Genome-wide association study and regional heritability mapping of protein efficiency and performance traits in Swiss Large White pigs

Esther Oluwada Ewaoluwagbemiga[1,2], Audald Lloret-Villas[2], Adéla Nosková[2], Hubert Pausch[2] and Claudia Kasper[1*]

[1]Animal GenoPhenomics, Agroscope, Tioleyre 4, 1725 Posieux, Switzerland

[2]Animal Genomics, Department of Environmental Systems Science, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland

*Corresponding author

E-mail addresses:

E.O.E: esther.ewaoluwagbemiga@usys.ethz.ch

A.L.V: avillas@usys.ethz.ch

A.N: adela.Noskova@usys.ethz.ch

H.P: hubert.pausch@usys.ethz.ch

C.K: claudia.kasper@agroscope.admin.ch

Abstract

Background

The improvement of protein efficiency (PE) is a key factor for a sustainable pig production as nitrogen excretion contributes substantially to environmental pollution. Protein efficiency has been shown to be clearly heritable and genetically correlated with some performance traits, such as feed conversion ratio (FCR) and average daily feed intake (ADFI). The study aimed to identify genomic regions associated with these traits through genome-wide association studies (GWAS) and regional heritability mapping (RHM) using imputed whole genome sequence variants for more than 1,000 Swiss Large White pigs.

Results

The genomic-based heritability estimates using ~15 million SNPs were moderate, ranging from 0.33 to 0.47. Using GWAS, no significant SNPs were found at the genome-wide Bonferroni and false-discovery rate (FDR) thresholds for any of the traits, with the exception of ADFI, where 52 significant SNPs were found on chromosome 1 at the FDR threshold. No region was found to be significant at the Bonferroni threshold using RHM. Regional heritability mapping found two suggestive regions for PE on chromosomes 2 and 9 located between 31 and 32Mb and between 2 and 3Mb, respectively, one suggestive region for ADG on chromosome 5 between 104 and 105Mb, and four suggestive regions for ADFI on chromosomes 1 (270-271Mb, 272-273Mb and 273-274Mb) and 14 (133-134Mb).

Conclusions

Our study identified suggestive regions for PE and the performance traits with RHM, except for FCR. However, the apparent difficulty in detecting significant regions probably reflects the relatively small sample size used in this study rather than a lack of true associations with PE. The finding of this study

helps to understand the polygenicity of PE and may help in the design of breeding for protein-efficient pigs in a genomic prediction.

Background

Efficient livestock production is gaining importance due to the increasing global demand for meat that has led to an increased environmental pollution. A key pollutant in livestock production is nitrogen, which forms harmful compounds such as nitrate, ammonia and nitrous oxide [1, 2, 3]. Although the environmental impact of pig production is lower than that of beef production, approximately 50% of the dietary protein consumed by pigs is excreted as waste [4, 5]. Methods such as reducing dietary nitrogen [6, 7] and selection to increase protein efficiency (PE; the proportion of total dietary protein intake retained in the carcass) in pigs [8] have been proposed to reduce the contribution of animal-based food production to environmental pollution. Varying heritability estimates between 0.21 and 0.59 have been recently reported for PE and related traits (e.g., nitrogen digestible coefficient), depending on the breed, fattening phase and diet type [8, 9, 10]. The heritability (h^2) estimates for PE from these studies indicate that this trait can be genetically improved and thus presents a promising target towards a more sustainable pig production through reduced nitrogen excretion. Performance traits such as feed conversion ratio (FCR), average daily feed intake (ADFI), and average daily gain (ADG) are also important, considering their economic and environmental impacts. Several studies have reported moderate heritabilities for these traits [8, 11]. Genetic correlations (±SE) of -0.55 ± 0.14, -0.53 ± 0.14, and -0.19 ± 0.19 have also been previously reported between PE and FCR, ADFI, and ADG, respectively, in Swiss Large White pigs [8].

Genome-wide association studies (GWASs) in pigs have reported loci associated with several important traits, such as meat quality [12], performance [13, 14], carcass [15], body composition [16], and efficiency-related traits [17]. However, despite the environmental importance of nutrient

65 efficiency traits, such as PE and nitrogen excretion, to date, only the study of Shirali et al. [18] has
66 identified genomic regions associated with nitrogen-excretion traits in pigs. The study used 315 pigs
67 from Pietrain grand-sires and grand-dams from a three-way cross, with pigs genotyped for 88
68 microsatellite markers on 10 (of 18) chromosomes [18]. Their study identified three quantitative trait
69 loci (QTL) associated with total nitrogen excretion throughout the 60 – 140 kg live body weight (BW)
70 growth period on chromosomes (SSC) 2, 4, and 7 [18]. Three additional QTL were found for another
71 excretion trait – average daily nitrogen excretion – on SSC 6, 9, and 14 for the same growth period
72 [18]. However, the study by Shirali et al. [18] is limited by its very small sample size and the use of
73 a small number of markers that are not evenly distributed across the genome.

74 As FCR, ADG, and ADFI have either direct or indirect impacts on efficiency and production costs, a
75 number of studies have identified QTL for FCR, ADG and ADFI [19, 20, 21, 22]. The majority of
76 these QTL were found in Duroc and Landrace pigs, with only a few QTL identified in Large White
77 pigs [23]. However, although FCR and RFI may be correlated with PE as reported in the studies by
78 Ewaoluwagbemiga et al. [8] and Saintilan et al. [11], it has been suggested that selection for improved
79 FCR and RFI with the aim of reducing nutrient excretion is clearly less efficient than direct selection
80 for the nutrient efficiency trait itself (e.g., PE) in poultry [24].

81 Besides GWAS, regional heritability mapping (RHM) is another approach used to identify genotype–
82 phenotype associations [25, 26]. Unlike GWAS, RHM has been proposed to have the ability to detect
83 regions that contribute to the genetic variance of a trait, but individually have too small an effect to
84 be detected by GWAS, because it integrates multiple SNP effects by analysing regions rather than
85 single SNPs [27]. Complex traits are typically influenced by many genes (i.e., are polygenic), with
86 many genetic variants having too small effect sizes to be detected at the Bonferroni-corrected or false-
87 discovery rate (FDR) threshold of GWAS [28, 29], giving RHM some advantage over GWAS. RHM

88 is done by dividing the genome into small segments or regions, building a genomic relationship
89 matrix (GRM) by using all the SNPs in each region, followed by estimating the variance of the trait
90 explained by each region [27]. RHM has been applied by Resende et al. [30], who detected 26 QTL
91 altogether associated with 7 traits in Eucalyptus, whereas GWAS detected only 13 QTL. Sutera et al
92 [31] found 5 QTL associated with fat percentage in sheep by using RHM, but RHM has, so far, been
93 applied rarely to livestock body composition.

94 The aim of this study was therefore to investigate the genetic basis of PE and performance traits in
95 Swiss Large White pigs. We did that by estimating genomic heritability and performing GWAS and
96 RHM using a low-pass sequence data.

97 Methods

98 Animals and phenotypes

99 We analysed a total of 1,036 pigs, which were previously included in several nutrition experiments
100 and one genetic study. The data set is described in detail in Ewaoluwagbemiga et al. [8]. Briefly, all
101 experiments were carried out at Agroscope Posieux in Switzerland. Pigs had *ad libitum* access to
102 isocaloric diets that differed in crude protein or fibre content, leading to five dietary treatment groups
103 including the control. The control group had no reduction in dietary crude protein, and diets were
104 formulated according to the Swiss feeding recommendations for pigs[1]; the diets of the protein-
105 restricted groups contained 80% of the crude protein and digestible essential amino acids content of
106 the control diets. In all experiments, pigs were fed a grower diet from approximately 20 to 60 kg live
107 BW and a finisher diet from 60 kg to slaughter at 100 kg. Pigs were slaughtered at about 100 kg BW

[1] Fütterungsempfehlungen und Nährwerttabellen für Schweine (Feeding recommendations and nutrient tables for pigs). Agroscope, Posieux, Switzerland. Retrieved 31 January 2017 from https://www.agroscope.admin.ch/agroscope/fr/home/services/soutien/aliments-pour-animaux/apports-alimentaires-recommandes-pour-les-porcs.html

in all experiments except one, where 52 pigs and 41 pigs were kept until 120 kg and 140 kg live BW, respectively, and fed another specially formulated finisher diet from 100 to 140 kg [32]. Every week, pigs were weighed individually, and, once a pig reached a live BW of approximately 20 kg, it was allocated to grower-finisher pens and the experimental treatments were started. This was done until a maximum number of 12 (or 24 or 48) pigs per pen (depending on the pen layout; minimum 1m^2 per pig and maximum 12 pigs/feeder) was reached. Pigs remained in their pen until slaughter.

