Investigating pedigree- and SNP-associated components of heritability in a wild population of Soay sheep

James, Caelinn; Pemberton, Josephine M.; Navarro, Pau; Knott, Sara

Published in:
Heredity

DOI:
10.1038/s41437-024-00673-6

First published: 10/02/2024

Document Version
Publisher's PDF, also known as Version of record

the geneticssociety

www.nature.com/hdy

ARTICLE OPEN

Check for updates

Investigating pedigree- and SNP-associated components of heritability in a wild population of Soay sheep

Caelinn James[1,2]✉, Josephine M. Pemberton[1], Pau Navarro[3,4] and Sara Knott[1]

Estimates of narrow sense heritability derived from genomic data that contain related individuals may be biased due to the within-family effects such as dominance, epistasis and common environmental factors. However, for many wild populations, removal of related individuals from the data would result in small sample sizes. In 2013, Zaitlen et al. proposed a method to estimate heritability in populations that include close relatives by simultaneously fitting an identity-by-state (IBS) genomic relatedness matrix (GRM) and an identity-by-descent (IBD) GRM. The IBD GRM is identical to the IBS GRM, except relatedness estimates below a specified threshold are set to 0. We applied this method to a sample of 8557 wild Soay sheep from St. Kilda, with genotypic information for 419,281 single nucleotide polymorphisms. We aimed to see how this method would partition heritability into population-level (IBS) and family-associated (IBD) variance for a range of genetic architectures, and so we focused on a mixture of polygenic and monogenic traits. We also implemented a variant of the model in which the IBD GRM was replaced by a GRM constructed from SNPs with low minor allele frequency to examine whether any additive genetic variance is captured by rare alleles. Whilst the inclusion of the IBD GRM did not significantly improve the fit of the model for the monogenic traits, it improved the fit for some of the polygenic traits, suggesting that dominance, epistasis and/or common environment not already captured by the non-genetic random effects fitted in our models may influence these traits.

Heredity; https://doi.org/10.1038/s41437-024-00673-6

INTRODUCTION

The "genetic architecture" of a trait is a broad term used to describe the characteristics of genetic variants that contribute to the trait's phenotypic variation. Such characteristics include the number of loci contributing towards variation, the amount of variation attributable to each causal locus, the location of these loci in the genome, and how rare or common causal alleles are in the population (Mackay 2001). It is important to understand the genetic architecture underpinning traits of interest; in disease research this can be used to inform clinical diagnosis and prognosis as well as identify potential treatments; in livestock and crop breeding, it allows for more informed selective breeding strategies to improve trait yield; in evolutionary genetics we can use knowledge of the genetic architecture of a trait to understand how evolutionary processes may act on that trait and what micro-evolutionary dynamics are occurring in the population.

Historically, heritability estimation has been a key metric when investigating the genetic architecture of a trait. Narrow sense heritability (h^2) is the proportion of phenotypic variation which is explained by additive genetic variation in the population, and traditionally was estimated using correlations between family members in twin studies or through parent-offspring regression (Falconer and Mackay 1996). More recently the animal model (Meyer 1989) has become widely used, and has the advantage of using relationships between all individuals and across generations to increase power and better estimate the additive genetic component. The proportion of the variance from the additive genetic component can be estimated by an animal model using relatedness inferred from either the pedigree (h^2_{ped} – obtained using the numerator relationship matrix A (Henderson 1975)) or genetic markers such as SNPs (h^2_{GRM} – obtained using a genomic relationship matrix (GRM) (Yang et al. 2010)). The pedigree captures variance due to identity-by-descent and provides estimates of expected relatedness rather than realised relatedness. The pedigree can also be prone to error, especially when it is derived from observational data, but also in cases when it is derived from genotype data and paternities still cannot be accurately or uniquely assigned. A conventional GRM captures identity-by-state, and the resulting estimates are dependent on which SNPs are used to calculate the GRM; if neither the causal SNP nor any SNPs in linkage disequilibrium (LD) with the causal SNP have been genotyped, the h^2_{GRM} estimate may be under-estimated (Yang et al. 2015). This can be avoided by using high density genotyping arrays or whole genome sequencing, however in some wild populations it has been shown that increasing the number of SNPs above a certain threshold does not affect GRM-based heritability estimates (Bérénos et al. 2014; Perrier et al. 2018; James et al. 2022).

Inclusion of related individuals when estimating h^2_{GRM} can result in overestimation by inadvertently capturing effects of

[1]Institute of Ecology and Evolution, School of Biological Sciences, The University of Edinburgh, Edinburgh, UK. [2]Scotland's Rural College (SRUC), The Roslin Institute Building, Easter Bush, Midlothian, UK. [3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, UK. [4]MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK. Associate editor: Joram Mwacharo. ✉email: caelinnj@gmail.com

Received: 29 May 2023 Revised: 19 January 2024 Accepted: 24 January 2024
Published online: 10 February 2024

common environment, epistasis and dominance (Kang et al. 2010; Zaitlen et al. 2013). However, removal of related individuals is not optimal for smaller study populations such as wild populations, as this reduces sample sizes. In 2013, Zaitlen et al. proposed a partitioning method that simultaneously estimates both h^2 and h^2_{GRM} (referred to as h^2_g by the authors) of a trait in a population that contains related individuals. The method involves running an animal model simultaneously fitting both a GRM and an additional GRM which is thresholded in such a way that relatedness estimates below a specified threshold are set to 0. Both GRMs jointly model variance due to identity-by-descent (IBD), whilst the non-thresholded GRM additionally models variance due to identity-by-state (IBS). This method has so far been applied to a range of traits in humans (Xia et al. 2016; Hill et al. 2018) as well as in wild passerine birds (Silva et al. 2017) and farmed salmon (Kokkinias 2022). The extent to which thresholded GRMs explain trait variance differs across these studies, presumably due to differences in relatedness structure, trait architecture, trait heritability and other variables, so it is of interest to explore the outcome of the approach in a wide range of populations and traits.

