GenTORE

Genomic management Tools to Optimise Resilience and Efficiency

Grant agreement n°: 727213

H2020 - Research and Innovation Action

D4.3

Genomic prediction of efficiency in multiple admixed breeds

Due date: M48

Actual submission date: M48

Project start date: 1st June 2017 **Duration:** 60 months

Workpackage concerned: WP4

Concerned workpackage leader: Mogens Sandø Lund

Lead Beneficiary: DLO

Dissemination level:

☒ **PU:** Public (must be available on the website)

☐ **CO:** Confidential, only for members of the consortium (including the Commission Services)

☐ **CI:** Classified, as referred to in Commission Decision 2001/844/EC

Table of content

1. Summary

An analysis protocol, based on readily available BLUP software packages, has been developed to allow for heterogeneous SNP (co)variances in genomic genotype by environment interaction models. The developed approach has been applied to the efficiency trait age at slaughter, an indirect measure of growth efficiency, of Irish dairy and beef crossbred animals. This deliverable reports genetic parameters for age at slaughter and the accuracy of genomic prediction reaction norm models for this trait. As a next step, the proposed models will be tested in dual purpose Fleckvieh cattle by GenTORE partner LfL (Germany) within Task 4.2.

2. Introduction

The reduction of the number of days from birth until the target weight at slaughter is reached, represents a sustainable option to increase efficiency of beef cattle production on animal and herd level as well as reduce the environmental impact. In the presence of genotype by environment interaction (GxE) selection of efficient and resilient animals is important. In this deliverable, we have developed genomic GxE models to allow for heterogeneous SNP (co)variances across the genome based on readily available BLUP software packages. Applications in simulated data have shown a slight increase of accuracy of genomic breeding values (GEBV) with heterogeneous compared to homogeneous SNP (co)variances. The new approach was applied in genomic prediction GxE analysis for the efficiency trait age at slaughter in an Irish dairy and beef crossbred cattle population. This deliverable evaluated the extent of GxE for age at slaughter and compared the accuracies of a genomic reaction norm model for age at slaughter, modelling either homogeneous or heterogeneous SNP (co)variances across the genome.

3. Results

Results of the developed methods applied to the efficiency trait age at slaughter in an Irish dairy and beef crossbred cattle population are presented in the Annex. The report is written in the format to be submitted to a peer-reviewed journal.

4. Conclusions

The genetic analysis of age at slaughter in an Irish dairy and beef crossbred population reveals large genetic variation for this trait. Genomic reaction norm models resulted in higher accuracies for young selection candidates compared to the expected accuracy of parent average breeding values without genomic information. The use of genomic information for age at slaughter will be beneficial to increase efficiency and reduce environmental impact of beef cattle production. Results show the existence of GxE for age at slaughter to some extent. Unlike results based on simulated data, the genomic reaction norm model allowing for heterogeneous SNP (co)variances did not increase the accuracy of genomic breeding values for age at slaughter.

5. Partners involved in the work

DLO (NL), TEAGASC (Ireland)

6. Annexes

Annex 1: Accuracy of genomic reaction norm model for age at slaughter in an Irish dairy and beef crossbreed population
Annex 2: presentation from EAAP 2020
Annex 3: presentation from Interbull Meeting 2021

Working title:
Accuracy of genomic reaction norm model for age at slaughter in an Irish dairy and beef crossbreed population

Birgit Gredler-Grandl, Jeremie Vandenplas, Mario Calus

Wageningen University & Research Animal Breeding and Genomics
PO Box 338
6700 AH Wageningen
The Netherlands

Introduction

Efficiency and environmental impact of livestock productions system is of increasing importance. To increase efficiency and reduce the environmental footprint of animal production systems, management, but also breeding strategies (e.g. Neeteson et al., 2013; Veerkamp et al., 2013) have been proposed. Breeding strategies to increase efficiency and reduce the environmental impact can focus on the selection of direct (e.g. feed efficiency or methane emission) or indirect traits. In beef cattle, breeding for reduced number of days from birth until the target carcass weight at slaughter is reached, could be a sustainable option to increase efficiency on animal and herd level and reduce environmental impact. Berry et al. (2017) have investigated the potential of genetic selection for a younger age at slaughter, exploiting large amount of both phenotypic and genetic variation for this trait.

In the presence of genotype by environment interaction (GxE), selection of efficient and resilient animals is important. GxE is typically modelled by a multi-trait approach, where the same trait measured in different discrete environments is considered as a different, but genetically correlated trait (e.g. Falconer, 1952). Alternatively, GxE can be modelled with reaction norm models (Kolmodin et al., 2002; Calus and Veerkamp, 2003), where the breeding values are modelled as a function of the environment defined as a continuous variable. Both approaches can be implemented as genomic prediction models by replacing the pedigree based relationship matrix by the genomic relationship matrix. A variety of studies show the advantage of genomic GxE models, e.g. genomic reaction norm models, resulting in higher accuracy of genomic breeding values (GEBV) (Silva et al. 2014, Mota et al., 2020). In addition, specific SNP-by-environment interaction can be studied (e.g Carvalheiro et al., 2019; Zhang et al., 2019) and GEBV of non-phenotyped animals can be estimated for different

environments. Both, genomic multi-trait models or reaction norm models, implicitly assume the same (co)variance matrix for every SNP. Since certain regions in the genome may contain QTL, the assumption of equal (co)variances across the genome may be violated. In genomic prediction models without considering GxE, this assumption can be relaxed by the implementation of a Bayesian variable selection model (de los Campos et al., 2013). However, to our knowledge no implementation of a Bayesian genomic prediction reaction norm model using variable selection to differentiate SNP (co)variances across the genome is described in literature. Once SNP-specific variances are obtained (e.g. from a Bayesian model), it has been shown for univariate models that these can be used in a ridge regression BLUP (or SNP-BLUP) model to obtain breeding values that are equivalent to those from the full Bayesian model (Calus et al., 2014). This allows to compute GEBV using standard BLUP software currently used for large scale routine genomic evaluations, but still requires currently not available software to estimate the heterogeneous SNP (co-)variances, which may be computationally very demanding.

Wang et al. (2012) described the single-step GWAS approach as an alternative, which involves the computation of BLUP solutions for each SNP followed by derivation of SNP variances from those BLUP solutions for each SNP. Every next iteration uses the SNP-specific variances from the previous iteration, where in the first iteration equal variances are assumed for all SNP. The limitations of this approach are that SNP variances may become very large (Garcia et al., 2018) and that there is no formal criteria to determine convergence of the estimated variance components.

To overcome the limitations described above, we have developed an alternative approach to allow for heterogeneous SNP (co)variances in genomic GxE models that can be implemented using standard BLUP software currently used for large scale routine genomic evaluations. The approach has been tested in simulated data resulting in slightly higher accuracies of GEBV estimated with both, genomic multi-trait and reaction norm models (Gredler-Grandl and Calus, 2020; Gredler-Grandl and Calus, 2021). The objective of this study was to apply this approach in genomic GxE analysis for age at slaughter in an Irish dairy and beef crossbred cattle population. We evaluated the extent of GxE for age at slaughter and compared the accuracies of genomic reaction norm model for age at slaughter, modelling either homogeneous or heterogeneous SNP (co)variances across the genome.

Material and methods

Phenotype data

Phenotypic data consisted of a pre-edited data set of 2,845,504 individual records for age at slaughter of an Irish crossbreed dairy and beef cattle population. The corresponding pedigree file comprised 20,236,559 animals. Slaughter records were provided for 700,894 bulls, 1,255,825 steers and 888,785 heifers. The animals were born between the years 2000 and 2018 and were kept in 31,869 herds. The data set consisted of dairy and beef purebred and crossbred animals. The number of purebred animals per breed (i.e. animals with a gene proportion of 100% of the corresponding breed determined by pedigree information) is as follows: Holstein (n=268,255), Limousine (n=13,644), Charolais (n=12,104), Simmental (n=3,607), Aberdeen Angus (n=2,471), Saler (n=1,916), Aubrac (n=1,583), Hereford (n=1,358), Shorthorn (n=562), Blonde d'Aquitaine (n=482), Parthenaise (n=148), Jersey (n=55), and Belgian Blue (n=31). The remaining animals were classified as crossbred animals. All pre-editing steps are described in detail by Berry et al. (2017). Briefly, the main editing criteria were: only animals with known sire and dam were used; at least 5 animals per contemporary group (CG), carcass weight within the range of ≥ 100 kg and ≤800 kg, age at slaughter ≥12 months and ≤36 months. Animals of the same gender and similar age purchased into the same herd in close period of time (within 10 d of each other) were assigned to the same CG. This resulted in 260,547 CG with an average number of 10.92 animals per CG (SD = 9.66, min = 5, max = 257).

