



This is a repository copy of *Genomic prediction in the wild: A case study in Soay sheep*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/180418/>

Version: Published Version

---

**Article:**

Ashraf, B., Hunter, D.C., Bérénos, C. et al. (5 more authors) (2022) Genomic prediction in the wild: A case study in Soay sheep. *Molecular Ecology*, 31 (24). pp. 6541-6555. ISSN 0962-1083

<https://doi.org/10.1111/mec.16262>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Genomic prediction in the wild: A case study in Soay sheep

Bilal Ashraf<sup>1,2</sup> | Darren C. Hunter<sup>1,3</sup> | Camillo Bérénos<sup>4</sup> | Philip A. Ellis<sup>4</sup> |  
Susan E. Johnston<sup>4</sup> | Jill G. Pilkington<sup>4</sup> | Josephine M. Pemberton<sup>4</sup> | Jon Slate<sup>1</sup> 

<sup>1</sup>School of Biosciences, University of Sheffield, Sheffield, UK

<sup>2</sup>Department of Anthropology, Durham University, Durham, UK

<sup>3</sup>School of Biology, University of St Andrews, St Andrews, UK

<sup>4</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

## Correspondence

Jon Slate, School of Biosciences, University of Sheffield, Sheffield, UK.  
Email: j.slate@sheffield.ac.uk

## Funding information

H2020 European Research Council, Grant/Award Number: Wild Evolutionary Genomics; Natural Environment Research Council, Grant/Award Number: NE/M002896/1

## Abstract

Genomic prediction, the technique whereby an individual's genetic component of their phenotype is estimated from its genome, has revolutionised animal and plant breeding and medical genetics. However, despite being first introduced nearly two decades ago, it has hardly been adopted by the evolutionary genetics community studying wild organisms. Here, genomic prediction is performed on eight traits in a wild population of Soay sheep. The population has been the focus of a >30 year evolutionary ecology study and there is already considerable understanding of the genetic architecture of the focal Mendelian and quantitative traits. We show that the accuracy of genomic prediction is high for all traits, but especially those with loci of large effect segregating. Five different methods are compared, and the two methods that can accommodate zero-effect and large-effect loci in the same model tend to perform best. If the accuracy of genomic prediction is similar in other wild populations, then there is a real opportunity for pedigree-free molecular quantitative genetics research to be enabled in many more wild populations; currently the literature is dominated by studies that have required decades of field data collection to generate sufficiently deep pedigrees. Finally, some of the potential applications of genomic prediction in wild populations are discussed.

## KEYWORDS

genetic architecture, genomic prediction, genomic selection, molecular quantitative genetics

## 1 | INTRODUCTION

A major aim of evolutionary quantitative genetics is to measure and understand heritable genetic variation, and to explain the maintenance of that variation in the face of natural and sexual selection and genetic drift (Walsh & Lynch, 2018). One way that researchers have tried to answer these questions is by conducting long-term ecological studies of natural populations (Charmantier et al., 2014; Kruuk et al., 2008). There are now several studies, mostly of vertebrates, where individual life-histories of entire cohorts have been collected for several decades, spanning 10s of generations

of the focal organism (Clutton-Brock & Sheldon, 2010). Typically, the pedigree of the population has been determined, allowing researchers to use either (i) quantitative genetics (Kruuk, 2004) and/or (ii) gene mapping approaches (Slate et al., 2010) to study how selection and evolution have shaped diversity. Both approaches have led to genuine breakthroughs in our understanding of how genetic variation and selection have combined to shape biodiversity, but they also both have limitations (especially when applied to natural populations). Proponents of both approaches are aware of these limitations, and they are actively seeking solutions (Charmantier et al., 2014).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

The most obvious limitations of a quantitative genetics approach are that (i) the loci explaining trait variation cannot be identified, and (ii) several generations of data have to be collected before pedigree-based analyses are possible. Genomic solutions to understanding heritable genetic variation also suffer from problems. Whether conducting pedigree-based linkage mapping or a pedigree-free genome-wide association study (GWAS) it is hard to find quantitative trait loci (QTL) that reach strict genome-wide statistical significance, and in linkage mapping the effect sizes of reported QTL are almost certainly substantially overestimated (Slate, 2013; Slate et al., 2010). GWAS studies of wild populations have had some notable successes, although usually only when large effect loci are segregating (Comeault et al., 2014; Johnston et al., 2011). Because linkage mapping or GWAS studies require very high statistical significance thresholds ( $p < 1 \times 10^{-7}$  is not uncommon) to distinguish true positives from false ones, they are biased towards detecting causal loci of large effect. If a trait is polygenic then much of the genetic variation will remain unmapped. The most (in)famous example of this is human height, where in a sample size of ~one-quarter of a million people, 697 genomewide significant loci only explained 16% of the heritability (Wood et al., 2014). Studies of natural populations typically involve sample sizes of hundreds, or at most, a few thousand individuals, and therefore statistical power to detect loci of medium-small effect is low; for an excellent example see (Kardos et al., 2016). The problem may be exacerbated if phenotypes are hard to measure in the field or if environmental heterogeneity is a further source of phenotypic variation. Put simply, even in the longest-running studies in the wild, the power to detect most causal genes is very low. This makes it impossible to describe trait architectures or to understand the relationship between genotypes and phenotypes. How then, can evolutionary quantitative geneticists use genomics tools to understand and predict the genetic architecture and microevolution of their focal traits?

One possible solution to the low power/high polygenicity problem is to switch focus away from the hunt for individual significant loci, and instead try to understand the genetic component of phenotypic variation by using information from all of the typed SNPs. Two conceptually different approaches can be taken. The first involves running an “animal model” with the genetic relationship matrix estimated from markers rather than a pedigree. It was recognised several decades ago, before the necessary genotyping technology was available, that allele-sharing at markers could be used to estimate relatedness between pairs of individuals, and that relatedness coefficients could then be regressed on phenotypic similarity to predict heritabilities (Ritland, 1996; Ritland & Ritland, 1996). As genotyping technology improved, methods using marker-based relatedness became more tractable. In fact, marker-based approaches are potentially more accurate than pedigree-based methods because the realised relatedness between individuals rather than the expected relatedness can be estimated e.g. (Hayes et al., 2009; Visscher et al., 2006). Using a genomic relationship matrix (GRM) estimated from marker data, it is possible to estimate quantitative genetic parameters in a mixed model best linear unbiased predictor (BLUP)

framework (VanRaden, 2008); that is, to perform genomic best linear unbiased prediction (GBLUP). GBLUP has been used to estimate trait heritabilities in natural populations (Bérénos et al., 2014; Robinson et al., 2013; Santure et al., 2013; Silva et al., 2017), humans (Yang et al., 2010) and agriculturally important organisms (VanRaden et al., 2009). An extension of the model, where the GRM is estimated from a subset of markers, has been used to partition genetic variance into specific parts of the genome, e.g. in chromosome-partitioning (Visscher et al., 2007) or regional heritability mapping (Nagamine et al., 2012) studies.

In addition to providing estimates of parameters such as a trait's heritability, individual breeding values can also be estimated from GBLUP models. For a historical background and accessible summary of these developments see (Hill, 2014). An assumption of GBLUP is that the distribution of all of the marker effects are sampled from a single normal distribution; in other words, all markers explain some of the variation and the trait is highly polygenic. Of course, this assumption is unrealistic—many quantitative traits are highly polygenic, but not all of the markers will contribute to phenotypic variation. Furthermore, traits with nonpolygenic architectures are frequently of interest to breeders, medical geneticists and evolutionary geneticists studying natural populations. A potential improvement on GBLUP methods then, is to use methods that can accommodate different genetic architectures. This is achievable with a second major approach, whereby marker effects are estimated, and can come from more than one (not necessarily normal) distribution. The approach was first described almost 20 years ago (Meuwissen et al., 2001), and has since revolutionised animal and plant breeding (de los Campos et al., 2013). The concept is deceptively simple. Usually two steps are involved. In a first stage, marker effects are estimated, usually by Bayesian MCMC methods, in a “training population” of individuals with known phenotypes that are typed at many markers. Next, another panel of individuals with unknown phenotypes (the “validation” or “test” population) is genotyped at the same markers, and their genomic estimated breeding values (GEBVs) are predicted from their genotypes and the previously estimated effect sizes of each marker. The approach works, provided that the markers are in sufficiently strong linkage disequilibrium (LD) with the unknown loci that cause trait variation. If marker density is sufficiently high that adjacent markers are in LD with one another, then they should also be in LD with unknown causal loci. All of the approaches for estimating genomic breeding values from phenotypes and marker data have collectively become known as genomic prediction (and their application in breeding programs, as genomic selection). In medical genetics a similar approach, termed the polygenic risk score, is frequently used to predict phenotypes or disease risk (Wray et al., 2013). The main distinction between genomic prediction and polygenic risk scores is that the latter usually only uses SNPs that surpass a certain (usually quite stringent) significance threshold in a GWAS, rather than all available SNPs (Khera et al., 2018). This is only possible because studies of humans often involve hundreds of thousands of subjects, and therefore there is power to detect multiple causal variants of small effect. In wild populations that is unrealistic, so this

manuscript focuses on genomic prediction rather than polygenic risk scores.