Here, we use the method of Zaitlen et al. (2013) to simultaneously estimate h^2 and h^2_{GRM} for a selection of polygenic and monogenic traits in a wild population of related Soay sheep. The effect of additive genetic variation segregating at the population level is fitted using the non-thresholded IBS GRM, whilst the family-associated variation (such as epistasis, dominance, common environmental factors not captured by our non-genetic random effects fitted in the model, and any additional additive genetic variation segregating within families) is captured using the thresholded IBD GRM.

In addition, we explored a modified version of the approach, in which the thresholded IBD GRM was replaced with a GRM calculated from SNPs that had a minor allele frequency (MAF) below a certain threshold (but not thresholding relatedness), allowing us to examine how much of the genetic variation is captured by rare alleles and whether a rare-allele GRM can be used as a substitute for a relatedness thresholded GRM. We analysed a mixture of monogenic and polygenic traits to better understand how extremes of the genetic architecture of a trait affect the outcome of our heritability partition.

METHODS

Phenotypic data

The Soay sheep (*Ovis aries*) of the St. Kilda archipelago is a primitive breed of sheep that has been the focus of a longitudinal, individual-based study since 1985 (Clutton-Brock and Pemberton 2003). Individuals are ear-tagged at first capture (usually two to ten days after birth) to allow re-identification, regularly recaptured in order to measure, amongst others, various morphometric and life history traits across an individual's lifespan, and, after death, skeletal remains are collected and measured.

We focused on five polygenic morphometric traits across three different age classes (neonate, lamb and adult). Birth weight was measured in neonates. Three traits (weight, foreleg length and hindleg length) were measured on live individuals during an August catch up for lambs and adults. As adults are often recaptured across different years, the live traits have repeated measurements for many individuals. The remaining two traits (metacarpal length and jaw length) are *post mortem* measures taken from skeletal remains. Both birth and August weight are measured to the nearest 0.1 kg, 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). Several studies have demonstrated that these traits are under polygenic control (Bérénos et al. 2015; Ashraf et al. 2021; Hunter et al. 2022; James et al. 2022).

Neonates were defined as individuals who were caught and weighed between two and ten days after birth. For live traits, lambs were defined as individuals whose morphometric data was recorded in the August of their birth year, and for *post mortem* traits in lambs were individuals who died before 14 months of age. Similarly, adults were defined as individuals with phenotypic data recorded at least two years after birth for the live traits, or if they died after 26 months of age for the *post mortem* traits. We did not include yearling data in our analyses as these are individuals who have not yet reached their adult size and thus have small sample sizes available due to first winter mortality.

In addition to the polygenic traits, we analysed four monogenic traits (male horn type, female horn type, coat colour and wild/self coat pattern) to examine whether the different genetic architectures underpinning polygenic and monogenic traits yield different results when partitioning the heritability. These traits are well characterised in this population and the causal gene for each trait has been identified (see Johnston et al. (2013), Gratten et al. (2007), Gratten et al. (2010) and Supplementary Table 1 for additional information on the underlying genetics of horn type, coat colour and coat pattern respectively). Male and female horn types were analysed separately as while both are controlled by the same locus, in males the "normal horned" allele is dominant, so only two phenotype classes (normal horned and scurred) are observed, whilst in females there is codominance and three phenotypes are observed (normal horned, scurred and polled). As the monogenic traits investigated here do not change over ontogeny, we did not analyse these traits by age class.

Table 1 lists the number of individuals and records per trait.

Genetic data

Data were available for 8557 sheep genotyped on the Ovine SNP50 Illumina BeadChip, which gave genotypes for 38,130 autosomal variants remaining after quality control (MAF > 0.001, SNP locus genotyping success >0.99, individual genotyping success >0.95, IBS with another individual <0.9) that are polymorphic in the population. A subset of 438 SNPs are used to recover the pedigree using Sequoia (Huisman 2017), alongside observational data. 188 individuals have also been genotyped on the Ovine Infinium HD SNP BeadChip, which has a much higher density of variants than the SNP50 BeadChip. After quality control (SNP locus genotyping success >0.99 and individual genotyping success >0.95), 430,702 polymorphic SNPs were retained. This has allowed for the remaining genotypes to be imputed to this higher density using AlphaImpute, which combines shared haplotype and pedigree information for phasing and imputation (Hickey et al. 2012) (see Stoffel et al. (2021) for information on our imputation procedure). We used imputed genotype "hard" calls rather than genotype probabilities in downstream analyses. Genotypes with a probability of <0.99 were excluded, resulting in, 419,281 autosomal SNPs remaining for 8557 individuals (4035 females, 4452 males). We have previously shown that imputation does not affect heritability estimates for these traits in this population (James et al. 2022).

Overview of population structure

The Soay sheep population consists of individuals with a range of relationships. Currently, 10,979 individuals are included in the pedigree with an additional 379 genotyped individuals having no known connection to the pedigree, and the average pedigree pairwise relatedness is 0.0079. Full siblings are much rarer than half-siblings. In total, there are 17,948 parent-offspring pairs in the pedigree, 686 full sibling pairs, 124,739 half-sibling pairs, and 29,141 grandparent-grandchild pairs.

Using the imputed genotype data, Fhat3 inbreeding values (Yang et al. 2011) ranged between −0.23 and 0.65, with a median of 0 and a mean of 0.02. 90% of individuals have an inbreeding estimate lower than 0.04.

