The impact of SNP density on quantitative genetic analyses of body size traits in a wild population of Soay sheep

Caelinn James[1*], Josephine M. Pemberton[1], Pau Navarro[2], Sara Knott[1]

[1] Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh

[2] MRC Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh

[*] Corresponding author

Corresponding author mailing address: Caelinn James, Office 133, Ashworth Laboratories, Kings Buildings, Charlotte Auerbach Road, Edinburgh EH9 3FL

Corresponding author email address: c.james-5@sms.ed.ac.uk

Running head: Impact of SNP density on quantitative analyses

Abstract

Understanding the genetic architecture underpinning quantitative traits in wild populations is pivotal to understanding the processes behind trait evolution. The 'animal model' is a popular method for estimating quantitative genetic parameters such as heritability and genetic correlation and involves fitting an estimate of relatedness between individuals in the study population. Genotypes at genome-wide markers can be used to estimate relatedness; however, relatedness estimates vary with marker density, potentially affecting results. Increasing density of markers is also expected to increase the power to detect quantitative trait loci (QTL). We estimated heritability and performed genome-wide association studies (GWAS) on five body size traits in an unmanaged population of Soay sheep using two different SNP densities: a dataset of 37,037 genotyped SNPs, and an imputed dataset of 417,373 SNPs. Heritability estimates did not differ between the two SNP densities, but the high-density imputed SNP dataset revealed five new SNP-trait associations that were not found with the lower density dataset. Conditional GWAS analyses after fitting the most significant SNPs revealed two more novel SNP-trait associations.

Introduction

Investigating the genetic architecture behind heritable traits is key to understanding the biological diversity of wild populations. If we know the number of loci influencing a trait and their effect size, we can better understand the evolutionary processes that underpin traits, improve inferences about trait evolution (Barton and Keightley 2002), and understand micro-evolutionary dynamics that occur due to environmental change. Most quantitative genetic research in animals is carried out in artificial populations; either domestic, agricultural or laboratory. Such populations experience controlled environmental conditions which make it easier to account for environmental factors when studying the effect of genetic variants on phenotypic variation. However, given that environmental factors can influence the phenotype of a quantitative trait (Charmantier et al. 2014), and that the presence of genotype-by-environment interactions can cause additive genetic variance to differ between environmental conditions, studies on artificial populations arguably cannot be fully extrapolated to wild populations (Kruuk et al. 2008). Therefore, it is important to also study quantitative traits in wild populations in their natural habitats. There is a wealth of quantitative genetics research in human populations (for examples, see (Kang et al. 2010; Locke et al. 2015; Manolio et al. 2009; Xia et al. 2016; Xia et al. 2021; Yang et al. 2010; Zaitlen et al. 2013)), but humans also arguably experience a more buffered environment than wild populations.

A popular method to decompose phenotypic variation in wild populations into genetic variance and environmental variance is the 'animal model', originally developed by animal breeders (Henderson 1984; Kruuk 2004; Wilson et al. 2010). As part of the model, genetic relatedness is fitted, which is often derived from a pedigree. Pedigrees can be constructed using field observations, assigning parentage using genetic markers, or a mixture of both (Pemberton 2008). However, wild pedigrees are often short, incomplete, and contain errors: observational data may be inaccurate due to incorrect parent-offspring assumptions, and if the genetic markers chosen are not sufficiently discriminatory they may result in misassigned parentage. Erroneous pedigree links can bias results of analyses using animal models; for example, misidentification of sires in cattle resulted in decreased heritability estimates for milk yield, fat yield, and milk-fat ratio (Van Vleck 1970).

In place of a pedigree, genotypes at multiple polymorphic loci can be used to estimate relatedness for use in an animal model. This has the advantage of not relying on recovering a pedigree and thus is not affected by incomplete or incorrect familial links (though knowledge of pedigree relationships is still valuable, for example so that maternal effects can be fitted). Relatedness estimated using genotype data is also potentially more precise than that from pedigrees – for example, with a pedigree it is presumed that full-sibs have a relatedness of 0.5, however the exact relationship varies depending on which DNA segments each sib has inherited (Visscher et al. 2006). Despite this greater accuracy, genotype-based relatedness estimates can still vary depending on which variants in the population have been genotyped, and the density of the genotyped variants. Increasing the density of genotyped polymorphisms means they are more likely to be in linkage disequilibrium (LD) with causal variants for the trait of interest, either by being physically closer to the causal variants or by matching the allele frequency of the causal variants more accurately. Thus, in species such as humans where genotyping is commonly of unrelated individuals and LD is generally low, the estimated heritability of a trait increases with SNP density due to an increase in the number of causal variants being in LD with genotyped SNPs (for instance, heritability of human height was estimated to be 0.45 when using 294,831 SNPs (Yang et

al. 2010) and 0.56 when using ~17 million imputed SNPs (Yang et al. 2015)). However, increasing the number of genotyped markers means larger, denser genotyping arrays with prices increasing with density. For commonly studied species, high-density arrays are more affordable due to high demand, but for more niche species, including wild populations, large genotyping arrays are often unaffordable. Genotyping-by-sequencing, e.g. ddRAD (Peterson et al. 2012) is a potentially useful alternative for upscaling SNP density, though the combination of bioinformatics and samples sizes required in quantitative genetic research means that this approach is not yet in widespread use.

As an alternative to expensive high-density genotyping, genotype imputation can be used to increase the number of variants analysed (Burdick et al. 2006). Imputation involves predicting genotypes at untyped SNPs in a 'target' population using a subset of the study population – or more generally a "reference" population – genotyped at a higher density, either through a high density SNP array or by genotyping-by sequencing. The genotypes at these untyped SNPs for individuals in the target population are inferred using their genotypes at typed markers and taking advantage of existing linkage disequilibrium (LD) between SNPs. Pedigree information can also be used to increase the accuracy of the imputation by identifying haplotype blocks that are identical by descent (Burdick et al. 2006).

The Soay sheep *(Ovis aries)* of St Kilda are a primitive, unmanaged breed of sheep that have been the focus of a longitudinal, individual-based study since 1985 (Soay sheep: Dynamics and selection in an island population 2003). As part of the study, life history and environmental data is collected, DNA samples are collected, and a pedigree has been constructed using observation and genetic parentage inference. 7630 sheep have been genotyped on the Ovine SNP50 Illumina Beadchip, on which 37,037 SNPs are autosomal and polymorphic in this population.

To date, quantitative genetic analyses of the Soays have been performed using either the pedigree or the 50K SNP data. Bérénos et al. (2014) (Bérénos et al. 2014) investigated the difference in quantitative genetic parameter estimates when using the pedigree or a genomic relationship matrix (GRM) constructed from the 50K SNP data. The authors estimated heritability, maternal genetic effects and genetic correlations for body size traits (weight, foreleg length, hindleg length, metacarpal length and jaw length) across four age groups. The additive genetic variance and the heritability estimates using the GRM were lower than when using the pedigree to estimate relatedness, with the SNPs explaining 84% of the additive genetic variance of the pedigree on average, though for the majority of the traits the standard errors of the pedigree-based and SNP-based heritability estimates were within one standard error of each other. Genetic correlations were found to differ little between analyses using the different relatedness estimates. A SNP rarefaction analysis of the heritability estimates was conducted on the adult traits, and it was found that heritability estimates asymptoted when 50% of the SNPs were used to estimate relatedness.

