THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Accuracy of genomic prediction within and across populations for nematode resistance and body weight traits in sheep

Citation for published version:
Riggio, V, Abdel-Aziz, M, Matika, O, Moreno, CR, Carta, A & Bishop, SC 2014, 'Accuracy of genomic prediction within and across populations for nematode resistance and body weight traits in sheep', *Animal*, vol. 8, no. 4, pp. 520-528. https://doi.org/10.1017/S1751731114000081

Digital Object Identifier (DOI):
10.1017/S1751731114000081

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Animal

Publisher Rights Statement:
This is a PDF file of an unedited manuscript that has been accepted for publication. The publisher version is available at: http://dx.doi.org/10.1017/S1751731114000081

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 27. Sep. 2023

Please cite this article as: Riggio, V, Abdel-Aziz, M, Matika, O, Moreno, CR, Carta, A & Bishop, SC 2014, 'Accuracy of genomic prediction within and across populations for nematode resistance and body weight traits in sheep' , *Animal* (2014), http://dx.doi.org/10.1017/S1751731114000081

1 # Accuracy of genomic prediction within and across populations for nematode
2 resistance and body weight traits in sheep

3

4 V. Riggio [1,a], M. Abdel-Aziz [2,a], O. Matika [1], C.R. Moreno [3], A. Carta [4], and S.C.
5 Bishop [1]

6

7 *[1] The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, Midlothian*
8 *EH25 9RG, Scotland, UK*
9 *[2] Department of Animal and Fish Production, College of Agriculture and Food*
10 *Sciences, King Faisal University, Al-Ahsa, 31982, Saudi Arabia*
11 *[3] INRA, UR631, Station d'Amélioration Génétique des Animaux, BP 27, F-31326,*
12 *Castanet-Tolosan, France*
13 *[4] Settore Genetica e Biotecnologie, AGRIS Sardegna, Olmedo, Sassari 07040, Italy*

14

15 *[a] Equal contributors*

16

17 Corresponding author: Valentina Riggio. Email: valentina.riggio@roslin.ed.ac.uk

18

19 Short title: Genomic predictions for sheep nematodes and weight

20

21 ## Abstract

22 Genomic prediction utilizes SNP chip data to predict animal genetic merit. It has the
23 advantage of potentially capturing the effects of the majority of loci that contribute to
24 genetic variation in a trait, even when the effects of the individual loci are very small.
25 To implement genomic prediction, marker effects are estimated with a training set

including individuals with marker genotypes and trait phenotypes; subsequently genomic estimated breeding values (GEBV) for any genotyped individual in the population can be calculated using the estimated marker effects. In this study we aimed to: i) evaluate the potential of genomic prediction to predict GEBV for nematode resistance traits and body weight in sheep, within and across populations; ii) evaluate the accuracy of these predictions through within-population cross-validation; and iii) explore the impact of population structure on the accuracy of prediction. Four datasets comprising 752 lambs from a Scottish Blackface population, 2,371 from a Sarda x Lacaune backcross population, 1,000 from a Martinik Black-Belly x Romane backcross population, and 64 from a British Texel population were used in this study. Traits available for the analysis were faecal egg count for *Nematodirus* and *Strongyles* and body weight at different ages or as average effect, depending on the population. Moreover, immunoglobulin A was also available for the Scottish Blackface population. Results show that GEBV had moderate to good within-population predictive accuracy, whereas across-population predictions had accuracies close to zero. This can be explained by our finding that in most cases the accuracy estimates were mostly due to additive genetic relatedness between animals, rather than linkage disequilibrium (LD) between SNP and QTL. Our results, therefore, suggest that genomic prediction for nematode resistance and body weight may be of value in closely related animals, but that with the current SNP chip genomic predictions are unlikely to work across breeds.

Keywords: genomic prediction, population structure, nematode resistance, body weight, sheep

51 **Implications**

52 Genomic prediction utilizes SNP chip data to predict animal genetic merit. Using data
53 from several populations, our results suggest that genomic prediction may be of
54 value for nematode resistance and body weight in closely related animals, but with
55 current technologies it is unlikely to work across populations. Genetic relatedness
56 between animals and population structure affect these estimates and need to be
57 taken into consideration before considering implementation.

58

59 **Introduction**

60 Traditional genetic improvement has relied on the use of phenotypes together with
61 the knowledge of the pedigree of each animal to estimate its breeding value. This
62 has led to genetic gains in most farmed species; especially with 'easy-to-measure'
63 production traits. However, the efficiency decreases when traits are difficult to
64 measure, have a low heritability, or cannot be quickly, inexpensively and correctly
65 measured. An example is nematode resistance, assessed using indicator traits such
66 as faecal egg count (FEC), which is critically important for the sheep industry.

67 To overcome this issue, there has long been an interest in using simply inherited
68 genetic markers to increase the rate of genetic gain (Dekkers and Hospital, 2002).
69 However, for many quantitative traits, such as production and health traits, a large
70 number of loci appear to affect the trait, with each of them individually explaining only
71 a limited proportion of the total genetic variance (Hayes and Goddard, 2001, Sanna
72 *et al.*, 2008, Kemper *et al.*, 2011). Genomic selection (GS) has the advantage of
73 potentially capturing the effects of the majority of loci that contribute to genetic
74 variation, even when the effects of the individual loci are very small (Hayes *et al.*,
75 2009a). With GS, first marker effects are estimated with a training set (TS) which

includes individuals with marker genotypes and trait phenotypes; genomic estimated breeding values (GEBV) of any genotyped individual in the population can then be calculated using the estimated marker effects (Habier *et al.*, 2007). The resulting GEBV, therefore, exploit associations between markers and QTL through linkage disequilibrium (LD) and linkage, along with the capture of pedigree relationships between animals (Habier *et al.*, 2007).

Accessing sufficient animals to both train and validate GEBV remains challenging in practice, and cross-validation with individuals from the same population is often used to assess the accuracy of the GEBV (Habier *et al.*, 2007). However, validation studies can be also performed using separate phenotyped and genotyped populations (Hayes *et al.*, 2009a, Luan *et al.*, 2009, Su *et al.*, 2010), with an accuracy which depends on the genetic relationship of the validation set to the TS (Habier *et al.*, 2007, Habier *et al.*, 2010). This is possible because markers used in the statistical models to estimate marker effects also capture additive genetic relationships between individuals (Cockerham, 1969, Ritland, 1996), therefore, even if markers are not in LD with QTL, the accuracy of GEBV will still be non-zero. However, animals more closely related to those included in the TS are expected to obtain more reliable predictions (Habier *et al.*, 2007, Legarra *et al.*, 2008, Sonesson and Meuwissen, 2009).