4.39% of non-diagonal relatedness estimates in the GRM of the total genotyped population were above 0.05. Of the non-diagonal estimates above 0.05, 0.98% were higher than 0.4, suggesting they were likely parent-offspring pairs, full siblings, or half-siblings where the non-shared parents were closely related. 7.64% of non-zero estimates were between 0.2 and 0.4, which would include half-siblings where the non-shared parents were not closely related, grandparent-offspring pairs, and avuncular relationships. 24.11% of non-zero estimates were between 0.1 and 0.2, which would include first cousins and great-grandparent-great-offspring pairs, as well as more complex relationships such as grand-avuncular pairs or half-avuncular pairs. The remaining estimates above 0.05 include more distant relationships, such as half-cousins and double second cousins.

1.44% of non-diagonal relatedness estimates in the thresholded GRM of the total genotyped population were above 0.1. Of the non-diagonal estimates above 0.1, 2.99% were higher than 0.4, whilst 23.35% of estimates were between 0.2 and 0.4.

The proportion of non-zero relatedness estimates within each relatedness interval remained similar across the samples of individuals for each trait.

Table 1. Number of individuals and records, fixed and random effects fitted in each trait and age class model in addition to the genomic relationship matrices or A matrix.

Age	Trait type	Trait	Number of individuals	Number of records	Fixed effects	Random effects
Neonate	Polygenic	Birth weight	2975	2975	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	2424	2424	Sex	Year of birth
					Litter size	Mother ID
					Population size	
					Age (days)	
		Foreleg	2512	2512	Sex	Year of birth
					Litter size	Mother ID
					Population size	
					Age (days)	
		Hindleg	2577	2577	Sex	Year of birth
					Litter size	Mother ID
					Population size	
					Age (days)	
		Metacarpal	2117	2117	Sex	Year of birth
					Litter size	Mother ID
					Age at death (months)	
		Jaw	2172	2172	Sex	Year of birth
					Litter size	Mother ID
					Age at death (months)	
Adult		Weight	2092	3860	Sex	Year of capture
					Population size	Permanent environment
					Age (years)	
		Foreleg	1936	3594	Sex	Year of capture
					Population size	Permanent environment
					Age (years)	
		Hindleg	2027	3481	Sex	Year of capture
					Population size	Permanent environment
					Age (years)	
		Metacarpal	987	987	Sex	Year of birth
					Age at death (years)	
		Jaw	1057	1057	Sex	Year of birth
					Age at death (years)	
Monogenic		Male horn type	3291	3291	N/A	N/A
		Female horn type	2973	2973	N/A	N/A
		Coat colour	7319	7319	N/A	N/A
		Coat pattern	7319	7319	N/A	N/A

Variance component analysis

We used animal models to partition the phenotypic variance for each trait into genetic and non-genetic variance components. For each trait, we ran multiple models:

$$\text{Model 1}: y = X\beta + \sum_r Z_r u_r + Wg + \varepsilon$$

Where **y** is the vector of phenotypic values; **X** is a design matrix linking individual records with the vector of fixed effects **β, Z_r** is an incidence matrix that relates a non-genetic random effect to the individual records; **u_r** is the associated vector of non-genetic random effects; **g** is the vector of additive genetic random effects of the whole study population with **W** the incidence matrix; and **ε** is the vector of residuals. It is assumed that

Table 2. Number of SNPs used in the MAF-thresholded GRMs at each threshold, and percentage of the total number of genotyped SNPs they represent.

MAF threshold	Number of SNPs	Percentage of total SNPs
0.1	112671	26.872%
0.05	60303	14.382%
0.01	13460	3.210%
0.005	7692	1.835%
0.001	1686	0.402%

$\mathbf{g} \sim MVN(0, M_g\sigma_g^2)$, where σ_g^2 is the additive genetic variance and M_g is the GRM for the whole study population. This model was run once per trait for the polygenic traits. For the monogenic traits, this model was run three times: once with all SNPs being used to construct the GRM, once with all SNPs on the same chromosome as the causal gene being used to construct the GRM, and once with all SNPs within 1 Mb upstream or downstream of the causal gene being used to construct the GRM. For monogenic traits, these models were designed to eliminate noise from genetic variation from non-causal loci, given that the only variants affecting these traits should be those within the previously identified causal regions.

$$\text{Model 2}: y = X\beta + \sum_r Z_r u_r + Wg + Wk_t + \varepsilon$$

The terms in this model are the same as in Model 1, with the inclusion of $\mathbf{k_t}$, the vector of extra genetic random effects associated with relatives with a genomic relatedness higher than threshold t. It is assumed that $\mathbf{k_t} \sim MVN(0, M_{kt}\sigma_{kt}^2)$, where σ_{kt}^2 is the kinship genetic variance and M_{kt} is the kinship GRM with relationships equal to or less than t being set to 0. This model was run twice per trait, once at $t = 0.05$ and once at $t = 0.1$. Both thresholds capture parent-offspring, full sibling and half-sibling relationships but differ as to the proportion of more distantly related pairs of individuals that are retained.

$$\text{Model 3}: y = X\beta + \sum_r Z_r u_r + Wg + Wg_{MAF} + \varepsilon$$

The terms in this model are the same as in Model 1, with the inclusion of $\mathbf{g_{MAF}}$, the vector of additive genetic random effects of SNPs with a MAF under a specified threshold in the whole study population. It is assumed that $\mathbf{g_{MAF}} \sim MVN(0, M_{MAF}\sigma_{MAF}^2)$, where σ_{MAF}^2 is the additive genetic variance of SNPs with a MAF below a set threshold, and M_{MAF} is the GRM calculated from the SNPs with a MAF below the threshold for the whole study population and range of relationships. For polygenic traits, this model was run five times, with MAF thresholds varying between 0.1 and 0.001 (see Table 2 for the full list of MAF thresholds and the number of SNPs that remained for each threshold).