Meuwissen et al. (2001) described two Bayesian genomic prediction models, Bayes A and Bayes B. In Bayes A, the marker effects are all nonzero, but are drawn from a scaled-inverse  $\chi^2$  distribution, allowing some markers to have large effects. In Bayes B, there are two distributions of marker effects, with a proportion of markers assumed to have zero effect and the remainder with effects drawn from a scaled-inverse  $\chi^2$  distribution. Following the development of Bayes A and Bayes B, further refinements, such as the marker effects coming from different types of distribution, have been made (Gianola, 2013; Habier et al., 2013). Methods that are flexible enough to capture different genetic architectures are potentially the most useful (Daetwyler et al., 2013). For example, methods that model marker effects as a mixture of different distributions can accommodate large and small effect loci (Erbe et al., 2012; Zhou et al., 2013). However, the main principle of using a training population to estimate marker effects, which are then used to predict phenotypes in a genotyped test population is common to all of these methods.

The accuracy of genomic prediction can be estimated by correlation between GEBV and phenotype, divided by the square root of the heritability (Meuwissen et al., 2013). Accuracy depends on a number of factors such as the heritability ( $h^2$ ) of the trait (Daetwyler et al., 2008; Goddard, 2009; Hayes et al., 2009), the training population sample size (Habier et al., 2013; Meuwissen et al., 2001), the marker density (Habier et al., 2009; Meuwissen & Goddard, 2010), the underlying genetic architecture (i.e., number and effect size of loci) of the trait (Daetwyler et al., 2010) and the statistical approach used (Cossa et al., 2010; Daetwyler et al., 2013; de los Campos et al., 2013; Meuwissen et al., 2001). Empirical investigation of these factors on genomic prediction accuracy have been described in the animal and plant breeding literature (Hayes, Bowman, et al., 2009; Hayes et al., 2010; Heffner et al., 2011; Zhao et al., 2012).

While genomic prediction has become a widely used tool in animal and plant breeding, it remains rare in quantitative evolutionary genetics studies of wild populations (but see Bosse et al., 2017; Gienapp et al., 2019; McGaugh et al., 2021; Stocks et al., 2019), even though there are a number of long-term longitudinal studies where suitable phenotypic and genomic data have been collected. Therefore, the main motivation of this study was to investigate the feasibility and accuracy of genomic prediction in a wild population of Soay sheep that has been the focus of a long-term study since 1985. Previous genetic studies of Soay sheep (Bérénos et al., 2015; Gratten et al., 2007, 2010; Johnston et al., 2011) have demonstrated that some phenotypes have a simple Mendelian basis (e.g., coat colour, coat pattern), others are highly polygenic (e.g. adult weight), and some are intermediate with many genes of small effect and a small number of loci with quite large effect contributing to genetic variation (e.g., horn length). Thus, a second aim of the study was to assess the accuracy of genomic prediction in different types of trait. Finally, the third aim of the study was to compare the accuracy of different genomic prediction methods (Daetwyler et al., 2013; de los Campos, Hickey, et al., 2013; Meuwissen et al., 2013). Thus, we compared five

different methods: (i) GBLUP, (ii) BayesA & (iii) BayesB - the models first introduced by (Meuwissen et al., 2001), (iv) Bayesian Lasso - a model similar to GBLUP and BayesA in the sense that it assumes all SNPs explain some variance and (v) BayesR - a model that assumes SNPs either have zero effect or come from a mixture of  $>1$  Gaussian distribution, potentially with some large effect loci.

## 2 | MATERIALS AND METHODS

### 2.1 | Study population

The Soay sheep is a primitive feral breed that resides on the St Kilda archipelago, off the NW of Scotland. Since 1985, the population on the largest island, Hirta (57°48'N, 8°37'W), has been the subject of a long-term individual based study (Clutton-Brock et al., 2004). Briefly, the majority of the sheep resident in the village Bay area of Hirta are ear-tagged and weighed shortly after birth and followed throughout their lifetime. Ear punches and blood samples suitable for DNA analysis are collected at tagging. Every year during August, adult sheep and lambs are captured and morphological measurements are taken. Winter mortality is monitored, with the peak of mortality occurring at the end of winter/early spring, and ca. 80% of all deceased sheep are found. To date, extensive life history data have been collected for over 10,000 sheep. Field work was carried out according to UK Home Office procedures and is licensed under the UK Animals (Scientific Procedures) Act of 1986 (licence no. PPL60/4211). More details of the natural history of the study system can be found elsewhere (Clutton-Brock et al., 2004).

### 2.2 | Genetic data

Genotyping of the population was performed using the Illumina Ovine SNP50 beadchip array, developed by the International Sheep Genomics Consortium (ISGC) (Kijas et al., 2012). Genotyping was performed at the Wellcome Trust Clinical Research Facility Genetics Core (Edinburgh, UK). Details about the genotype calling and quality control of the data are available elsewhere (Bérénos et al., 2014, 2015; Johnston et al., 2011, 2016). Briefly, pruning of SNPs and individuals was performed using `PLINK v 1.9` (Chang et al., 2015). SNPs were retained if they had a minor allele frequency of at least 0.01, a call rate of at least 0.98 and a Hardy-Weinberg equilibrium test  $p$ -value  $>.00001$ . Individuals were retained if they were typed for at least 0.95 of SNPs. Only autosomal SNPs were analysed, as most genomic prediction software cannot distinguish between autosomal and sex-linked loci; the sheep X chromosome represents ~5% of the total genome. Pruning was performed on a per-trait basis to ensure consistency of cutoff values between traits. Because SNP pruning was performed on a per-trait basis, the proportion of SNPs retained after pruning varied slightly between traits (fewest SNPs = 35,885 for coat colour and coat pattern; most SNPs = 36,437 for horn length).

TABLE 1 Summary of sample sizes and genetic architecture of the eight morphological traits

Trait	N	Architecture	Important loci/genes	Previous heritability	Heritability* (this study)	Reference
Weight	1168	Polygenic	-	0.30	0.32 (0.23–0.43)	Béréños et al. (2014, 2015)
Jaw length	897	Polygenic	-	0.50	0.59 (0.44–0.71)	Béréños et al. (2014, 2015)
Foreleg length	1126	Polygenic	Chr16: s23172., Chr19: s74894.1	0.25	0.53 (0.43–0.64)	Béréños et al. (2014, 2015)
Hindleg length	1139	Polygenic	Chr16: s23172.1	0.45	0.50 (0.38–0.59)	Béréños et al. (2014, 2015)
Metacarpal length	890	Polygenic	Chr16: s23172.1, Chr19: s74894.1	0.50	0.62 (0.52–0.75)	Béréños et al. (2014, 2015)
Male horn length	472	Oligogenic	Relaxin-like receptor 2 (RXFP2)	0.30	0.42 (0.24–0.58)	Johnston et al. (2011, 2013)
Coat colour	4737	Mendelian (dark coat caused by dominant allele)	Tyrosinase-related protein 1 (TYRP1) locus	-	-	Gratten et al. (2007)
Coat pattern	4737	Mendelian (wild-type coat caused by dominant allele)	Agouti signalling protein (ASIP)	-	-	Gratten et al. (2010)

Note: N is the number of phenotyped and genotyped individuals for each trait. Genes that explain all or most of the variation in oligogenic or Mendelian traits and SNPs that explain more than 10% of the additive genetic variance of polygenic traits are all reported. References indicate publication where estimation of heritability and linkage mapping and/or GWAS were performed. \*Heritability estimates unique to this study are from a MCMCglmm animal model using a pedigree-derived relationship matrix and phenotypes preadjusted for terms reported in Table S1. These estimates are expected to be greater than the published estimates, due to the removal of some nongenetic sources of variation. However, they are appropriate estimates when predicting the accuracy of GEBVs.

