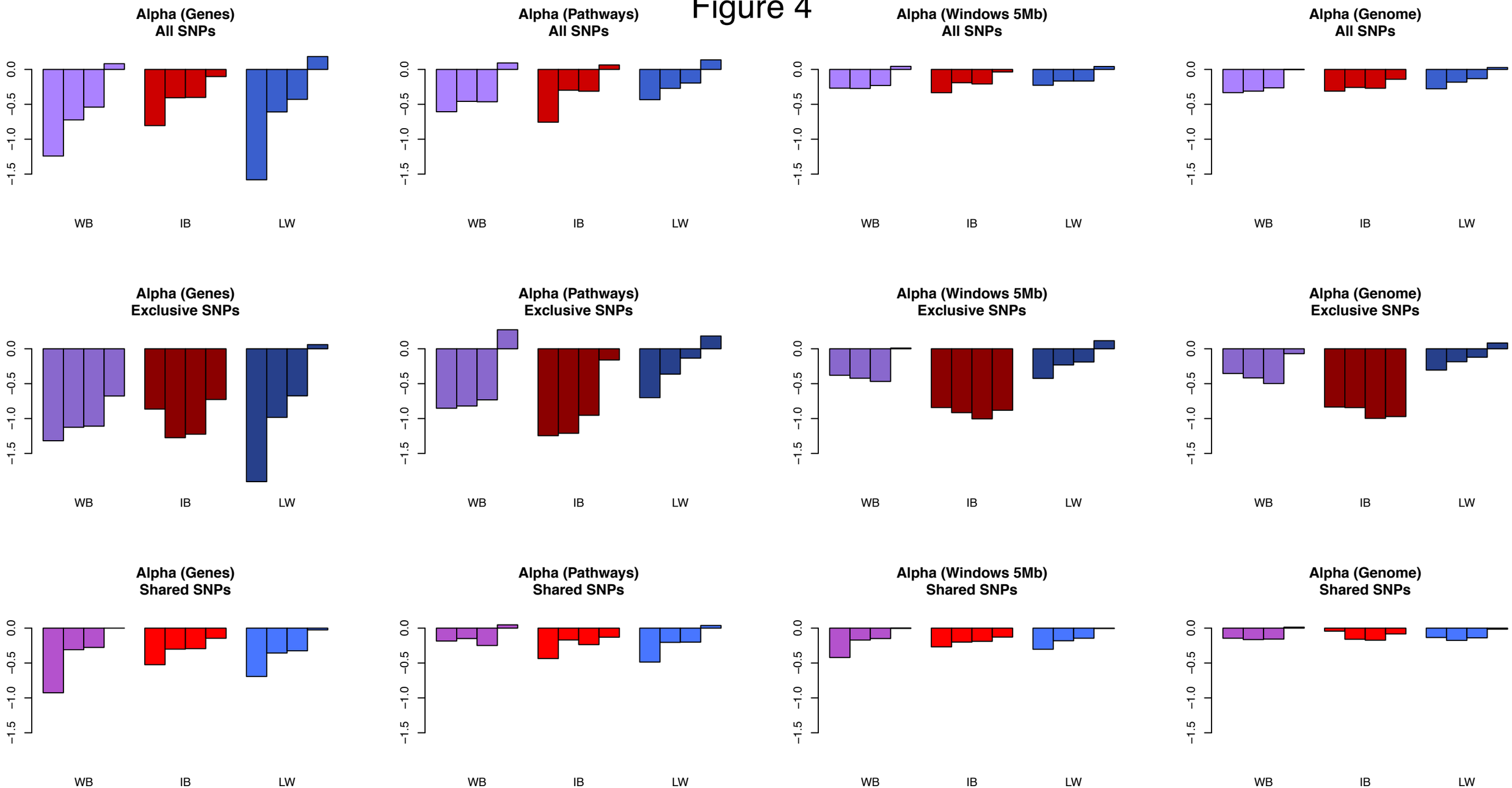

Alpha (Genes)
All SNPs
Alpha (Pathways)
All SNPs
Alpha (Windows 5Mb)
All SNPs
Alpha (Genome)
All SNPs
Alpha (Genes)
Exclusive SNPs
Alpha (Pathways)
Exclusive SNPs
Alpha (Windows 5Mb)
Exclusive SNPs
Alpha (Genome)
Exclusive SNPs
Alpha (Genes)
Shared SNPs
Alpha (Pathways)
Shared SNPs
Alpha (Windows 5Mb)
Shared SNPs
Alpha (Genome)
Shared SNPs
0.0
−0.5
−1.0
−1.5
WB
IB
LW