For the monogenic traits, in addition to running the same models as for the polygenic traits, we also ran the model for each threshold where both GRMs were only constructed from SNPs on the chromosome containing the causal gene, and again for each threshold where both GRMs were only constructed from SNPs within the region spanning 1 Mb either side of the causal gene (See Table 3 for the number of SNPs that remained for each threshold for each trait for both the single chromosome and region based GRMs). We did not run the model for any thresholds that resulted in a GRM being computed for less than 10 SNPs. These models were designed to eliminate noise from genetic variation from non-causal loci, given that the only variants affecting these traits should be those within the previously identified causal regions.

$$\text{Model 4}: y = X\beta + \sum_r Z_r u_r + Wped + \varepsilon$$

The terms in this model are the same as in Model 1, with the exception of **ped** being the vector of pedigree-based effects of the whole study population. It is assumed that $\mathbf{ped} \sim MVN(0, A_g\sigma_g^2)$, where σg^2 is the additive genetic variance and A_g is the relationship matrix for the whole study population. Again, this model was run once per trait.

Fixed and non-genetic random effects were only fitted for the polygenic traits, with the effects fitted differing between traits and age classes (see Table 1). For each trait, the same individuals were analysed across all models.

Table 3. Number of SNPs used in the MAF-thresholded GRMs at each threshold for each trait for both the GRMs constructed from SNPs on the focal chromosome and SNPs within 1 Mb either side of the focal gene.

MAF threshold	Number of SNPs (chromosome) – Horn type	Number of SNPs (region) – Horn type	Number of SNPs (chromosome) – Coat colour	Number of SNPs (region) – Coat colour	Number of SNPs (chromosome) – Coat pattern	Number of SNPs (region) – Coat pattern
No threshold	15517	422	41917	294	13626	142
0.1	4125	66	12997	*102 (Failed to converge)*	3127	*33 (Failed to converge)*
0.05	2109	24	7120	*89 (Failed to converge)*	1432	*16 (Failed to converge)*
0.01	284	*Not run due to an insufficient number of SNPS*	1764	*29 (Failed to converge)*	327	*10 (Failed to converge)*
0.005	177	*Not run due to an insufficient number of SNPS*	847	*Not run due to an insufficient number of SNPS*	215	*Not run due to an insufficient number of SNPS*
0.001	79	*Not run due to an insufficient number of SNPS*	111	*Not run due to an insufficient number of SNPS*	45	*Not run due to an insufficient number of SNPS*

Male and female horn are controlled by the same locus, so the number of SNPs in each GRM is the same for these two traits. Italics indicates that the model either was not run or failed to converge.

The GRMs were computed using GCTA (Yang et al. 2011), which computes the genetic relationship between individuals i and j as

$$A_{ij} = \frac{1}{N}\sum_z A_{ijz} = \begin{cases} \frac{1}{N}\sum_z \frac{(s_{iz}-2p_k)(s_{jz}-2p_z)}{2p_z(1-p_z)}, i \neq j \\ 1 + \frac{1}{N}\sum_z \frac{s_{iz}^2-(1+2p_z)s_{iz}+2p_z^2}{2p_z(1-p_z)}, i = j \end{cases}$$

where s_{iz} is the number of copies of the reference allele for SNP z of the individual i, p_z is the frequency of the reference allele for the SNP z, and N is the number of SNPs (Yang et al. 2010). The models described above were run using DISSECT (Canela-Xandri et al. 2015). This method of calculating the GRM weights SNP effects by their MAF; if two individuals share a rare allele, they are more likely to be related than two individuals that share the common allele of the same SNP. Giving more weight to rare alleles means that individuals that are genotypically similar due to IBD are likely to have a higher relatedness estimate than those who are similar due to IBS and are unrelated, so this estimator works to our advantage.

To deal with multiple measurements per individual for the adult live traits, we used a repeatability model by fitting ID as a random effect to ensure that uncertainty was correctly propagated through all estimations (Mrode 2014). Although DISSECT does not currently have the option to automatically analyse repeated measures, it is possible to modify input files to allow for a repeated measures model (see James et al. (2022) for method details).

For models 1 and 4, we estimated the narrow sense heritability (h^2) by dividing the additive genetic variance (the variance associated with the GRM or pedigree respectively) by the total estimated phenotypic variance (yielding h^2_{GRM} and h^2_{ped} respectively). For models 2 and 3, we estimated three heritabilities: to avoid confusion when comparing with models 1 and 4, we refer to them as h^2_{pop}, the additive genetic variance explained by the full GRM (equivalent to h^2_{GRM} from model 1); h^2_{kin}, the additive genetic variance explained by the thresholded GRM; and h^2_{pk}, which is estimated as the sum of h^2_{pop} and h^2_{kin} (equivalent to h^2 and h^2_{ped}). For the monogenic traits, we refer to the heritability estimates from models 1 and 3 when only including SNPs on the same chromosome as the causal gene as $h^2_{GRM\_Chrom}$, $h^2_{pop\_Chrom}$, $h^2_{kin\_Chrom}$, and $h^2_{pk\_Chrom}$. Likewise, for the heritability estimates when only including SNPs within 1 Mb of the causal gene, we shall refer to them as $h^2_{GRM\_Region}$, $h^2_{pop\_Region}$, $h^2_{kin\_Region}$, and $h^2_{pk\_Region}$.

We performed log-likelihood ratio tests (LRT) between models 2 and 3 and model 1 at each threshold for each trait. For models where the inclusion of the thresholded GRM was determined to improve the fit, h^2_{pk} estimates were compared to h^2_{ped} estimates to determine whether the model including the thresholded GRM was able to estimate h^2_{ped} accurately, whilst the h^2_{pop} estimates were compared to h^2_{GRM} estimates to determine if the h^2_{GRM} estimates were biased due to the presence of related individuals.

RESULTS

Polygenic traits

For the neonate and lamb traits, the inclusion of a thresholded GRM in the model only significantly improved the fit of the model for lamb jaw length for both relationship-thresholded models (model 2, t = 0.05 and t = 0.1) (Fig. 1A, Supplementary Tables 2–4).