## 2.3 | Phenotypic data

We studied phenotypic data collected for eight different traits: foreleg length ( $n = 1126$ ), hindleg length ( $n = 1139$ ), weight ( $n = 1168$ ), metacarpal length ( $n = 890$ ), jaw length ( $n = 897$ ), horn length (normal-horned males only;  $n = 472$ ), coat colour ( $n = 4737$ ) and coat pattern ( $n = 4737$ ). An aim of the study was to investigate whether comparisons between the alternative genomic prediction methods were sensitive to the underlying genetic architecture. Therefore, we chose to study traits that had been the focus of previous gene mapping studies and were known to have variable architectures (summarised in Table 1). Among the body size traits, foreleg length, hindleg length, weight and horn length were measured on live animals, whereas metacarpal length and jaw length were taken from cleaned post-mortem skeletal material. Coat colour and coat pattern are independent discrete traits that are recorded at the time of capture and remain fixed throughout life (Clutton-Brock et al., 2004; Gratten et al., 2007, 2010). Further details of how the morphological traits are measured can be found elsewhere (Béréños et al., 2014, 2015; Johnston et al., 2011). Analyses of live-capture morphological data were restricted to animals captured in August, that were at least 28 months old, to remove most complications of growth and give consistency with previous genetic studies of these traits (Béréños et al., 2014, 2015). Male horn length included measurements taken outside of August, as many males are captured and measured during the rut in November. Post-mortem measurements were likewise restricted to animals who were at least 28 months old when they died.

## 2.4 | Modelling nongenetic effects

The quantitative traits being investigated are known to vary between sexes and age classes and are influenced by environmental effects (Béréños et al., 2014). Because some of the software being tested does not allow the inclusion of repeated measurements or fitting of random effects, it was necessary to fit models with nongenetic effects prior to performing the genomic prediction. Therefore, mixed effects models were run that adjusted phenotypes for fixed effects (sex and age of the animal) and non-genetic random effects (birth year, capture year and animal identity). Traits recorded on live animals had repeated measurements, so individual identity was fitted to remove any permanent environmental effect on phenotype. Traits recorded on skeletal material were recorded just once per animal. Models were run using the lmer function in the R v3.6.0 package lme4 v1.1-23 (Bates et al., 2015). The random effect of individual identity was extracted from each model and retained as the phenotype to be analysed in the genomic prediction models. The two Mendelian traits are categorical traits that are fixed throughout lifetime, unaffected by environmental conditions and with the same penetrance in each sex; therefore no phenotypic adjustments were necessary. Details of the model outputs are summarised in Table S1.

## 2.5 | Running genomic prediction models

A major aim of this study was to compare different methods for obtaining genomic estimated breeding values (GEBVs). Below we briefly discuss these methods, which rely on subtly different assumptions about the distribution of marker effect sizes, and outline how each model was run.

### 2.5.1 | BayesA

The BayesA method was one of the two models in the first genomic prediction paper (Meuwissen et al., 2001). All markers have a nonzero effect size, and the distribution of marker effects follows a scaled-inverse  $\chi^2$  distribution. Here, BayesA was implemented in the R package BGLR v1.0.8 (Perez & de los Campos, 2014). The analyses were performed with default parameter settings except the number of iterations was set to 120,000 (default = 5,000) with a burnin of 20,000 (default = 1,000) and a thinning interval of 100 (default = 10).

### 2.5.2 | BayesB

The BayesB method of genomic prediction was also introduced in Meuwissen et al. (2001). In this model, the assumption that all markers have a nonzero effect is removed. Instead, the BayesB method assumes a mixture of two distributions, such that some markers have zero effect and others, following a scaled-inverse  $\chi^2$  distribution, have nonzero effects, with some potentially of large effect. In reality, many loci explain zero genetic variance. BayesB analyses were also performed in the BGLR package with default parameter settings except the number of iterations, burnin length and thinning interval, which were the same values as used in the BayesA analyses.

### 2.5.3 | Bayesian Lasso (BayesL)

This method of genomic prediction is similar to BayesA in that all SNP effects come from a single distribution. However, BayesL models marker effects as following a double exponential distribution rather than a t distribution. A consequence of using this distribution is that, relative to the BayesA method, SNPs with small effect sizes are more strongly shrunk and SNPs with larger effect sizes are less strongly shrunk (de los Campos et al., 2009). BayesL was also implemented in the R package BGLR, using default parameters except the number of iterations, burnin length and thinning interval, which were the same values as for the BayesA and BayesB analyses.

### 2.5.4 | GBLUP

This method assumes that marker effects are drawn from a normal distribution. Rather than estimate the effect of each SNP, GBLUP

utilises the genomic relationship matrix (GRM) which is estimated from the proportion of alleles that are identical-by-state (IBS) between every pair of individuals. Here, the GRM was calculated from all typed markers. Genomic best linear unbiased predictions (GBLUPs) of breeding values are obtained by examining the covariance between phenotypic similarity and pairwise relatedness. GRMs have previously been used to estimate the heritability of traits in the Soay sheep population (Béréanos et al., 2014, 2015), but not to obtain GEBVs. As with the BayesA, BayesB and BayesL analyses, BGLR was used to obtain genomic EBVs.

### 2.5.5 | BayesR

BayesR models SNP effects as a mixture of  $>1$  normal distribution and an additional component of zero variance (Erbe et al., 2012). We implemented the method within the BayesR v1 software package (Moser et al., 2015) and used the default priors of four mixture components with variances of 0, 0.0001, 0.001 and 0.01 of the phenotypic variance. Dirichlet priors for the number of pseudo-observations (SNPs) in each distribution were set to 1, 1, 1 and 5. Priors for the genetic and residual variances were chosen as a scaled inverse-chi squared distribution with scaling parameter estimated from previous pedigree-based animal models (see references in Table 1) and degrees of freedom set to 10. Note that the default parameters give very similar values for genomic estimated breeding values, but we find that some samples of the MCMC chain return zero or very low estimates of additive genetic and residual variance under the default settings (see Table S2). As with all analyses run in BGLR, the BayesR analyses were run for a total of 120,000 iterations with a burnin of 20,000 and a thinning interval of 10.

Because BayesR reports the genetic variance explained by each SNP, SNP effect sizes can be aggregated (e.g., across the genome, across individual chromosomes) to describe the genetic architecture of the trait. Detailed descriptions of the trait architectures are not the main goal of this work, but we do report them in the Supporting Information. This is partially motivated by the knowledge that many of our focal traits already have well-described architectures (Table 1) and so it is useful to examine whether BayesR is assigning amounts of variance to loci that are consistent with genome wide association studies. We also compare descriptions of trait architectures between BayesR models using the default parameters and models using priors informed by previous quantitative genetic studies of those traits (Table S2).

## 2.6 | Assessing prediction accuracy and bias

Genomic prediction accuracy was assessed by cross-validation; each data set was randomly split into a training population containing 95%, 50% or 10% of individuals and a test population containing the remaining 5%, 50% or 90% of individuals. Twenty replicates

were generated for every combination of trait, model and training population size (8 traits  $\times$  5 methods  $\times$  3 training sizes  $\times$  20 replicates = 2400 models). For the six continuous traits, prediction accuracy was estimated as the correlation between the GEBV and an individual's phenotype, divided by the square root of the heritability (i.e.,  $r_{\text{GEBV},y}/h$ , where  $h^2$  is the heritability and  $y$  is the phenotype). We also report the correlation between the GEBV and phenotype, without dividing by the square root of the heritability (Table S3). Accuracy was averaged across the 20 replicates. The phenotypes were the values preadjusted for age, sex, year of birth and year of death using the same models as described in the section on *Modelling non-genetic effects*. Because the phenotypes were adjusted for random and fixed effects, the heritability of each trait was estimated from the adjusted phenotypes. These estimates of heritability will be greater than previously published estimates, because nongenetic sources of variation have already been removed, but they are more appropriate for measuring the observed and expected accuracy of the GEBVs. Trait heritabilities were estimated using Animal models run in the R package `MCMCGLMM` v2.29 (Hadfield, 2010). In these models the additive genetic variance was estimated from a relationship matrix determined by the population pedigree; the pedigree was constructed from a combination of behavioural observations and genotype data from 315 SNPs. Pedigree construction methods are described elsewhere (Bérénos et al., 2014). `MCMCglmm` models were run for 120,000 iterations with a burnin of 20,000 and sampling every 100th iteration. In addition to recording the correlation between GEBVs and phenotypes, a linear regression was performed to estimate the slope between the two variables. A slope of 1 is indicative of there being no bias in GEBVs (Meuwissen et al., 2001; Moser et al., 2015). Slopes greater than one would suggest that between-individual differences in GEBVs underestimate the difference in phenotypic values.