Linkage disequilibrium is high in Soay sheep (Bérénos et al. 2014), which may explain the results of the rarefaction analysis: the same causal variations can be represented by multiple genotyped SNPs, and so many genotyped SNPs are not providing unique information towards the additive genetic variance. However, given that heritability estimates when using SNP-based estimates of relatedness were lower than when using pedigree-based estimates, it is also possible that there are variants that contribute to variation in these traits that aren't in perfect LD with the genotyped SNPs. This may be depressing estimates of additive genetic variance in comparison to pedigree-based estimates. Increasing the

density of genotyped SNPs may therefore increase the SNP-based heritability estimates of these traits to closer match the pedigree-based estimates.

Bérénos et al. (2015) (Bérénos et al. 2015) focused on the same traits as (Bérénos et al. 2014) in adults only, and partitioned the genetic variance for each trait, first by chromosome and then by 150 SNP windows, and performed GWAS. For the three leg length phenotypes, a disproportionately high amount of variance was explained by two SNPs on chromosome 16 (s23172.1) and chromosome 19 (s74894.1). For the remaining traits, the proportion of additive genetic variance explained by a chromosome was proportional to its length, suggesting that variation in these traits is influenced by many genetic variants of small effect, and no SNP-trait associations were discovered. The authors performed a two-step GWAS analysis; first they modelled the traits by fitting fixed and random effects, and then they extracted the residuals from the models and tested for association between SNPs and the residuals. Three traits (weight, foreleg length and hindleg length) are measured in the same individual across multiple years – for these traits, the authors analysed the mean residual values.

Since the studies from 2014 and 2015, 1895 more individuals have been both phenotyped and genotyped on the Ovine SNP50 Illumina Beadchip. In addition, 188 individuals were genotyped on the Ovine Infinium High Density chip with 600K attempted SNPs, which has enabled the genotypes of the remaining individuals to be accurately imputed to that higher density using LD and pedigree data (Stoffel et al. 2021).

In this study we performed a direct comparison of heritability estimates and GWAS associations between the lower density SNP data and the imputed high density SNP data in the Soay population using an increased sample size and compared the results to (Bérénos et al. 2015; Bérénos et al. 2014). We focused on the same five traits as the previous studies in neonates, lambs and adults. Unlike the 2015 study (Bérénos et al. 2015), we performed GWAS by fitting fixed and random effects in the same step as testing for SNP-trait associations. This has the advantage of correctly propagating error throughout the analysis, reducing the chance of false positive results. We also carried out a two-step GWAS approach similar to that of (Bérénos et al. 2015), focusing on the adult traits using the 50K SNP data, to investigate whether any SNP-trait associations identified using our approach were due to the increased population size or due to the different methodology (single-step vs. two-step GWAS).

Our aims were as follows:

1) To determine whether the increased density of SNPs changes the heritability estimates of the traits using the same individuals for both the low SNP density and the high SNP density analysis. Given the previous rarefaction analysis (Bérénos et al. 2014), our prediction was that there would be no change.

2) To determine whether the imputed SNP data enables the identification of new SNP-trait associations via GWAS. We predicted we would find more associations, either due to increased power to detect small effect size SNPs through increased LD with the imputed SNPs or due to tagging of new causal variants.

3) To examine how a single-step GWAS methodology compares the two-step approach previously used on the Soay population, to investigate whether any novel SNP-trait associations identified

154 with the 50K SNP data were due to increased population size or due to the difference in
155 methods.

METHODS

Phenotypic data

We focused on five body size traits in three age groups: neonates, lambs, and adults. Of the five traits, three (weight, foreleg length and hindleg length) are live measures, recorded in April for neonates and in August for lambs and adults. The remaining two traits (metacarpal length and jaw length) are *post mortem* measures taken from skeletal material. The sheep are ear-tagged when they are first captured which allows for reidentification for life. Both birth and August weight are measured to the nearest 0.1kg, whilst the remaining traits are all measured to the nearest mm. A detailed description of trait measurements can be found in (Beraldi et al. 2007).

We defined neonates as individuals who were caught and weighed between two and ten days after birth – birth weight was the only trait recorded for this age group. Lambs were classed as individuals who had phenotypic data recorded in the August of their birth year for the live traits, and as individuals who died before 14 months of age for the *post mortem* measures. Individuals were classed as adults if they had August phenotypic data recorded at least two years after birth, or if they died after 26 months of age for *post mortem* measures. Unlike (Bérénos et al. 2014), we chose not to analyse yearling data due to the small sample sizes in comparison to the other age classes, which is due to high first winter mortality.

Genetic data

Most of the sheep in our study population have been genotyped using the Ovine SNP50 Illumina BeadChip, which targets 54,241 SNPs across the sheep genome. After removing SNPs which failed quality control standards (minor allele frequency (MAF) > 0.001, call rate > 0.99, deviation from Hardy-Weinberg Equilibrium P > 1e-05), 39,368 polymorphic variants remained for 7630 individuals (3643 female, 3987 male). See Bérénos et al. 2014 (Bérénos et al. 2014) for information on genetic sampling protocol and marker characteristics).

Of these 7630 individuals, 188 have also been genotyped using the Ovine Infinium HD SNP BeadChip which types 606,066 SNPs, of which 417,373 were autosomal and polymorphic in the Soays. This has allowed for the low density genotypes to be imputed to the higher density using AlphaImpute, which combines shared haplotype and pedigree information for phasing and genotype imputation (Hickey et al. 2012) (see Stoffel et al. (2021)(Stoffel et al. 2021) for information on imputation). We used imputed genotype “hard” calls (rather than genotype probabilities) in downstream analyses. After filtering SNPs that failed quality control standards (minor allele frequency (MAF) > 0.001, call rate > 0.99, deviation from Hardy-Weinberg Equilibrium P > 1e-05) and imputation QC, 343,889 autosomal SNPs remained for 7362 individuals (3529 females, 3736 males).

Both the 50K SNP data and the imputed SNP data are mapped to the OAR_v3.1 genome assembly.

Narrow sense heritability estimation

We used animal models to partition the phenotypic variance for each trait in each age class into genetic and non-genetic variance components. Fixed and random effects were fitted for all models, with the

194 effects differing between traits and age classes (Table 1). We implemented these analyses in DISSECT
195 (Canela-Xandri et al. 2015) using the following model:

$$y = X\beta + \sum_r Z_r u_r + Wg + \varepsilon$$

196 where y is the vector of phenotypic values; X is a design matrix linking individual records with the vector
197 of fixed effects β, Z_r is an incidence matrix that relates a random effect to the individual records; u_r is the
198 associated vector of non-genetic random effects; g is the vector of additive genetic random effects with
199 W the incidence matrix; and ε is the vector of residuals. It is assumed that g ~ *MVN*(0, $M\sigma_g^2$), where σ_g^2 is
200 the additive genetic variance and M is the genomic relationship matrix (GRM). For each trait in each age
201 class, we ran this model twice: first with M being a GRM calculated from the 50K genotype data, and
202 second with M being a GRM calculated from the imputed SNP genotypes. The GRMs were computed
203 using DISSECT(Canela-Xandri et al. 2015), and the genetic relationship between individuals i and j is
204 computed as:

$$A_{ij} = \frac{1}{N} \sum_{k=1}^{N} \frac{(s_{ik} - 2p_k)(s_{jk} - 2p_k)}{2p_k(1 - p_k)}$$

205

206 where s_{ik} is the number of copies of the reference allele for SNP k of the individual i, p_k is the frequency
207 of the reference allele for the SNP k, and N is the number of SNPs.

208 The narrow sense heritability was estimated by dividing the additive genetic variance (the variance
209 explained by the GRM) by the total estimated phenotypic variance (the sum of the variance explained by
210 the GRM and other fitted random effects after fitting fixed effects).