At present, the accuracy of GEBV has been evaluated in experiments involving several livestock species, such as dairy (Harris *et al.*, 2008, Hayes *et al.*, 2009b) and beef (Saatchi *et al.*, 2011) cattle populations, chicken (González-Recio *et al.*, 2009), and sheep (Daetwyler *et al.*, 2010b, Daetwyler *et al.*, 2012a, Daetwyler *et al.*, 2012b, Duchemin *et al.*, 2012). Apart from the study of Kemper *et al.* (2011), the use of high density genomic information to select for nematode resistance in sheep has received

101 less attention. Therefore, the aims of this study were to: i) evaluate the potential of
102 GS to predict GEBV for nematode resistance traits, as well as body weight, both
103 within and across populations; ii) evaluate the accuracy of these predictions through
104 within-population cross-validation; and iii) explore the impact of population structure
105 within population, by decomposing the accuracy of genomic prediction into
106 component parts.

107

108 Material and methods

109 Four datasets comprising 752 lambs from a Scottish Blackface (SBF) population,
110 2,371 ewes from a Sarda x Lacaune (SAR) backcross population, 1,000 lambs from
111 a Martinik Black-Belly x Romane (MBR) backcross population, and 64 lambs from a
112 British Texel (BT) population were used in this study. As shown in the principal
113 components plot of the SNP chip markers reported in Supplementary Figure S1, the
114 four populations are genetically distant. Genomic predictions were conducted firstly
115 within population, using the SBF data. This was because of the availability of both
116 pedigree and SNP marker data, along with several traits, allowing us to potentially
117 explore a variety of trait architectures as well as contributions of LD and linkage to
118 genomic predictions. Secondly, an evaluation of across-population prediction was
119 conducted using all four populations, albeit with limited phenotypes common across
120 datasets.

121 *Phenotype data*

122 *SBF data:* The SBF lambs were bred over a period of three years (2001-2003), with
123 traits measured including lamb weights (16 and 24 weeks, and average animal effect
124 from a repeatability model excluding pedigree) and faecal egg counts (FEC) for
125 *Nematodirus* and *Strongyles* collected at 16, 20 and 24 weeks of age, and their

126 average animal effects as well as plasma IgA (on 737 out of the 752 lambs). The
127 population comprised F2 and double backcross lambs from two originally different
128 lines, bred from 10 sires (half-sib family size = 11-146). More details on the data
129 structure and the phenotypes are given in Riggio *et al.* (2013). Fecal samples were
130 collected from the rectum of each lamb at the time of weighing and used for FEC
131 assays, using the modified McMaster technique as described by Gordon and
132 Whitlock (1939) and Bairden (1991). The activity of plasma IgA against a somatic
133 extract of third-stage larvae from *Teladorsagia* was measured by indirect ELISA, as
134 described by Strain *et al.* (2002), using blood samples collected at 24 weeks of age.
135 The relative IgA activity was calculated according to the formula suggested by Sinski
136 *et al.* (1995). The average animal effects were estimated by fitting a repeatability
137 model to trait values across the different time points, and then standardized to a
138 mean of 0 and a standard deviation of 1. FEC and IgA measurements were all right-
139 skewed. Therefore, prior to analysis, FEC measurements were log-transformed by
140 *ln*(FEC+x), where x is a constant used to avoid the zero values, whereas IgA
141 measurements were cube-root transformed.

142 *Other populations:* Phenotypes available on BT lambs were for FEC at 20 weeks for
143 *Strongyles* and *Nematodirus*, and body weight at 24 weeks. A detailed description of
144 the data was given in Matika *et al.* (2011). The phenotype available for the two
145 remaining populations (SAR and MBR) was the “average animal effect” for
146 *Strongyles* FEC. A detail description of the animals in the MBR population was given
147 in Sallé *et al.* (2012), and for the SAR population in Sechi *et al.* (2009).

148 *Genotype data*

149 All animals from the four populations were genotyped using the 50k SNP chip. The
150 SNP genotypes data were subjected to quality control (QC) measures, specific for

151 each population (see Supplementary Material S1). After QC, 42,841 SNPs were
152 available for the SBF and BT populations, 44,859 for the SAR, and 42,469 for the
153 MBR. Out of these SNPs, 38,991 were in common among the four populations and
154 therefore used for further analyses.

155 *Assessment of GEBV predictive value*

156 *SBF data:* For the analysis within population, validation sets were obtained by
157 masking the phenotype (i.e., setting the phenotype as "unknown") for a defined
158 number of individuals from the TS. The individuals whose phenotype was masked
159 were selected in two different ways. The first way was through random selection: five
160 non-overlapping cross-validation sets were created by randomly selecting 150 (152
161 for the fifth subset) lambs at a time, masking each phenotype only once. The second
162 way was to select individuals belonging to specific families, to test the extent to which
163 results differed depending on how related families were to the remaining families
164 forming the TS.

165 Data were first analysed without fitting any polygenic or genomic effect, to correct for
166 fixed effects. The following model was fitted:

167 $$y_{ijlmn} = \mu + S_i + K_j + L_l + G_m + A_n + \beta DB + e_{ijlmn}$$

168 where, y_{ijlmn} is the phenotype of the nth individual, S_i is the effect of the sex (male and
169 female), K_j is the effect of the year of birth (2001 to 2003), L_l is the effect of the litter
170 size (single or multiple), G_m is the effect of management group (two levels,
171 corresponding to those born in the first 2 weeks of the lambing season and those
172 born subsequently), A_n is the effect of age of dam (1 to 4 years), DB is a covariate
173 effect of day of birth and β its regression coefficient, and e_{ijlmn} is the residual error.

174 The resulting adjusted phenotypes or residuals (y*) were then analysed using the
175 ASReml package (Gilmour *et al.*, 2009), fitting the model:

176 $$y^* = \mu + Zg + e,$$

177 where y* is a vector of the adjusted phenotypic records, **Z** is a design matrix, g is a
178 vector of random additive genomic effects distributed as $N(0,\sigma_g^2\mathbf{G})$, σ_g^2 is the additive
179 genetic variance, **G** is the genomic relationship matrix, and e is the vector of
180 residuals. The **G** matrix was constructed using the method of VanRaden (2008). The
181 genetic variance/covariance matrix and GEBV (i.e., $\hat{g}$) of the SBF lambs in the TS
182 were estimated by utilizing both phenotype and genotype information. The predicted
183 genomic breeding values (PGEBV), i.e. GEBV calculated without phenotypic
184 information on the individual, were estimated fitting the model described above but
185 masking the phenotypes of each subset in turn. Thus, in addition to its GEBV, after
186 analysing each randomisation, every individual had a PGEBV obtained from marker
187 data alone from random masking of phenotypes, with a similarly obtained PGEBV
188 following masking of families.