For lamb jaw length, the h^2_{kin} estimates were 0.084 and 0.083 for t = 0.05 and t = 0.1 respectively (S.E. 0.048 and 0.047 respectively) whilst h^2_{pop} estimates were 0.215 and 0.229 (S.E. 0.043 and 0.039), suggesting that family-associated variance explains 8% of phenotypic variance and 26.5–28.1% of the genetic variance underpinning lamb jaw length, depending on the threshold used. For this trait, the inclusion of the MAF-thresholded GRMs (model 3) led to non-convergence of the model. Estimates of h^2_{pk} ($h^2_{pop} + h^2_{kin}$) were higher than that of h^2_{GRM} for the two relatedness-thresholded models (h^2_{GRM}: 0.254 (S.E. 0.036), h^2_{pk} t = 0.05: 0.300, t = 0.1: 0.312), with the h^2_{pk} estimate for the higher threshold being the highest. The two h^2_{pk} estimates were similar to the estimate of h^2_{ped} (0.317 (S.E. 0.062)), with the h^2_{pk} estimate at t = 0.1 falling within the standard error around the estimate of h^2_{ped}. The standard errors around the estimates of h^2_{pop} overlapped with that of h^2_{GRM}, suggesting that the h^2_{GRM} estimate for this trait does not suffer from major biases from family-associated affects such as dominance and epistasis (Table 4).

Of the adult traits, the inclusion of a relatedness-thresholded GRM was significant for August weight, foreleg length and jaw length at t = 0.1 (Fig. 1B–D, Table 4, Supplementary Tables 2 and 4). For adult August weight, the estimate of h^2_{kin} was 0.098 (S.E. 0.047) and the estimate of h^2_{pop} was 0.167 (S.E. 0.048), suggesting that family-level genetic variance explains 9.8% of the phenotypic variation of this trait and made up 37.1% of the underlying genetic variance. For adult foreleg length, the estimate of h^2_{kin} was 0.112 (S.E. 0.041) and the estimate of h^2_{pop} was 0.194 (S.E. 0.045), suggesting that family-level genetic variance explains 11.2% of phenotypic variation and 36.6% of underlying genetic variance for adult foreleg length. For adult jaw length, the estimate of h^2_{kin} was 0.169 (S.E. 0.074) and the estimate of h^2_{pop} was 0.427 (S.E. 0.079), suggesting that family-level variance explained 16.9% of the phenotypic variation and 28.4% of the underlying genetic variance for this trait. Estimates of h^2_{pk} of all three traits fell within the standard errors around their respective h^2_{ped} estimates, and for adult August weight, the standard error around the estimate of h^2_{pop} overlapped with that of h^2_{GRM}. However, for adult foreleg length and adult jaw length the standard errors around the estimates of h^2_{pop} and h^2_{GRM} did not overlap, suggesting that there may be some effect of family-associated variance such as dominance or epistasis biasing the h^2_{GRM} estimates (Table 4).

The MAF-thresholded models (model 3) did not yield a significant change in the additive genetic variance explained for any trait (Supplementary Tables 3 and 4).

Monogenic traits

Inclusion of either the relatedness-thresholded GRM (Model 2, t = 0.05 and t = 0.1) or the MAF-thresholded GRM (Model 3) was not significant for any of the monogenic traits we investigated (Supplementary Tables 5–7). These results are somewhat surprising, as dominance is known to play a role in all four of these traits (Dolling 1961; Ryder et al. 1974; Kinsmann 2001; Coltman and Pemberton 2003; Gratten et al. 2007; Gratten et al. 2010; Johnston et al. 2013).

Re-running model 3 focusing on smaller SNP windows (focal chromosome and 1 Mb either side of the causal gene) did not improve the model fit for any of the MAF thresholds across any of the traits. For male and female horn type, the regional model was only run for MAF = 0.1 and MAF = 0.05 due to the fact that less than 10 SNPs remained when applying the more stringent thresholds. For coat colour and coat pattern, the regional model was not run for MAF = 0.005 and MAF = 0.001 for the same reasons, however the model at the remaining thresholds failed to converge.

Interestingly, for male horn type, coat colour, and coat pattern, estimates of $h^2_{GRM\_Chrom}$ and $h^2_{GRM\_Region}$ from the chromosome and regional models that did run were lower than estimates of h^2_{GRM} when the whole genome SNP data was used to construct the GRM. The standard error around the estimate of h^2_{GRM} did not overlap with those of $h^2_{GRM\_Chrom}$ and $h^2_{GRM\_Region}$, suggesting that these estimates are significantly different from each other. For female horn type, $h^2_{GRM\_Chrom}$ and $h^2_{GRM\_Region}$ estimates for the chromosomal and regional models that did run were slightly higher than h^2_{GRM}, though the standard errors around the estimates did overlap.

DISCUSSION

As far as we are aware, Zaitlen et al. (2013)'s method has been used to estimate heritability in populations containing related individuals in humans (Zaitlen et al. 2013), one livestock population (Kokkinias 2022) and two wild populations of birds (Silva et al. 2017). Population structure and sample size both differ between the different studies, as well as the proportion of h^2_{pk} attributed to h^2_{kin}. Zaitlen et al. (2013) focused on a sample of