Genotype data

Genotypes of 51,924 animals were available. The genotypes were imputed to a high density (HD) SNP chip level of 734,159 autosomal SNP. The given genotypes were a subset of a data set described by Twomey et al. (2019), where genotypes were available from six different genotyping panels ranging from 17,137 SNP (International Dairy and Beef version 1 SNP chip) up to 777,962 SNP (Illumina bovine High-Density BeadChip). The imputation to HD was carried out with the software FImpute2 (Sargolzaei et al., 2014) following a two-step approach described in detail by Twomey et al. (2019). First, all animals genotyped with a low-density SNP chip panel were imputed to the Bovine SNP50 density (50k) and in a second step all 50k genotypes were imputed to HD level based on a multi-breed reference population.

For further analyses in this study, SNP genotypes with MAF <0.005 were discarded resulting in a final set of 662,011 SNP. Three pairs of identical twins have been identified in the genotyped data set, where one twin has been randomly removed. 29,558 genotyped animals

had phenotypes for age at slaughter. The minimum required number of genotyped animals per CG was set to three, leaving 14,665 genotyped and phenotyped animals. The genotyped animals in the final data set were born between the years 2009 and 2016, kept in 2,041 herds and assigned to 3,146 CG. The final data set included purebred animals of Holstein (n=1,909), Limousine (n=202), Charolais (n=157), Aberdeen Angus (n=21), Belgian Blue (n=1), Hereford (n=23), Simmental (n=20), Saler (n=13), Aubrac (n=2), Blonde d'Aquitaine (n=4), and Parthenaise (n=3). To infer population structure a Principal Component Analysis implemented in calc_grm (Calus and Vandenplas, 2016) has been carried out on the genomic relationship matrix .

Model

To allow for heterogeneous SNP (co)variances across the genome in genomic GxE models a protocol consisting of several steps has been developed. Initially, the data set of interest is split in two subsets, similar to an approach where data is split into a subset for QTL discovery, and a subset where those QTL are used to upweight the contribution of those SNP to the explained variance (Moghaddar et al. 2019). Subset 1 is used to estimate SNP effects $\hat{\alpha}$ using a model that assumes equal (co)variances for all SNP. This model could be a genomic relationship matrix based REML (GREML) analysis followed by backsolving of $\hat{\alpha}$, or a random regression on SNP genotypes. SNP specific variances are then computed as $2p_k(1-p_k)\hat{\alpha}_k^2$. For multivariate models SNP specific covariances between traits *i* and *j* can equivalently be computed as $2p_k(1-p_k)\hat{\alpha}_{k_i}\hat{\alpha}_{k_j}$, where $\hat{\alpha}_{i_k}(\hat{\alpha}_{j_k})$ is the estimated effect for SNP *k* for trait *i* (*j*). The same principle can be applied for reaction norm models, where SNP specific covariances between coefficients (intercept (0), slope (1), etc.) *i* and *j* can be computed as $2p_k(1-p_k)\hat{\alpha}_{k_i}\hat{\alpha}_{k_j}$, where $\hat{\alpha}_{k_i}(\hat{\alpha}_{k_j})$ is the estimated effect for SNP *k* for coefficient *i* (*j*). The model applied to subset 1 then considers these SNP specific variances as weights to compute a weighted SNP (co)variance matrix. In the following, the steps of the analysis protocol are described in detail applied to the crossbred dairy and beef cattle data set described above. A genomic reaction norm model allowing for heterogeneous SNP (co)variances across the genome (HET) was compared to a model with homogeneous SNP (co)variances (HOM).

Estimation of Contemporary group effects

Estimated contemporary group effects were used as continuous environmental descriptor in the reaction norm model. The CG effects for age at slaughter were estimated in a univariate BLUP analysis with the MiXBLUP software package (ten Napel et al., 2020):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

Where $\mathbf{y}$ is the vector of phenotypic observations for age at slaughter in days; $\mathbf{b}$ is the vector of fixed effects; a is the vector of additive genetic effects, which are assumed to follow a normal distribution with $N(0, \mathbf{A}\sigma_a^2)$, where $\mathbf{A}$ is the pedigree based relationship matrix and σ_a^2 is the additive genetic variance; $\mathbf{e}$ is a vector for the random residual, which is assumed to follow a normal distribution with $N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is the identity matrix and σ_e^2 is the variance of the random residual. $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating the observations to the fixed and random effects in the model. The fixed effects included in the model were CG, the interactions between gender (bull, steer, or heifer) and carcass weight and gender and carcass fat, parity of the dam (classes 1 to 4, and 5+), herd source, which is whether the animal was born in a beef or dairy herd, class effects of the general heterosis coefficient and recombination loss. Classes for heterosis were: 0%, >0% and ≤10%, >10% and ≤20%, >20% and ≤30%, >30% and ≤40%, >40% and ≤50%, >50% and ≤60%, >60% and ≤70%, >70% and ≤80%, >80% and ≤90%, >90 and <100% and 100%. Recombination loss classes were 0%, >0% and ≤5%, >5% and ≤10%, >10% and ≤15%, >15 and ≤20%, >20% and ≤25%, >25% and ≤30%, >30% and ≤35%, >35% and ≤40%, >40% and ≤45%, >45% and <50% and 50%. The additive genetic and residual variances used in the BLUP analysis for age at slaughter were 471.9 d and 875.905 d, respectively (A. Twomey, personal communication, December 11, 2020). YD for age at slaughter for each animal were calculated as the sum of the estimated breeding value and the residual term estimated with MiXBLUP. For further analyses, CG effects were standardized to have a mean of 0 and sd of 1. Descriptive statistics for YD, original and standardized CG effects are reported in Table 1. The frequency distribution of CG is shown in Figure 1.

Assigning animals to subsets

To assign the animals to subsets 1 and 2 a K-means clustering approach (Saatchi et al., 2011) has been applied to a genomic relationship matrix of the herds. First, average genotypes per herd for each SNP were calculated. Using calc_grm (Calus and Vandenplas, 2016) the genomic relationship matrix of the herds was computed based on the first method of VanRaden (2008) as $\mathbf{G_{herds}} = \frac{\mathbf{ZZ'}}{2\sum p_k(1-p_k)}$, where $\mathbf{G_{herds}}$ is the genomic relationship matrix, $\boldsymbol{Z}$ is the

incidence matrix containing average genotypes for all herds for all SNP and p_k is the allele frequency of SNP *k* in the genotyped animals. A dissimilarity matrix between all herds was calculated based on elements of the genomic relationship matrix as follows:

$$d_{ij} = 1 - \frac{g_{ij}}{\sqrt{g_{ii} * g_{jj}}}$$

where d_{ij} is a measure of genomic distance between herd *i* and herd *j*, g_{ij} is the genomic relationship between herd *i* and herd *j*, g_{ii} and g_{jj} are diagonal elements of the matrix $\mathbf{G_{herds}}$ representing the genomic relationship coefficient of herd *i* (*j*) with itself. For the K-means clustering the companion program *kmeanscluster* of calc_grm (Calus and Vandenplas, 2016) has been used. The number of iterations and the number of clusters were set to 20 and 12, respectively. Ideally, each major breed is associated with at least two clusters, such that one cluster could be used in subset 1 and one for subset 2. We have set the number of clusters to 12 to ensure that all main breeds are represented at least by three clusters. For subset 1 four clusters (clusters 1, 2, 6, and 8, Table 2) comprising a large number animals representing the main breeds (Holstein, Limousine, Charolais and Angus) with highest importance for cross breeding have been chosen. The remaining eight clusters (0, 3, 4, 5, 7, 9, 10, and 11) were used in subset 2. The number of animals in subsets 1 and 2 were 5,326 and 9,342, respectively. Due to small sample size and similar breed composition (Holstein and Aberdeen Angus) clusters 9 and 10 were combined resulting in total seven clusters for subset 2. In subset 2 a seven-fold cross validation has been applied, where each cluster has been used as validation set once for genomic prediction.

Analysis subset 1

Genomic reaction norm model

A univariate genomic reaction norm model has been applied to age at slaughter following Ni et al. (2019) and Chung et al. (2020) in subset 1 using the software package mtg2 (Lee et al., 2016):

$$\mathbf{y} = \mathbf{1}\mu + \boldsymbol{\beta_0} + \mathbf{Q}\boldsymbol{\beta_1} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of YD for age at slaughter for all animals, μ is an overall mean, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are the vectors of intercept and second order of regression coefficients for the random genetic effects, $\mathbf{1}$ is a vector of ones, $\boldsymbol{Q}$ is a (diagonal) incidence matrix storing the squared estimated CG effects describing the environment for each animal, and $\boldsymbol{e}$ is the vector of random

residuals. It is assumed that $\begin{bmatrix}\boldsymbol{\beta_0}\\ \boldsymbol{\beta_1}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{G_{VR}}\begin{bmatrix}\sigma^2_{\beta 0} & \sigma_{\beta 0\beta 1}\\ \sigma_{\beta 0\beta 1} & \sigma^2_{\beta 1}\end{bmatrix}\right)$, where $\mathbf{G_{VR}}$ is a genomic relationship matrix of the animals in subset 1 using the first method of VanRaden (2008) and $\mathbf{e}$ is assumed to follow $N(\mathbf{0}, \mathbf{I}\sigma^2_e)$.