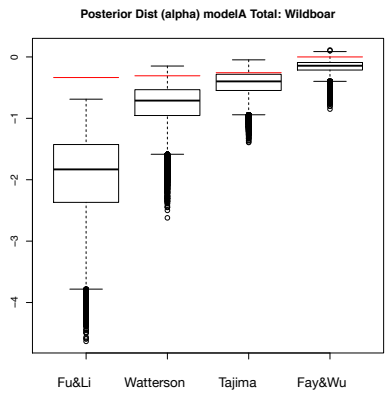
Figure 5

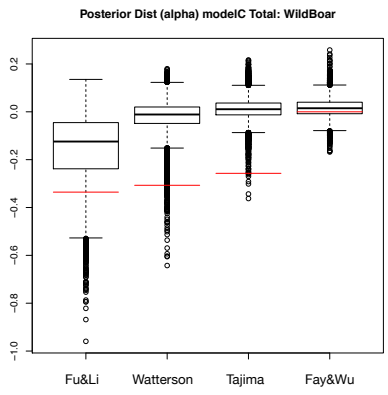
model A
(only deleterious)

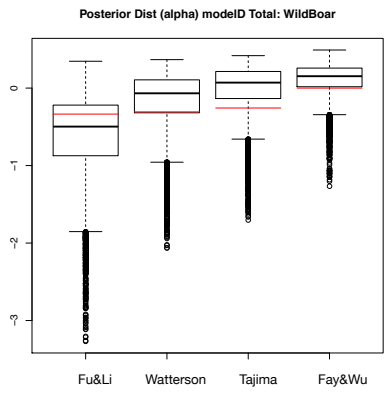
model C
(deleterious plus beneficial)

model D
(deleterious plus beneficial, discrete distribution)