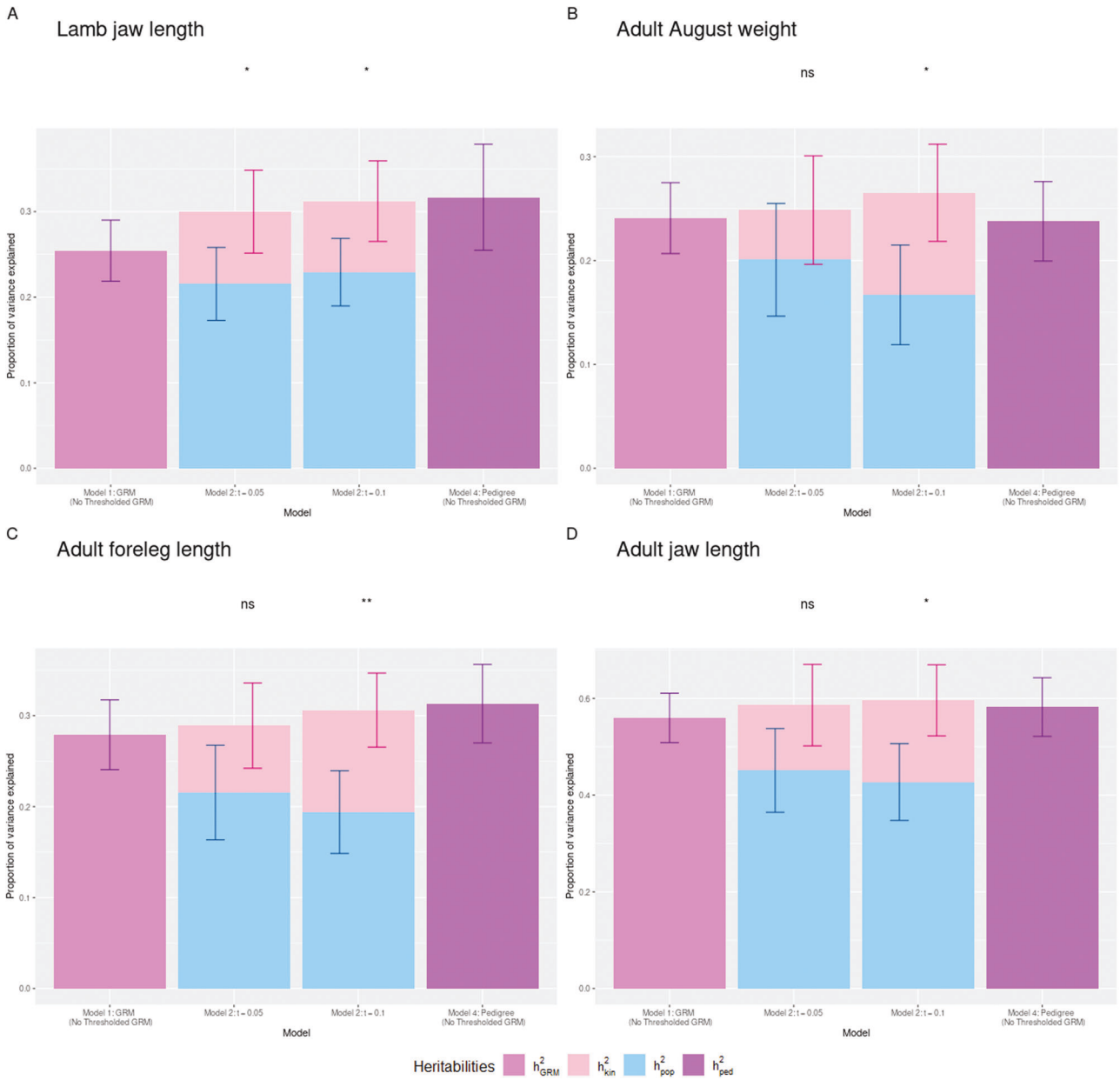

Fig. 1 Heritability estimates for traits for which inclusion of a thresholded GRM was significant. Estimates displayed are **A** lamb jaw length, **B** adult August weight, **C** adult foreleg length and **D** adult jaw length for models 1, 2 and 4. h is the h estimate from model 1 (no thresholded GRM fitted), h is the additive genetic variance explained by the full GRM (equivalent to h from model 1), h is the additive genetic variance explained by the thresholded GRM, h is the sum of h and h, and h is the h estimate when using the pedigree. For each trait, light purple represents the estimate of h for model 1, blue represents the estimate of h for model 2, pink represents the estimate of h for model 2, and purple represents the estimate of h for model 4. Stars represent the p value when performing loglikelihood ratio tests between models 1 and 2: ns means $p > 0.05$, *$0.01 < p < 0.05$, **$p < 0.01$.

~38,000 Icelandic individuals including sibling, half-sibling, parent-offspring, grandparent-grandchild and avuncular relationships, and found that h^2_{kin} made up 25–67% of h^2_{pk} estimates across their focal traits. Kokkinias (2022) focused on ~5000 individuals in a commercial salmon breeding program which contained four breeding lines, with most individuals in a breeding line being closely related (usually either full siblings or half siblings) – they found that h^2_{kin} made up 0–10% of h^2_{pk} estimates. Silva et al. (2017) focused on both a sample of 700–1400 Norwegian house sparrows and a sample of ~800 Swedish collared flycatchers, with the sample of house sparrows being more related to each other than the sample of collared flycatchers were. h^2_{kin} made up 19–100% of h^2_{pk} in the house sparrow data and 2–74% of h^2_{pk} estimates in the collared flycatcher data. The Soay sheep population is comprised of related individuals with very complex family structures – both males and females are promiscuous meaning full-siblings are much rarer than half-siblings, and the population is uniformly inbred. For the traits in which inclusion of a thresholded GRM improved model fit, we found that h^2_{kin} estimates made up 26.5–37.1% of h^2_{pk} estimates.

It is possible that the difference between these results is because of population relatedness structure; the salmon population contains no individuals without any relatives and the degree of relatedness between relatives was high, thus the GRM and the thresholded GRM were probably similar. In comparison, the human and bird data comprised a more complex set of

Table 4. The h^2 estimates and their standard errors for the traits in which the inclusion of the thresholded GRM significantly improved model fit for the comparison of model 1 (with no thresholded GRM) and model 2.