For the two categorical traits (coat colour/pattern), accuracy was determined by measuring the area under the curve (AUC) in a receiver operating characteristic (ROC) curve. ROC curves are well established tools for assessing how well genetic markers predict categorical traits such as disease presence/absence (Wray et al., 2010). ROC curve analysis were performed with the R package `PROC` v1.16.2 (Robin et al., 2011).

## 2.7 | Comparisons between observed and expected accuracy

It is possible to predict the accuracy of genomic prediction if information about the effective population size, genome size, recombination rate, trait heritability, and the genomic architecture are available (Daetwyler et al., 2010; Goddard et al., 2010). Here we used equation 2 of Daetwyler et al. (2010):

$$r_{\hat{g}\hat{g}} = \sqrt{\frac{N_p h^2}{N_p h^2 + \min(N_{\text{QTL}}, M_e)}}$$

where  $r_{\hat{g}\hat{g}}$  is the correlation between the true breeding value  $g$  and the GEBV  $\hat{g}$ .  $r_{\hat{g}\hat{g}}$  is equivalent to the  $r_{\text{GEBV},y}/h$ , the measure of accuracy we use in this paper (Meuwissen et al., 2013).  $N_p$  is the number of individuals in the training set,  $h^2$  is the trait heritability and  $N_{\text{QTL}}$  is the number of independent loci contributing to the trait. The other parameter,  $M_e$ , is the estimate for the number of independent chromosomal segments, given by  $M_e = 2N_e L / \log(4N_e L)$  where  $N_e$  is the effective population size and  $L$  is the genome linkage map length, measured in Morgans. The length of the Soay sheep genome is 33.04 Morgans (Johnston et al., 2016), and the effective population size is around 194 – see the supplementary material of (Kijas et al., 2012) – giving an  $M_e$  of 1263. We used the pedigree-estimated heritabilities of the traits adjusted for fixed and random effects (Table 1). We estimated the number of causal loci as 3000 (weight and jaw length), 1000 (foreleg, hindleg, metacarpal), 100 (horn length) or 1 (coat colour and coat pattern). 100 loci may seem like a large number of loci for horn length, given one locus explains most of the additive genetic variance (Johnston et al., 2011, 2013), but the BayesR analysis suggests that the remaining additive genetic variation is likely to be determined by many loci (Table S4). Note that while the true number of QTL can only be crudely guessed, the predicted accuracy is a function of whichever is smaller of  $N_{\text{QTL}}$  and  $M_e$ ; it is likely that for polygenic traits of Soay sheep  $M_e$  is lower than  $N_{\text{QTL}}$ , making the expected accuracy insensitive to guesses of the true number of QTL.

## 3 | RESULTS

### 3.1 | Genomic prediction accuracy

The main finding is that genomic prediction in the Soay sheep population is accurate regardless of trait architecture or of the method used (Figure 1, Table S2), even when training sizes are relatively modest (a few hundred individuals). For the continuous traits, accuracy was typically high (~0.70) when most of the phenotyped animals were included in the training set (Table S2, Figure 1a). Unsurprisingly, the accuracy declines when the training size is decreased. The phenotypes of the two categorical traits were predicted with very high accuracy, even when only 10% of animals were used in the training data set (Table S2, Figure 1b). As expected, for all traits, the accuracy was generally lower when the training set was smaller (Table S2, Figure 1).

### 3.2 | Genomic prediction bias

If genomic prediction estimates of GEBVs are unbiased, then it is expected that regressions of GEBVs on phenotypes should be equal to 1.0 (Meuwissen et al., 2001). There was a tendency for regression coefficients to be a little <1.0, meaning GEBVs overestimated between-individual variation, but this was largely driven by the models with smallest training sizes (Figure 2, Table S4). In the 50% and 95% training population data sets, there was relatively little bias with regression coefficients close to 1.0 for most traits and most models.



**FIGURE 1** Accuracy of genomic prediction for (a) quantitative and (b) Mendelian traits. For quantitative traits, accuracy is measured as  $r_{\text{GEBV},y}/h$ , the mean correlation between GEBVs and the phenotype, divided by the square root of the heritability. For Mendelian traits, accuracy is determined as area under curve from ROC analysis. All plots show the mean and SE obtained from 20 replicates of training sizes comprising 95%, 50% or 10% of the available individuals

Taking an average across 360 analyses (20 replicates  $\times$  3 training sizes  $\times$  6 quantitative traits) BayesL had the mean regression coefficient closest to 1.0 (Table S3, top row). However, of all of the methods, it had by far the largest standard deviation of regression coefficient, and therefore may be the most prone to severe bias. Of the remaining methods, BayesR had the mean regression coefficient closest to 1.0 and with the lowest standard deviation, although all of the other methods performed quite similarly with respect to bias (Table S4).

### 3.3 | Comparison of models

Among most of the continuous traits, there was little difference in the accuracy (Table S2, Figure 1a). This is not surprising as all of the models either make an assumption that the trait is polygenic, or

can explicitly model the trait as polygenic. BayesR and BayesB were perhaps marginally more accurate than BayesA, BayesL and GBLUP. For horn length, the most oligogenic of the continuous traits, and the two single-locus Mendelian traits, coat colour and coat pattern, BayesR and BayesB were the most accurate (Table S2, Figure 1b), presumably because they allow for the modelling of both zero effect and major effect loci. BayesA, was the next best performing model, with BayesL and GBLUP being considerably less accurate than all others, especially with smaller training populations. Thus, across all types of trait architecture, and considering both accuracy and bias, BayesR and BayesB appeared to be the most reliable methods of those examined here. Previous studies of humans and livestock have reached similar conclusions (Erbe et al., 2012; Moser et al., 2015). From this point, we mostly focus on results from the BayesR analyses, as that is the most accurate approach, and because it allows a detailed investigation of trait genetic architecture.