189 *Across populations:* Two combined datasets were used for across population
190 predictions, with SBF, SAR and MBR making the first set (4,123 individuals) and SBF
191 and BT making the other (816 lambs). In the former data, two populations were used
192 as TS to predict the third one (i.e., SAR and MBR to predict SBF; SBF and SAR to
193 predict MBR; and SBF and MBR to predict SAR). Moreover, to test for the impact of
194 cross-family links on GEBV, two analyses were conducted in which a few half-sib
195 family members were allocated to the TS and used as a connection with the rest of
196 the half-sib family members in the validation set. In these analyses, either one or 10

197 lambs from each half-sib family from the SBF data were randomly chosen to be in the
198 TS.

199 *Accuracy and predictive values of PGEBV*

200 Genomic prediction accuracies were calculated for each validation set (both within
201 and across populations). Firstly, the Pearson correlations of PGEBV with the
202 adjusted phenotypes ($r_{\hat{g}\hat{y}}$) were calculated and the accuracy ($r_{\hat{g}g}$) for each validation
203 set was estimated by dividing $r_{\hat{g}\hat{y}}$ by the the square root of the heritability of each trait
204 for that specific validation set:

205 $Accuracy = \frac{r_{\hat{g}\hat{y}}}{\sqrt{h_y^2}}$ (Legarra *et al.*, 2008).

206 The accuracy for each trait was then obtained by averaging the estimates across
207 validation groups.

208 The sampling properties of the prediction accuracies were explored by repeating the
209 overall within-SBF cross-validation analysis, described above, 10 times and
210 calculating the accuracy separately for each replicate. For each replicate, a new
211 randomisation was performed so that the individuals comprising each of the groups
212 were different. The standard error of the accuracy was then estimated as the
213 empirical standard deviation of the 10 accuracy values. This exercise was performed
214 for the average animal effect for *Strongyles* FEC, as an example trait.

215 Two further sets of analyses were performed using SBF data, alone. Firstly, we
216 calculated the correlation between GEBV and PGEBV. This case represents a
217 situation where progeny's performance is predicted from markers before the
218 availability of phenotypes. Secondly, the cross validation prediction accuracy analysis

219 was also performed using pedigree-based EBVs, rather than genomic EBVs. This
220 addresses the question of how, in this population, the accuracy of genomic
221 predictions compares to the accuracy of pedigree-based predictions.

222 *Exploring contribution of population structure in the Scottish Blackface data*

223 To explore the contribution of population structure to the accuracies of the genomic
224 predictions, several analyses were performed. Firstly, to determine the effectiveness
225 of the **G** matrix in capturing additive genetic effects relative to the **A** matrix, we
226 analysed the SBF data fitting both the **G** matrix and the pedigree-based numerator
227 relationship matrix **A** using the following model:

228 $$y^* = \mu + Zv + Zg + e,$$

229 where the effects are as defined above, with *v* being an additional vector of additive
230 polygenic effects normally distributed as $N(0,\mathbf{A}\sigma_a^2)$, with **A** being the numerator
231 relationship matrix.

232 Secondly, the contribution of population and genome structure to genomic prediction
233 accuracies of the SBF population was assessed by fitting chromosome-specific **G**
234 matrices. Following the methodology of Daetwyler *et al.* (2012a), 26 chromosome
235 specific **G** matrices were calculated, using only the SNPs on each chromosome.
236 Each chromosome was then fitted instead of the overall **G** matrix. To measure the
237 proportion of the total genetic variance explained by each chromosome, we also
238 carried out an analysis fitting each chromosome and the **G** matrix consisting of all
239 SNPs minus those in that specific chromosome (which corresponds to fitting all
240 chromosomes simultaneously). The following model was then fitted:

241 $$y^* = \mu + Zg_{chr} + Zg_{rest} + e,$$

242 where g_{ch} and g_{rest} are the vectors of additive genomic effects unique to the
243 chromosome under investigation and to all remaining chromosomes, respectively.
244 The terms g_{ch}, g_{rest} and e were assumed to be normally distributed: N(0,$\mathbf{G_{ch}}\sigma_{gch}^2$) and
245 N(0,$\mathbf{G_{rest}}\sigma_{grest}^2$), respectively. Here, $\mathbf{G_{ch}}$ is the genomic matrix for one chromosome
246 and $\mathbf{G_{rest}}$ is the genomic matrix estimated from the rest of the genome excluding the
247 unique fitted chromosome markers.

248 Insight into the components contributing to the accuracy can be gained by regressing
249 the difference in phenotypic variance explained by individually vs. simultaneously
250 fitted chromosomal **G** matrices on chromosome length (Yang *et al.*, 2011, Daetwyler
251 *et al.*, 2012a). This was given by this equation:

252 $$\sigma^2_{c(sep)} - \sigma^2_c = b_0 + b_1 L_c + e$$

253 where $\sigma^2_{c(sep)}$ is variance explained by each chromosome analysed individually and
254 σ^2_c the variance when the chromosome are analysed jointly, with b_0 being the
255 intercept which represents the component due to relatedness amongst animals
256 rather than tagged QTL, and b_1 the slope that relates genetic variance to
257 chromosome length (Lc), i.e. tagged QTL. We calculated the proportion of the
258 genetic variance explained by the population structure (i.e. additive genetic
259 relatedness as opposed to QTL tagged by the SNP chip) by dividing b_{0d} (intercept of
260 the difference) with the intercept from regressing the variance explained by
261 individually fitted chromosomes on chromosome length (b_{0i}).
262

263 **Results**

264 *Accuracy and predictive values of PGEBV*

265 *SBF data:* Correlations between PGEBV and adjusted phenotypes, with
266 corresponding accuracies for each trait, for the cross-validation groups in the SBF
267 population are reported in Table 1, together with the accuracies estimated using
268 pedigree-based EBV. Correlations varied between groups, ranging from marginally
269 negative (-0.027 in group 1 for *Nematodirus* FEC at 16 weeks) to positive and
270 moderate (0.382 in group 5 for IgA). Moderate accuracies $\left(r_{\hat{g}g}\right)$ were observed,
271 generally between 0.42 and 0.68, with the exception of the accuracy for *Nematodirus*
272 FEC at 16 weeks (0.10), this being the trait with the lowest heritability. Accuracies
273 using pedigree-based EBV ranged from 0.27 to 0.52, and were slightly lower than the
274 genomic EBV accuracies for 9 of the 12 traits. The empirical standard error of the
275 accuracy for *Strongyles* FEC average animal effect, estimated as the standard
276 deviation of the accuracies across the 10 replicated cross validation, was 0.04.
277 Correlations between GEBV and PGEBV (Table 2), representing the relationship
278 between genomic EBVs predicted with and without individual data were all strong
279 and positive. The average value across all traits was 0.76.