211 In adults, there are multiple records for August weight, foreleg length and hindleg length for the same
212 individual due to individuals being caught across multiple years. For these traits we used a repeatability
213 model in order that uncertainty was correctly propagated through all estimations (Mrode 2014). To
214 implement a repeatability model in DISSECT, we edited the input files so that each measurement had its
215 own row in the genotype and covariate files. Individual ID was replaced with a unique capture reference
216 number, and individual permanent environment was fitted as a random effect (see Supplementary
217 Methods for a more detailed explanation).

218 Sample sizes and total number of phenotypic measurements for all traits are shown in Table 1, with
219 effects fitted in all models.

220

221 *Genome wide association analysis*

222 We also conducted genome-wide association analyses using DISSECT (Canela-Xandri et al. 2015). We
223 fitted the same fixed and random effects for each trait and age class as for the heritability estimation
224 (Table 1). To account for population structure, when testing SNPs on a given chromosome for

225 association with the phenotype, a GRM calculated from the remaining autosomes (referred to as Leave
226 One Chromosome Out GRM (Yang et al. 2014)) was fitted. Input files for repeated-measure traits were
227 reformatted as above. Our significance threshold was corrected for multiple testing using the SimpleM
228 (Gao et al. 2008) method, which accounts for linkage disequilibrium between markers in order to
229 calculate the effective number of independent tests.

230 We estimated the variance explained by SNPs that passed the significance threshold using the equation

$$V(SNP) = 2pq\alpha^2$$

231 where p and q are the major and minor allele frequencies of the SNP, and α is the estimated SNP effect.
232 We then calculated the proportion of additive genetic variance explained by each SNP by dividing by the
233 total additive genetic variance estimated for that trait.

234 For any trait for which several SNPs in the same region were associated with variation in the trait and
235 thus had strong support for at least one QTL in the region, we carried out conditional analysis to
236 understand if the region could harbour potentially several independent QTL, or if further QTL could be
237 uncovered elsewhere in the genome. To that aim, the genotypes of the SNP with the smallest
238 association p value from each associated region (hereafter called the "top SNP") were added to the
239 GWAS model as a fixed covariate and removed from the GRMs and genotype data. The GWAS analysis
240 was re-run accounting for those associations to try and reveal novel peaks either in the same regions or
241 elsewhere in the genome.

242

243 *Two-step GWAS analysis*

244 To investigate whether any novel SNP associations identified (since (Bérénos et al. 2015)) by performing
245 GWAS on the adult traits using the 50K SNP data were due to the increased population sample or due to
246 the change in methodology, we also performed a two-step GWAS, focusing on adults only and using the
247 50K SNP data. We performed mixed model analyses using ASReml-R (Butler et al. 2017) for each trait
248 fitting the same fixed and random effects as in our single-step analyses, including whole-genome
249 relatedness (in the form of a GRM) and, for repeated-measure traits, permanent environment. We then
250 extracted the residuals from the mixed models and performed GWAS with the residuals as the trait
251 phenotypes using DISSECT (Canela-Xandri et al. 2015). For repeated measure traits, we used the mean
252 residual value for each individual. We used the Bonferroni correction calculated in (Bérénos et al. 2015)
253 to determine the significance threshold for our two-step GWAS in order to best compare with previous
254 GWAS performed on the Soay sheep.

255

256 RESULTS

257 *Heritability estimation*

258 Neonates

259 In neonates, the heritability of birth weight was 0.058 (S.E. 0.021) when using the 50K SNPs to calculate
260 relatedness, and 0.049 (S.E. 0.020) when using the imputed SNPs (Figure 1, Supplementary Table 1). The
261 two estimates were within one standard error of each other, meaning the difference between the two
262 estimates was not significant.

263 Lambs

264 In lambs, the heritability estimates for the live August measures were lower than those for the *post*
265 *mortem* measures (Figure 1, Supplementary Table 1). Across all the traits, heritability estimates were
266 similar when using the 50K SNP data and the imputed SNP data, with the biggest difference being 0.016
267 for jaw length. For all traits, estimates were within one standard error of each other, meaning that the
268 small differences in heritability estimates between the two SNP densities were not significant.

269 Adults

270 As observed in lambs, heritability estimates for live measures in adults were lower than those of the
271 *post mortem* measures. Across all traits, heritability estimates were higher in adults than in lambs.
272 Estimates obtained using the 50K SNPs and using the imputed SNPs were similar and were within one
273 standard error of each other (Figure 1, Supplementary Table 1), meaning that the imputed SNPs
274 provided no additional information to partition the variation into genetic and environmental variance.

275 Estimates for all variance components are listed in Supplementary Table 1.

276

277 *GWAS*

278 *50K SNP data*

279 To correct for multiple testing, we calculated the effective number of tests to be 20082 using the
280 SimpleM method (Gao et al. 2008), giving a genome-wide significance threshold of $2.49e^{-06}$ for the 50K
281 SNP data.

282 For weight in neonates (birth weight), lambs and adults (August weight), no SNPs were found to have an
283 association p value smaller than this threshold, suggesting that any variants that influence weight
284 variation are either of small effect or were not tagged by SNPs in the 50K SNP data (Figure 2A,
285 Supplementary Figure 1B and 1G).

286 For all three leg length measures in lambs, we found associations with the same region on chromosome
287 16. SNP s23172.1 was the SNP with the lowest p value for lamb foreleg and hindleg, explaining 0.45%
288 and 0.54% of the genetic variance for each trait respectively (Supplementary Table 2, Supplementary
289 Figure 1C and 1D). For lamb metacarpal, SNP 22142.1 in the same chromosome 16 region had the
290 lowest p value and explained 0.98% of the genetic variance. There was also a single SNP on chromosome
291 3 (OAR3_100483326.1) and a cluster of SNPs on chromosome 19 that had p values smaller than the
292 genome-wide significance threshold and were associated with variation in lamb metacarpal length,

293 explaining 1.95% and 2.32% of the genetic variance respectively (Supplementary Table 2, Supplementary
294 Figure 1E).

295 The two regions on chromosomes 16 and 19 that were associated with lamb metacarpal length variation
296 were also significantly associated with all three leg length measures in adults, with SNP s22142.1 on
297 chromosome 16 and SNP s74894.1 on chromosome 19 respectively explaining 0.74% and 1.96% of the
298 genetic variation in adult foreleg, 0.83% and 1.44% of the genetic variation in adult hindleg, and 1.08%
299 and 1.96% of the genetic variation in adult metacarpal length. There were other regions of the genome
300 also associated with variation in the adult leg length traits; a region on chromosome 11 was significant
301 across all three adult leg length traits, with the most significant SNP explaining 2.84%, 2.35% and 2.72%
302 of the genetic variance in adult foreleg, hindleg and metacarpal respectively (Figure 2B, Supplementary
303 Table 2, Supplementary Figure 1H and 1J). For adult foreleg, a SNP on chromosome 7 and two on
304 chromosome 9 were also associated, with the most significant SNPs in each region explaining 1.17% and
305 2.99% of the genetic variance respectively for this trait (Supplementary Table 2, Supplementary Figure
306 1H).

307 In lambs, there were no associations with jaw length found (Supplementary Figure 1F). In adults, a
308 region on chromosome 20 was associated with jaw length variation, with 2.69% of the genetic variance
309 for this trait being explained by SNP s07047.1 (Supplementary Table 2, Supplementary Figure 1K).

310 *Imputed data*

311 Using the SimpleM method (Gao et al. 2008), we calculated the number of effective tests to be 48635,
312 giving a genome-wide significance threshold of 1.03e-06.