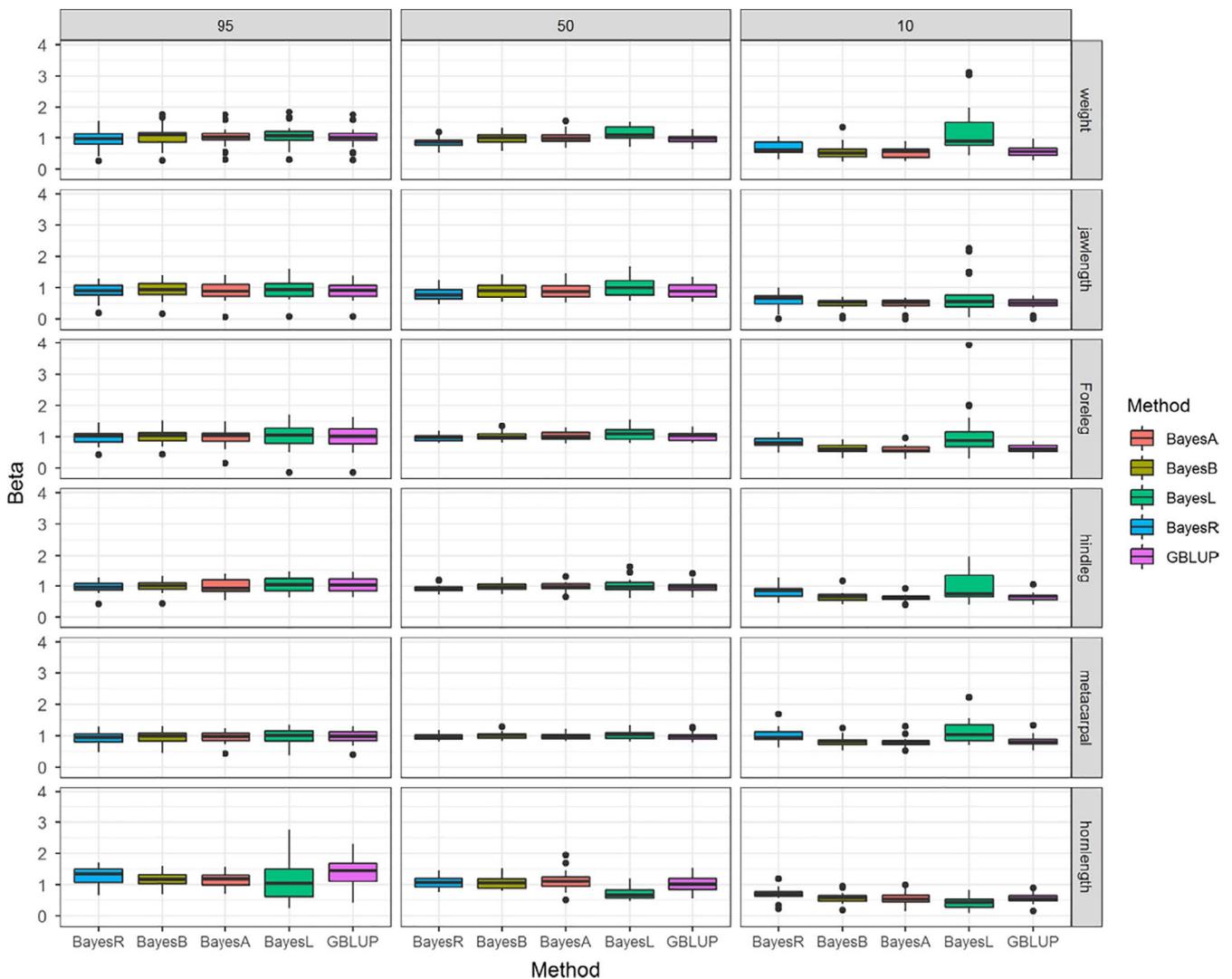


FIGURE 2 Boxplots showing the mean and distribution of regression coefficients when GEBVs were regressed on phenotypes for the six continuous traits. Each boxplot summarises 20 replicates per trait/training size/method combination

### 3.4 | Comparison between default and informed priors for BayesR models

GEBVs estimated from BayesR models running the default parameters and BayesR models using priors informed by previous molecular genetic studies were almost identical; for all eight traits the correlation in GEBVs between the two models was  $>0.996$  (Table S5). Genetic architectures were similar between the two models, although there was a tendency for the heritability to be a little greater when using the model that used informed priors (Table S5). For one trait, horn length, the default model tended to have a large autocorrelation between estimates of  $V_A$  from consecutive samples of the MCMC and it also frequently estimated there to be no additive genetic variance for the trait (almost 10% of MCMC samples), despite the trait having a high heritability.

### 3.5 | Comparisons between observed and expected accuracy

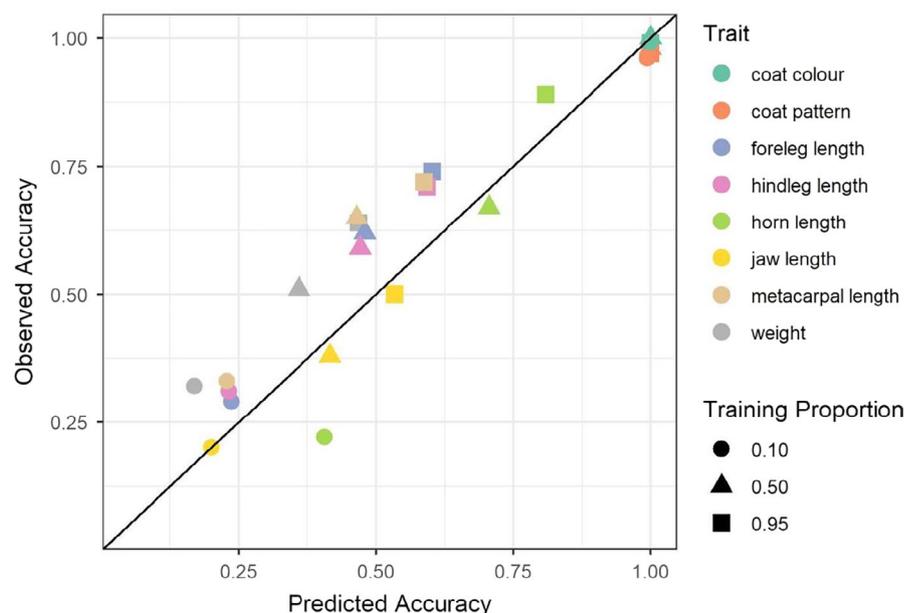
Overall, for all three training population sizes (95%, 50% and 10%), there was a good fit between the observed and expected accuracy of genomic prediction (Figure 3); for the leg length and weight traits, the observed accuracy was a little higher than expected.

### 3.6 | Description of trait architectures by BayesR

Although a description of trait architectures was not a primary objective of this work, BayesR provides estimates of many parameters of interest, including the trait heritability and the number and effect sizes of SNPs contributing to phenotypic variation. The trait architectures are summarised in the Supporting Information. Because most of the traits considered here have been the focus

of previous studies, it is possible to compare the BayesR outputs with other approaches such as GWAS and chromosome partitioning analyses. Trait architectures of the eight traits are summarised in Table S5, chromosome partitioning plots are presented in Figure S1, and Manhattan plots of posterior inclusion probabilities of each SNP having a nonzero effect size are presented in Figure S2. Broadly, the genetic architectures described by BayesR are consistent with those reported previously. Estimates of trait heritability were similar to estimates made using pedigrees (for comparisons see Table 1 and Table S5), with the six continuous traits all having BayesR estimated heritabilities between 0.39 and 0.64. Coat colour and coat pattern had lower heritabilities, despite both being determined by a single locus; this is expected as both *TYRP1* (coat colour) and *ASIP* (coat pattern) have one allele that is completely dominant to the other, so some nonadditive (dominance) genetic variance at the causal loci is well documented (Gratten et al., 2007, 2010).

BayesR returns an estimate of the number of SNPs that explain the phenotypic variation. We urge caution against taking these estimates as definitive (and note that the 95% confidence intervals span a ~10-fold range for most traits; Table S5). Any estimate is certain to be imprecise because full genome sequences have not been used, the sample size make it hard to distinguish between zero effect and small effect loci, and distinguishing between multiple tightly linked causal SNPs and one causal SNP is very difficult without doing functional genomics. However, the estimated number of SNPs was greatest for the traits that were thought to be highly polygenic (weight, jaw length), and was least for the Mendelian (coat colour and coat pattern) and major locus (horn length) quantitative traits. The estimated number of causal SNPs was intermediate for those quantitative traits where moderate effect size QTL had been discovered in previous GWAS studies. Thus, within-population, between-trait comparisons in the number of SNPs may give an indication of which traits are relatively more/less polygenic, provided the same set of markers is used for each trait.



**FIGURE 3** Observed and predicted accuracy of BayesR genomic prediction in Soay sheep

Chromosome partitioning plots (Figure S1) were broadly consistent with previous work using chromosome-wide relatedness matrices (figure 1 of Bérénos et al., 2015). However, in that study, two traits (adult weight and jaw length) had significant correlations between chromosome length and the proportion of variation explained by the chromosome, but in this study neither trait showed a significant relationship. Nonetheless, the chromosomes that tended to explain the most variation in the earlier study, also explained the most variation here. An exception was for jaw length, where Chromosome 20 explained more than 30% of the additive genetic variance here, but did not in the earlier study. However, no SNP on Chromosome 20 had a posterior inclusion probability  $>0.5$  (Figure S2b), so there is no compelling evidence of a major locus QTL on that chromosome. Instead, posterior inclusion probabilities plots indicate at least two regions of chromosome 20 contribute to jaw length variation (Figure S2b).

The two Mendelian traits, coat colour and coat pattern, are characterised by one allele being dominant to the other (at *Typr1* that dark allele is dominant to the light allele and at *ASIP* the wild allele is dominant to the self allele). Despite neither causal SNP being included on the ovine SNP chip, the distribution of GEBVs for both traits was trimodal (Figure S3), indicating that the linked SNPs successfully distinguished between the three possible genotypes (heterozygotes and the alternative homozygotes). In other words, it was possible to distinguish between phenotypically identical animals that are either heterozygous or homozygous for the dominant allele.