Piglets were weaned at an average age of 27 ± 2 days after birth by removing the sow and were fed a standard starter diet with crude protein levels following the recommendation. At 22.3 (± 1.6) kg, pigs were placed in pens equipped with automatic feeders (single-spaced automatic feeder stations with individual pig recognition system by Schauer Maschinenfabrik GmbH & Co. KG, Prambachkirchen, Austria) and stayed on the starter diet. The automatic feeder recorded all visits and feed consumption per visit, from which the total feed intake of each pig was calculated. The protein content of feed was monitored during production by near-infrared spectroscopy for each 500 kg batch. To obtain more accurate data on feed composition at the time of consumption, a sample was taken from each automatic feeder station each week, and the crude protein content was determined by wet-chemistry methods.

Phenotype data

The phenotypes were derived as reported in Ewaoluwagbemiga et al. [8]. Total and average daily feed (ADFI) were recorded, and average daily gain (ADG) and the feed conversion ratio (FCR) were calculated as follows:

$$ADG = \frac{live\ BW\ (kg)\ slaughter - live\ BW\ (kg)\ start}{age\ (days)\ slaughter - age\ (days)\ start}$$

129 $$FCR = \frac{ADFI}{ADG}$$

130 where $live\ BW\ (kg)\ slaughter$ and $age\ (days)\ slaughter$ are the live pre-slaughter body weight
131 in kg and the age in days at slaughter, respectively, and $live\ BW\ (kg)\ start$ and $age\ (days)\ start$
132 are the exact body weight in kg and the age in days at the start of the grower phase, respectively. To
133 measure PE, the left carcass half, including the whole head and tail, was scanned with a dual-energy
134 X-ray absorptiometry (DXA; GE Lunar i-DXA, GE Medical Systems, Glattbrugg, Switzerland) to
135 determine the lean tissue content, which was used in the equation of Kasper et al. [33] to estimate the
136 protein content retained in the carcass. This method of estimating carcass protein content using DXA
137 yields a highly precise and accurate phenotype with an R^2 between 0.983 – 0.998 [33, 34]. PE was
138 calculated as

139 $$protein\ efficiency = \frac{CP_{carcass}\ (g)\ slaughter - CP_{carcass}\ (g)\ start}{CP_{feed}\ intake\ (\mathrm{g})}$$

140 The protein content of pigs at the start of this experiment ($CP_{carcass}\ (g) start$) was estimated from a
141 sample of 38 piglets (12 females, 12 castrated males and 14 entire males). These 38 piglets were
142 slaughtered at an average of 20.98 ± 1.85 kg BW in a previous experiment, and their carcass protein
143 content was chemically determined [32]. The average protein content per kg carcass for each sex
144 (female, entire male, castrated male) was used to estimate $CP_{carcass}\ (g)\ start$ for the pigs by
145 multiplying the actual live BW of pigs when they entered the experiment (i.e., at approximately 20
146 kg body weight) with the protein content per kg carcass of piglet, as previously determined from the
147 38 piglets [32].

148 ## Genotype data and imputation

149 DNA was extracted from blood, and the sampled pigs were genotyped on three different platforms,
150 namely the Affymetrix 600K axiom porcine genotyping array, whole-genome sequence data at an
151 intended 4-fold coverage, and low-pass sequence data at an intended 1-fold coverage with Gencove.
152 Thus, the three genotyping/sequencing platforms comprised of (i) 258 pigs genotyped at 600K
153 obtained with the Axiom Porcine Genotyping array; (ii) 297 pigs sequenced at an intended read depth
154 of 4×; and (iii) 492 pigs sequenced at an intended read depth of 1×. The array genotyping data was
155 imputed to whole genome sequence level with a reference panel consisting of 421 pigs (Landrace and
156 Large White) that were sequenced at a coverage ranging between 4× and 37.5× [23, 36]. For pigs
157 sequenced at 4× coverage, imputation of sporadically missing genotypes was done using Beagle (v
158 4.1 [36]). Finally, the low-pass genetic data sequenced at an intended depth of 1× was imputed by
159 Gencove using their loimpute pipeline v0.1.5 [37]. Therefore, the imputed array data finally contained
160 29,469,425 SNPs, the sequenced 4× genetic data contained 30,179,303 SNPs and the sequenced 1×
161 genetic data contained 45,100,556 SNPs (including 13,361,070 non-variant sites). Eight pigs without
162 phenotypes and three pigs with a mis-match between pedigree and genomic-based relationship matrix
163 were excluded from the further analyses. PLINK (v1.9) [38, 39] was used to merge the three different
164 SNP panels based on their physical positions according to the Sus Scrofa 11.1 assembly [40]. After
165 merging, there were 23,171,650 intersecting biallelic SNPs (including indels) and 1,036 individuals.

166 ## Genome-wide association study

167 Prior to GWAS, we tested for outliers in the phenotypes, and removed individuals with phenotypes
168 not in the range of $\mu \pm 3\sigma$. This resulted in 1025, 1033, 1034, and 1024 individuals remaining for PE,
169 ADG, ADFI, and FCR, respectively. For each trait, we removed SNPs with minor allele frequency
170 (MAF) < 5% and SNPs that deviated from Hardy-Weinberg equilibrium ($P<0.0001$). After quality

control, 15,269,953, 15,192,400, 15,200,584, and 15,220,328 SNPs were included for PE, ADFI, ADG, and FCR, respectively.

The residuals for each trait were used as phenotypes in GWAS by adjusting for environmental effects using the linear model in R software v 4.2.1 [41]. The environmental effects included the fixed effects from a model selection step prior to estimating genetic parameters as described in Ewaoluwagbemiga et al. [8]. In brief, the fixed effects included year (factor variable), treatment (factor variable), sex (factor variable), slaughter weight, ambient temperature in the barn at the start of the experiment, slaughter age, interaction of slaughter weight and sex, interaction of treatment and sex, interaction of treatment and slaughter age, and interaction of year and slaughter age.

The GWAS was performed with GCTA using the fastGWA method [42], where SNP effects were tested using a linear mixed effects model approach, incorporating the genomic relationship matrix (GRM) to account for relatedness in the sampled population. The linear mixed effects model fitted to the data was

$$Y_j = \mu + b_{ij}M_i + a_j + e_{ij}$$

where Y_j is a vector of residuals of phenotypes corrected for environmental effects; μ is the overall mean; b_{ij} are the marker genotypes, coded as 0, 1, and 2, of the ith SNP for the jth individual; M_i is the additive effect of the ith SNP; a_j is the random polygenic effect of the jth individual following the distribution $a_j \sim N(0,G\sigma^2_a)$, G is the GRM and σ^2_a is the additive genetic variance; e_{ij} is the random residual effect with $e_{ij} \sim N(0,I\sigma^2_e)$, I is an identity matrix and σ^2_e is the residual variance. We used the Bonferroni-corrected significance threshold at an alpha level of 0.05 and a 5% FDR [43] to correct for multiple testing. We inspected the quantile-quantile (QQ) plots for the inflation of small p-values, which could indicate false-positive association signals. The genomic inflation factors were calculated

to compare the deviation of the distribution of the observed to the distribution of the expected chi-square test statistics.