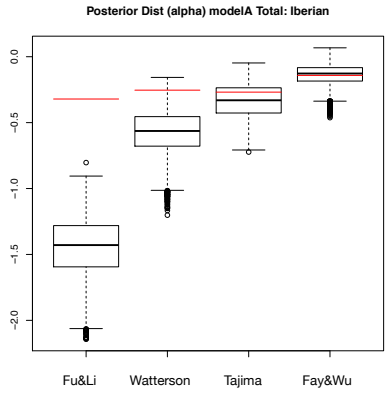
WB

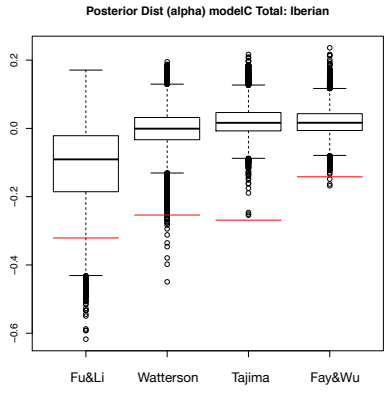


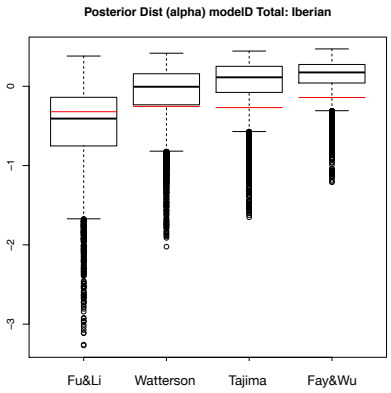




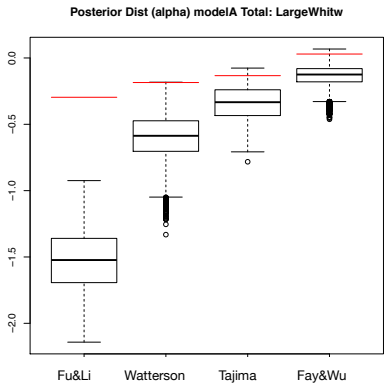
IB

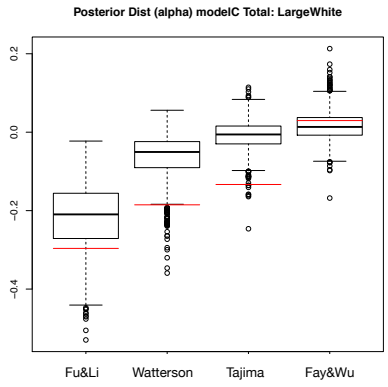






LW





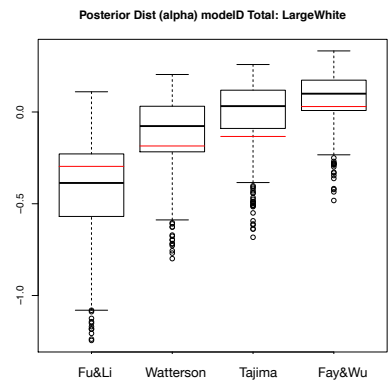


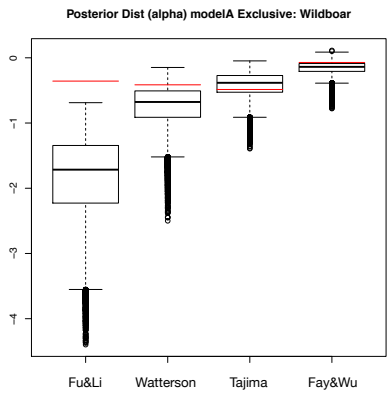
Figure 6

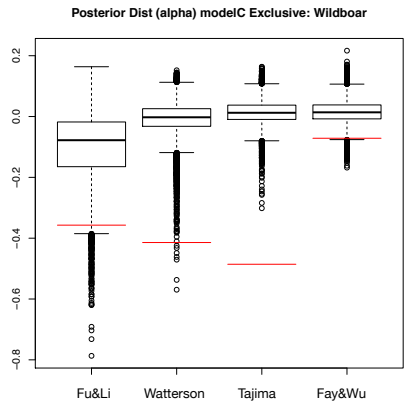
model A
(only deleterious)

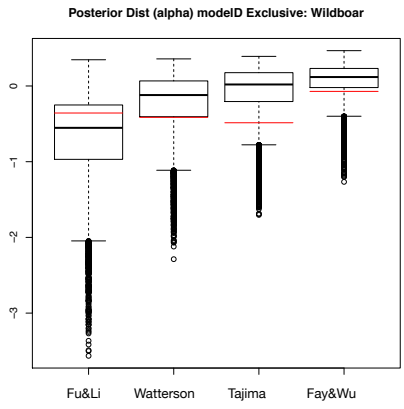
model C
(deleterious plus beneficial)

model D
(deleterious plus beneficial, discrete distribution)

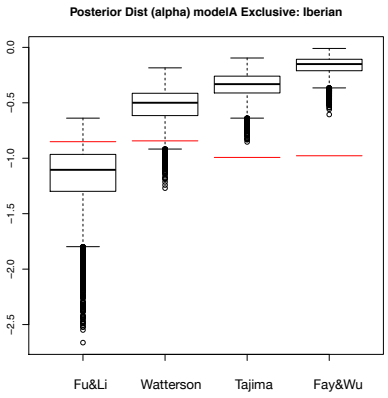
WB

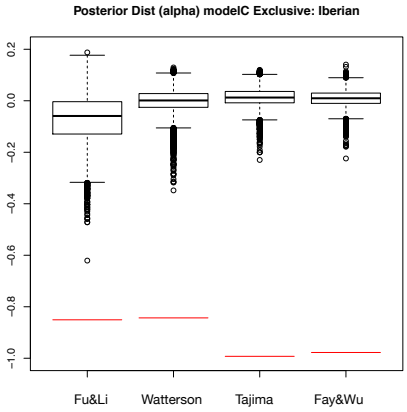


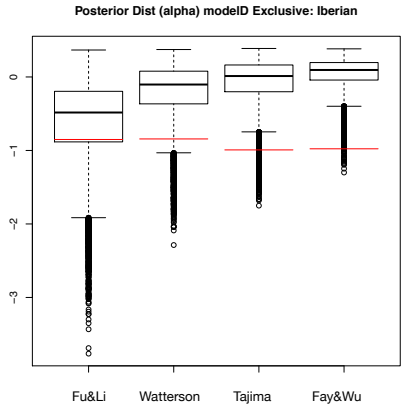
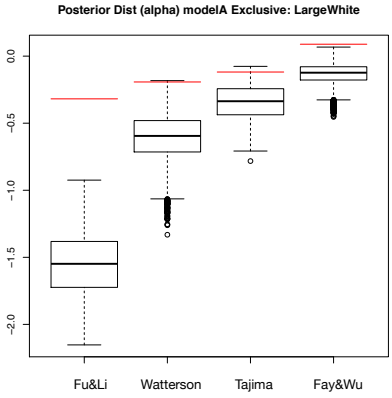
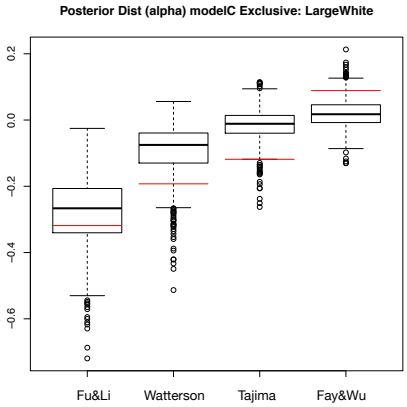
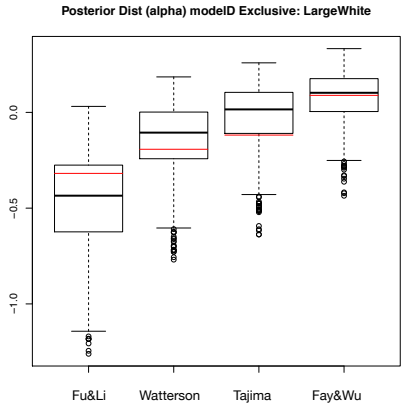




IB







LW