## 4 | DISCUSSION

Attempts to use genomic prediction in an ecological context remain very rare compared to its application in breeding or medicine, although some examples are now described (Bosse et al., 2017; Gienapp et al., 2019; McGaugh et al., 2021; Stocks et al., 2019). As far as we are aware this is the first genomic prediction study in a wild population that considers either multiple traits with different genetic architectures or different genomic prediction methods. Comparisons between studies are not straightforward because accuracy depends on both biological (e.g. genome size, effective population size, heritability) and technical (e.g. marker density, training population size) factors (Daetwyler et al., 2008, 2010; de los Campos, Vazquez, et al., 2013; Goddard, 2009), which inevitably will vary between studies. Certainly, the accuracy reported here is comparable to situations in plant and animal breeding where genomic selection is routinely used for trait improvement (Hayes, Bowman, et al., 2009; Lin et al., 2014) and is probably greater than has been observed for morphological traits in commercial sheep populations (Auvray et al., 2014; Brito et al., 2017). Therefore, the prospects for genomic prediction in this population are promising.

The relatively high accuracy of genomic prediction in Soay sheep is perhaps not especially surprising, as the values are mostly close to, or marginally greater, than expectations. If the effective population size had been overestimated, the predicted accuracy may have

been underestimated, compared to if the correct  $N_e$  had been used. However, overestimating  $N_e$  would be expected to affect the predicted accuracy of all traits, not just a subset of them. It is notable that previous GWAS studies (Bérénos et al., 2015) of the three leg traits identified QTL of reasonably large effect (up to 15% of the additive genetic variance), despite the traits being reasonably polygenic. These findings are supported by the BayesR analyses of genetic architecture (see Table S5). The accuracy of jaw length GEBVs were a little lower than for the other traits, although not much lower than predicted (Figure 3). Jaw length is highly polygenic and, unlike the leg length traits, no QTL have previously been identified in GWAS studies (Bérénos et al., 2015). In general, the traits with larger effect size loci had the greatest accuracy (see Supporting Information Part 4, Figure S4, Table S6), so perhaps the greater than expected accuracy of leg length but not jaw length is attributable, at least in part, to these relatively large effect loci.

The accuracy of genomic prediction is sensitive to the size of the training population. Unsurprisingly, the lowest accuracy was obtained for the traits with fewest records (horn length, jaw length, metacarpal length), when using training sets comprising just 10% of the animals. It is perhaps remarkable that the accuracy of horn length genomic prediction was as large as it was (0.22) when just 47 (10% of the total) animals were used in the training set. When 448 animals (95% of the total) were used in the training set, the accuracy was as high as 0.89. Thus, in populations with reasonably high linkage disequilibrium between SNPs, relatively modest increases in sample sizes can facilitate successful genomic prediction. The accuracy of genomic prediction is also sensitive to the accuracy with which the phenotype is measured. Improvements are likely if nongenetic sources of variation can be adjusted for, or if those effects can be included directly in the models. Here, we preadjusted phenotypes after accounting for sex, age, birth year and capture year. The accuracy of the phenotypic measurement can potentially be improved if repeated measurements are available. Some of our traits had multiple measurements per individual (weight, foreleg length, hindleg length), while others were only recorded once (e.g., jaw length and metacarpal length). Here, there was no obvious difference in the accuracy of GEBVs of traits that were measured repeatedly and GEBVs of traits that were measured once, although that may be because the repeat measures traits were recorded on live animals (which are harder to measure accurately) and the single measure traits were recorded on immobile material (bones).

### 4.1 | Conclusions and future directions

In this population the accuracy of genomic prediction was comparable to that seen in applied animal and plant breeding programs. In part, this is because Soay sheep have a history of isolation on a small island and a relatively small effective population size. This means LD extends several megabases across much of the genome (Bérénos et al., 2014; Feulner et al., 2013), and the number of SNPs (~38K)

on a medium density chip is sufficient to tag unknown causal loci. Future work will examine whether a higher marker density (400–500K SNPs) results in a further improvement in accuracy, although we note that in livestock populations, improvements in GEBV accuracy between 50–60 K SNP chips and whole genome sequences are often tiny (Frischknecht et al., 2018; Heidaritabar et al., 2016; Veerkamp et al., 2016). While other study systems used in evolutionary and ecological research may not have LD extending as far as it does in Soay sheep, the era of whole-genome sequencing hundreds or thousands of individuals is upon us. Therefore, an insufficient marker density will soon become a problem of the past, and genomic prediction should be feasible in other populations. Excitingly, this means that quantitative genetic studies are no longer restricted to wild populations with multigenerational pedigrees. This is especially important for long-lived organisms where sampling individuals across several generations may be a decades-long endeavour or for small organisms that are hard to track in the wild such as insects. Instead, cross-sectional rather than vertical sampling of study populations is possible, which greatly expands the range of organisms that could be studied. Genomic methods for describing genetic architectures, like the BayesR approach, mean that many different aspects of genomic architecture (e.g., a trait's heritability, additive genetic variance, and the contribution of individual loci and individual chromosomes etc.) can be estimated, even in the absence of a pedigree.

Of course, genomic prediction analyses usually seek to estimate breeding values (and by extension, phenotypes) which are properties of the individual rather than of a population (such as, e.g., a trait's heritability). There are several obvious areas where GEBVs can be used to address long-standing problems in evolutionary ecology. First, being able to estimate an individual's breeding value and/or phenotype before it is expressed means it should be possible to predict (and test) how different individuals will respond to a future environmental event. A good example of this comes from a genomic prediction study of ash trees experiencing an outbreak of ash dieback, a disease caused by the fungus *Hymenoscyphus fraxineus* (Stocks et al., 2019). Being able to identify which young plants are most resistant to dieback could be an essential tool in re-establishing populations devastated by disease. Second, GEBVs can be used to better understand the "invisible fraction" problem in evolutionary biology, highlighted more than three decades ago (Grafen, 1988). The invisible fraction refers to the idea that the individuals in a population who reach reproductive age, may not be representative of the entire population if mortality up to reproduction is nonrandom (Hadfield, 2008; Hemmings & Evans, 2020). Because estimating GEBVs in a test population requires only genomic data and no phenotypes, it should be possible to compare GEBVs between all of the individuals born in a cohort and the subset of them that reach reproductive age. Thus, genomic prediction at a fitness or survival-related trait should allow direct testing for, and quantification of, invisible fractions.

Perhaps the most obvious application of genomic prediction in wild populations will be to explore microevolutionary trends and

in particular revisit the "evolutionary stasis" problem (Merilä et al., 2001). This refers to the observation that traits are frequently observed to be heritable and under directional selection, yet they do not always evolve in response to that selection in the expected way. Detailed descriptions and explanations of the possible explanations for stasis are available elsewhere (Kruuk et al., 2001; Merilä, Kruuk, & Sheldon, 2001a, b; Merilä et al., 2001). Quantitative genetic approaches to understand evolutionary stasis have used EBVs derived from animal models applied to pedigreed populations (Coltman et al., 2003; Garant et al., 2004; Wilson et al., 2007), but a number of problems and biases associated with this approach have been highlighted (Hadfield et al., 2010; Postma, 2006). Notably, pedigree-derived predicted breeding values are more strongly correlated with the phenotype than true breeding values are (Postma, 2006). This is especially true for individuals without many phenotyped relatives in the pedigree. This is problematic for studies looking at microevolutionary trends in breeding values, as it means temporal changes in EBVs may reflect environmental effects on phenotypes rather than genetic ones. Similarly, in a pedigree-based approach, those individuals that lack phenotypic records and have few relatives have EBVs very close to the population mean of true breeding values (Hadfield et al., 2010). In studies that explore temporal trends in breeding values, these relatively uninformative individuals may be clustered towards either or both ends of a time series. Both problems are largely overcome with GEBVs generated from a genomic prediction test population, because each individual's genome rather than its phenotype or number of phenotyped relatives is used to predict its breeding value. It has been shown through simulations that genomic prediction is a more accurate method for estimating EBVs than pedigree-based methods when the focal individual is not closely related to phenotyped individuals (Clark et al., 2012). Of course, other problems highlighted in discussions of using breeding values to study microevolutionary trends, such as failing to incorporate the uncertainty in GEBVs (Hadfield et al., 2010), must still be addressed, but that is relatively straightforward if a Bayesian solution is used. Thus, it seems likely that genomic prediction can pave the way for new analyses of microevolutionary trends.

In conclusion, this study shows that genomic prediction can reliably measure individual breeding values in the Soay sheep population, and that high accuracy does not appear to be restricted to traits with specific underlying genetic architectures. Approaches that can model zero effect, small effect and large effect loci seem to be the most consistently reliable. Finally, we anticipate that similar studies will soon be possible in many other previously understudied organisms, paving the way towards both applied evolutionary quantitative genetics research and a re-exploration of some classic, yet unresolved, problems.