Regional heritability mapping

Regional heritability mapping was performed using GCTA software [42]. For this analysis, each chromosome (SSC1 to SSC18) was divided into 1Mb-window regions (i.e., 0-1Mb, 1-2Mb, 2-3Mb…...), and the variance was estimated for each region. The linear mixed effects model below was used to test the effect of all SNPs within each genomic region, which included the random regional genomic effect and the random genomic effect of the rest of the genome excluding a specific region:

$$Y_j = \mu + b_{ij}M_i + a_k + a_u + e_{ij}$$

where Y_j is a vector of residuals of phenotypes corrected for environmental effects; μ is the overall mean; b_{ij} are the marker genotypes, coded as 0, 1, and 2, of the ith SNP for the jth individual; M_i is the additive effect of the ith SNP; a_k is the random regional additive genomic effect of the kth region following the distribution $a_k \sim N(0, G_k\sigma^2_k)$ and a_u is the random polygenic effect of the rest of the SNPs following the distribution $a_u \sim N(0, G_u\sigma^2_u)$. G_k is the regional GRM, σ^2_k is the regional variance, G_u is the GRM that excludes a region and σ^2_u is the additive genetic variance excluding a region. Regional and genome heritability were estimated as $h^2_k = (\sigma^2_k/ \sigma^2_p)$ and $h^2_u= (\sigma^2_u/ \sigma^2_p)$, respectively, where σ^2_p is the sum of the regional variance (σ^2_k), genome variance (σ^2_u), and residual variance (σ^2_e). The statistical significance of the variance of a region was tested using the likelihood ratio test (LRT), which compares the log likelihood of the full model (including regional and genome variance) with the reduced model (including only genome variance). This was done by specifying the –reml-lrt 1 option in GCTA, which gives the LRT and p-value of the first genetic variance.

For a total of 2,146 regions, a Bonferroni correction for multiple testing was applied at 0.05 alpha level. It should be noted that the thresholds presented for GWAS and RHM, in addition to the Bonferroni correction, differ between the methods. Instead of an FDR, we set a suggestive threshold for RHM following the procedure described in [31]. The suggestive threshold implies that, at every genome scan, one false positive is expected [31]. The thresholds applied in the current study were thus at p-values of 4.66×10^{-5} ($-\log_{10}(p) = 4.34$) and 9.32×10^{-4} ($-\log_{10}(p) = 3.03$) for the genome-wide 5% significance and the suggestive threshold, respectively.

Results

GWAS

The Bonferroni and FDR threshold for all traits, at the alpha level of 0.05, was 3.29×10^{-9} and 3.00×10^{-8}, respectively, except for ADFI where the FDR threshold was 1.65×10^{-7}. No significant SNP was found at the Bonferroni and FDR thresholds for PE (Figure 1), FCR (Figure 2), and ADG (Figure 3). For ADFI, there was no significant SNP at the Bonferroni threshold, but there were 52 significant SNPs at the FDR threshold on SSC1 (Figure 4). Although no significant associations were found for PE, FCR and ADG, the genomic heritability for these traits, including ADFI, ranged from 0.33 to 0.47 using all available SNPs for each trait (Table 1), with ADG having a slightly higher genomic heritability (0.47) than pedigree-based heritability (0.45). The 52 siginificants SNPs at the FDR threshold for ADFI explained 3% of the total variation observed for this trait (Table 2). The QQ plots of the GWAS analyses (Figure 5) and the genomic inflation factor, which was close to 1 for all the traits, suggested that the influences of population structure and inflation have been sufficiently corrected for.

235 **Regional heritability mapping**

236 A total of 2,146 genomic regions were subjected to RHM, with the genomic regions including
237 between 6,687 and 6,721 SNPs on average per region for all traits. For all traits, no region was found
238 significant at the genome-wide level (Figure 1 – Figure 4). Two regions reached the suggestive
239 threshold for PE on SSC9 (2Mb – 3Mb) and SSC2 (31Mb – 32Mb) (Table 2 and Figure 1); no region
240 reached the suggestive threshold for FCR (Figure 2), one region on SSC5 reached the suggestive
241 threshold for ADG (Table 2 and Figure 3), and four regions reached the suggestive threshold for
242 ADFI on SSC1 (270 – 271Mb), SSC1 (272 – 273Mb), SSC1 (273 – 274Mb), and SSC14 (133 –
243 134Mb) (Table 2 and Figure 4). For PE, the suggestive region identified by RHM on SSC9 was
244 present in the top 50 SNPs (ordered from the lowest p-value) of GWAS, and 7 of the top 10 SNPs by
245 GWAS were seen on SSC2 (~80Mb) and SSC9 (2 and 3Mb) (Table S1). For ADG, the suggestive
246 region identified by RHM was on SSC5 (104 – 105Mb), and GWAS identified a SNP on SSC5
247 (~102Mb), which was the second top SNP by GWAS (Table S2). For ADFI, the suggestive regions
248 identified by RHM on SSC1 were present in the top 50 SNPs of GWAS, and 9 of the top 10 SNPs by
249 GWAS were seen on the 273 and 274Mb region (Table S3). The heritability captured by each of the
250 suggestive regions was between 0.04 – 0.05 for PE, 0.04 - 0.09 for ADFI, and the single region at the
251 suggestive threshold for ADG had a heritability of 0.26. The heritability captured by the suggestive
252 regions for ADFI was similar to the heritability captured by the siginificant SNPs with GWAS at the
253 FDR threshold.

254 **Table 1: Descriptive statistics and pedigree and genomic heritability estimates for protein**
255 **efficiency and performance traits**

Trait	*N*	Mean ± SD	Min - Max	h^2_{ped} (SE)	h^2_{geno} (SE)
PE	1025	0.39 ± 0.03	0.28 – 0.49	0.54 (0.10)	0.42 (0.05)
FCR	1024	2.68 ± 0.21	1.99 – 3.77	0.39 (0.12)	0.33 (0.04)
ADG	1033	0.85 ± 0.11	0.51 – 1.20	0.45 (0.11)	0.47 (0.04)
ADFI	1034	2.26 ± 0.31	1.30 – 3.14	0.53 (0.12)	0.43 (0.05)