Trait	h^2_{GRM}	h^2_{pop}	h^2_{kin}	h^2_{pk}	P value	h^2_{ped}
Lamb jaw length t = 0.05	0.254 (0.036)	0.216 (0.043)	0.084 (0.048)	0.300	0.041	0.317 (0.062)
Lamb jaw length t = 0.1	0.254 (0.036)	0.229 (0.039)	0.083 (0.047)	0.312	0.038	0.317 (0.062)
Adult August weight t = 0.1	0.241 (0.034)	0.167 (0.048)	0.098 (0.047)	0.265	0.014	0.238 (0.038)
Adult foreleg length t = 0.1	0.279 (0.038)	0.194 (0.045)	0.112 (0.041)	0.306	0.004	0.313 (0.043)
Adult jaw length t = 0.1	0.441 (0.041)	0.380 (0.064)	0.075 (0.060)	0.455	0.011	0.445 (0.047)

h^2_{GRM} is the h^2 estimate from model 1 (no thresholded GRM fitted), h^2_{pop} is the additive genetic variance explained by the full GRM (equivalent to h^2_{GRM} from model 1), h^2_{kin} is the additive genetic variance explained by the thresholded GRM, h^2_{pk} is the sum of h^2_{pop} and h^2_{kin}, and h^2_{ped} is the h^2 estimate when using the pedigree. *P* value indicates the *p* value from the log-likelihood ratio test for the inclusion of the thresholded GRM.

relationships resulting in a bigger difference between the GRM and thresholded GRM. Xia et al. (2016) performed a similar analysis to Zaitlen et al. (2013) in a population of Scottish humans, but including environmental relationship matrices in the models alongside the thresholded GRM and performed stepwise model selection to determine which combination of matrices resulted in the best model fit for each of their traits. They first did this on a sample of 10 K individuals, primarily nuclear families and unrelated individuals, and then on a sample of 20 K individuals, made up of the original 10 K plus an additional 10 K individuals that were mainly related to the original 10 K individuals. Of the 16 traits studied, the relatedness-thresholded GRM was retained after model selection for 10 traits when using the 10 K data and for 14 when using the 20 K data. The authors suggest that the increase in sample size and inclusion of more distant relationships allowed the model to separate the effect of family-associated genetic variance (h^2_{kin}) from family-associated environmental variance for the four traits for which the relatedness-thresholded GRM was retained only when using the 20 K data.

We identified three traits that are potentially influenced by family-associated variance. For weight and foreleg length, this effect was only observed in adulthood, whilst for jaw length the effect was found in both lambs and adults.

As jaw length is a skeletal trait, it is measured *post-mortem* and thus there is no overlap between individuals who have a recorded lamb jaw length and those who have a recorded adult jaw length. This suggests that the effect of family-level genetic variance on jaw length is not dependent on any age-related mortality factors such as juvenile survival and is potentially consistent throughout life, affecting both juvenile jaw length and total potential (adult) jaw length.

Weight and foreleg length are live measures, meaning there is an overlap between individuals measured as adults and individuals measured as lambs (and in the case of weight, individuals measured as neonates). In addition, individuals can be caught across multiple years as adults, meaning that adult August weight and adult foreleg length might have repeated measures. There are therefore two main potential reasons as to why our models showed an effect of family-level genetic variance in adulthood but not in juveniles or neonates. Firstly, h^2 estimates for birth weight, lamb August weight and lamb foreleg length are low ($h^2 < 0.15$), so these traits are likely primarily controlled by environmental factors (including maternal effects). Thus, any family-level genetic variation affecting these traits may also be small. Secondly, it is possible that any family-level genetic variation affecting weight and foreleg length only affect the total potential (adult) weight and foreleg length, rather than juvenile weight and foreleg length. This means the effect would only be picked up among individuals who have reached their potential for these traits (i.e. individuals who have stopped growing – adults).

The standard errors around the h^2_{pop} estimates for adult foreleg length and adult jaw length at t = 0.01 did not overlap with those for their respective h^2_{GRM} estimates, suggesting that any family-associated variance that is influencing these traits is significantly impacting estimates of h^2_{GRM} when related individuals are included in the analysis. It is interesting that, despite there being a genetic correlation of 0.944 (S.E. 0.019) between adult foreleg length and adult hindleg length (Bérénos et al. 2014), inclusion of a thresholded GRM at t = 0.1 is significant for adult foreleg length and not adult hindleg length. However, in a recent genome-wide association study of this population, we found two peaks on chromosomes 7 and 9 that were significantly associated with adult foreleg length but not with adult hindleg length (James et al. 2022). It is possible that family-associated variation such as epistasis and dominance are acting upon these two regions, resulting in the inclusion of the thresholded GRM being significant for adult foreleg length but not adult hindleg length. Further investigation, with larger sample sizes that may become available in the future, would shed more light on the commonalities and differences in the determinants of these traits.

Inclusion of either the relatedness-thresholded GRM models (model 2) or the MAF-thresholded models (model 3) did not improve model fit for any of the monogenic traits we investigated. From our results, this model is not well suited for detecting any dominance in monogenic traits. It is possible that this is because the minor allele frequencies of the causal variants are too high to influence the thresholded GRMs – especially the MAF-thresholded GRMs, given that the minor allele frequencies are all higher than our most lenient threshold (Supplementary Table 1). Interestingly, for three of the traits (male horn type, coat colour and coat pattern), estimates of $h^2_{GRM\_Chrom}$ and $h^2_{GRM\_Region}$ were lower than that of h^2_{GRM}, suggesting that there may be some additive genetic variance influencing these traits located elsewhere in the genome, and that these traits may not be truly monogenic. Further investigation using methods such as regional genomic relationship mapping (Nagamine et al. 2012) or haplotype heritability mapping (Shirali et al. 2018) may therefore provide new insights to the genetic architecture underpinning these traits.