Initial analyses with a first order random regression on the estimated CG effects did not yield meaningful results, while the variance in YD appeared to be largest at intermediate CG values, supporting a random regression on the squared CG effects. To evaluate the fit of the genomic reaction norm model to the data, a reduced model without considering the regression on squared CG effects was compared using a likelihood ratio test (LRT) based on the LRT statistic D = -2 * log(likelihood) for the reduced model + 2 * log(likelihood) for the alternative model. For convenience, both models were considered being significantly different if the test statistic was above the 5% critical value (2.71) from a mixture χ^2 distribution with 0 and 1 degrees of freedom (Self and Liang, 1987). The genetic variance σ^2_q for a specific environment q (CG effect) was calculated as $\sigma^2_q = \sigma^2_{\beta 0} + \sigma^2_{\beta 1} q^2 + 2\sigma_{\beta 0\beta 1} q$. The heritability ($h^2_q$) of a specific environment q was calculated as $h^2_q = \frac{\sigma^2_q}{\sigma^2_q + \sigma^2_e}$.

Calculation of SNP specific weights

Allowing for heterogeneous SNP variances (HET), SNP specific weights for each SNP *k* for each coefficient *i* of the reaction norm model (i.e. intercept β_0 and the quadratic regression coefficient β_1) were calculated as

$$D_{k_i} = \sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$$

where D_{k_i} is diagonal element *i* of diagonal matrix $\mathbf{D_k}$ that stores the weights for SNP *k*, p_k is the allele frequency of SNP k, and $\hat{\alpha}_k$ is the estimated effect of SNP *k* for coefficient *i.* The SNP effects $\hat{\alpha}_k$ for intercept and quadratic regression coefficient were obtained by backsolving based on the GEBV for β_0 and β_1 obtained from the genomic reaction norm model. SNP effects were calculated following the approach described in Bouwman et al. (2017) implemented in the companion program *compute_SNP_effects* of *calc_grm* (Calus and Vandenplas, 2016).

Analysis subset 2

In subset 2, the following SNP-BLUP (VanRaden, 2008) model was applied

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\boldsymbol{\gamma}_\mathbf{0} + \mathbf{QZ}\boldsymbol{\gamma}_\mathbf{1} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of YD of animals in the training set of each cross validation run, μ is an overall mean, $\mathbf{Z}$ is a matrix including the centered genotypes for each SNP, $\mathbf{Q}$ is a diagonal matrix storing the squared estimated CG effects (environmental descriptor) for each animal, $\boldsymbol{\gamma}_\mathbf{0}$ and $\boldsymbol{\gamma}_\mathbf{1}$ are vectors of estimated SNP effects for random intercept and quadradic regression coefficient, respectively, and $\mathbf{e}$ is a random residual term. For HET the following (co)variance matrix is used for SNP *k*:

$$Var([\boldsymbol{\gamma}_\mathbf{0}, \boldsymbol{\gamma}_\mathbf{1}]') = \mathbf{D}_k * \boldsymbol{G} * \mathbf{D}_k$$

where $\mathbf{G}$ is the estimated genetic (co)variance matrix between intercept and quadratic regression coefficient obtained from the reaction norm model in the analysis of subset 1. For HOM, homogeneous SNP variances for intercept and quadratic regression coefficient are provided by $\sigma_g^2/2\sum p_k(1-p)$, where σ_g^2 is the genetic variance for either intercept or quadratic regression coefficient estimated in subset 1. The GEBV for validation animals were calculated as $\mathbf{GEBV} = \mathbf{1}\hat{\mu} + \mathbf{Z}\hat{\boldsymbol{\gamma}}_\mathbf{0} + \mathbf{ZQ}\hat{\boldsymbol{\gamma}}_\mathbf{1}$.

The accuracies of GEBV for animals in the validation set were obtained as the correlation coefficient between the observed YD and predicted GEBV divided by the square root of heritability: $\hat{r} = cor(\mathbf{YD}, \mathbf{GEBV})/\sqrt{h^2}$. In order to calculate accuracies for different environments, three environments (env1 – env3) have been defined by the <20% quantile, >=20% and <=80% quantile, and >80% quantile of the environmental descriptor (CG effect). The number of validation animals per environment for each cross validation is shown in Table 4. The bias of GEBV (inflation) was evaluated by the coefficient of the regression of YD on GEBV.

Results

Descriptive statistics

Summary statistics for age at slaughter for all phenotyped and genotyped animals with phenotypes are presented in Table 1. The arithmetic means for age at slaughter for all phenotyped and genotyped animals with phenotypes were 745.6 days (SD 145) and 746.7 days (SD 123.5), respectively, which is in concordance with Berry et al. (2017).

The population structure of the genotyped animals evaluated by a PCA is illustrated in Figure 2. The majority of animals (n=12,310) are classified as crossbred animals in the pedigree. The most important breeds regarding contribution to crossbred animals are Limousine, Holstein, Charolais and Aberdeen Angus as shown in the density distribution of proportion of genes of those breeds for all animals (Figure 3). The average proportion of genes for Limousine, Holstein, Charolais and Aberdeen Angus are 0.27 (SD 0.31), 0.26 (SD 0.38), 0.23 (SD 0.29) and 0.07 (SD 0.17), respectively.

The number of clusters in the K-means clustering analysis was set to 12. Table 2 shows the number of animals and herds per cluster as well as the average proportion of genes per breed within cluster. The number of animals per cluster varied between 242 and 1,847 (mean 1,222, SD 603.7) and the number of herds per cluster were in the range of 32 and 323 (mean 314, SD 106.9) herds. Different breeds and different genetic lines (sires) within breed are associated with different clusters. The structure of clusters is visualised in Figure 4, where Holstein animals were associated with cluster 2, 3 and 10 with an average Holstein proportion of 0.98, 0.87 and 0.68, respectively. Limousine animals were distributed across several clusters, i.e. 1, 7, 5, 0 with average gene proportions of 0.63, 0.49, 0.41, and 0.30 respectively. Charolais was mainly associated with clusters 8, 11, 0 and 5 with average gene proportions of 0.49, 0.48, 0.45, and 0.33. The highest gene proportion of Aberdeen Angus (0.59 and 0.40) was found for clusters 6 and 9. Subset 1 consisted of clusters 1, 2, 6 and 8 representing the main breeds Limousine, Holstein, Aberdeen Angus and Charolais (Table 2).

LRT and Variance components estimated in the reaction norm model

The log(likelihood) for the reaction norm model and the reduced model (accounting for intercept only) were -23,318.0 and -23,326.7, respectively. The calculated test statistic D is 17.4, which is higher than the 5% critical value of 2.71 from a mixture χ^2 distribution with 0 and 1 degrees of freedom. The reaction norm model accounting for the environment is considered being significantly different from the reduced model.

The estimated genetic variances for the intercept and quadratic regression coefficient in subset 1 were 780.51 (SE 99.43) and 14.32 (SE 9.39), respectively. The covariance between intercept and quadratic regression coefficient was -85.09 (SE 29.37), leading to a genetic correlation of -0.80. The estimate for the residual variance was 1,773.5. The heritability for age at slaughter across environments (CG) is shown in Figure 5. Dependent on the regression function modelled in the reaction norm model, the shape of the curve follows a parabola, with highest heritability estimates for intermediate environments, and lowest estimates for extreme

environment with either very low or very high CG effects for age at slaughter. The average heritability across all CG was 0.264 (SD 0.044). The highest observed heritability was 0.306, the lowest 0.134. The average heritability for the 0-20% (env1), 20-80% (env2), and 80-100% (env3) quantiles of CG effects were 0.231 (SD 0.035), 0.295 (SD 0.011), and 0.214 (SD 0.037), respectively.

Each 50iest environment along the environmental gradient (CG) has been selected, resulting in 21 environments. Figure 6 shows the full genetic correlation matrix between those environments. The correlations were between 0.89 and 1. The lowest genetic correlations (0.89) were observed between intermediate and extreme environments (CG).

Accuracy and bias of GEBV with HOM and HET

Table 3 reports accuracy and bias of GEBV of validation animals in the cross validation analysis estimated with a genomic reaction norm model for both modelling homogeneous and heterogeneous SNP (co)variances across the genome. Here, accuracies were calculated across all environments that occur within a CV. Accuracy for HOM ranged between 0.185 and 0.582 across all cross validation sets. The highest accuracy (0.582) was observed for cross validation 2, where the main breeds regarding contributing gene proportion are Holstein and Aberdeen Angus. The average accuracy for HOM across all cross validations was 0.317. The slope of the regression of YD on GEBV indicate inflation of the GEBV where values greater than 1 indicate underestimation of the variance of GEBV and values lower than 1 overestimation. In all cross validations, an overestimation can be observed, except for CV2, where bias was 1.143 (Table 3). For HET, accuracies were in general slightly lower than for HOM, except for CV2, where HET accuracy was slightly higher (0.587 for HET and 0.582 for HOM, Table 3). The highest accuracy was again achieved for CV2. Similar as for HOM, overestimation of GEBV was obtained for HET. For CV2, bias was close to 1 with 1.108.