280 Lower correlation estimates between phenotype and PGEBV were obtained when all
281 members in one sire family were predicted from the remaining sire families in the
282 SBF data (Table 3). However, differences were observed in relationship connectivity
283 between families. For example, nematode resistance indicator trait results (i.e., both
284 IgA and FEC) showed that the families which were more closely related to the
285 remaining families in the TS were those with more accurate PGEBV. In particular, the
286 half-sib family sired by ram 22 (i.e., Fam22), which is the most highly related to the
287 remaining TS families (data not shown) showed the highest correlations. However,
288 different results were found for body weight, suggesting that not only relatedness is

289 important but other factors (such as trait heritability or markers in LD with mutations
290 affecting the trait) may play a part.

291 *Across populations:* The correlations between PGEBV and adjusted phenotype for
292 the *Strongyles* average animal effect were -0.054, -0.030 and 0.005 for SBF vs.
293 (MBR plus SAR), MBR vs. (SBF plus SAR) and SAR vs. (SBF plus MBR) datasets,
294 respectively. The correlations between PGEBV and adjusted phenotypes for the BT
295 data vs. SBF were -0.012, -0.010 and 0.067 for *Strongyles* and *Nematodirus* FEC at
296 20 weeks and for body weight at 24 weeks, respectively. In both analyses, the
297 predictions for genetically distant groups were usually close to zero. However, when
298 one or 10 lambs from each sire family from the SBF data were randomly chosen and
299 included in the TS, the correlations between PGEBV and y* were slightly higher, and
300 always positive with 0.129 and 0.070 for SBF vs. (MBR plus SAR plus 10SBF) and
301 SBF vs. (MBR plus SAR plus 100SBF), respectively.

302 *Exploring contribution of population and genome structure*

303 The results of the analysis in the SBF data, fitting either the **A** or **G** matrix alone, or
304 both together, are reported in Supplementary Table S1. For some traits the
305 heritability estimates were either completely explained by the **G** matrix (i.e., IgA and
306 *Nematodirus* FEC at 20 weeks) or the **A** matrix (*Strongyles* FEC at 20 weeks and
307 *Nematodirus* FEC at 16 weeks) when the analysis was done fitting both **G** and **A**
308 matrices. However, for the other FEC traits (both *Strongyles* and *Nematodirus*) there
309 was a contribution from both matrices. In general there was little discernible pattern
310 in these results. Moreover, the relative partitioning of genetic variation between the **A**
311 and **G** matrices may be expected to vary as the number and size of families varies,
312 thus it is difficult to draw general conclusions from these results.

313 For the SBF population, heritability estimates were also obtained either fitting only
314 one chromosome or when simultaneously fitting one chromosome plus the whole **G**
315 matrix (results not shown). Although similar trends were observed, the proportions of
316 genetic variation accounted for when fitting only one chromosome were always
317 overestimated. However, in both cases it is possible to identify the chromosomes that
318 explain most of the genetic variation of the traits.

319 We tested the hypothesis that fitting all $\mathbf{G}_{\mathbf{ch}}$ (i.e., chromosome-wide genomic
320 matrices) simultaneously would result in each chromosome explaining a fraction of
321 the total genetic variance proportional to its length, consistent with the polygenic
322 assumptions underlying GBLUP. Whilst there was a weak tendency for this to be the
323 case for most traits (as an example, Figure 1), the majority of the captured genetic
324 variation appeared to be independent of chromosome length. This can be seen in
325 Table 4 which reports intercept, slope, and R^2 for the three regressions (i.e., by fitting
326 each chromosome individually, by fitting all chromosomes simultaneously, and the
327 difference between the two) as well as the proportion of genetic variance explained
328 by relatedness for all traits considered. These proportions (ranging from 0.39 to 0.98,
329 with an average of 0.77) suggest that in most cases our accuracy estimates are
330 mostly due to additive genetic relatedness, rather than LD between SNP and QTL.
331 The **A**-matrix-derived heritabilities were compared to accuracies and proportion of
332 genetic variance explained by relatedness (b_{0d}/b_{0i}) for all nematode resistance
333 indicator traits (results not shown). Amongst the *Strongyles* FEC and IgA results
334 there was little discernible relationship between these variables. The *Nematodirus*
335 traits were more variable, however they tended to have lower heritabilities and
336 relatively large genetic effects (i.e. QTL) had previously been observed on some of

337 the smaller chromosomes (see Discussion) suggesting that the polygenic inheritance
338 assumption was inappropriate for the *Nematodirus* traits.

339

340 **Discussion**

341 One of the objectives of the current study was to understand the dynamics of
342 applying genomic selection to hard-to-measure traits using field data. We assumed
343 two scenarios, with the first scenario having young animals selected from markers
344 before their phenotypes can be measured and secondly, where we break the
345 assumption that the animals of the TS and the validation sets are from the same
346 population i.e., we explore situations where the animals vary from being closely
347 related to unrelated. Therefore, we explored the possibility of using genomic
348 predictions within and across populations; whilst prediction accuracies within a
349 population were good, with a small empirical standard error, our results highlighted
350 the difficulties of prediction using genetically distant individuals.

351 We also reported prediction accuracies estimated by using both the **G** and the **A**
352 relationship matrix. The accuracies estimated with the **G** matrix were usually higher
353 that those with the **A** matrix, suggesting an advantage in using genomic information
354 for predictions, even when pedigree knowledge is available. The one case where the
355 accuracies estimated with the **A** matrix was substantially better, *viz*. *Nematodirus*
356 FEC at 16 weeks, was for a trait for which heritability estimate was mostly explained
357 by the **A** matrix (Supplementary Table S1).