313 When performing GWAS using the imputed SNP data, we were able to recover significant SNPs in the
314 same locations for all traits as those we found using the 50K SNP data, with the exception of the
315 association between s48811.1 on chromosome 7 and adult foreleg length (Supplementary Table 2). Of
316 the 84 SNP-trait associations that we identified with the 50K SNP data, only 78 were significant using the
317 imputed SNP data – the remaining 6 SNPs were no longer significant due to the increased multiple
318 testing burden (which leads to a more stringent significance threshold) between the 50K SNP data and
319 the imputed SNP data ($2.49e^{-06}$ and $1.03e^{-06}$ respectively).

320 We also identified 845 new SNP-trait associations using the imputed SNP data (Supplementary Table 2).
321 The majority of new associations were in the same regions as the SNPs identified using the 50K SNP
322 data, but we also found new associations: a SNP on chromosome 7 was associated with birth weight
323 (Figure 2A, Supplementary Table 2), and two SNPs on chromosome 13 were associated with adult
324 hindleg length (Figure 2B, Supplementary Table 2).

325 Manhattan and QQ plots for all traits can be found in Supplementary Figure 1.

326 *Conditional analysis*

327 For any trait that had at least two SNPs on the same region associated with variation in that trait, we
328 fitted the genotype of the SNP with the lowest p value in each region in the GWAS model and removed
329 the SNP from the genotype file. For traits that had multiple SNP associations on more than one
330 chromosome, we fitted the genotypes of the SNP with the lowest p value from each associated
331 chromosome simultaneously. We performed conditional analysis on all three leg length traits in both

332 lambs and adults, as well as on adult jaw length (See Supplementary Table 2 for all SNPs that were fitted
333 for each trait).

334 Of the traits we performed conditional analysis on, only two traits had further significant SNPs after
335 fitting the SNPs with the lowest p value; lamb metacarpal length and adult foreleg length. After fitting
336 the genotypes for s22142.1 on chromosome 16 and s74894.1 on chromosome 19, three SNPs were
337 associated with lamb metacarpal length, both when using the 50K SNP data and when using the imputed
338 SNP data (Figure 3A, Supplementary Table 3). All three SNPs were on chromosome 3 – when we
339 performed GWAS without fitting SNP genotypes, we found an association between this same region on
340 chromosome 3 and lamb metacarpal length, however only one SNP had previously reached the
341 significance threshold and thus had not been fitted in the conditional analysis.

342 After fitting the genotypes for the SNPs with the smallest p value on chromosomes 9, 11, 16 and 19 for
343 adult foreleg length, two further SNPs were found to be significant when performing conditional analysis
344 with the 50K SNP data; one SNP on chromosome 7 (s48811.1) and one SNP on chromosome 11
345 (s72261.1) (Figure 3B, Supplementary Table 3). The SNP on chromosome 7 was found to be associated
346 with adult hindleg length when using the 50K SNP data, but was not fitted in the conditional analysis due
347 to being the only SNP on this chromosome to pass the significance threshold. The SNP on chromosome
348 11 is located 1.2Mb upstream of the SNP from chromosome 11 that was fitted during the conditional
349 analysis. When performing the conditional analysis with the imputed SNP data, three new SNPs were
350 found to be associated with adult foreleg length, all in the same region of chromosome 7 previously
351 identified when we performed GWAS (Figure 3B, Supplementary Table 3.)

352 *Two-step GWAS*

353 To compare our results using the 50K SNP data to previous GWAS of adult traits in the Soays (Yang et al.
354 2010), we also performed a two-step GWAS. For our two-step analysis, we used the significance
355 threshold previously calculated in (Bérénos et al. 2015) ($1.35e^{-6}$), higher than the significance threshold
356 we calculated using the SimpleM method for the same data.

357 Across all 5 traits, we recovered the SNP-trait associations identified by (Bérénos et al. 2015). However,
358 we were unable to recover any of the novel SNP-trait associations we had found when performing our
359 single-step GWAS on the 50K SNP data, with the exception of the association between chromosome 16
360 and adult foreleg (however Bérénos et al. noted that SNPs in this region approached significance in their
361 analysis). Our QQ plots using the two-step method also matched the QQ plots of (Bérénos et al. 2015).
362 In both, the observed p values were higher than the expected p values, causing the majority of points in
363 the plots to fall below the x=y line (Supplementary Figure 3).

364

365 DISCUSSION

366 Heritability

367 Our results corroborate previous findings that all five body size traits we studied in Soay sheep are
368 influenced by genetic variation in the population (Bérénos et al. 2014), that *post mortem* measures
369 (metacarpal length and jaw length) have higher heritability estimates than live measures (weight,
370 foreleg length and hindleg length), that leg measures have higher heritability than weight (Beraldi et al.
371 2007; Bérénos et al. 2014; Wilson et al. 2006), and that heritability estimates increase with age (Bérénos
372 et al. 2014; Wilson et al. 2006).

373 The heritability estimates for the 50K data were very similar to those estimated using a GRM based on
374 the 50K data in a smaller sample of the same population of sheep by Bérénos et al. (2014) (Bérénos et
375 al. 2014), with estimates for the same trait falling within one standard error of each other. The biggest
376 difference was in adult metacarpal length with a heritability difference of 0.049 (estimates were 0.643
377 (0.049) and 0.594 (0.047) for our and Bérénos et al.'s results respectively). Given that we used the same
378 models as Bérénos et al., it is likely that the small differences between heritability estimates for each
379 trait may be due to our increased sample sizes.

380 Comparing the heritabilities estimated using the imputed SNP data against the estimates using the 50K
381 SNP data, we found little difference between the two SNP densities in any traits in any age class. The
382 additional genotypes at the imputed SNPs do not give any additional information on additive genetic
383 variation for these traits. This result is not surprising given the previous rarefaction analysis showing that
384 the heritability of these body size traits in adults asymptoted when about half the 50K SNP data was
385 used (Bérénos et al. 2014). There is high LD between nearby SNPs in the Soay sheep genome, which
386 suggests that most, if not all, of the causal variants tagged by the imputed SNP data may have already
387 been tagged by the 50K SNPs. The high LD was reflected when calculating GWAS significance thresholds
388 – whilst the number of SNPs between the 50K SNP data and the imputed SNP data increased by a factor
389 of ten, the number of effective tests only doubled (39K SNPs, 20082 effective tests and 401K SNPs and
390 48635 effective tests respectively).

391 For some of the traits we have analysed there is still a difference in heritability estimated using SNP data
392 versus heritability estimated using pedigree – for example, the highest SNP-based heritability estimate
393 for lamb metacarpal length (the estimate using the imputed SNP data) gave an estimate 59% of Bérénos
394 et al.'s pedigree-based estimate (Bérénos et al. 2014). Given that our SNP-based heritability estimates
395 were similar when using the 50K SNP data as when using the imputed SNP data, and the results of
396 Bérénos et al.'s rarefaction analysis (Bérénos et al. 2014), we believe it is unlikely that increasing the
397 density of genotyped SNPs that are common in the population will increase heritability estimates of
398 these traits. It is possible instead that the difference in heritability estimates obtained from pedigree
399 and genomic data is due to rare familial variants that do not segregate widely in the population, as well
400 as due to dominance and epistasis.