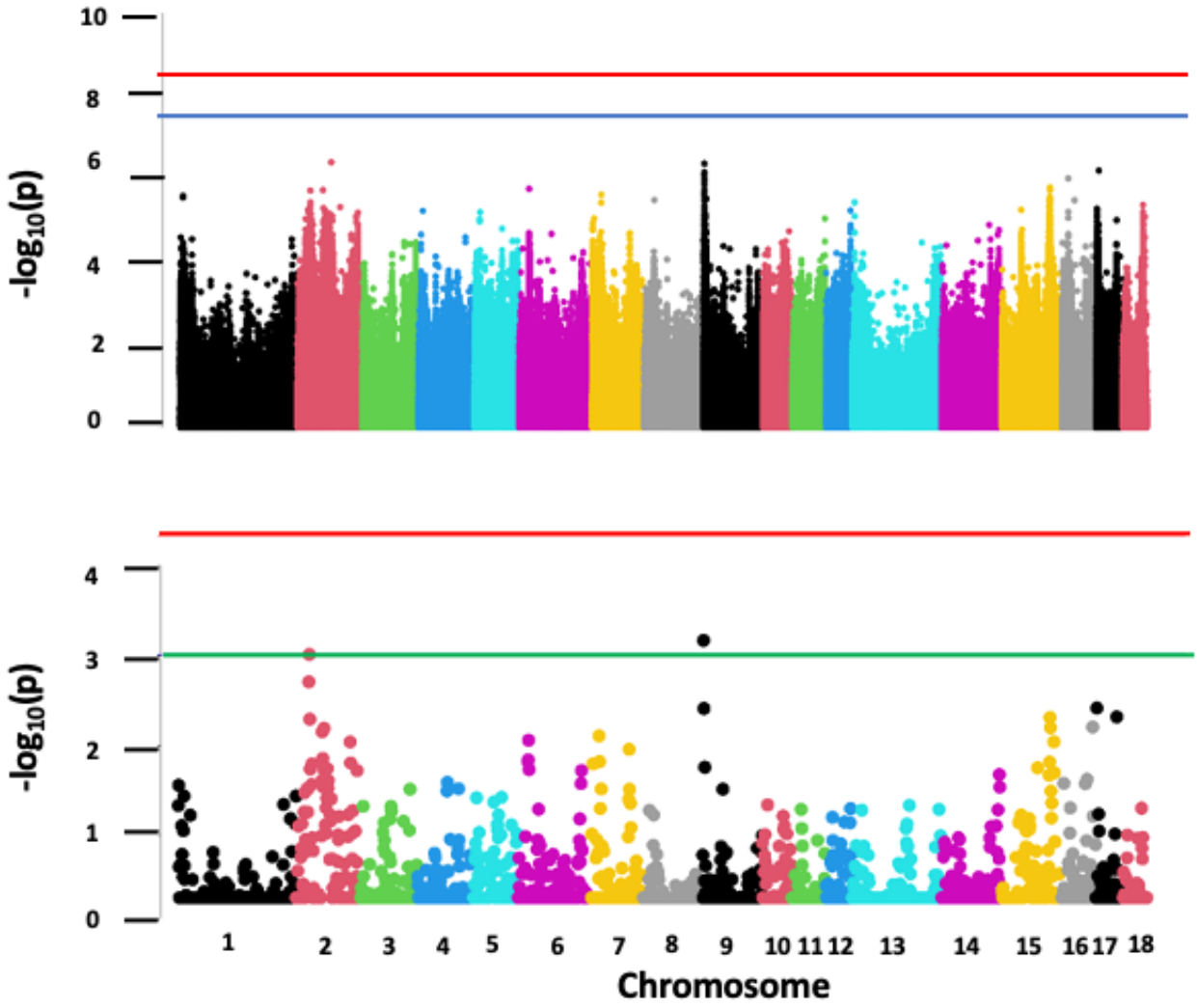
256 PE**:** Protein efficiency; FCR: feed conversion ratio; ADG: average daily gain; ADFI: average daily feed intake

257 **Table 2: SNPs at the FDR threshold for GWAS and at the suggestive threshold for RHM**

	Trait	SSC	Number of SNPs in a region	SNP position (bp)	$-\log_{10}(p)$	h^2
GWAS						
	PE	-	-	-	-	-
	ADG	-	-	-	-	-
	ADFI	1	52	270,589,693 - 273,545,902	6.80 - 7.96	0.03
RHM	PE	9	13,440	2,000,246 - 2,998,788	3.17	0.04
	PE	2	5399	31,000,420 - 31,998,421	3.02	0.05
	ADG	5	1911	104,000,464 - 104,517,994	3.21	0.26
	ADFI	1	14,224	273,000,002 - 273,999,851	4.16	0.04
	ADFI	1	12,799	272,000,033 - 272,999,968	3.24	0.04
	ADFI	14	12,993	133,000,013 - 133,999,879	3.86	0.09
	ADFI	1	11,327	270,000,182 - 270,999,980	3.03	0.03

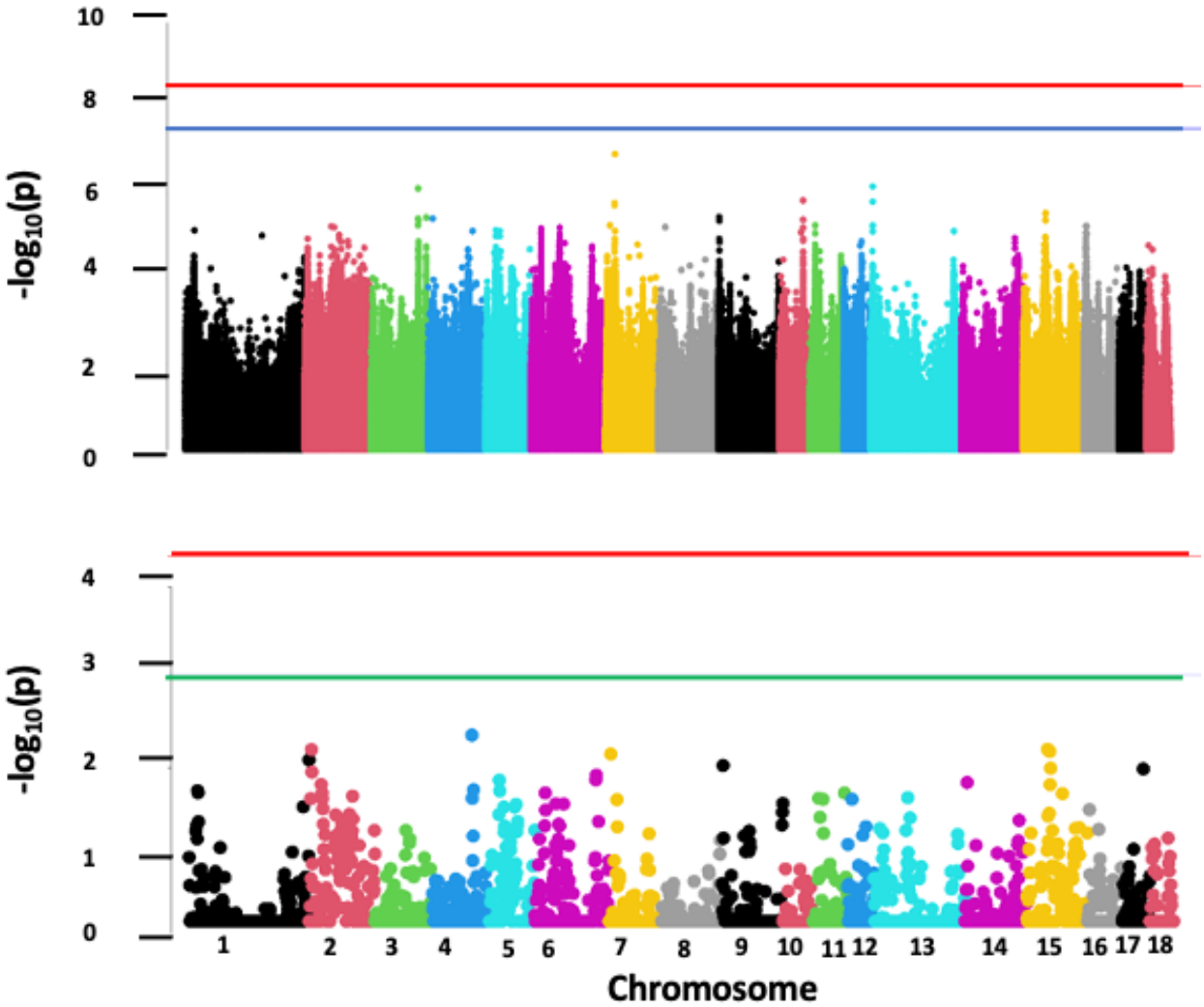
258 The FDR threshold ($-\log_{10}(p)$) for ADFI is 6.79, and the suggestive threshold ($-\log_{10}(p)$) for all traits with RHM is 3.03