358 Although several studies on GEBV accuracy/reliability estimated from real data have
359 been reported in the literature for cattle with GEBV reliabilities ranging from 18 to
360 78% (Harris *et al.*, 2008, Hayes *et al.*, 2009b, VanRaden *et al.*, 2009), fewer are
361 reported for sheep. Our GEBV accuracies are similar to others obtained using a

362 medium-density markers chip of 15 to 79% for wool traits in Merino sheep (Daetwyler
363 *et al.*, 2010b), and 7 to 31% for carcass and meat quality traits in multi-breed sheep
364 data (Daetwyler *et al.*, 2012b). In a study on the Lacaune dairy sheep breed using
365 different genomic methods, Duchemin *et al.* (2012) reported accuracies varying from
366 0.4 to 0.6, according to the traits (i.e. milk yield, fat content, and somatic cell scores),
367 with minor differences among genomic approaches. These authors also showed that
368 the inclusion of molecular information, as compared with traditional schemes,
369 increased accuracies of EBV of young males at birth from 18 up to 25%, according to
370 the trait (Duchemin *et al.*, 2012). However, it has to be considered that the accuracy
371 of the GEBV depends on the size of the population and on the heritability of the trait.
372 For low heritability traits, a very large number of records will be required in the TS to
373 subsequently achieve high accuracies of GEBV in unphenotyped animals. If we
374 consider our SBF population, where the effective population size (Ne) is ~500 (Kijas
375 *et al.*, 2012), then according to the formula suggested by Daetwyler *et al.* (2010a) to
376 achieve an accuracy of 0.6, we would need ~ 30,000 individuals for a trait with very
377 low heritability (e.g., *Nematodirus* FEC at 16 weeks), and ~ 5,000 for a trait with
378 moderate heritability (e.g., IgA).

379 The current study explored the contributions of LD and relatedness to the accuracies
380 of genomic predictions. The heritability estimates obtained either fitting only one
381 chromosome or when simultaneously fitting one chromosome plus the whole **G**
382 matrix showed that nematode resistance in sheep is a complex trait with
383 contributions from many regions in the genome affecting these traits. However, with
384 the exception of *Nematodirus* FEC at 16 weeks (Supplementary Figure S2; Riggio *et*
385 *al.*, 2013), the results favour a polygenic mode of inheritance, which is largely
386 captured by additive relationships between animals. This is illustrated by the results

387 when a chromosome at a time was fitted, that overestimated the proportion of genetic
388 variance explained as opposed to when one chromosome and the **G** matrix were
389 simultaneously fitted. As highlighted by Daetwyler *et al.* (2012a), if the only
390 contribution of the SNP to the accuracy of genomic prediction was through LD with
391 QTL, and assuming a polygenic model, then a **G** matrix constructed from only the
392 SNP on one chromosome should capture genetic variation in proportion to its length,
393 assuming that there is no population stratification. However, this was not the case in
394 our study. It was therefore clear that a large proportion of the accuracy of genomic
395 prediction in the SBF population, at the current SNP density, is due to population
396 structure, i.e. relatedness between animals. In other words, only a small proportion of
397 the accuracy was due to LD between SNP and QTL.

398 This proposition was tested formally using the regression approach suggested by
399 Yang *et al.* (2011). The intercept (b_{0d}) of the difference between the variance for each
400 chromosome when analysed individually or simultaneously was highly significant for
401 all traits ($P<0.0001$), with the exception of body weight at 24 weeks ($P=0.09$). On the
402 other hand, the slope (b_{1d}) of the difference was significant only for some of the traits.
403 These values show the importance of the relatedness in our SBF population,
404 suggesting that most of our accuracy is probably captured by additive relatedness.
405 The ratio b_{0d}/b_{0i} is a measure of the proportion of genetic variance explained by such
406 relatedness (Yang *et al.*, 2011), and with the exception of NFEC16, this measure was
407 high (0.59-0.98) and therefore accounted for most of the variation in our SBF GEBV
408 predictions. Of interest is the observation that accuracy and the component due to
409 relatedness were largely independent of the **A**-matrix-derived heritability estimates
410 (results not shown).

411 The impact of relatedness has been previously studied, and differences in accuracies
412 have been ascribed to the number of relatives in the TS and the degree of additive-
413 genetic relationships with training individuals (Habier *et al.*, 2010). Legarra *et al.*
414 (2008) analysed accuracies of GEBV for individuals either related or unrelated to the
415 TS in a mouse population, concluding that markers were able to recover family
416 information to some extent. Our choice of predicting all members of a single sire
417 family from the remaining sire families in the SBF data was designed to reduce the
418 upward biases of accuracies resulting from within-family prediction when half-sib
419 families are randomly split between TS and validation sets. In this case we showed
420 that the closer the individuals in the validation set are to the TS, the higher the
421 accuracy. This is probably due in part to the fact that genomic predictions across
422 closely related individuals capture linkage effects, whereas those across distantly
423 related animals require LD between SNP and QTL. However, it should be noted that
424 although we used distinct sire families with the SBF data, these families were in most
425 part, also closely related.

426 We also estimated the accuracy achieved when predicting breeding values across
427 populations. These across-population accuracies were very low, sometimes even
428 negative. These low estimates may be explained by extension from our previous
429 results. Firstly, much of the accuracy in the SBF dataset was due to additive genetic
430 relationships between animals, as captured by the marker IBS relationships. This will
431 not be possible in distant populations. Secondly, the component of accuracy due to
432 LD between SNP and QTL is also likely to be low in distant breeds, as the linkage
433 phase between SNP and QTL will differ randomly in different breeds. The more
434 distant the relationship between individuals, the shorter the genomic distance over
435 which phase will be consistent. This outcome is reinforced by the finding that the

436 accuracy achieved for across-population prediction was somewhat higher when a
437 small number of animals from the population to be predicted were included in the TS.

438 It has been suggested that the use of a different method (i.e., BayesSSVS; Verbyla
439 *et al.*, 2009) could increase across-breed prediction, as it assigns SNP to either a
440 distribution with very small variance (i.e. near 0) or one with a larger variance in the
441 prediction model, unlike GBLUP which assumes that all SNP effects are sampled
442 from distributions with the same variance (Daetwyler *et al.*, 2012a). However, this
443 suggestion pre-supposes that the same gene variants are segregating in different
444 populations, and that the SNP density is sufficient for there to be consistent LD
445 between marker and QTL in (some of) the different populations. It has been
446 suggested that the number of SNP needed to predict unrelated individuals is equal to
447 10NeL, where L is the length of the genome in Morgans (Meuwissen, 2009). In the
448 SBF population, with Ne of ~500 (Kijas *et al.*, 2012) and L of approximately 27
449 Morgans, predictions for unrelated individuals would require at least 135,000 SNP.
450 This marker density may be achievable with the forthcoming high density sheep SNP
451 chip.