The accuracy and bias of GEBV were also assessed across the environmental gradient, where three different environments (env1-env3) representing the 0-20%, 20-80%, and 80-100% quantiles of the CG effects, were defined. Accuracies for HOM and HET follow the same pattern across environments (Table 4). For CV1, CV4, CV5 and CV6, highest accuracies were observed for env2, representing intermediate environments. For CV2, CV3, and CV7 the highest accuracies were observed for animals in env1, representing extreme environments with low CG effects for age at slaughter. In general, accuracies were lowest for env3 (except for CV2). The highest accuracy across environments has been obtained for CV2 (0.748 for HOM in env1). Within HET, accuracies follow a similar pattern as for HOM: accuracy was

highest for intermediate environments (CV1, CV4, CV6) and lowest for env3 (CV1, CV3, CV6 and CV7). Across environments, accuracies for HET and HOM are almost equal for intermediate environments (env2, CV1, CV2, CV7). As shown in Table 4, GEBV were less biased in env2 (both HOM and HET), however, still over- and underestimation occurred.

Discussion

The objective of this study was to evaluate accuracies of a genomic reaction norm model allowing for heterogeneous SNP (co)variances applied to age at slaughter in an Irish dairy and beef crossbred population. The trait age at slaughter is of high relevance in terms of efficiency on the animal, herd and production system level. Reduction in age at slaughter can help to reduce methane emission and thereby contribute to address current environmental challenges in livestock production (Cromie, 2019). Age at slaughter is a new trait. Currently, no genetic and genomic evaluation for age at slaughter is implemented in Ireland (Cromie, 2019). The results in this study exploit large usable phenotypic and genetic variation for age at slaughter (Table 1). Average heritability across the environmental gradient was 0.264. This is in agreement with results presented by Berry et al. (2017), who estimated heritabilities of 0.26 and 0.23 based on a 2-step and 1-step linear mixed model, respectively.

Without genomic information, the best criteria to select young animals for replacement for age at slaughter will be the parent average (PA) estimated breeding value. Following Dekkers (1992) and Bijma (2012) we have derived the expected theoretical accuracy of selection based on the PA (Appendix), assuming a heritability of 0.264 (average across environments) and considering 2 offspring per dam as well as either 10, 100 and 1,000 offspring per sire. The accuracy of selection of an average animal in the Irish dairy and crossbred data set is expected to be 0.14 (for 10 offspring per sire) and 0.23 for offspring of a sire with 1,000 offspring (Table 1 in Appendix). The accuracies of GEBV obtained in this study were between 0.176 and 0.587 across all CV and independent of environment (Table 3) and between 0.212 and 0.548 for intermediate environments (Table 4). Comparing accuracies of selection based on PA and GEBV (Tables 3 and 4), shows a clear benefit of using genomic information for age at slaughter.

When GxE exists, selection of robust and resilient animals with flat reaction norms across the environmental gradient is desirable to ensure good quality performance across different environments (Strandberg et al., 2000). Estimated genetic parameters in subset 1 suggest the

existence of GxE interactions for age at slaughter to some extent. The average heritability of extreme environments, i.e. low and high CG effects for age at slaughter, is lower (0.231 for env1 and 0.214 for env3) compared to intermediate environments (0.295 for env2). Accuracies of GEBV were higher for intermediate environments (CV1, CV4, CV5, and CV6). In general, this would be expected given the higher average heritability of age at slaughter for env2 compared to extreme environments and most of the data was observed for intermediate environments (Table 4).The lowest genetic correlations between intermediate environments and extreme environments were 0.89 indicating that estimated breeding values of animals may be different in changing environments. Nevertheless, this range of observed genetic correlations between environments, i.e. all being 0.89 or greater, indicates that a single breeding program is sufficient to serve the range of environments included in this data (Mulder et al., 2006).

In this study, SNP effects for intercept and quadratic regression coefficient were derived based on a genomic reaction norm model in subset 1, assuming equal (co)variances for each SNP across the genome. Alternatives, as proposed by Wang et al. (2012) with the single-step GWAS, are to estimate SNP effects by an iterative approach, where equal (co)variances are assumed for all SNP in the first iteration and every next iteration uses SNP variances estimated in the previous iteration. The limitation of this approach is, that SNP effects may become very large leading to inflated SNP variances. In addition, there is no formal criteria for convergence. Zhang et al. (2016) use a maximum of 10 iterations. Recently, Zhang et al. (2019) applied single-step GWAS to evaluate genomic regions associated with the intercept and slope of a reaction norm model for fertility traits in Danish Holstein. To avoid very large SNP effects, they have used only three iterations following suggestions by Wu et al. (2018). Bayesian approaches to derive SNP effects to upweight SNP in a following GBLUP or single-step SNPBLUP analysis, have been suggested by Karaman et al. (2018), Su et al. (2018), and Karaman et al. (2020). However, no such Bayesian implementation is available for genomic reaction norm models. The analysis protocol allowing for heterogeneous SNP (co)variances developed in this study is based on readily available software allowing quick and large scale implementations.

The current approach has been previously tested in simulated data (Gredler-Grandl and Calus (2020) and Gredler-Grandl (2021) resulting in a small increase in accuracy of GEBV with models allowing for heterogeneous SNP (co)variances. An increase in accuracy between 0.01 and 0.03 has been observed when applied in genomic multi-trait and reaction norm models.

However, in the current study, when applied to a real data set, no advantage in accuracy of GEBV estimated with HET was observed. The accuracies of HET are either slightly lower or equal to HOM (Tables 3 and 4). Based on this result, we hypothesize that the genetic architecture of age at slaughter is highly polygenic with no really big QTL, therefore leaving little advantage of the proposed method. Until now, age at slaughter has not undergone strong selection, but indirect selection by selection for other correlated traits may have occurred, which may have caused fixation of any big QTL previously. Furthermore, the heterogeneous nature of the data set with different breeds and crossbred animals may make it difficult to take advantage of the proposed method. If QTL affecting age at slaughter exist, but are different or have different effects across breeds, polygenicity will be increased across breeds. Depending on the genetic architecture, the accuracy of GEBV for other traits may still benefit from modelling heterogeneous SNP (co)variances.

Conclusions

The genetic analysis of age at slaughter in an Irish dairy and beef crossbred population reveals large genetic variation for this trait. Genomic reaction models resulted in higher accuracies for young selection candidates compared to the expected accuracy of PA breeding values without genomic information. Thus, using genomic information has the potential to considerably increase the rate of genetic gain for age at slaughter, and thereby to increase efficiency and reduce environmental impact of beef cattle production. The genetic parameters obtained from the genomic reaction norm model support the existence of GxE for age at slaughter to some extent, indicating that breeding values of an animal may change along the environmental gradient. Compared to homogeneous SNP (co)variances, the genomic reaction norm model allowing for heterogeneous SNP (co)variances across the genome did not result in higher accuracy of GEBV.

Acknowledgement

We thank the Irish Cattle Breeding Federation for the interest in the study and the permission to use the data set. Alan Twomey and Donagh Berry (both Teagasc, Ireland) are acknowledged for preparation of the data and helpful discussions. This work has been carried out within the project GenTORE – Genomic management Tools to Optimize Resilience and Efficiency. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grand Agreement No 727213.


GENTORE

References

Berry, D.P., Cromie, A.R., and Judge, M.M. 2017. Rapid Communication: Large exploitable genetic variability exists to shorten age at slaughter in cattle. J. Animal Sci., 95:4526-4532.

Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129:345-358.

Bouwman, A.C., Hayes, B.J., and Calus, M.P.L. 2017. Estimated allele substitution effects underlying genomic evaluation models depend on the scaling of allele counts. Genet. Sel. Evol. 49, 79.

Calus M.P.L., Veerkamp R.F. 2003. Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. Journal of Dairy Science86, 3756–64

Calus, M. P. L., Schrooten, C., and Veerkamp, R.F. 2014. Genomic prediction of breeding values using previously estimated SNP variances. Genet. Sel. Evol. 46:52.

Calus M., and Vandenplas, J. 2016. Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices. ABGC, Wageningen UR Livestock Research.

Carvalheiro, R., Costilla, R., Neves, H.H.R., Albuquerque, L.G., Moore, S., and Hayes, B.J. 2019. Unravelling genetic sensitivity of beef cattle to environmental variation under tropical conditions. Genet. Sel. Evol. 51, 29.

Chung, Y., Lee, S.H., Lee, H.-K., Lim. D., van der Werf, J.H., and Lee, S.H. 2020. THI modulation of genetic and non-genetic variance components for carcass traits in Hanwoo cattle. Front. Genet. 11:576377.

Cromie, A. 2019. BDGP & BEEP; key initiatives that are helping deliver greater profit and sustainability for our suckler beef farmers. Teagasc National Beef Conference, 10th December, 2019, Claremorris, Ireland. Proceedings p18-23: https://www.teagasc.ie/media/website/publications/2019/Teagasc-National-Beef-Conference-2019.pdf Visited 11. May, 2021.

Dekkers, J. C. M. 1992. Asymptotic response to selection on best linear unbiased predictors of breeding values. Anim. Prod. 54:351-360.

de los Campos, G., Hickey, J.M, Pong-Wong, R., Daetwyler, H.D., and Calus, M.P.L. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.