#### ACKNOWLEDGEMENTS

We thank the National Trust for Scotland (NTS) for permission to carry out fieldwork on St Kilda and to QinetiQ and Eurest for logistical support. We are grateful to the many volunteers who have helped with field data collection on St Kilda over the last three

decades. Genotyping was carried out at the Wellcome Trust Clinical Research Facility Genetics Core. This work was funded by a Natural Environment Research Council (NERC) grant, (NE/M002896/1) awarded to JS and JMP. SNP genotyping was mostly funded by a European Research Council (ERC) grant (Wild Evolutionary Genomics) awarded to JMP.

#### AUTHOR CONTRIBUTIONS

Coordination and collection of field data: Jill Pilkington and Josephine Pemberton. SNP genotype collection and quality control: Camillo Bérénos, Philip Ellis and Susan Johnston. Data analysis: Bilal Ashraf, Darren Hunter and Jon Slate. Manuscript writing: Jon Slate, Bilal Ashraf with input from all authors.

#### DATA AVAILABILITY STATEMENT

Genotype and phenotype data, along with files showing which individuals were allocated to training and test populations can be found on Dryad with doi: <https://doi.org/10.5061/dryad.h44j0zpkq>.

#### ORCID

Jon Slate  <https://orcid.org/0000-0003-3356-5123>

#### REFERENCES

- Auvray, B., McEwan, J. C., Newman, S. A., Lee, M., & Dodds, K. G. (2014). Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. *Journal of Animal Science*, *92*(10), 4375–4389. <https://doi.org/10.2527/jas.2014-7801>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bérénos, C., Ellis, P. A., Pilkington, J. G., Lee, S. H., Gratten, J., & Pemberton, J. M. (2015). Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, *24*(8), 1810–1830. <https://doi.org/10.1111/mec.13146>
- Bérénos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology*, *23*(14), 3434–3451. <https://doi.org/10.1111/mec.12827>
- Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., Gosler, A. G., McMahon, K., Poissant, J., Verhagen, I., Groenen, M. A. M., van Oers, K., Sheldon, B. C., Visser, M. E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, *358*(6361), 365–368. <https://doi.org/10.1126/science.aal3298>
- Brito, L. F., Clarke, S. M., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., Dodds, K. G., Sargolzaei, M., & Schenkel, F. S. (2017). Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genetics*, *18*(1), 7. <https://doi.org/10.1186/s12863-017-0476-8>
- Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., & Lee, J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Charmantier, A., Garant, D., & Kruuk, L. E. B. (2014). *Quantitative genetics in the wild*. Oxford University Press.
- Clark, S. A., Hickey, J. M., Daetwyler, H. D., & van der Werf, J. H. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genetics Selection Evolution*, *44*, 4. <https://doi.org/10.1186/1297-9686-44-4>
- Clutton-Brock, T., Pemberton, J., Coulson, T., Stevenson, I. R., & MacColl, A. D. C. (2004). The sheep of St Kilda. In T. Clutton-Brock, & J. Pemberton (Eds.), *Soay sheep: Dynamics and selection in an island population* (p. 383). Cambridge University Press.
- Clutton-Brock, T., & Sheldon, B. C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, *25*(10), 562–573. <https://doi.org/10.1016/j.tree.2010.08.002>
- Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C., & Festa-Bianchet, M. (2003). Undesirable evolutionary consequences of trophy hunting. *Nature*, *426*(6967), 655–658. <https://doi.org/10.1038/nature02177>
- Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L., & Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *American Naturalist*, *183*(5), 711–727. <https://doi.org/10.1086/675497>
- Crossa, J., Campos, G. D. L., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., & Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*(2), 713–724. <https://doi.org/10.1534/genetics.110.118521>
- Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G., & Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*, *193*(2), 347–365. <https://doi.org/10.1534/genetics.112.147983>
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, *185*(3), 1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, *3*(10), e3395. <https://doi.org/10.1371/journal.pone.0003395>
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, *193*(2), 327–345. <https://doi.org/10.1534/genetics.112.143313>
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K., & Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, *182*(1), 375–385. <https://doi.org/10.1534/genetics.109.101501>
- de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C., & Sorensen, D. (2013). Prediction of complex human traits using the genomic best linear unbiased predictor. *Plos Genetics*, *9*(7), e1003608. <https://doi.org/10.1371/journal.pgen.1003608>
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A., & Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, *95*(7), 4114–4129. <https://doi.org/10.3168/jds.2011-5019>
- Feulner, P. G. D., Gratten, J., Kijas, J. W., Visscher, P. M., Pemberton, J. M., & Slate, J. (2013). Introgression and the fate of domesticated genes in a wild mammal population. *Molecular Ecology*, *22*(16), 4210–4221. <https://doi.org/10.1111/mec.12378>
- Frischknecht, M., Meuwissen, T. H. E., Bapst, B., Seefried, F. R., Flury, C., Garrick, D., Signer-Hasler, H., Stricker, C., Bieber, A., Fries, R.,

- Russ, I., Sölkner, J., Bagnato, A., & Gredler-Grandl, B. (2018). Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss Cattle. *Journal of Dairy Science*, 101(2), 1292–1296. <https://doi.org/10.3168/jds.2017-12890>
- Garant, D., Kruuk, L. E. B., McCleery, R. H., & Sheldon, B. C. (2004). Evolution in a changing environment: A case study with great tit fledging mass. *American Naturalist*, 164(5), E115–E129. <https://doi.org/10.1086/424764>
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*, 194(3), 573–596. <https://doi.org/10.1534/genetics.113.151753>
- Gienapp, P., Calus, M. P. L., Laine, V. N., & Visser, M. E. (2019). Genomic selection on breeding time in a wild bird population. *Evolution Letters*, 3(2), 142–151. <https://doi.org/10.1002/evl3.103>
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 136(2), 245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- Goddard, M. E., Hayes, B. J., & Meuwissen, T. H. E. (2010). Genomic selection in livestock populations. *Genetics Research*, 92(5–6), 413–421. <https://doi.org/10.1017/s0016672310000613>
- Grafen, A. (1988). On the uses of data on lifetime reproductive success. In T. Clutton-Brock (Ed.), *Reproductive success. Studies of individual variation in contrasting mating systems* (pp. 454–471). University of Chicago Press.
- Gratten, J., Beraldi, D., Lowder, B. V., McRae, A. F., Visscher, P. M., Pemberton, J. M., & Slate, J. (2007). Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. *Proceedings of the Royal Society B-Biological Sciences*, 274(1610), 619–626
- Gratten, J., Pilkington, J. G., Brown, E. A., Beraldi, D., Pemberton, J. M., & Slate, J. (2010). The genetic basis of recessive self-colour pattern in a wild sheep population. *Heredity*, 104, 206–214. <https://doi.org/10.1038/hdy.2009.105>
- Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2009). Genomic selection using low-density marker panels. *Genetics*, 182(1), 343–353. <https://doi.org/10.1534/genetics.108.100289>
- Habier, D., Fernando, R. L., & Garrick, D. J. (2013). Genomic BLUP Decoded: A look into the black box of genomic prediction. *Genetics*, 194(3), 597–607. <https://doi.org/10.1534/genetics.113.152207>
- Hadfield, J. D. (2008). Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B-Biological Sciences*, 275(1635), 723–734. <https://doi.org/10.1098/rspb.2007.1013>
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22. <https://doi.org/10.18637/jss.v033.i02>
- Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. B. (2010). The misuse of BLUP in ecology and evolution. *American Naturalist*, 175(1), 116–125. <https://doi.org/10.1086/648604>
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433–443. <https://doi.org/10.3168/jds.2008-1646>
- Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., & Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. *Plos Genetics*, 6(9), e1001139. <https://doi.org/10.1371/journal.pgen.1001139>
- Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, 91(1), 47–60. <https://doi.org/10.1017/s0016672308009981>
- Heffner, E. L., Jannink, J. L., Iwata, H., Souza, E., & Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science*, 51(6), 2597–2606. <https://doi.org/10.2135/cropsci2011.05.0253>
- Heidaritabar, M., Calus, M. P., Megens, H. J., Vereijken, A., Groenen, M. A., & Bastiaansen, J. W. (2016). Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *Journal of Animal Breeding and Genetics*, 133(3), 167–179. <https://doi.org/10.1111/jbg.12199>
- Hemmings, N., & Evans, S. (2020). Unhatched eggs represent the invisible fraction in two wild bird populations. *Biology Letters*, 16(1), 20190763. <https://doi.org/10.1098/rsbl.2019.0763>
- Hill, W. G. (2014). Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. *Genetics*, 196(1), 1–16. <https://doi.org/10.1534/genetics.112.147850>
- Johnston, S. E., Béréños, C., Slate, J., & Pemberton, J. M. (2016). Conserved genetic architecture underlying individual recombination rate variation in a wild population of soay sheep (*Ovis aries*). *Genetics*, 203(1), 583–598. <https://doi.org/10.1534/genetics.115.185553>
- Johnston, S. E., Gratten, J., Béréños, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, 502(7469), 93–95. <https://doi.org/10.1038/nature12489>
- Johnston, S. E., McEWAN, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., Pemberton, J. M., & Slate, J. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, 20(12), 2555–2566. <https://doi.org/10.1111/j.1365-294X.2011.05076.x>
- Kardos, M., Husby, A., McFarlane, S. E., Qvarnstrom, A., & Ellegren, H. (2016). Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*, 16(3), 727–741. <https://doi.org/10.1111/1755-0998.12498>
- Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics*, 50(9), 1219–1224. <https://doi.org/10.1038/s41588-018-0183-z>
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., & Dalrymple, B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, 10(2), e1001258. <https://doi.org/10.1371/journal.pbio.1001258>
- Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London, Series B*, 359, 873–890. <https://doi.org/10.1098/rstb.2003.1437>
- Kruuk, L. E. B., Merila, J., & Sheldon, B. C. (2001). Phenotypic selection on a heritable size trait revisited. *American Naturalist*, 158(6), 557–571. <https://doi.org/10.1086/3233585>
- Kruuk, L. E. B., Slate, J., & Wilson, A. J. (2008). New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology Evolution and Systematics*, 39, 525–548. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173542>
- Lin, Z., Hayes, B. J., & Daetwyler, H. D. (2014). Genomic selection in crops, trees and forages: a review. *Crop & Pasture Science*, 65(11), 1177–1191. <https://doi.org/10.1071/cp13363>
- McGaugh, S. E., Lorenz, A. J., & Flagel, L. E. (2021). The utility of genomic prediction models in evolutionary genetics. *Proceedings of the Royal Society B: Biological Sciences*, 288(1956), 20210693. <https://doi.org/10.1098/rspb.2021.0693>