259 PE: Protein efficiency; FCR: feed conversion ratio; ADG: average daily gain; ADFI: average daily feed intake



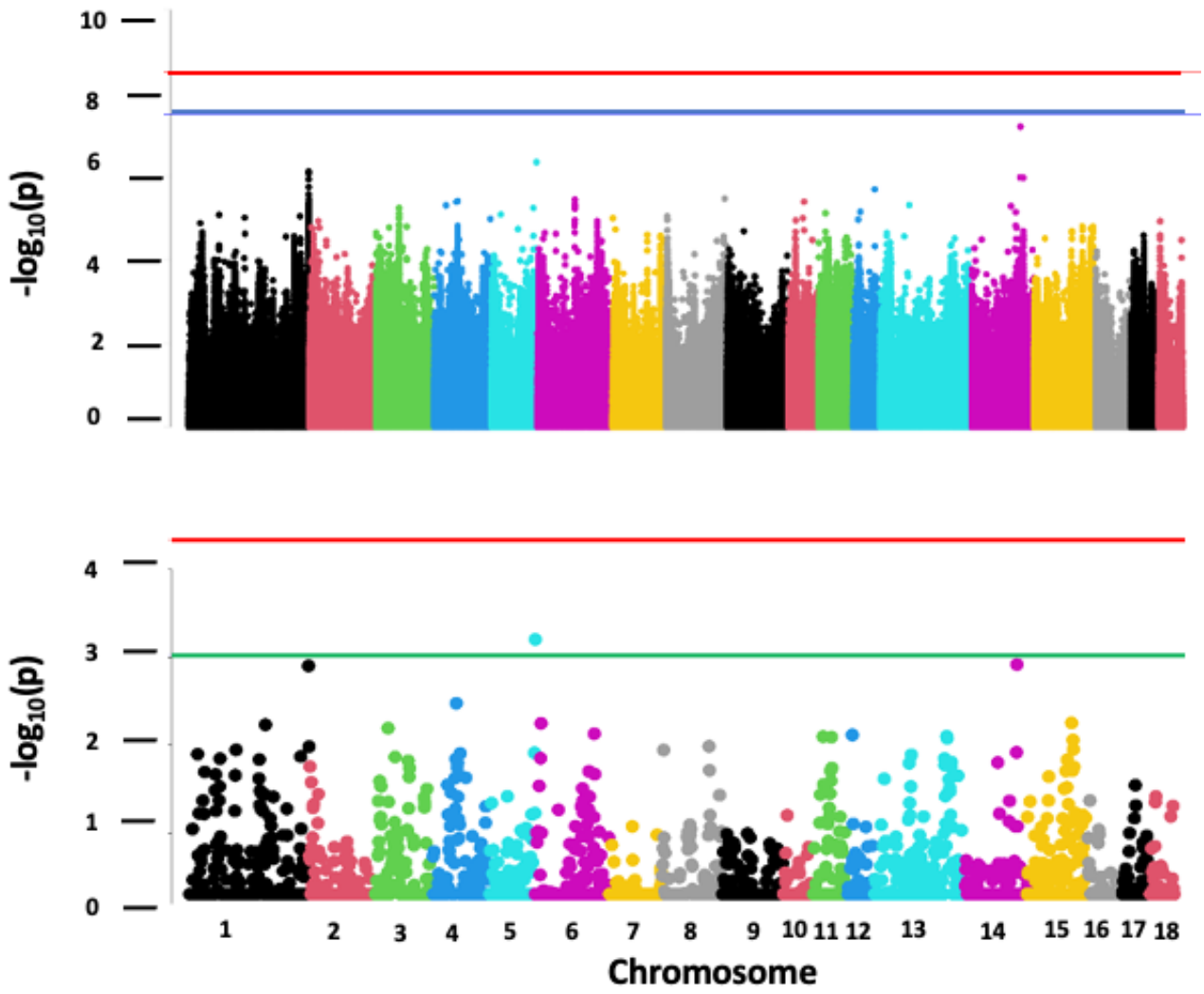
260

261 **Figure 1. Manhattan plot of the genome-wide association analysis (above) and regional**
262 **heritability mapping (below) of protein efficiency. The *x*-axis and the *y*-axis represent the**
263 **chromosomes and the observed -log$_{10}$(*P*-value), respectively. The red line is the Bonferroni**
264 **threshold, the blue line in the Manhattan plot is the False discovery rate (FDR) threshold, and**
265 **the green line in the regional heritability mapping plot is the suggestive threshold.**



266

267 **Figure 2. Manhattan plot of the genome-wide association analysis (above) and regional**
268 **heritability mapping (below) of feed conversion ratio. The *x*-axis and the *y*-axis represent the**
269 **chromosomes and the observed -log$_{10}$(*P*-value), respectively. The red line is the Bonferroni**
270 **threshold, the blue line in the Manhattan plot is the False discovery rate (FDR threshold), and**
271 **the green line in the regional heritability mapping plot is the suggestive threshold.**



272

273 **Figure 3. Manhattan plot of the genome-wide association analysis (above) and regional**
274 **heritability mapping (below) of average daily gain. The *x*-axis and the *y*-axis represent the**
275 **chromosomes and the observed -log$_{10}$(*P*-value), respectively. The red line is the Bonferroni**
276 **threshold, the blue line in the Manhattan plot is the False discovery rate (FDR threshold), and**
277 **the green line in the regional heritability mapping plot is the suggestive threshold.**

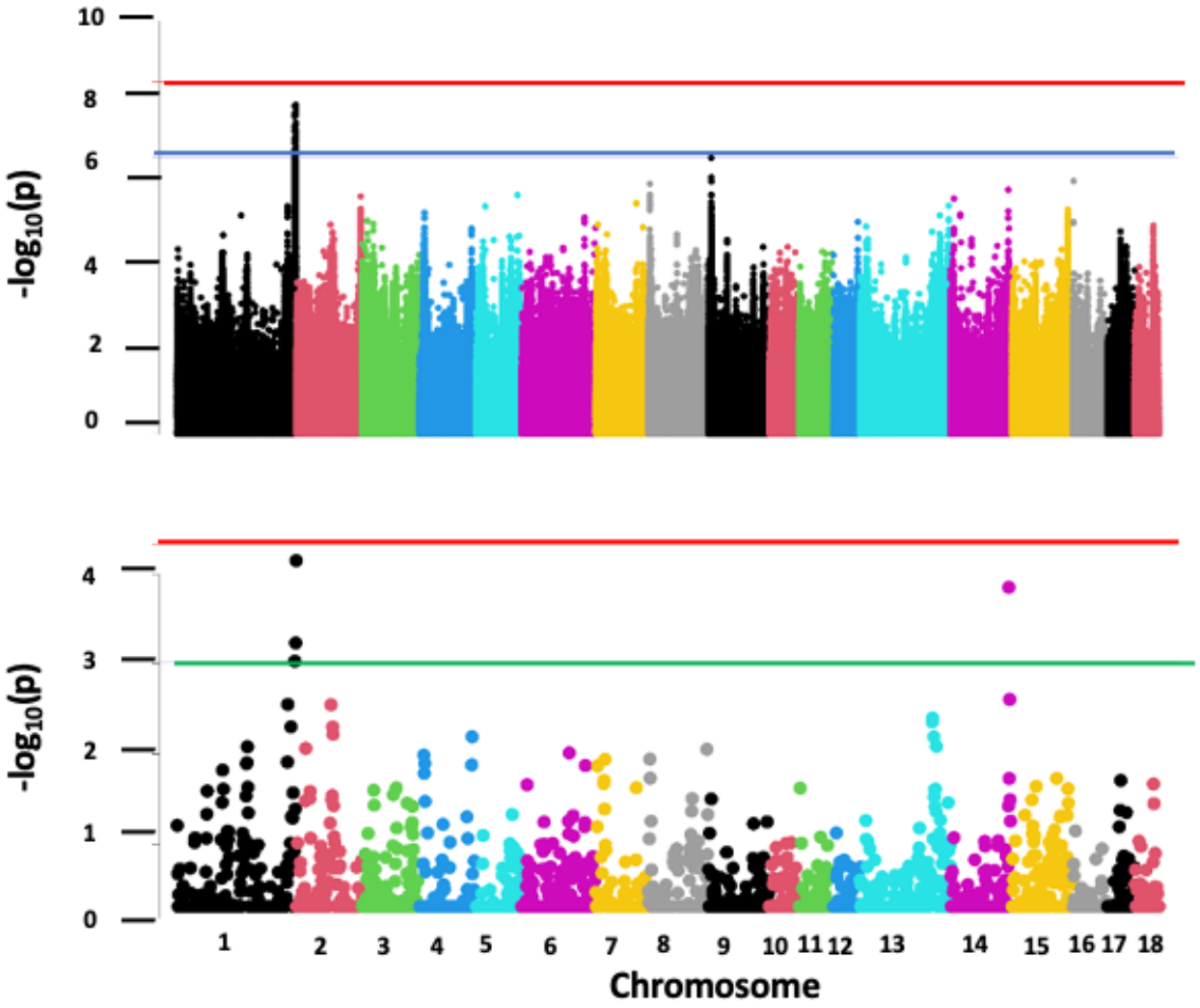


Figure 4. Manhattan plot of the genome-wide association analysis (above) and regional heritability mapping (below) of average daily feed intake. The *x*-axis and the *y*-axis represent the chromosomes and the observed -$\log_{10}$(*P*-value), respectively. The red line is the Bonferroni threshold, the blue line in the Manhattan plot is the False discovery rate (FDR threshold), and the green line in the regional heritability mapping plot is the suggestive threshold.