Falconer, D. S. 1952. The problem of environment and selection. Am. Nat. 86:293-298.

Garcia, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta, and D. A. L. Lourenco. 2018. Development of genomic predictions for harvest and carcass weight in channel catfish. Genet. Sel. Evol. 50:66.

Gredler-Grandl, B., and Calus. M. (2020). Genomic models considering GxE: modelling heterogeneous SNP variances. Book of abstracts No. 26, p192. 71st Annual Meeting of European Federation of Animal Sciences, Virtual meeting, 1st – 4th December, 2020.

Gredler-Grandl, B., and Calus., M. (2021). Genomic GxE approaches modelling heterogeneous SNP variances: applied to simulated data. Interbull Open Meeting, Virtual meeting, 25th – 30th April 2021.

Karaman, E., Lund, M.S., Anche, M.T., Janss, L., and Su. G. 2018. Genomic prediction using multi-trait weighted GBLUP accounting for heterogeneous variances and covariances across the genome. G3. 8(11):3549-3558.

Karaman, E., Lund, M.S., Su, G. 2020. Multi-trait single-step genomic prediction accounting for heterogeneous (co)variances over the genome. Heredity 124:274-287.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta Agric. Scand. A Anim. Sci. 52:11-24.

Lee, S.H. and van der Werf, J.H. 2016. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32, 1420-1422.

Moghaddar, N., M. Khansefid, J. H. J. van der Werf, S. Bolormaa, N. Duijvesteijn, S. A. Clark, A. A. Swan, H. D. Daetwyler, and I. M. MacLeod. 2019. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Genet. Sel. Evol. 51:72.

Mota, L.F.M., Fernandes Jr., G.A., Herrera, A.C., Scalez, D.C.B., Espigolan, R., Magalhaes, A.F.B., Carvalheiro, R., Baldi, F., and Albuquerque, L.G. 2020. Animal Genetics 51, 210-223.

Mulder, H. A., R. F. Veerkamp, B. J. Ducro, J. A. M. van Arendonk, and P. Bijma. 2006. Optimization of dairy cattle breeding programs for different environments with genotype by environment interaction. J. Dairy Sci. 89:1740-1752.

Neeteson-van Nieuwenhoven, A.-M., P. Knap, and S. Avendaño. 2013. The role of sustainable commercial pig and poultry breeding for food security. Anim. Front. 3:52-57.

Ni, G., van der Werf, J.H., Zhou, X., Hyppoenen, E., Wray, N.R., and Lee, S.H. 2019. Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. Nat. Commun. 10:2239.

ten Napel, J., J. Vandenplas, M. Lidauer, I. Stranden, M. Taskinen, E. Mäntysaari, M. P. L. Calus, and R. F. Veerkamp. 2020. MiXBLUP, the Mixed-model Best Linear Unbiased Prediction software for PCs for large genetic evaluation systems – Manual V2.2 - 2020 - 05, Wageningen, the Netherlands www.mixblup.eu

Twomey, A.J., Berry, D.P., Evans, R.D., Doherty, M.L., Graham, D.A., and Purfield, D.C. 2019. Genome-wide association study of endo-parasite phenotypes using imputed whole genome sequence data in dairy and beef cattle. Genet. Sel. Evol. 51:51.

Saatchi, M., McClure, M.C., McKay, S.D., Rolf, M.M., Kim, J., Decker, J.E., Taxis, T.M., Chapple, R.H., Ramey, H.R., Northcutt, S.L., Bauck, S., Woodward, B., Dekkers, J.C.M., Fernando, R.L., Schnabel, R.D., Garrick, D.J., and Taylor, J.F. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation.

Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15:478.

Self, S. G. and K. Y. Liang. 1987. Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. 82:605-610.

Silva, F.F., Mulder, H.A., Knol, E.F., Lopes, M.S., Guimaraes, S.E.F., Lopes, P.S., Mathur, P.K., Viana, J.M.S., and Bastiaansen, J.W.M. 2014. Sire evaluation for total number of born in pigs using a genomic reaction norms approach. J. Anim. Sci. 92:3825-3834.

Strandberg, E., Kolmodin, R., Madsen, P., Jensen, J., Jorjani, H. (2000). Genotype by environment interaction in Nordic dairy cattle studied by reaction norms. Interbull Bulletin 25:41-45.

Su, G., Christensen, O. F., Janss, L., and Lund, M. S. (2014). Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *J. Dairy Sci.* 97, 6547–6559.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.

Veerkamp, R. F., J. E. Pryce, D. Spurlock, D. P. Berry, M. Coffey, P. Lovendahl, C. Van der Linde, J. Bryant, F. Miglior, Z. Wang, M. Winters, N. Krattenmacher, N. Charfedinne, J. Pedersen, and Y. de Haas. 2013. Selection on feed intake or feed efficiency: A position paper from gDMI breeding goal discussions. Interbull Bull. 47:15-22.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73-83.

Wu, P., Yang, Q., Wang, K., Zhou, J., Ma, J., Tang, Q., et al. 2018. Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. Genomics 110(3):171-179.

Zhang, X., D. Lourenco, I. Aguilar, A. Legarra, and I. Misztal. 2016. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. Front. Genet. 7:151.

Zhang, Z., Kargo, M., Liu, A., Thomasen, J.R., Pan, Y., and Su, G. 2019. Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. Heredity 123:202-214.

Table 1. Descriptive statistics for age at slaughter for all phenotyped animals (age all) and genotyped animals (age geno), yield deviations (YD), and contemporary group effects of the genotyped animals on the original (CG orig) and standardized scale (CG stand)

Item	N	mean	SD	MIN	MAX
Age all (d)	2,845,504	745.6	145.0	426	1095
Age geno (d)	14,665	746.7	123.5	427	1094
YD (d)	14,665	-3.5	52.0	-278.7	339.4
CG orig (d)	3,146	100.9	92.9	-195.1	441.4
CG stand	3,146	0	1	-3.2	3.8

Table 2. The number of animals and herds per cluster and the average breed proportion per cluster

Cluster	N animals	N herds	HO[1]	LM[1]	CH[1]	AA[1]	HF[1]	BB[1]
0	1,717	143	0.039	0.303	0.446	0.027	0.020	0.021
1	1,636	218	0.042	0.625	0.116	0.025	0.022	0.021
2	1,720	156	0.979	0.002	0.002	0.001	0.001	0.000
3	1,407	300	0.873	0.038	0.001	0.044	0.020	0.003
4	559	118	0.125	0.234	0.102	0.082	0.175	0.111
5	1,847	41	0.045	0.406	0.334	0.023	0.024	0.035
6	722	106	0.066	0.102	0.074	0.587	0.050	0.017
7	1,574	323	0.074	0.486	0.092	0.055	0.039	0.040
8	1,248	271	0.067	0.131	0.488	0.072	0.042	0.034
9	242	32	0.291	0.099	0.087	0.399	0.048	0.011
10	274	45	0.684	0.044	0.053	0.100	0.059	0.023
11	1,719	288	0.043	0.231	0.484	0.034	0.029	0.024

[1] HO=Holstein, LM = Limousine, CH=Charolais, AA = Aberdeen Angus, HF = Hereford, BB = Belgian Blue

Table 3 Number of animals in the training (Train) and validation (Val) sets, accuracy and bias of GEBV estimated with genomic reaction norm models allowing for homogeneous (HOM) and heterogeneous (HET) SNP (co)variances across the genome

	CV1	CV2	CV3	CV4	CV5	CV6	CV7
Number of animals in training and validation set							
Train	7,623	8,826	7,767	7,495	8,782	7,934	7,625
Val	1,719	516	1,574	1,847	559	1,407	1,717
Cluster for Val	11	9+10	7	5	4	3	0
Accuracy of GEBV[1]							
HOM	0.286	0.582	0.185	0.305	0.227	0.315	0.320
HET	0.267	0.587	0.176	0.282	0.183	0.301	0.301
Bias of GEBV[1]							
HOM	0.664	1.143	0.411	0.581	0.460	0.648	0.719
HET	0.584	1.108	0.356	0.505	0.334	0.565	0.644

[1] Accuracy and bias have been calculated across all environments that occur within the CV data set

Table 4. Number of animals in the validation set per environment (env1-env3), accuracy and bias of GEBV estimated for different environments (env1-env3) with genomic reaction norm models allowing for homogeneous (HOM) and heterogeneous SNP (co)variances across the genome

	CV1	CV2	CV3	CV4	CV5	CV6	CV7
Number of animals in the validation set for each environment							
env1	290	131	328	340	127	311	241
env2	1,015	354	964	1,158	336	882	989
env3	414	31	282	349	96	214	487
Accuracy of GEBV HOM							
env1	0.357	0.748	0.318	0.285	0.234	0.278	0.452
env2	0.363	0.539	0.212	0.322	0.262	0.379	0.375
env3	0.182	0.590	0.093	0.308	0.240	0.013	0.224
Accuracy of GEBV HET							
env1	0.266	0.745	0.328	0.245	0.117	0.278	0.480
env2	0.377	0.548	0.221	0.304	0.239	0.375	0.384
env3	0.140	0.644	0.068	0.282	0.292	0.003	0.143
Bias of GEBV HOM							
env1	0.507	1.210	0.565	0.333	0.479	0.411	0.905
env2	1.303	1.308	0.696	0.805	0.634	1.042	1.148
env3	0.242	0.575	0.129	0.450	0.294	0.019	0.316
Bias of GEBV HET							
env1	0.352	1.144	0.548	0.269	0.208	0.375	0.912
env2	1.297	1.293	0.701	0.726	0.581	0.980	1.142
env3	0.173	0.584	0.080	0.379	0.297	0.004	0.188