401

402 GWAS

403 Body size traits have been the focus of many kinds of analyses in Soay sheep (Ashraf et al. 2021; Bérénos
404 et al. 2015; Bérénos et al. 2014; Ozgul et al. 2009; Pemberton et al. 2017; Regan et al. 2017), and several

405 SNPs have already been identified as being associated with variation in these traits. A 2014 study aiming
406 to find SNP-trait associations for these body size traits in adults identified QTL for leg length measures
407 on chromosomes 16 and 19 (s23172.1 and s74894.1 respectively) (Bérénos et al. 2015). A more recent
408 study comparing genomic prediction methods in Soays using the 50K SNP data identified s48811.1 on
409 chromosome 7 and s50107.1 on chromosome 9 as having a probability higher than 0.9 of having a non-
410 zero effect on adult foreleg length (Ashraf et al. 2021). We were able to identify all four of these
411 associations in our GWAS, alongside associations that have not previously been identified in this
412 population. Use of the imputed SNP data allowed us to discover two more associations with loci that
413 were not genotyped in the 50K SNP data, suggesting that future identification of polymorphisms
414 influencing trait variation in the Soay sheep may benefit from using the imputed data.

415 Performing a two-step analysis confirmed that the novel SNP-trait associations we were able to identify
416 using the 50K SNP data were due to being able to fit the fixed and random effects for each trait whilst
417 performing GWAS all in a single step, rather than the increased population sample. Given the increase in
418 SNP-trait associations when using the single-step methodology, we suggest that a two-step GWAS is
419 redundant with the availability of software like DISSECT which is able to fit fixed and random effects
420 whilst performing GWAS. As we have shown, although DISSECT does not currently have the option to
421 automatically run a repeated measures GWAS, it is possible to modify input files to allow for repeated
422 measures.

423 The imputed SNP data revealed SNP-trait associations in two regions of the genome that were not
424 discovered using the 50K SNP data; one between a region on chromosome 7 and birth weight, and one
425 between a region on chromosome 13 and adult hindleg length. (Supplementary Table 2). When
426 examining the Manhattan plot for the 50K data for each trait (Figure 2A and 2B, Supplementary Table 2)
427 it is clear that there was a small cluster of SNPs just under the significance threshold in the 50K analyses.
428 The additional (imputed) SNPs may have matched the allele frequency of the underlying causal variants
429 more accurately, resulting in a smaller association p value.

430 We performed conditional analysis on all three leg length traits in both lambs and adults, as well as on
431 adult jaw length. For each trait, we simultaneously fitted the genotype for the SNP with the lowest p
432 value for any chromosome that had at least two SNPs found to be associated with the trait (see
433 Supplementary Table 2 for a list of SNPs fitted for each trait). We found that, with the exception of
434 s72261.1 on chromosome 11 associated with adult foreleg length, SNPs that were significant in the
435 GWAS analysis were no longer significant in the conditional analysis when a significant SNP on the same
436 chromosome was fitted (Figure 3, Supplementary Table 3). S72261.1 is located more than 1Mb away
437 from the SNP on chromosome 11 that was fitted during the conditional analysis of adult foreleg, and we
438 feel confident in stating that the effect that this SNP has on the trait is independent of the effects of the
439 SNP fitted as a covariate. We suggest that any future work looking to pinpoint the exact location of the
440 genetic variants affecting body size traits in Soay sheep primarily focus on the regions around the SNPs
441 listed in Supplementary Table 2, but keep in mind the results of our conditional analyses when
442 examining chromosome 11 and adult foreleg due to the potential for multiple causal variants on this
443 chromosome.

444 Across all traits for all age classes, the QQ plots showed deviation from the expected distribution of test
445 statistics under the null hypothesis (x=y line) for a wide range of test statistics, including low values,
446 indicative of underlying population structure not accounted for by the GRMs. Preliminary analysis

447 suggests that there is a low level of population stratification associated with sire ID, with the first 20
448 principal components of the GRM accounting for just under 10% of the observed genetic variation.
449 Preliminary work using LMM-OPS (Conomos et al. 2018), a method that splits genetic sample structure
450 into familial relatedness (alleles inherited that are identical by descent from recent common ancestors)
451 and population structure (alleles shared due to distant common ancestors) appears to reduce the
452 genomic inflation for the lower range of test statistics, while increasing the significance of more robust
453 associations. However further investigation of the application of that methodology to our population is
454 still required.

455 In order to have sufficient power to detect associations between markers and a trait of interest, GWAS
456 primarily requires two factors: i) a very high density of genotyped SNPs, and ii) a large number of
457 individuals that have been genotyped and phenotyped (Santure and Garant 2018). For intensively
458 studied organisms, both are achievable; such populations tend to have more individuals accessible to
459 collect data from, high density genotyping can be done at a lower cost due to higher demand, and, as in
460 humans, data from different populations can be combined to create larger sample sizes. GWA studies of
461 humans are the most obvious example of this; studies often have study populations made up of
462 hundreds of thousands of individuals and human SNP chips commonly genotype hundreds of thousands
463 of variants (for example, see (Ishigaki et al. 2020; Wood et al. 2014; Wu et al. 2021)). In comparison, wild
464 study population samples are much smaller – often struggling to reach one thousand individuals – and
465 the number of SNPs genotyped is much lower (for example, see (Malenfant et al. 2018; Perrier et al.
466 2018; Silva et al. 2017)). Analyses of wild populations therefore generally lack the power of more
467 intensively studied study organisms. Here, we have increased power by increasing the number of
468 genotyped markers via imputation. Despite high LD in the Soay sheep population, use of imputed data
469 has allowed us to identify two new SNP-trait associations, including an association with birth weight,
470 which had yet to be associated with any QTL in the Soay population. We have therefore shown that for a
471 given sample size, more information can be obtained by increasing the density of markers for those
472 individuals have been phenotyped. We suggest that, where possible, analyses of wild populations
473 impute SNP data in order to increase power and obtain results that may otherwise remain
474 undiscovered.

475 Although we have discovered new SNP-trait associations, it is likely that there are still causative variants
476 that remain undetected. GWAS lacks power to detect rare causative variants and variants with very
477 small effect sizes (Yang et al. 2010). Also, GWAS power drops when the same amount of phenotypic
478 variation is a consequence of multiple variants in the same region as opposed to a single variant
479 (Nagamine et al. 2012). Regional mapping methods have been developed that partition trait variance
480 into regions by simultaneously fitting a whole genome and a regional GRM, with the regions either being
481 defined as fixed SNP windows (Nagamine et al. 2012) or haplotype blocks (Shirali et al. 2018). Such
482 methodologies have the potential to identify regions of the genome that contain variants associated
483 with a trait that are unable to be identified by GWAS either due to being rare, or individually having
484 small effects on trait variation. Genomic prediction, which simultaneously estimates all marker effects
485 drawn from multiple distributions, can also be used to study the genetic architecture of traits by
486 estimating the posterior inclusion probability of a SNP having a non-zero effect on a trait. Genomic
487 prediction has already been used on adult body size traits in Soays, and has identified several of the
488 SNPs we identified through our GWAS approach (Ashraf et al. 2021). Ultimately, we believe that it is

important to use a variety of methodologies when studying the genetic architecture of complex traits, as different analyses have different strengths and may be able to identify different QTL.

References

Ashraf B, Hunter DC, Bérénos C, Ellis PA, Johnston SE, Pilkington JG, Pemberton JM, Slate J. 2021. Genomic prediction in the wild: A case study in soay sheep. Molecular Ecology. 1– 15.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. Nature reviews Genetics. 3(1):11-21.

Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM, Pemberton JM. 2007. Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. 61(6):1403-1416.

Bérénos C, Ellis PA, Pilkington JG, Lee SH, Gratten J, Pemberton JM. 2015. Heterogeneity of genetic architecture of body size traits in a free-living population. 24(8):1810-1830.

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM. 2014. Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. 23(14):3434-3451.

Burdick JT, Chen W-M, Abecasis GR, Cheung VG. 2006. In silico method for inferring genotypes in pedigrees. Nature genetics. 38(9):1002-1004.