452 In summary, we have applied genomic prediction techniques to nematode resistance
453 and body weight data and found GEBV which, at first sight, appeared to have
454 moderate to good within-population predictive accuracy, despite a relatively limited
455 training set. However, much of the accuracy achieved appears to be a result of the
456 markers capturing additive genetic relationships between animals in the population.
457 This is reinforced by the observations that (i) the accuracy tends to drop when
458 predictions are across more distantly related animals in the same population, (ii)
459 across-population predictions have accuracies close to zero and (iii) some across-
460 population accuracy can be recovered by including a small number of animals from

461 the target population in the training set. These results suggest that genomic
462 prediction for nematode resistance and body weight may be of value in closely
463 related animals, but with the current SNP chip genomic predictions are unlikely to
464 work across breeds.

465

466 **Acknowledgements**

467 These results are obtained through the EC-funded FP7 Project 3SR-245140. French
468 SNP data were funded by the SHEEPSNPQTL ANR project. Funding from the
469 Regional Government of Sardinia contributed to the collection of Sardinian SNP and
470 phenotype data. We also wish to acknowledge funding contributions from
471 EADGENE_S, the BBSRC Institute Strategic Programme Grant at The Roslin
472 Institute and the Scottish Government's Strategic Partnership for Animal Science
473 Excellence (SPASE) initiative.

474

475 **References**

476 Bairden K 1991. Ruminant parasitic gastroenteritis: some observations on epidemiology and
477 control. PhD Thesis, University of Glasgow.
478 Cockerham CC 1969. Variance of gene frequencies. Evolution 23, 72-84.
479 Daetwyler HD, Pong-Wong R, Villanueva B and Woolliams JA 2010a. The Impact of
480 Genetic Architecture on Genome-Wide Evaluation Methods. Genetics 185, 1021-1031.
481 Daetwyler HD, Kemper KE, van der Werf JHJ and Hayes BJ 2012a. Components of the
482 Accuracy of Genomic Prediction in a Multi-Breed Sheep Population. Journal of Animal
483 Science 90, 3375-3384.
484 Daetwyler HD, Swan AA, van der Werf JHJ and Hayes BJ 2012b. Accuracy of pedigree and
485 genomic predictions of carcass and novel meat quality traits in multi-breed sheep data
486 assessed by cross-validation. Genetics Selection Evolution 44, 33.
487 Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, van der Werf JHJ and Hayes
488 BJ 2010b. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-
489 breed sheep population. Animal Production Science 50, 1004-1010.
490 Dekkers JCM and Hospital F 2002. Multifactorial genetics: the use of molecular genetics in
491 the improvement of agricultural populations. Nature Reviews Genetics 3, 22-32.
492 Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM, Barillet F,
493 Robert-Granié C and Manfredi E 2012. Genomic selection in the French Lacaune dairy sheep
494 breed. Journal of Dairy Science 95, 2723-2733.

495 Gilmour AR, Gogel BJ, Cullis BR and Thompson R 2009. ASReml User Guide Release 3.0.
496 VSN Int. Ltd.
497 González-Recio O, Gianola D, Rosa GJM, Weigel KA and Kranis A 2009. Genome-assisted
498 prediction of a quantitative trait measured in parents and progeny: application to food
499 conversion rate in chickens. Genetics Selection Evolution 41, 3.
500 Gordon HM and Whitlock HV 1939. A new technique for counting nematode eggs in sheep
501 faeces. Journal Council for Scientific and Industrial Research Australia 12, 50.
502 Habier D, Fernando RL and Dekkers JCM 2007. The impact of genetic relationship
503 information on genome-assisted breeding values. Genetics 177, 2389-2397.
504 Habier D, Tetens J, Seefried FR, Lichtner P and Thaller G 2010. The impact of genetic
505 relationship information on genomic breeding values in German Holstein cattle. Genetics
506 Selection Evolution 42, 5.
507 Harris BL, Johnson DL and Spelman RJ 2008. Genomic selection in New Zealand and the
508 implications for national genetic evaluation. Proceedings of the Interbull Meeting. Sattler,
509 J.D. (ed). Niagara Falls, NY, 325-330.
510 Hayes BJ and Goddard ME 2001. The distribution of the effects of genes affecting
511 quantitative traits in livestock. Genetics Selection Evolution 33, 209-229.
512 Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009a. Invited review: Genomic
513 selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92, 433-443.
514 Hayes BJ, Bowman PJ, Chamberlain AJ, Verbyla K and Goddard ME 2009b. Accuracy of
515 genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution
516 41, 51.
517 Kemper KE, Emery DL, Bishop SC, Oddy H, Hayes BJ, Dominik S, Henshall JM and
518 Goddard ME 2011. The distribution of SNP marker effects for faecal worm egg count in
519 sheep, and the feasibility of using these markers to predict genetic merit for resistance to
520 worm infections. Genetics Research 93, 203-219.
521 Kijas JW, Lenstra JA, Hayes BJ, Boitard S, Porto Neto LR, San Cristobal M, Servin B,
522 McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J,
523 Dalrymple B and Consortium omotISG 2012. Genome-wide analysis of the World's sheep
524 breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology 10,
525 e1001258.
526 Legarra A, Robert-Granie C, Manfredi E and Elsen J-M 2008. Performance of genomic
527 selection in mice. Genetics 180, 611-618.
528 Luan T, Woolliams JA, Lien S, Kent M, Svendsen M and Meuwissen THE 2009. The
529 accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics
530 183, 1119-1126.
531 Matika O, Pong-Wong R, Woolliams JA and Bishop SC 2011. Confirmation of two
532 quantitative trait loci regions for nematode resistance in commercial British terminal sire
533 breeds. Animal 5, 1149-1156.
534 Meuwissen THE 2009. Accuracy of breeding values of 'unrelated' individuals predicted by
535 dense SNP genotyping. Genetics Selection Evolution 41, 35.
536 Riggio V, Matika O, Pong-Wong R, Stear MJ and Bishop SC 2013. Genome-wide association
537 and Regional Heritability Mapping to identify loci underlying variation in nematode
538 resistance and body weight in Scottish Blackface lambs. Heredity 110, 420-429.
539 Ritland K 1996. Estimators for pairwise relatedness and individual inbreeding coefficients.
540 Genetical Research 67, 175-185.
541 Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH,
542 Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel
543 RD, Garrick DJ and Taylor JF 2011. Accuracies of genomic breeding values in American