- Merilä, J., Kruuk, L. E. B., & Sheldon, B. C. (2001a). Cryptic evolution in a wild bird population. *Nature*, 412(6842), 76–79. <https://doi.org/10.1038/35083580>
- Merilä, J., Kruuk, L. E. B., & Sheldon, B. C. (2001b). Natural selection on the genetical component of variance in body condition in a wild bird population. *Journal of Evolutionary Biology*, 14(6), 918–929. <https://doi.org/10.1046/j.1420-9101.2001.00353.x>
- Merilä, J., Sheldon, B. C., & Kruuk, L. E. B. (2001). Explaining stasis: microevolutionary studies in natural populations. *Genetica*, 112, 199–222. <https://doi.org/10.1023/A:1013391806317>
- Meuwissen, T., & Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, 185(2), 623–631. <https://doi.org/10.1534/genetics.110.116590>
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Meuwissen, T., Hayes, B., & Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annual Review of Animal Biosciences*, 1(1), 221–237. <https://doi.org/10.1146/annurev-animal-031412-103705>
- Moser, G., Lee, S. H., Hayes, B. J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a bayesian mixture model. *Plos Genetics*, 11(4), e1004969. <https://doi.org/10.1371/journal.pgen.1004969>
- Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A. A., Pramstaller, P. P., Hastie, N., Wright, A. F., & Haley, C. S. (2012). Localising loci underlying complex trait variation using regional genomic relationship mapping. *PLoS One*, 7(10), 12. <https://doi.org/10.1371/journal.pone.0046501>
- Perez, P., & de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2), 483–495. <https://doi.org/10.1534/genetics.114.164442>
- Postma, E. (2006). Implications of the difference between true and predicted breeding values for the study of natural selection and microevolution. *Journal of Evolutionary Biology*, 19(2), 309–320. <https://doi.org/10.1111/j.1420-9101.2005.01007.x>
- Ritland, K. (1996). Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution*, 50(3), 1062–1073. <https://doi.org/10.2307/2410647>
- Ritland, K., & Ritland, C. (1996). Inferences about quantitative inheritance based on natural population structure in the yellow monkey-flower, *Mimulus guttatus*. *Evolution*, 50(3), 1074–1082. <https://doi.org/10.2307/2410648>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Muller, M. (2011). ROCR: An open-source package for R and S plus to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77. <https://doi.org/10.1186/1471-2105-12-77>
- Robinson, M. R., Santure, A. W., DeCauwer, I., Sheldon, B. C., & Slate, J. (2013). Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, 22, 3963–3980. <https://doi.org/10.1111/mec.12375>
- Santure, A. W., De Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular Ecology*, 22(15), 3949–3962. <https://doi.org/10.1111/mec.12376>
- Silva, C. N. S., McFarlane, S. E., Hagen, I. J., Rönnegård, L., Billing, A. M., Kvalnes, T., Kempainen, P., Rønning, B., Ringsby, T. H., Sæther, B.-E., Qvarnström, A., Ellegren, H., Jensen, H., & Husby, A. (2017). Insights into the genetic architecture of morphological traits in two passerine bird species. *Heredity*, 119(3), 197–205. <https://doi.org/10.1038/hdy.2017.29>
- Slate, J. (2013). From Beavis to beak color: a simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution*, 67(5), 1251–1262. <https://doi.org/10.1111/evo.12060>
- Slate, J., Santure, A. W., Feulner, P. G. D., Brown, E. A., Ball, A. D., Johnston, S. E., & Gratten, J. (2010). Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, 26(6), 275–284. <https://doi.org/10.1016/j.tig.2010.03.005>
- Stocks, J. J., Metheringham, C. L., Plumb, W. J., Lee, S. J., Kelly, L. J., Nichols, R. A., & Buggs, R. J. A. (2019). Genomic basis of European ash tree resistance to ash dieback fungus. *Nature Ecology & Evolution*, 3(12), 1686–1696. <https://doi.org/10.1038/s41559-019-1036-6>
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1), 16–24. <https://doi.org/10.3168/jds.2008-1514>
- Veerkamp, R. F., Bouwman, A. C., Schrooten, C., & Calus, M. P. (2016). Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. *Genetics Selection Evolution*, 48(1), 95. <https://doi.org/10.1186/s12711-016-0274-1>
- Visscher, P. M., Macgregor, S., Benyamin, B., Zhu, G. U., Gordon, S., Medland, S., Hill, W. G., Hottenga, J.-J., Willemsen, G., Boomsma, D. I., Liu, Y.-Z., Deng, H.-W., Montgomery, G. W., & Martin, N. G. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American Journal of Human Genetics*, 81(5), 1104–1110. <https://doi.org/10.1086/522934>
- Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G. U., Cornes, B. K., Montgomery, G. W., & Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *Plos Genetics*, 2(3), 316–325. <https://doi.org/10.1371/journal.pgen.0020041>
- Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.
- Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Clutton-Brock, T. H., Coltman, D. W., & Kruuk, L. E. B. (2007). Quantitative genetics of growth and cryptic evolution of body size in an island population. *Evolutionary Ecology*, 21(3), 337–356. <https://doi.org/10.1007/s10682-006-9106-z>
- Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Buchkovich, M. L., Croteau-Chonka, D. C., Day, F. R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A. U., ... Frayling, T. M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46(11), 1173–1186. <https://doi.org/10.1038/ng.3097>
- Wray, N. R., Yang, J., Goddard, M. E., & Visscher, P. M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *Plos Genetics*, 6(2), e1000864. <https://doi.org/10.1371/journal.pgen.1000864>
- Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14(7), 507–515. <https://doi.org/10.1038/nrg3457>
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7), 565–569. <https://doi.org/10.1038/ng.608>
- Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H. P., Longin, F. H., Ranc, N., & Reif, J. C. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics*, 124(4), 769–776. <https://doi.org/10.1007/s00122-011-1745-y>

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic Modeling With Bayesian Sparse Linear Mixed Models. *Plos Genetics*, 9(2), e1003264. <https://doi.org/10.1371/journal.pgen.1003264>

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Ashraf, B., Hunter, D. C., Béréanos, C., Ellis, P. A., Johnston, S. E., Pilkington, J. G., Pemberton, J. M., & Slate, J. (2021). Genomic prediction in the wild: A case study in Soay sheep. *Molecular Ecology*, 00, 1–15. <https://doi.org/10.1111/mec.16262>