The MAF-thresholded model was designed to identify whether our focal traits were influenced by rare alleles in the population. As the genetic architecture of some of our focal traits has not yet been fully characterised, it is unknown what proportion of the genetic variance is being contributed by SNPs with high or low MAFs. Using the MAF-thresholded model, we would expect to see

some of the genetic variance being attributed to the MAF-thresholded GRM if variance was being attributed to rare alleles. Given that the MAF-thresholded model was not significant for any of our focal traits, it is possible that none of these traits are influenced by variants with a low enough MAF to be detected by this model.

Overall, we have demonstrated that the method proposed by Zaitlen et al. (2013) can be useful for estimating heritability in wild population samples containing relatives, as well as showing how the method is affected by different genetic architectures of traits (monogenic versus polygenic). However, we found that the model including the thresholded GRMs was not a significantly better fit for the Soay sheep data when estimating heritability for most of our traits. On the other hand, there are indications that some traits previously thought to be monogenic in the Soay sheep may be influenced by genetic variation outwith the causal gene.

DATA AVAILABILITY

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

REFERENCES

Ashraf B, Hunter DC, Bérénos C, Ellis PA, Johnston SE, Pilkington JG et al. (2021) Genomic prediction in the wild: A case study in Soay sheep. Mol Ecol 31(24):6541–6555

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. Evolution 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. Mol Ecol 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. Mol Ecol 23(14):3434–3451

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. Nat Commun 6(1):10162

Clutton-Brock TH, Pemberton JM (2003). Soay Sheep: Dynamics and Selection in an Island Population. Cambridge, Cambridge University Press

Coltman DW, Pemberton JM (2003) Appendix 2 - Inheritance of coat colour and horn type in Hirta Soay sheep. Soay Sheep: Dynamics and Selection in an Island Population. Cambridge, Cambridge University Press. p. 321–327

Dolling C (1961) Hornedness and polledness in sheep. 4. Triple alleles affecting horn growth in the Merino. Aust J Agric Res 12(2):353–361

Falconer DS, Mackay TF (1996) Introduction to Quantitative Genetics. Longman, Harlow

Gratten J, Beraldi D, Lowder BV, McRae AF, Visscher PM, Pemberton JM et al. (2007) Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism in a free-living population of Soay sheep. Proc Biol Sci 274(1610):619–626

Gratten J, Pilkington JG, Brown EA, Beraldi D, Pemberton JM, Slate J (2010) The genetic basis of recessive self-colour pattern in a wild sheep population. Heredity 104(2):206–214

Henderson CR (1975) Use of Relationships Among Sires to Increase Accuracy of Sire Evaluation. J Dairy Sci 58(11):1731–1738

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. Genet Sel Evolut GSE 44(1):9

Hill WD, Arslan RC, Xia C, Luciano M, Amador C, Navarro P et al. (2018) Genomic analysis of family data reveals additional genetic effects on intelligence and personality. Mol Psychiatry 23(12):2347–2362

Huisman J (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Mol Ecol Resour 17(5):1009–1024

Hunter DC, Ashraf B, Bérénos C, Ellis PA, Johnston SE, Wilson AJ et al. (2022) Using genomic prediction to detect microevolutionary change of a quantitative trait. Proc Biol Sci 289(1974):20220330

James C, Pemberton JM, Navarro P, Knott S (2022) The impact of SNP density on quantitative genetic analyses of body size traits in a wild population of Soay sheep. Ecol Evol 12(12):e9639

Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM et al. (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature 502(7469):93–95

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42(4):348–354

Kinsmann DJJ (2001) Black sheep of Windermere: a history of the St Kilda or Hebredian sheep. Windy Hall Publications, Cumbria

Kokkinias P (2022) Optimising Genomic Breeding of Farmed Salmon. PhD Thesis, University of Edinburgh, Edinburgh

Mackay TFC (2001) The Genetic Architecture of Quantitative Traits. Annu Rev Genet 35(1):303–339

Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genet Sel Evol 21(3):317

Mrode RA (2014) Linear models for the prediction of animal breeding values. CABI Publishing, Wallingford

Nagamine Y, Pong-Wong R, Navarro P, Vitart V, Hayward C, Rudan I et al. (2012) Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. PloS One 7(10):e46501

Perrier C, Delahaie B, Charmantier A (2018) Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. Mol Ecol Resour 18(4):838–853

Ryder ML, Land RB, Ditchburn R (1974) Coat colour inheritance in Soay, Orkney and Shetland sheep. J Zool 173(4):477–485

Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS (2018) Haplotype Heritability Mapping Method Uncovers Missing Heritability of Complex Traits. Sci Rep 8(1):4982

Silva CNS, McFarlane SE, Hagen IJ, Rönnegård L, Billing AM, Kvalnes T et al. (2017) Insights into the genetic architecture of morphological traits in two passerine bird species. Heredity 119(3):197–205

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM (2021) Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nat Commun 12(1):2972

Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D et al. (2016) Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation. PLoS Genet 12(2):e1005804

Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47(10):1114–1120

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42(7):565–569

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 88(1):76–82

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S et al. (2013) Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. PLOS Genet 9(5):e1003520

ACKNOWLEDGEMENTS

We thank the National Trust for Scotland for permission to work on St Kilda and QinetiQ, Eurest and Kilda Cruises 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.

AUTHOR CONTRIBUTIONS

CJ conducted analyses and drafted the manuscript. JMP, PN and SK helped with analyses and interpretations of results. All authors contributed to revisions.

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. Most of the SNP genotyping was funded by a European Research Council Advanced Grant to JMP. PN is funded by BBSRC grant BBS/E/RL/230001A and acknowledges support from the MRC Human Genetics Unit program grant, "Quantitative traits in health and disease" (U. MC_UU_00007/10), and grant MC_PC_U127592696.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-024-00673-6.

Correspondence and requests for materials should be addressed to Caelinn James.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.