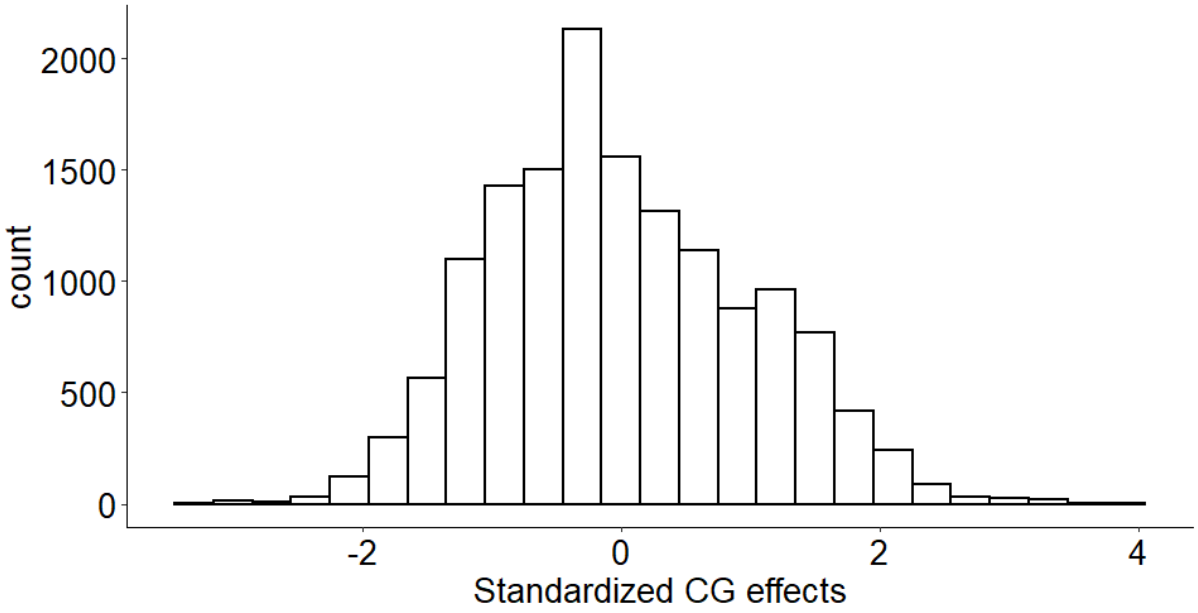


Figure 1. Frequency distribution of standardized CG effects

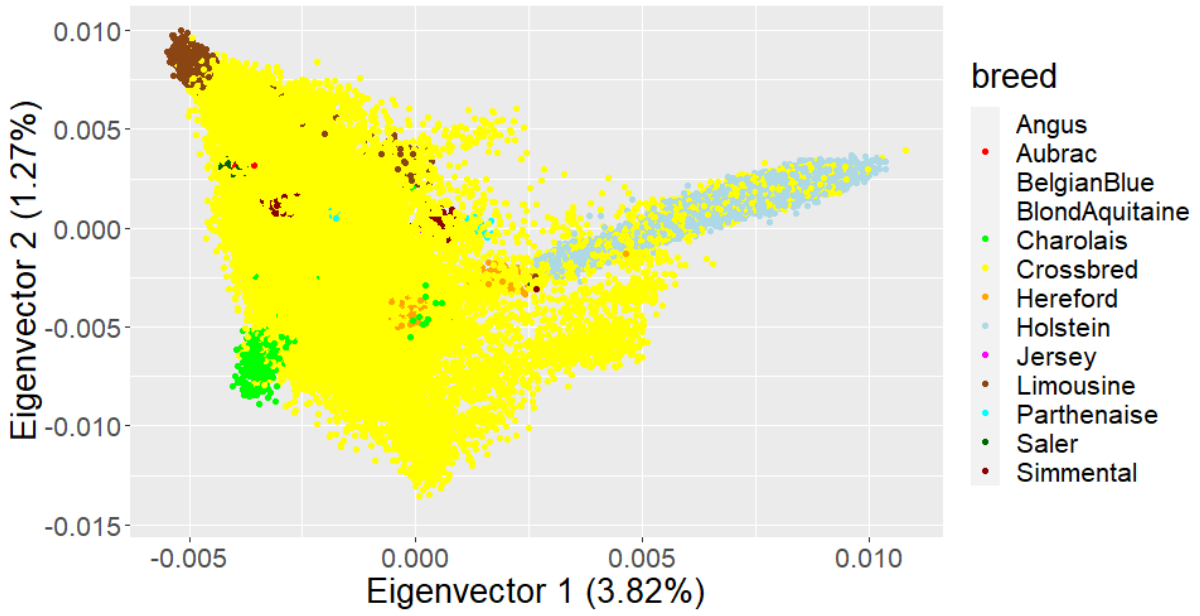


Figure 2. Population structure of genotyped animals (n=14,665) as determined by a PCA on the genomic relationship matrix. Yellow dots represent cross bred animals.

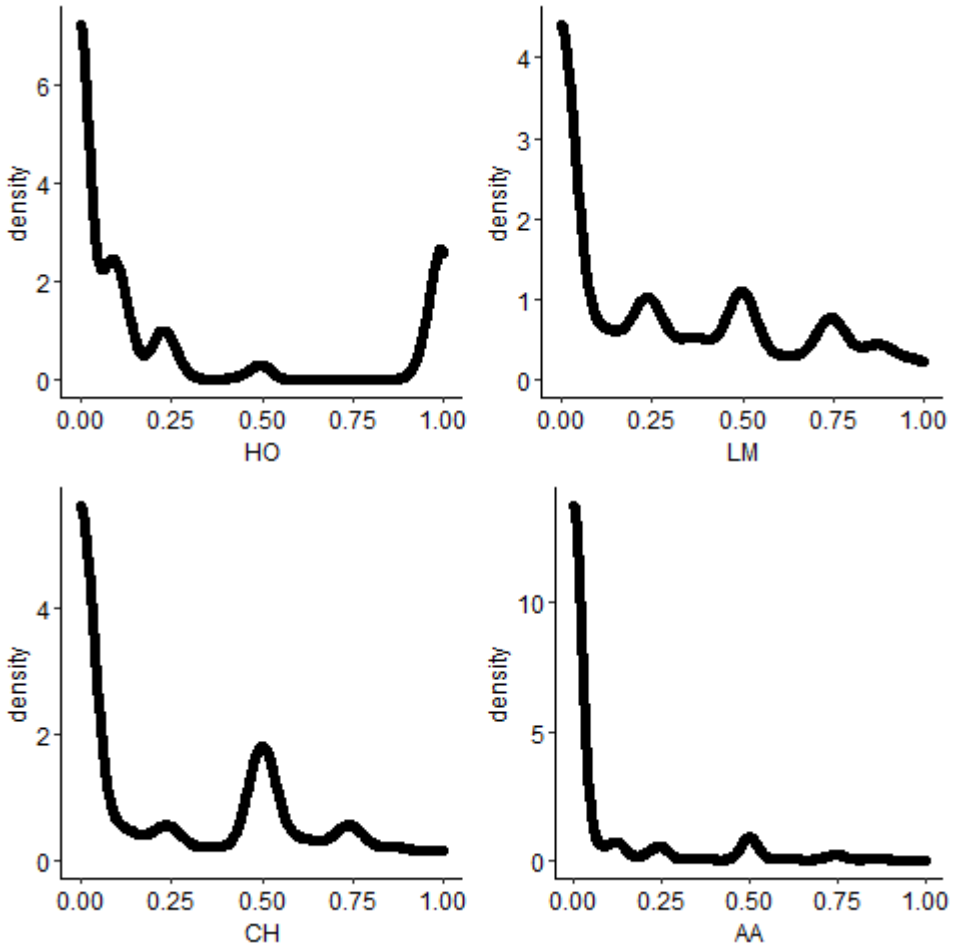


Figure 3. Density plot of gene proportion determined by pedigree information of the main breeds Holstein (HO), Limousine (LM), Charolais (CH) and Aberdeen Angus (AA)