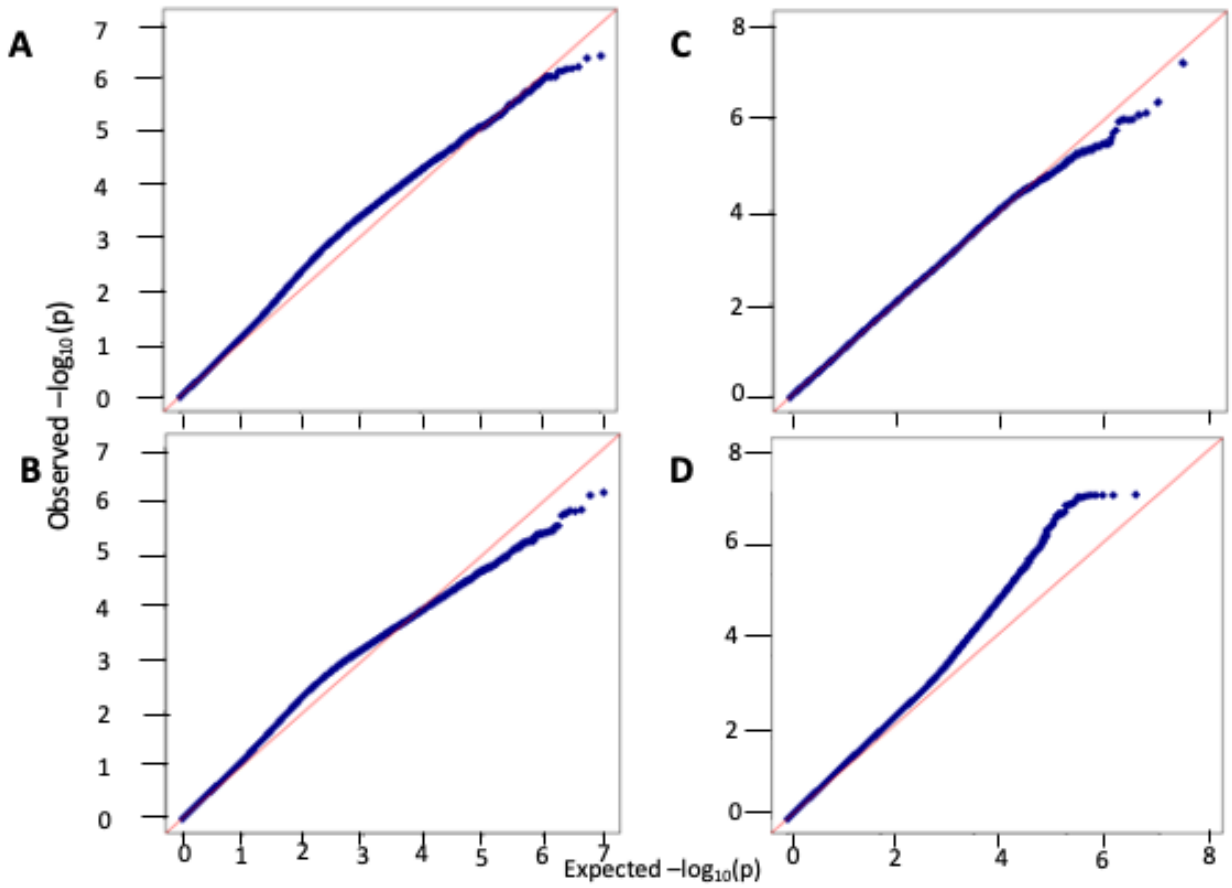


Figure 5. Quantile-Quantile plot for A) protein efficiency B) feed conversion ratio C) Average daily gain (ADG) and, C) Average daily feed intake (ADFI).

Discussion

This study aimed to identify genomic regions associated with PE, ADG, ADFI and FCR by using both genome-wide association and regional heritability mapping (RHM). The RHM approach identified two regions for PE at the suggestive threshold on SSC9 and SSC2, located at the 2–3 Mb and at 31–32Mb position, encompassing 13,440 and 5,399 SNPs, respectively. Similar to our study, Shirali et al. [23] found associations on SSC2 and SSC9 for total nitrogen excretion and average daily nitrogen excretion, respectively, during the 60 – 140kg growth stage, but these associations were not confirmed in our study. Our study also found four suggestive regions associated with ADFI located at 270–271Mb, 272–273Mb and 273–274Mb on SSC1, and at 133–134Mb on SSC14. Nosková et al. [23] conducted a multi-trait meta-GWAS in 5,109 Swiss Large White pigs with 60K array data imputed to sequence level and found QTL associated with ADFI on SSC1 at the 270 Mb – 272Mb position, confirming that this region is important for production traits in pigs. The associated chromosomal region also harbours QTL for ADFI in other pig populations [44, 45], possibly suggesting an ancestral origin of the QTL. Similar to our study, Onteru et al. [46] found QTL for ADFI on SSC14 located at 61Mb and at 107Mb for Yorkshire pigs, suggesting that this chromosome may also harbour associations with ADFI for Swiss Large White pigs.

The genomic heritability estimates for the suggestive regions by RHM ranged from 3% - 9% in our study, and similar range of regional heritability estimates have been found by other studies [30, 54, 55]. However, ADG showed a much a higher regional heritability of 26% at the suggestive region on SSC 5 in our study. The reason for this comparatively high regional heritability for ADG is unclear. Additionally, although it was expected that RHM has greater statistical power to identify regions associated with the traits than GWAS due to the combined effects of multiple variants contained in a region, the heritability estimate by RHM at the suggestive threshold was similar to that by GWAS in our study. The reason for the similar proportion of heritability observed between RHM and GWAS

311 may be due to linkage disequilibrium (LD), as genomic heritability depends on the LD structure
312 between causal variants and non-causal variants [47]. Another possible reason is that there are several
313 variants in the region with small effects and very few with an average effect, which dilutes the overall
314 heritability estimate for that region. A similar discrepancy was observed in the study of Resende et
315 al. [30], who observed that a region with a large single SNP effect detected in GWAS may not reach
316 significance in RHM, and vice versa, due to multiple small SNP effects in the RHM region that
317 cancels the overall effect of the region.

318 At the FDR threshold, GWAS found significant associations only for ADFI on SSC1. In general, the
319 Manhattan plots illustrating the GWAS and RHM results showed similar patterns for all traits, and so
320 the most significant SNPs found by GWAS were present in the regions found by RHM, suggesting
321 that our sample size was not large enough to detect statistical significance for these SNPs. However,
322 for ADFI, GWAS indicated some associations on SSC9 although they do not reach the FDR
323 threshold, but RHM did not show any association on this chromosome. Similar patterns were
324 observed in the study of Matika et al. [54], where GWAS found associations for muscle density at 8th
325 thoracic vertebra on sheep's chromosomes 1 and 16, but RHM did not find associations for
326 chromosome 16. Therefore, although RHM can detect additional regions not found by GWAS, for
327 example the additional suggestive region found by RHM on SSC14 for ADFI in our study, there may
328 be cases where RHM did not find associations identified by GWAS.