Butler DG, Cullis BR, Gilmour AR, G. GB, R. T. 2017. Asreml-r reference manual version 4. Hemel Hempstead, HP1 1ES, UK.: VSN International Ltd.

Canela-Xandri O, Law A, Gray A, Woolliams JA, Tenesa A. 2015. A new tool called dissect for analysing large genomic data sets using a big data approach. Nature Communications. 6(1):10162.

Charmantier A, Garant D, Kruuk LEB. 2014. Quantitative genetics in the wild. Oxford University Press.

Conomos MP, Reiner AP, McPeek MS, Thornton TA. 2018. Genome-wide control of population structure and relatedness in genetic association studies via linear mixed models with orthogonally partitioned structure. bioRxiv.409953.

Gao X, Starmer J, Martin ER. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. 32(4):361-369.

. Applications of linear models in animal breeding. 1984.

Hickey JM, Kinghorn BP, Tier B, van der Werf JH, Cleveland MA. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. Genetics, selection, evolution : GSE. 44(1):9.

Ishigaki K, Akiyama M, Kanai M, Takahashi A, Kawakami E, Sugishita H, Sakaue S, Matoba N, Low S-K, Okada Y et al. 2020. Large-scale genome-wide association study in a japanese population identifies novel susceptibility loci across different diseases. Nature genetics. 52(7):669-679.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. Nature genetics. 42(4):348-354.

Kruuk LEB. 2004. Estimating genetic parameters in natural populations using the "animal model". Philosophical transactions of the Royal Society of London Series B, Biological sciences. 359(1446):873-890.

Kruuk LEB, Slate J, Wilson AJ. 2008. New answers for old questions: The evolutionary quantitative genetics of wild animal populations. 39(1):525-548.

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. Nature. 518(7538):197-206.

Malenfant RM, Davis CS, Richardson ES, Lunn NJ, Coltman DW. 2018. Heritability of body size in the polar bears of western hudson bay. Molecular Ecology Resources. 18(4):854-866.

536 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon
537 LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. Nature.
538 461(7265):747-753.
539 Mrode RA. 2014. Linear models for the prediction of animal breeding values.
540 Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I, Campbell H, Wilson J, Wild S, Hicks
541 AA et al. 2012. Localising loci underlying complex trait variation using regional genomic
542 relationship mapping. PloS one. 7(10):e46501.
543 Ozgul A, Tuljapurkar S, Benton TG, Pemberton JM, Clutton-Brock TH, Coulson T. 2009. The dynamics of
544 phenotypic change and the shrinking sheep of st. Kilda. Science (New York, NY). 325(5939):464-
545 467.
546 Pemberton JM. 2008. Wild pedigrees: The way forward. Proceedings Biological sciences. 275(1635):613-
547 621.
548 Pemberton JM, Ellis PE, Pilkington JG, Bérénos C. 2017. Inbreeding depression by environment
549 interactions in a free-living mammal population. Heredity (Edinb). 118(1):64-77.
550 Perrier C, Delahaie B, Charmantier A. 2018. Heritability estimates from genomewide relatedness
551 matrices in wild populations: Application to a passerine, using a small sample size. Molecular
552 Ecology Resources. 18(4):838-853.
553 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest radseq: An inexpensive
554 method for de novo snp discovery and genotyping in model and non-model species. PloS one.
555 7(5):e37135.
556 Regan CE, Pilkington JG, Bérénos C, Pemberton JM, Smiseth PT, Wilson AJ. 2017. Accounting for female
557 space sharing in st. Kilda soay sheep (ovis aries) results in little change in heritability estimates.
558 Journal of Evolutionary Biology. 30(1):96-111.
559 Santure AW, Garant D. 2018. Wild gwas—association mapping in natural populations. Molecular Ecology
560 Resources. 18(4):729-738.
561 Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS. 2018. Haplotype heritability mapping method
562 uncovers missing heritability of complex traits. Scientific Reports. 8(1):4982.
563 Silva CNS, McFarlane SE, Hagen IJ, Rönnegård L, Billing AM, Kvalnes T, Kemppainen P, Rønning B, Ringsby
564 TH, Sæther BE et al. 2017. Insights into the genetic architecture of morphological traits in two
565 passerine bird species. Heredity (Edinb). 119(3):197-205.
566 Soay sheep: Dynamics and selection in an island population. 2003. Clutton-Brock TH, Pemberton JM,
567 editors. Cambridge: Cambridge University Press.
568 Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. 2021. Genetic architecture and lifetime dynamics
569 of inbreeding depression in a wild mammal. Nature communications. 12(1):2972.
570 Van Vleck LD. 1970. Misidentification in estimating the paternal sib correlation. Journal of Dairy Science.
571 53(10):1469-1474.
572 Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG.
573 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing
574 between full siblings. PLOS Genetics. 2(3):e41.
575 Wilson AJ, Pemberton JM, Pilkington JG, Clutton-Brock TH, Coltman DW, Kruuk LEB. 2006. Quantitative
576 genetics of growth and cryptic evolution of body size in an island population. Evolutionary
577 Ecology. 21(3):337.
578 Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LE, Nussey DH. 2010. An
579 ecologist's guide to the animal model. The Journal of animal ecology. 79(1):13-26.
580 Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z et al.
581 2014. Defining the role of common variation in the genomic and biological architecture of adult
582 human height. Nature genetics. 46(11):1173-1186.

583 Wu Y, Murray GK, Byrne EM, Sidorenko J, Visscher PM, Wray NR. 2021. Gwas of peptic ulcer disease
584 implicates helicobacter pylori infection, other gastrointestinal disorders and depression. Nature
585 communications. 12(1):1146-1146.
586 Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Hastie ND, Hayward C, Vitart V, Navarro
587 P et al. 2016. Pedigree- and snp-associated genetics and recent environment are the major
588 contributors to anthropometric and cardiometabolic trait variation. PLoS Genet.
589 12(2):e1005804.
590 Xia C, Canela-Xandri O, Rawlik K, Tenesa A. 2021. Evidence of horizontal indirect genetic effects in
591 humans. Nature human behaviour. 5(3):399-406.
592 Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, Robinson MR, Perry JR, Nolte IM, van Vliet-
593 Ostaptchouk JV et al. 2015. Genetic variance estimation with imputed variants finds negligible
594 missing heritability for human height and body mass index. Nature genetics. 47(10):1114-1120.
595 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG,
596 Montgomery GW et al. 2010. Common snps explain a large proportion of the heritability for
597 human height. Nature genetics. 42(7):565-569.
598 Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. 2014. Advantages and pitfalls in the application
599 of mixed-model association methods. Nature genetics. 46(2):100-106.
600 Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL. 2013. Using extended genealogy
601 to estimate components of heritability for 23 quantitative and dichotomous traits. PLOS
602 Genetics. 9(5):e1003520.