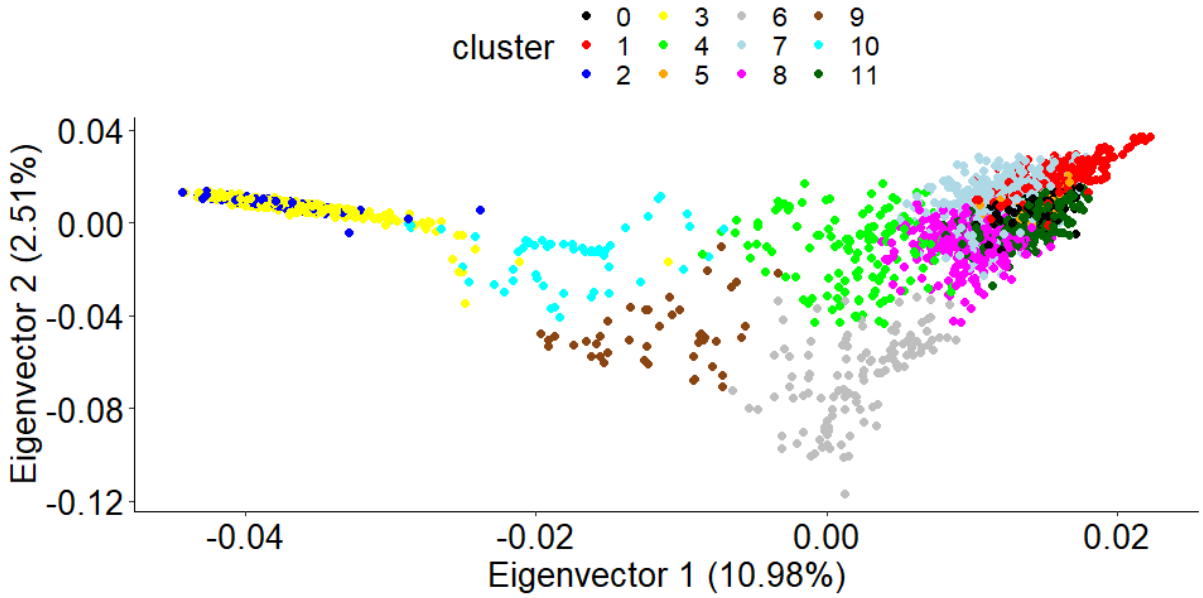


Figure 4. Herd structure as evaluated by a PCA on the genomic relationship matrix of the herds. Different colors represent the 12 clusters defined in the K-means clustering approach.

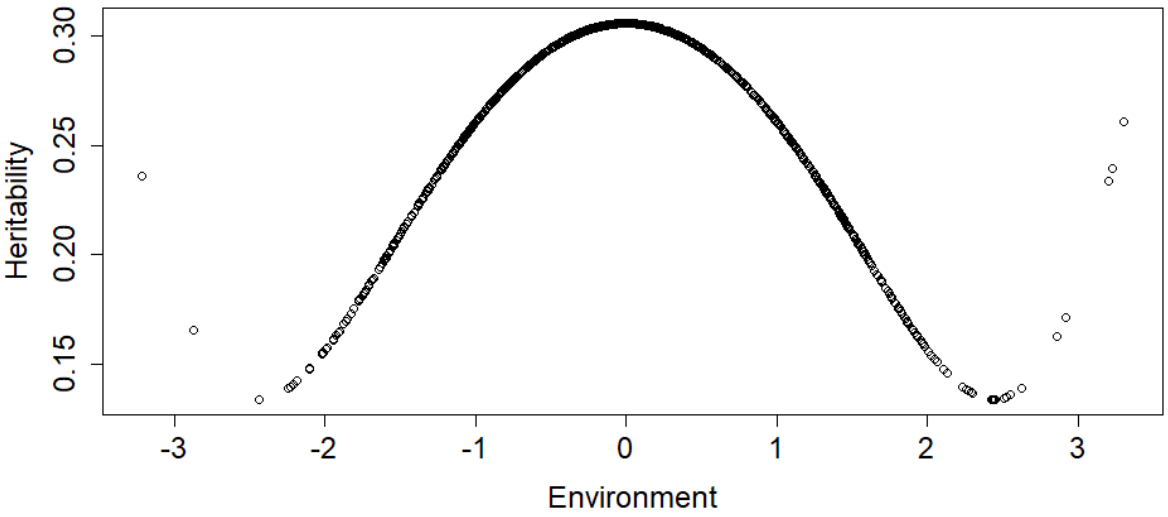


Figure 5. Heritability estimated with the genomic reaction norm model in subset 1 across different environments (CG effects)

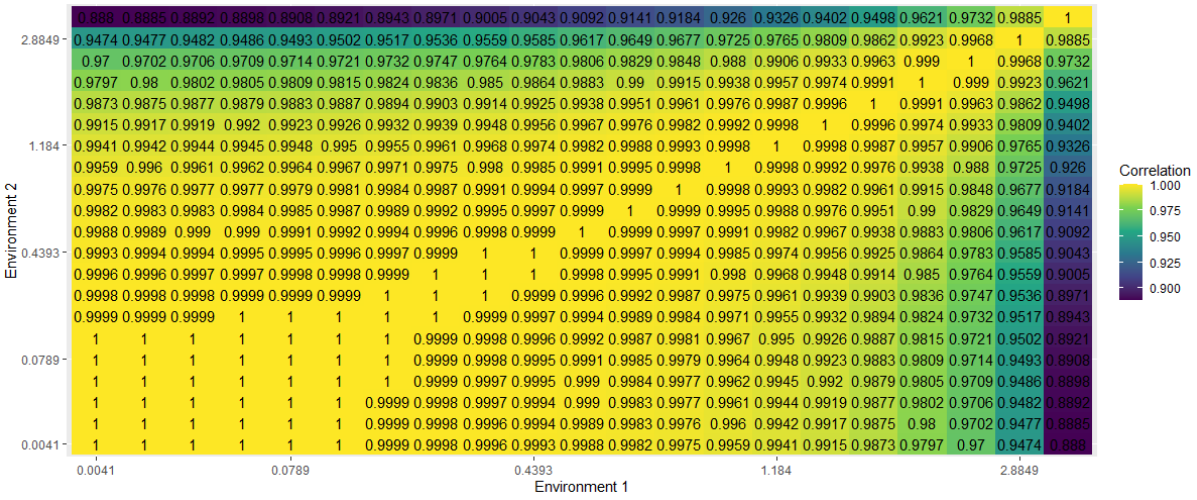


Figure 6: Genetic correlation matrix between environments in subset 1. Genetic correlation was calculated between each 50th environment. Environment is presented as the squared values of CG effects.

Appendix

If no genomic information is available, in most cases the parent average (PA) estimated breeding value (EBV) will be the best available criteria to select young animals for the trait age at slaughter. Here we derive the accuracy of selection based on PA, assuming a heritability of 0.264, reflecting an average environment, and considering that the EBV is based on 2 offspring for dams, and on 10, 100 or 1000 offspring for sires. Number of offspring being 2 for dams and 10 for sires reflect average values observed in the data. We computed equilibrium accuracies for PA ($\rho_{PA,\infty}$) (Dekkers, 1992; Bijma 2012), considering that the population is under selection. For this, we used formulas 8c, 9a, 9b and 10 in Bijma (2012). We assumed selected proportions of 0.2 for males and 0.75 for females. In reality the proportion of males selected in the population may be lower, but the derived selection intensity at best only partly applies to age at slaughter, either through direct selection for the trait or through indirect selection targeting correlated breeding goal traits. It should be noted that especially more intense selection in the males will reduce the resulting $\rho_{PA,\infty}$. With these assumed parameters, the contribution of the female to $\rho_{PA,\infty}$ was negligible, and therefore omitted from the calculations. The results are summarized in the Table below, and indicate that $\rho_{PA,\infty}$ is expected to be 0.14 for the average individual in the data analysed, and at best 0.23 for the offspring of a sire with an EBV of high accuracy (i.e. based on 1000 offspring).

Table 1. Description and values of all input and output parameters.

Symbol[1]	Description	$n_m = 10$	$n_m = 100$	$n_m = 1000$
p_m	Selected proportion males	0.2	0.2	0.2
i_m	Selection intensity males	1.400	1.400	1.400
k_m	Proportional reduction variance in males	0.781	0.781	0.781
n_m	Number of offspring per male	10	100	1000
$\rho_{m,0}$	Accuracy male (unselected population)	0.644	0.936	0.993
$\rho^2_{m,0}$	Reliability male (unselected population)	0.414	0.876	0.986
$\rho_{PA,0}$	Accuracy PA (unselected population)	0.322	0.468	0.497
$\rho^2_{PA,0}$	Reliability PA (unselected population)	0.104	0.219	0.247
$\rho_{m,\infty}$	Accuracy male (selected population)	0.591	0.918	0.991
$\rho^2_{m,\infty}$	Reliability male (selected population)	0.349	0.843	0.982
$\boldsymbol{\rho_{PA,\infty}}$	**Accuracy PA (selected population)**	**0.138**	**0.215**	**0.232**
$\rho^2_{PA,\infty}$	Reliability PA (selected population)	0.019	0.046	0.054

[1]Partly following notation in Bijma (2012). Subscript "0" denotes accuracies in an unselected population, while subscript "∞" denotes equilibrium accuracies in a population under selection.