544 Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution
545 43, 40.
546 Sallé G, Jacquiet P, Gruner L, Cortet J, Sauve C, Prevot F, Grisez C, Bergeaud JP, Schibler L,
547 Tircazes A, Francois D, Pery C, Bouvier F, Thouly JC, Brunel JC, Legarra A, Elsen JM,
548 Bouix J, Rupp R and Moreno CR 2012. A genome scan for QTL affecting resistance to
549 Haemonchus contortus in sheep. Journal of Animal Science 90, 4690-4705.
550 Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen W-M, Bonnycastle LL, Shen H, Timpson
551 N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L,
552 Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN,
553 Watanabe RM, Uda M, Tuomilehto J, Coresh J, Hirschhorn JN, Shuldiner AR, Schlessinger
554 D, Collins FS, Smith GD, Boerwinkle E, Cao A, Boehnke M, Abecasis GR and Mohlke KL
555 2008. Common variants in the GDF5-UQCC region are associated with variation in human
556 height. Nature Genetics 40, 198-203.
557 Sechi S, Salaris S, Scala A, Rupp R, Moreno C, Bishop SC and Casu S 2009. Estimation of
558 (co)variance components of nematode parasites resistance and somatic cell count in dairy
559 sheep. Italian Journal of Animal Science 8, 156-158.
560 Sinski E, Bairden K, Duncan JL, Eisler MC, Holmes PH, McKellar QA, Murray M and Stear
561 MJ 1995. Local and plasma antibodyresponses to the parasitic larval stages of the abomasal
562 nematode Ostertagia circumcincta. Veterinary Parasitology 59, 107-118.
563 Sonesson AK and Meuwissen THE 2009. Testing strategies for genomic selection in
564 aquaculture breeding programs. Genetics Selection Evolution 41, 37.
565 Strain SAJ, Bishop SC, Henderson NG, Kerr A, McKellar QA, Mitchell S and Stear MJ 2002.
566 The genetic control of IgA activity against Teladorsagia circumcincta and its association with
567 parasite resistance in naturally infected sheep. Parasitology 124, 545-552.
568 Su G, Guldbrandtsen B, Gregersen VR and Lund MS 2010. Preliminary investigation on
569 reliability of genomic estimated breeding values in the Danish Holstein population. Journal of
570 Dairy Science 93, 1175-1183.
571 VanRaden P 2008. Efficient methods to compute genomic predictions. Journal of Dairy
572 Science 91, 4414-4423.
573 VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF and
574 Schenkel FS 2009. Invited review: Reliability of genomic predictions for North American
575 Holstein bulls. Journal of Dairy Science 92, 16-24.
576 Verbyla KL, Hayes BJ, Bowman PJ and Goddard ME 2009. Accuracy of genomic selection
577 using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genetics
578 Research 91, 307-311.
579 Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade
580 M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling
581 H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME and
582 Visscher PM 2011. Genome partitioning of genetic variation for complex traits using
583 common SNPs. Nature Genetics 43, 519-525.

584

585

586 **Table 1** *Correlations between predicted genomic estimated breeding values and*
587 *adjusted phenotypes and accuracies* for the random cross-validation groups both*
588 *using the genomic relationship matrix and the pedigree-based relationship matrix in*
589 *the Scottish Blackface population*

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Genomic-based accuracy**	**Pedigree-based accuracy**
IgA	0.151	0.174	0.314	0.359	0.382	0.532	0.513
SFEC16	0.192	0.074	0.089	0.245	0.174	0.487	0.516
SFEC20	0.141	0.099	0.216	0.150	0.091	0.432	0.401
SFEC24	0.138	0.068	0.186	0.172	0.110	0.442	0.476
NFEC16	-0.027	0.059	0.071	0.034	-0.006	0.099	0.342
NFEC20	0.210	0.292	0.193	0.324	0.220	0.598	0.488
NFEC24	0.212	0.182	0.155	0.178	0.130	0.503	0.408
W16W	0.206	0.127	0.231	0.232	0.234	0.516	0.336
W24W	0.169	0.073	0.165	0.109	0.046	0.417	0.292
SFEC_av	0.319	0.179	0.254	0.303	0.175	0.540	0.442
NFEC_av	0.208	0.317	0.192	0.282	0.234	0.481	0.357
WW_av	0.149	0.147	0.195	0.136	0.057	0.684	0.270

590 IgA: Immunoglobulin-A; SFEC16, SFEC20, and SFEC24: faecal egg count at 16, 20 and 24 weeks for
591 *Strongyles*; NFEC16, NFEC20, NFEC24: faecal egg count at 16, 20 and 24 weeks for *Nematodirus*;
592 W16W and W24W: body weight at 16 and 24 weeks; SFEC_av, NFEC_av, WW_av: average animal
593 effect for *Strongyles* and *Nematodirus* faecal egg count and for body weight
594 *accuracy here is the average of the accuracies across validation sets, estimated as the correlation for
595 each validation set divided by the square root of its heritability
596
597

598 **Table 2** *Correlations between genomic estimated breeding values and predicted*
599 *estimated genomic breeding values for the random cross-validation groups in the*
600 *Scottish Blackface population*

	Group1	**Group2**	**Group3**	**Group4**	**Group5**	**average**
IgA	0.674	0.731	0.784	0.699	0.773	0.732
SFEC16	0.737	0.606	0.699	0.729	0.764	0.707
SFEC20	0.841	0.764	0.850	0.788	0.846	0.818
SFEC24	0.825	0.804	0.815	0.826	0.794	0.813
NFEC16	0.774	0.750	0.700	0.690	0.710	0.725
NFEC20	0.709	0.863	0.823	0.867	0.767	0.806
NFEC24	0.842	0.783	0.816	0.880	0.847	0.834
W16W	0.627	0.676	0.719	0.794	0.713	0.706
W24W	0.666	0.667	0.743	0.799	0.632	0.702
SFEC_av	0.811	0.697	0.777	0.769	0.795	0.770
NFEC_av	0.764	0.765	0.765	0.798	0.735	0.765
WW_av	0.661	0.779	0.828	0.830	0.750	0.770

601 IgA: Immunoglobulin-A; SFEC16, SFEC20, and SFEC24: faecal egg count at 16, 20 and 24 weeks for
602 *Strongyles*; NFEC16, NFEC20, NFEC24: faecal egg count at 16, 20 and 24 weeks for *Nematodirus*;
603 W16W and W24W: body weight at 16 and 24 weeks; SFEC_av, NFEC_av, WW_av: average animal
604 effect for *Strongyles* and *Nematodirus* faecal egg count and for body weight