329 In this study, the inability to detect more associations is likely due to sample size, which is a key
330 factor in detecting genotype-phenotype associations. According to Goddard and Hayes [48], the
331 number of animals required for a GWAS depends on the size of the QTL effects that one aims to find,
332 and the heritability of the trait. For instance, for a QTL that explains 3% of the genetic variation and
333 a heritability of 0.54, the number of animals required would be about 2,500. For this study, between

334 1,025 and 1,034 animals were available for the GWAS and at a genomic heritability of 0.42 for PE,
335 one could only detect QTL with large effects (e.g., >5%). Since our study had QTL effects ranging
336 from 0% to 1.4%, according to the calculation by Goddard and Hayes [48] a sample size of at least
337 7,000 individuals would be required to detect associations at the Bonferroni threshold. Achieving
338 such a sample size for PE is very challenging due to the difficulty of phenotyping this trait, as high-
339 throughput phenotyping tools for measuring PE are not yet available. For example, using a dual-X-
340 ray absorptiometry (DXA) scanner as in this study, it takes about 15 minutes to measure the lean meat
341 content of one pig carcass, from which the protein content is estimated. In addition, measurement
342 errors may affect the identification of QTL [49], but DXA was used in our study in order to reduce
343 measurement errors and to obtain highly accurate PE phenotypes [33]. In addition, we generated
344 whole-genome sequence data rather than array genotypes, which is expected to help identify more
345 QTL, but has not yet been shown to do so in GWAS. As in genomic prediction, the use of whole-
346 genome sequence data may have an advantage over high-density genotyping arrays, in that the first
347 allows for the inclusion of candidate or causal mutations [56]. Moreover, the use of whole-genome
348 sequence data can improve prediction accuracy if the correct prior information can be included in the
349 model (e.g., using Bayesian models) [57, 58]. However, in a simulation study, using whole-genome
350 sequence data rather than high-density genotyping array data did not automatically improve the
351 accuracy of genomic prediction models, especially when biological information was not included
352 [57]. This might be valid for GWAS as well, which might also explain why our study identified no
353 significant associations for the traits except for ADFI.

354 Our study found that, despite identifying only a few or even no significant associations with the traits,
355 the genomic heritability captured by using all available SNPs was considerable and yielded low
356 standard errors, ranging from 0.33 ± 0.04 to 0.47 ± 0.05. Considering the high heritability estimates
357 for both pedigree-based and genomic-based relationships, the identification of only two suggestive

regions for PE suggests that PE is highly polygenic and that many variants with small effects underlie this trait [50]; therefore, marker-assisted selection may not be possible because of the difficulty in identifying loci that contribute substantially to trait variation. The application of genomic selection might be a more promising approach to breed for protein efficient pigs. However, the accuracy of genomic prediction depends on the size of the reference population whose phenotypes are required, which is a limiting factor for PE in pigs. For difficult-to-measure and novel traits with limited sample size like PE, a cross-validation genomic prediction approach may be applied [51]. Alternatively, a multi-breed genomic selection approach can be used, requiring other breeds to be phenotyped for the same trait. Hayes et al. [52] reported up to 13% higher accuracy when using the multi-breed reference population than when a using a single-breed reference population in Bayesian models. Additionally, Raymond et al. [53] reported that the use of a multi-breed multi-genomic relationship matrix (i.e., fitting selected and unselected markers in separate GRMs) gave a higher accuracy than a multi-breed single-genomic relationship matrix.

Conclusions

The GWAS and RHM analysis did not show significant SNPs or regions at the Bonferroni threshold for any of the traits analysed, but identified significant SNPs at the FDR threshold for ADFI on SSC 1, which is the same region identified by Nosková et al. (2023) for the same breed of pig (i.e., Swiss Large White). Regional heritability mapping identified suggestive regions for PE, ADG and ADFI. Significant or close-to-significant SNPs by GWAS were also present in the suggestive regions identified by RHM, thus corroborating the evidence for a potential effect. We attribute the reason for the identification of very few to no significant SNPs to the relatively small sample size used in this study. However, the genomic heritability for these traits were moderate, ranging between 0.33 – 0.47. Our inability to identify major QTL influencing these traits, despite their genomic heritability, suggests that the traits are influenced by many genes with small effects. Our results clearly show that

PE is under genetic influence and can be considered in downstream analyses such as genomic prediction. However, a large reference population with phenotypes is required for genomic selection, and therefore, faster phenotyping method may be needed to easily determine the PE of hundreds to thousands of pigs.

Declarations

Ethics approval and consent to participate

The experimental procedure was approved by the Office for Food Safety and Veterinary Affairs (2018_30_FR) and all procedures were conducted in accordance with the Ordinance on Animal Protection and the Ordinance on Animal Experimentation.

Consent for publication

Not applicable

Competing interests

The authors report no conflicts of interest with any of the data presented.

Funding

This research was supported by the Fondation Sur-la-Croix to C.K.

Authors' contributions

EOE curated and analyzed the data, and drafted the manuscript. ALV, AN, and HP participated in the data curation and analysis. CK conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to Markus Neuditschko for comments on the manuscript. We thank Paolo Silacci and his team at Agroscope's Animal Biology Group for DNA extraction from blood.

Author's information

E.O.E.: 0000-0002-9455-8325

A.L.V: 0000-0001-6272-9639

A.N.: 0000-0002-2092-8661

H.P.: 0000-0002-0501-6760

C.K.: 0000-0001-7305-3996

References

1. Sutton MA, Bleeker A, Howard CM, Erisman JW, Abrol YP, Bekunda, M, et al. Our nutrient world. The challenge to produce more food & energy with less pollution. Centre for Ecology & Hydrology. 2013
2. Notarnicola B, Tassielli G, Renzulli PA, Castellani V, Sala S. Environmental impacts of food consumption in Europe. J Clean Prod. 2017;140:753–65.
3. Del Grosso SJ, Ogle SM, Nevison C, Gurung R, Parton WJ, Wagner-Riddle C, et al. A gap in nitrous oxide emission reporting complicates long-term climate mitigation. PNAS. 2022;119(31):e2200354119.
4. Millet S, Aluwé M, Van den Broeke A, Leen F, De Boever J, De Campeneere S. Pork production with maximal nitrogen efficiency. Animal. 2018;12(5):1060-67.
5. Kasper C, Ruiz-Ascacibar I, Stoll P, Bee G. Investigating the potential for genetic improvement of nitrogen and phosphorus efficiency in a Swiss large white pig population using chemical analysis. J Anim Breed Genet. 2020;137(6):545-58.
6. Andretta I, Pomar C, Rivest J, Pomar J, Radünz J. Precision feeding can significantly reduce lysine intake and nitrogen excretion without compromising the performance of growing pigs. Animal. 2016;10(7):1137-47.
7. Pomar C, Remus A. Precision pig feeding: a breakthrough toward sustainability. Anim Fron. 2019;9(2):52–9.
8. Ewaoluwagbemiga EO, Bee G, Kasper C. Genetic analysis of protein efficiency and its association with performance and meat quality traits under a protein-restricted diet. Genet Sel Evol. 2023;**55:**35.
9. Déru V, Bouquet A, Labussière E, Ganier P, Blanchet B, Carillier-Jacquin C, Gilbert H. Genetics of digestive efficiency in growing pigs fed a conventional or a high-fibre diet. J Anim Breed Genet. 2021;138(2):246-58.