603

Age	Trait	No. individuals	No. records	Fixed effects	Random effects
Neonate	Birth weight	2678	2678	Sex	Year of birth
				Litter size	Mother ID
				Population size year before birth	
				Age of mother (quadratic)	
				Ordinal date of birth	
				Age (days)	
Lamb	Weight	2228	2228	Sex	Year of birth
				Litter size	Mother ID
				Population size	Permanent environment
				Age (days)	
	Foreleg	2284	2284	Sex	Year of birth
				Litter size	Mother ID
				Population size	Permanent environment
				Age (days)	
	Hindleg	2349	2349	Sex	Year of birth
				Litter size	Mother ID
				Population size	Permanent environment
				Age (days)	
	Metacarpal	2059	2059	Sex	Year of birth
				Litter size	Mother ID
				Age at death (months)	
	Jaw	2113	2113	Sex	Year of birth
				Litter size	Mother ID
				Age at death (months)	
Adult	Weight	1152	3553	Sex	Year of capture
				Population size	Permanent environment
				Age (years)	
	Foreleg	1121	3331	Sex	Year of capture
				Population size	Permanent environment
				Age (years)	
	Hindleg	1135	3444	Sex	Year of capture
				Population size	Permanent environment
				Age (years)	
	Metacarpal	945	945	Sex	Year of birth
				Age at death (years)	
	Jaw	991	991	Sex	Year of birth
				Age at death (years)	

Table 1 Number of individuals and records, fixed and random effects fitted in each trait x age class model in addition to the GRM. The same individuals and records were used for both heritability estimates and for GWAS.

608 **Figure 1** Estimates of VA/VP for body size traits in neonates, lambs, and adult Soay sheep when using a
609 GRM calculated from the 50K SNP data (blue) compared with using a GRM calculated from the imputed
610 SNP data (yellow). Error bars represent standard error estimates.

611

612

613 **Figure 2** Manhattan plots for A) birth weight GWAS using 50K SNP data (left) and imputed SNP data
614 (right); and B) adult hindleg GWAS using 50K SNP data (left) and imputed SNP data (right). The red line
615 represents the significance threshold (2.49e-06 for the 50K SNP data and 1.03e-06 for the imputed SNP
616 data) – any SNPs above this threshold are considered to be significantly associated with variation in
617 their respective traits.

618

619

620 **Figure 3** Miami plots for A) lamb metacarpal GWAS using 50K SNP data (top left) and imputed SNP data
621 (top right), lamb metacarpal conditional analysis using 50K SNP data (bottom left) and imputed SNP data
622 (bottom right); and B) adult foreleg GWAS using 50K SNP data (top left) and imputed SNP data (top
623 right), lamb metacarpal conditional analysis using 50K SNP data (bottom left) and imputed SNP data
624 (bottom right). The red line represents the significance threshold (2.49e-06 for the 50K SNP data and
625 1.03e-06 for the imputed SNP data) – any SNPs above this threshold are considered to be significantly
626 associated with variation in their respective traits.

627

Data Availability Statement

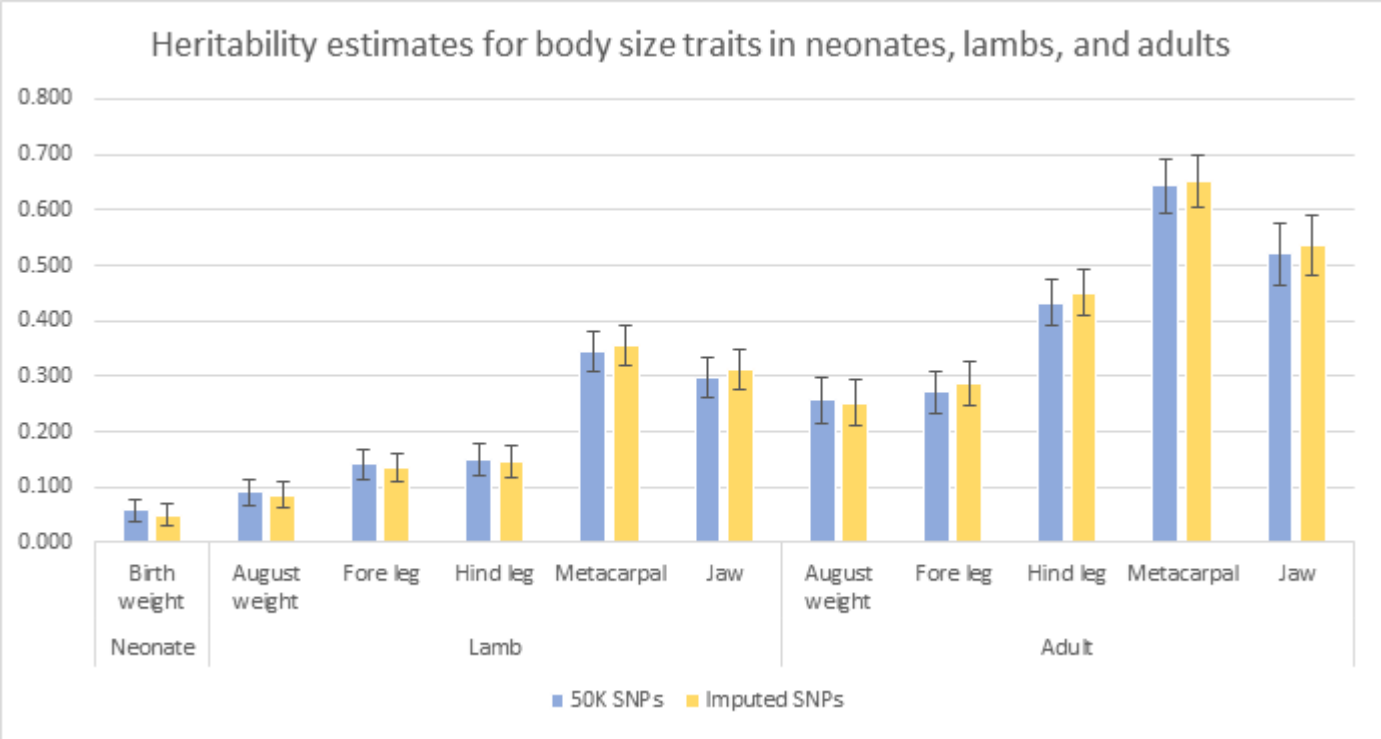
All data and scripts can be found at
https://github.com/CaelinnJames/Impact_of_SNPDensity_on_Soay_Sheep