605

606

607 **Table 3** *Correlations between predicted genomic estimated breeding values and*
608 *adjusted phenotypes for families in the Scottish Blackface population*

	Fam022	Fam058	Fam085	Fam161
IgA	0.324	0.087	0.174	0.119
SFEC16	0.198	0.023	0.179	0.055
NFEC16	0.108	-0.055	0.036	0.018
W16W	-0.072	0.162	0.291	0.124

609 IgA: Immunoglobulin-A; SFEC16, NFEC16, and W16W: *Strongyles* and *Nematodirus* faecal egg count
610 and body weight at 16 weeks
611

612 **Table 4** *Intercept, slope (i.e., proportion of phenotypic variance/Mb), and* R^2 *for the three regressions (i.e., by fitting each*
613 *chromosome individually, by fitting all chromosomes simultaneously, and the difference between the two) as well as the proportion*
614 *of genetic variance explained by relatedness* (b_{0d}/b_{0i}) *for all traits considered*

	Chromosome fitted individually			Chromosome fitted simultaneously			Difference			b_{0d}/b_{0i}
	R^2	Intercept	Slope	R^2	Intercept	Slope	R^2	Intercept	Slope	
IgA	0.26	0.058***	0.00025**	0.06	0.001	0.00010	0.34	0.056***	0.00015***	0.98
SFEC16	0.10	0.029**	0.00014	0.08	0.005	0.00011	0.02	0.024***	0.00003	0.84
SFEC20	0.10	0.041***	0.00009	0.00	0.012*	-0.00002	0.25	0.029***	0.00010**	0.71
SFEC24	0.06	0.039***	0.00006	0.02	0.008	0.00004	0.03	0.031***	0.00003	0.80
NFEC16	0.00	0.025**	-0.00002	0.00	0.015	-0.00002	0.00	0.010***	0.00000	0.39
NFEC20	0.44	0.063***	0.00020**	0.04	0.005	0.00005	0.56	0.058***	0.00015***	0.92
NFEC24	0.06	0.047***	0.00008	0.01	0.016*	-0.00003	0.28	0.032***	0.00011**	0.67
W16W	0.28	0.037***	0.00022**	0.00	0.009	-0.00001	0.46	0.028***	0.00024***	0.76
W24W	0.41	0.022***	0.00018***	0.00	0.009	-0.00001	0.28	0.013	0.00020**	0.59
SFECav	0.07	0.068***	0.00012	0.00	0.013	0.00001	0.17	0.056***	0.00011*	0.82
NFECav	0.07	0.079***	0.00015	0.02	0.011	0.00007	0.11	0.068***	0.00008	0.86
WWav	0.11	0.017**	0.00010	0.10	0.003	0.00008	0.01	0.015***	0.00002	0.85

615

616 IgA: Immunoglobulin-A; SFEC16, SFEC20, and SFEC24: faecal egg count at 16, 20 and 24 weeks for *Strongyles*; NFEC16, NFEC20, NFEC24: faecal egg

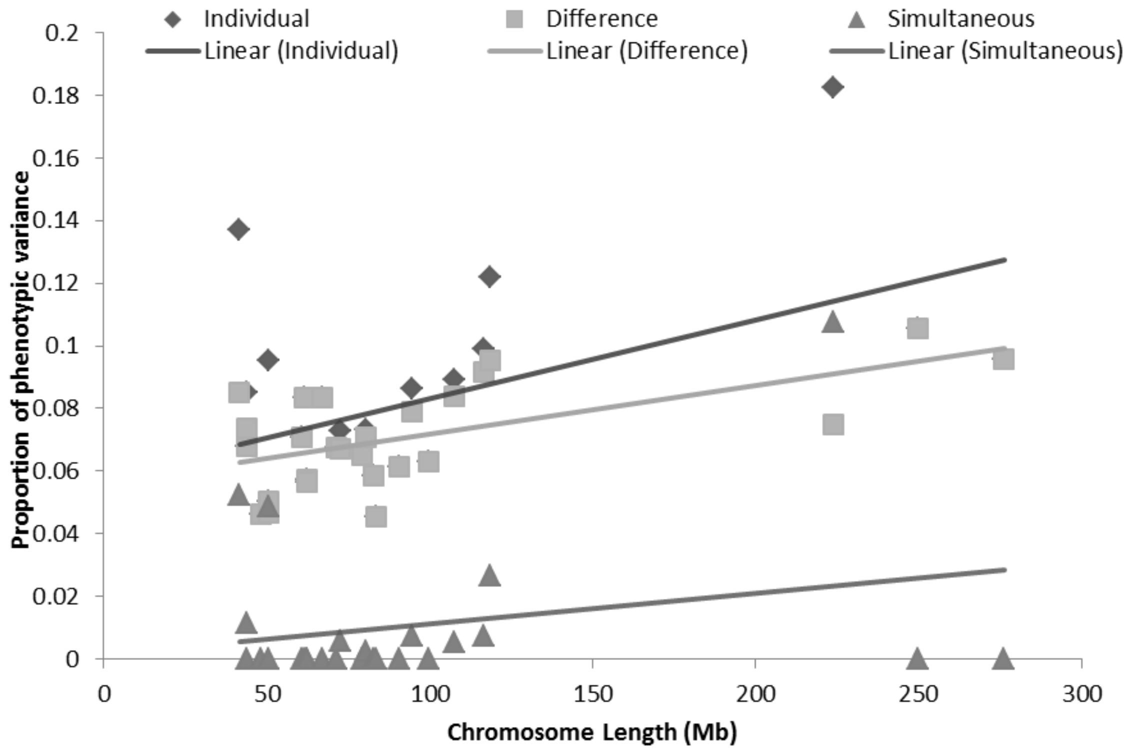
617 count at 16, 20 and 24 weeks for *Nematodirus*; W16W and W24W: body weight at 16 and 24 weeks; SFEC_av, NFEC_av, WW_av: average animal effect for

618 *Strongyles* and *Nematodirus* faecal egg count and for body weight

619 *P < 0.05; **P < 0.01; ***P < 0.001

620 **Figure 1** Proportion of phenotypic variance explained per chromosome for
621 Immunoglobulin-A (scattered points) and fitted regression (line). Chromosome fitted
622 individually (top regression) or simultaneously (bottom regression). Middle regression
623 results from plotting the difference between top and bottom regression.

624



625

626