10. Verschuren LM. Improving feed efficiency in pigs: bridging genetics and nutrition. Doctoral dissertation, Wageningen University and Research. 2021.
11. Saintilan R, Merour I, Brossard L, Tribout T, Dourmad JY, Sellier P, et al. Genetics of residual feed intake in growing pigs: relationships with production traits, and nitrogen and phosphorus excretion traits. J Anim Sci. 2013;91(6):2542-54.
12. Liu X, Xiong X, Yang J, Zhou L, Yang B, Ai H, et al. Genome-wide association analyses for meat quality traits in Chinese Erhualian pigs and a Western Duroc×(Landrace× Yorkshire) commercial population. Genet Sel Evol. 2015:47(1);1-11.
13. Tang Z, Xu J, Yin L, Yin D, Zhu M, Yu M, et al. Genome-wide association study reveals candidate genes for growth relevant traits in pigs. Front Genet. 2019:10;302.
14. Schmid M, Maushammer M, Preuß S, Bennewitz J. Mapping QTL for production traits in segregating Piétrain pig populations using genome-wide association study results of F2 crosses. Anim Genet. 2018;49(4):317-20.
15. Gilbert H, Riquet J, Gruand J, Billon Y, Feve K, Sellier P, et al. Detecting QTL for feed intake traits and other performance traits in growing pigs in a Piétrain–Large White backcross. Animal. 2010;4(8):1308-18.
16. Heidaritabar M, Bink MC, Dervishi E, Charagu P, Huisman A, Plastow GS. Genome-wide association studies for additive and dominance effects for body composition traits in commercial crossbred Piétrain pigs. J Anim Breed Genet. 2023; doi:10.1111/jbg.12768.
17. Fan B, Onteru SK, Du ZQ, Garrick DJ, Stalder, K. J., Rothschild MF. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. PloS one. 2011;6(2):e14726.
18. Shirali M, Duthie CA, Doeschl-Wilson A, Knap PW, Kanis E, van Arendonk JA, et al. Novel insight into the genomic architecture of feed and nitrogen efficiency measured by residual energy intake and nitrogen excretion in growing pigs. BMC Genet. 2013;14(1):121.
19. Ding R, Quan J, Yang M, Wang X, Zheng E, Yang H, et al. Genome-wide association analysis reveals genetic loci and candidate genes for feeding behavior and eating efficiency in Duroc boars. PLoS One. 2017;12(8):e0183244.
20. Ramayo-Caldas Y, Mármol-Sánchez E, Ballester M, Sánchez J. P, González-Prendes R, Amills M, et al. Integrating genome-wide co-association and gene expression to identify putative regulators and predictors of feed efficiency in pigs. Genet Sel Evol. 2019;51:1-17.
21. Fu L, Jiang Y, Wang C, Mei M, Zhou Z, Jiang Y, et al. A genome-wide association study on feed efficiency related traits in landrace pigs. Front Genet. 2020;11:692.
22. Fan B, Lkhagvadorj S, Cai W, Young J, Smith RM, Dekkers JCM, et al. Identification of genetic markers associated with residual feed intake and meat quality traits in the pig. Meat Sci. 2010;84(4):645-650.
23. Nosková A, Mehrotra A, Kadri NK, Lloret-Villas A, Neuenschwander S, Hofer A, et al. Comparison of two multi-trait association testing methods and sequence-based fine mapping of six additive QTL in Swiss Large White pigs. BMC Genom. 2023;24(1):1-14.
24. de Verdal H, Narcy A, Bastianelli D, Chapuis H, Même N, Urvoix S, et al. Improving the efficiency of feed utilization in poultry by selection. 2. Genetic parameters of excretion traits and correlations with anatomy of the gastro-intestinal tract and digestive efficiency. BMC Genet. 2011;12(1):71.
25. Uemoto Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Wilson JF, et al. The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. Front Genetics. 2013;4:232.

26. Cebamanos L, Gray A, Stewart I, Tenesa A. Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. Bioinformatics. 2014;30(8):1177-79.
27. Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, et al. Localising loci underlying complex trait variation using regional genomic relationship mapping. PloS one. 2012;7(10):e46501.
28. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565-69.
29. Hayes B. Overview of statistical methods for genome-wide association studies (GWAS). Genome-wide association studies and genomic prediction. 2013. p. 149-169.
30. Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, et al. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. New Phytol. 2017;213(3):1287-1300.
31. Sutera AM, Tolone M, Mastrangelo S, Di Gerlando R, Sardina MT, Portolano B, et al. (2021). Detection of genomic regions underlying milk production traits in Valle del Belice dairy sheep using regional heritability mapping. J Anim Breed Genet. 138(5), 552-561.
32. Ruiz-Ascacibar I, Stoll P, Kreuzer M, Boillat V, Spring P, Bee G. Impact of amino acid and CP restriction from 20 to 140 kg BW on performance and dynamics in empty body protein and lipid deposition of entire male, castrated and female pigs. Animal. 2017;11(03):394–404.
33. Kasper C, Schlegel P, Ruiz-Ascacibar I, Stoll P, Bee G. Accuracy of predicting chemical body composition of growing pigs using dual-energy X-ray absorptiometry. Animal. 2021;15(8):100307.
34. Suster D, Leury BJ, Ostrowska E, Butler KL, Kerton DJ, Wark JD, et al. Accuracy of dual energy X-ray absorptiometry (DXA), weight and P2 back fat to predict whole body and carcass composition in pigs within and across experiments. Livest Prod Sci. 2003;84(3):231-242.
35. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 2007;81(5):1084-1097.
36. Nosková A, Meenu B, Naveen KK, Danang C, Stefan N, Andreas H, et al. Characterization of a haplotype-reference panel for genotyping by low-pass sequencing in Swiss Large White pigs. BMC Genom. 2021;22:1-14.
37. Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ., King K, et al. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genom. 2021*;*22*:1-7.
38. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: A Tool Set for Whole-Genome and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81.
39. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4.
40. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K. An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience. 2020;9(6):giaa051.
41. R core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available from: https://www.R-project.org/. 2023.
42. Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 2011;88(1):76-82.
43. Hochberg Y, Benjamini Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. JR Stat Soc. 1995;57(1):289-300.
44. Jiao S, Maltecca C, Gray KA, Cassady JP. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: II. Genomewide association. J Anim Sci. 2014;92(7):2846-60.

45. Do DN, Ostersen T, Strathe AB, Mark T, Jensen J, Kadarmideen, HN. Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. BMC Genet. 2014;15:1-15.
46. Onteru SK, Gorbach DM, Young JM, Garrick DJ, Dekkers JC, Rothschild MF. Whole genome association studies of residual feed intake and related traits in the pig. PloS one. 2013;8(6):e61756.
47. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet. 2017;49(9):1304-1310.
48. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009;10(6):381-391.
49. Barendse W. (2011). The effect of measurement error of phenotypes on genome wide association studies. BMC Genom. 12, 1-12.
50. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. Nat Rev Genet. 2019;20(3):135-156.
51. Daetwyler HD, Swan AA, van der Werf JH, Hayes BJ. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet Sel Evol. 2012;44(1):1-11.
52. Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, and Goddard M.E. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol. 2009;41(1):1-9.
53. Raymond B, Bouwman AC, Wientjes YC, Schrooten C, Houwing-Duistermaat J, Veerkamp RF. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. Genet Sel Evol. 2018;50(1):1-14.
54. Riggio V, Matika O, Pong-Wong R, Stear MJ, Bishop SC. Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Hered. 2013;110(5):420-429.
55. Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL, et al. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. Genet Sel Evol. 2016;48:1-15.
56. Liu A, Lund MS, Boichard D, Karaman E, Fritz S, et al. Improvement of genomic prediction by integrating additional single nucleotide polymorphisms selected from imputed whole genome sequencing data. Heredity. 2020;124(1):37-49.
57. Pérez-Enciso M, Rincón JC, Legarra A. Sequence-vs. chip-assisted genomic selection: accurate biological information is advised. Genet. Sel. Evol. 2015;47(1):1-14.
58. Mollandin F, Rau A, Croiseau P. An evaluation of the predictive performance and mapping power of the BayesR model for genomic prediction. G3. 2021;11(11):jkab225.