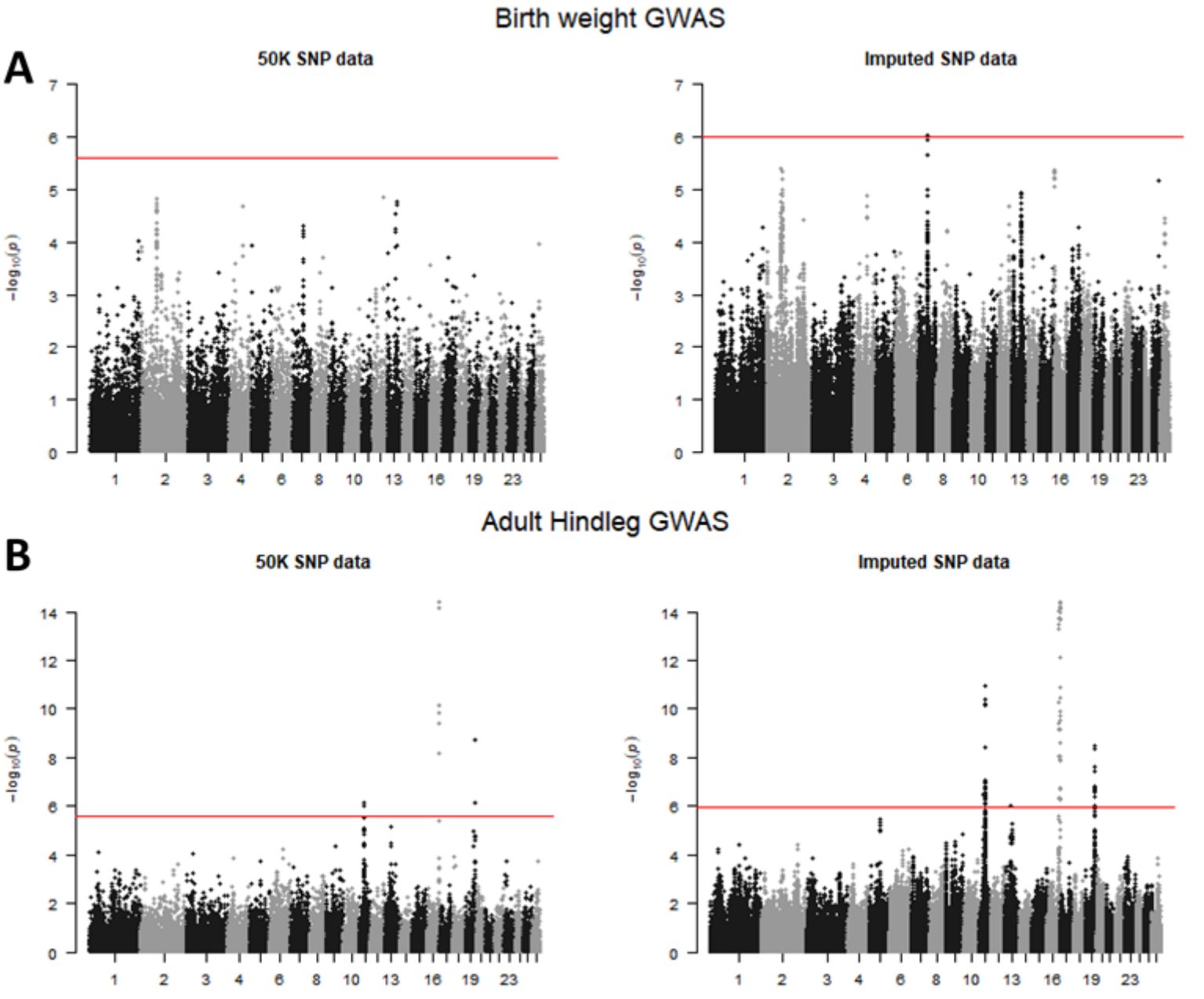
Acknowledgments

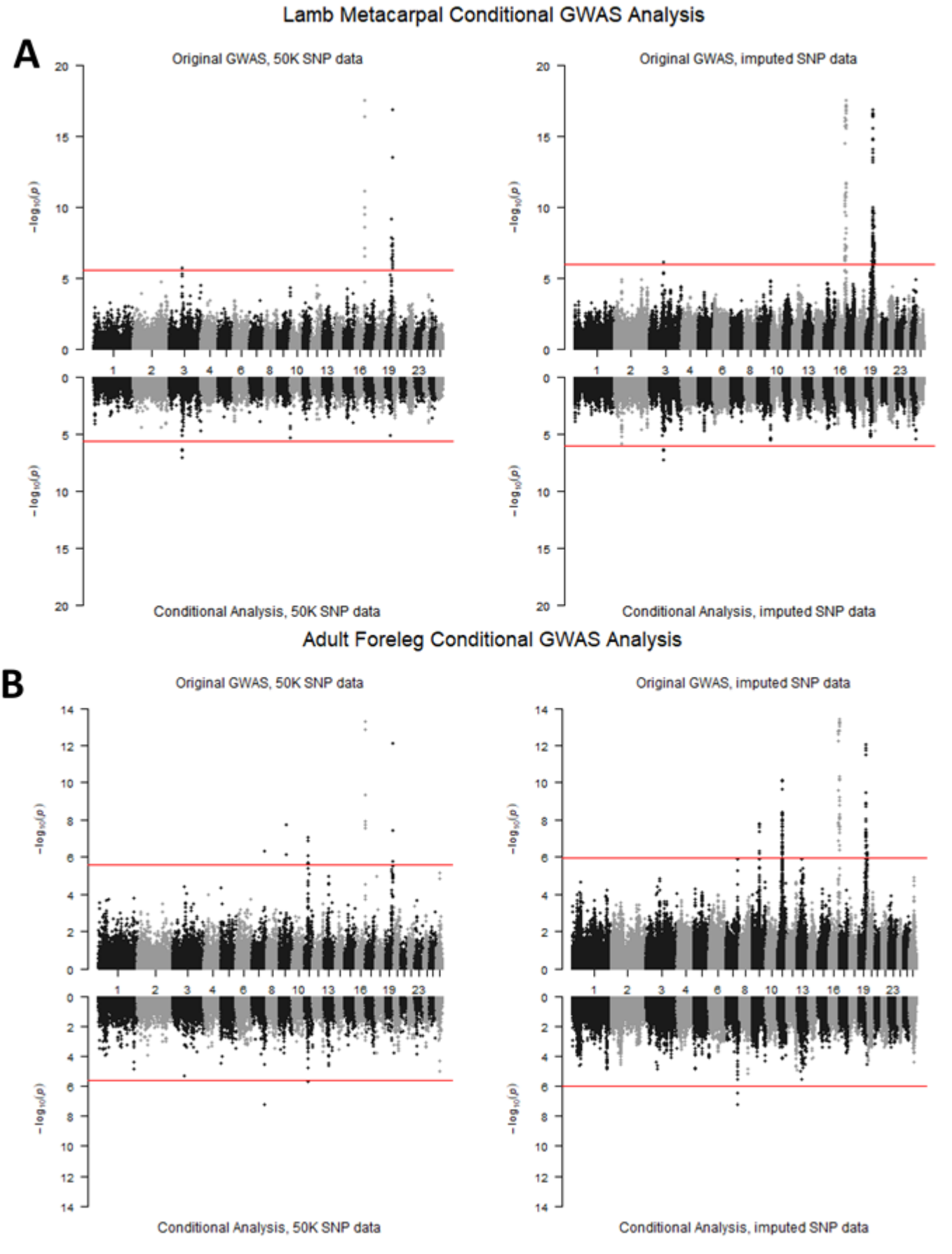
We thank the National Trust for Scotland and Scottish Natural Heritage for permission to work on St Kilda and QinetiQ and Eurest for logistics and other support on the island. We also thank all those who have been involved in the long-term project, including those who helped with field work on the island. We thank the Wellcome Trust Clinical Research Facility Genetics Core in Edinburgh for SNP genotyping.

Funding

This work was supported by a NERC Doctoral Training Partnership grant (NE/S007407/1). The long-term field project on St Kilda has been largely funded by the UK Natural Environment Research Council. The SNP genotyping was funded by a European Research Council Advanced Grant.


Heritability estimates for body size traits in neonates, lambs, and adults
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
Birth weight
August weight
Fore leg
Hind leg
Metacarpal
Jaw
August weight
Fore leg
Hind leg
Metacarpal
Jaw
Neonate
Lamb
Adult
50K SNPs
Imputed SNPs


Birth weight GWAS
A
50K SNP data
Imputed SNP data
$-\log_{10}(p)$
0
1
2
3
4
5
6
7
1
2
3
4
6
8
10
13
16
19
23
Adult Hindleg GWAS
B
50K SNP data
Imputed SNP data
$-\log_{10}(p)$
0
2
4
6
8
10
12
14
1
2
3
4
6
8
10
13
16
19
23


Lamb Metacarpal Conditional GWAS Analysis
A
Original GWAS, 50K SNP data
Original GWAS, imputed SNP data
Conditional Analysis, 50K SNP data
Conditional Analysis, imputed SNP data
Adult Foreleg Conditional GWAS Analysis
B
Original GWAS, 50K SNP data
Original GWAS, imputed SNP data
Conditional Analysis, 50K SNP data
Conditional Analysis, imputed SNP data
−log10(p)
1 2 3 4 6 8 10 13 16 19 23