Genomic models considering GxE: modelling heterogeneous SNP variances

Birgit Gredler-Grandl, Jeremie Vandenplas, Mario Calus (WLR/WUR)

EAAP virtual meeting 2020

GENTORE

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Genomic management **T**ools to **O**ptimise **R**esilience and **E**fficiency

GENTORE

- WP 4: Genomic indices for multi-breed selection in different environments
  - **Task 4.2: Genomic predictions across multiple environments**

Genomic GxE models

- Multi-trait approach:
  - Environments are divided in limited number of groups
  - Groups are considered different traits

- Reaction norm model:
  - Environment is defined with a continuous variable
  - Breeding values modelled as function of this variable

- Both can be implemented relying on:
  - Genomic relationships (GREML)
  - Random regression on SNP genotypes (RR-REML)

Assumptions SNP (co)variances

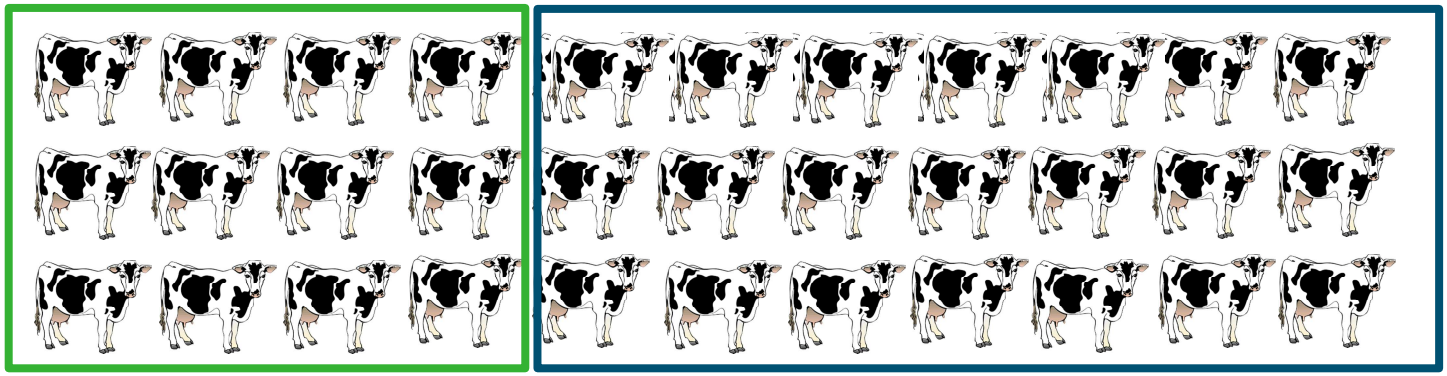
- GREML and RR-REML are equivalent
  - Homogeneous (co)variance assumed for all SNPs
- Certain regions in genome may harbour QTL → assumption of equal (co)variances is violated
- Can we model heterogeneous (co)variances?

Model heterogeneous SNP variances

- Make SNP (co)variances heterogeneous by **weighing**
  - (1) Weights derived from estimated SNP effects
  - (2) Re-compute SNP-effects using those weights

- Issue: computing (1) & (2) from the same data may inflate large SNP-effects

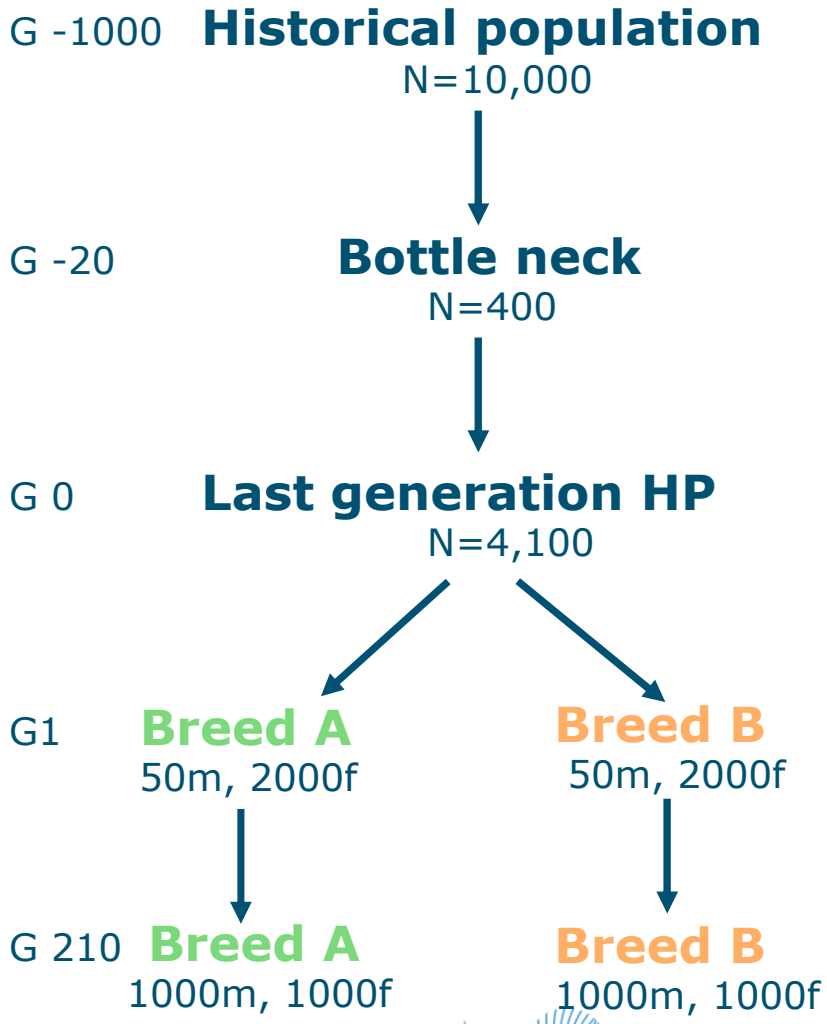
Proposed solution: split data in two



- Estimate SNP-effects assuming equal (co)variances for all SNP
- Calculate SNP specific weights within environment

- Estimate GEBV using the 2nd subset, applying weights on SNP (co)variance matrix within environment

Simulation (1) (QMSim, Sargolzaei and Schenkel, 2009)

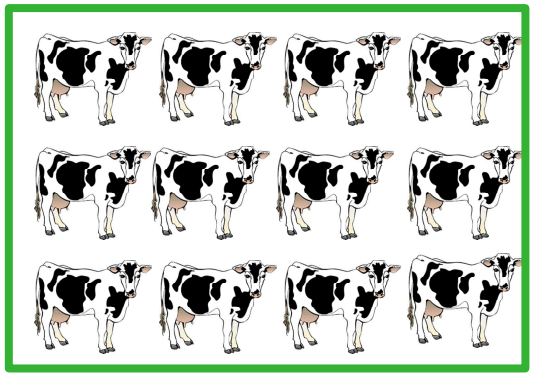


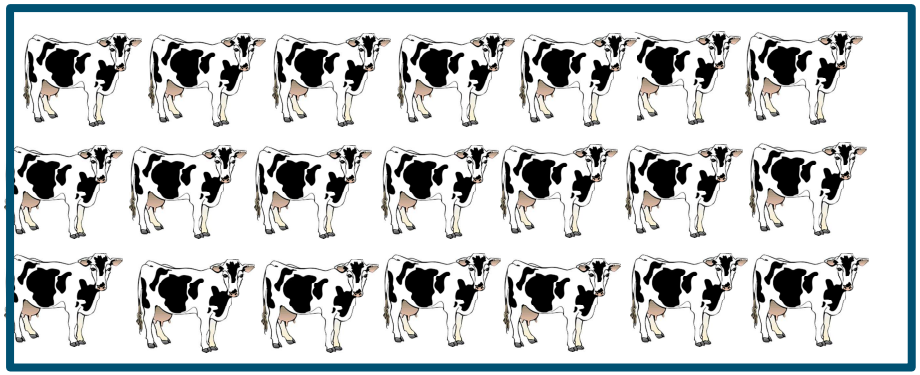
- Random mating and selection
- Genome 30 Chr
- 100 cM length
- 1700 markers per Chr
- 150 QTL per Chr
- ~ 51,000 markers
- ~ 4,500 QTL
- 5 replicates

Simulation of phenotypes

- Phenotypes follow multivariate and reaction norm models
- Input: environmental values, genetic & residual (co)variances
- QTL-effects are simulated for QTLs simulated in QMSim
- TBV: sum QTL and polygenic effect
- Phenotype: TBV*environmental value + polygenic effect + residual error

Validation study – bi-variate approach





Data set 1
Gen 205 + 206
4000 individuals
2000 Env1
2000 Env2

Data set 2 Training
Gen 207 + 208
4000 individuals
2000 Env1
2000 Env2

Data set 2 validation
Gen 209 + 210
4000 individuals
2000 Env1
2000 Env2

Validation study – bi-variate approach

Data set 1
Gen 205 + 206
4000 individuals

2000 Env1 | 2000 Env2

- GREML (mtg2)

$$y = \mu + Za + e$$

$$var(a) = \begin{bmatrix} \sigma^2_{a1} & \sigma_{a1a2} \\ \sigma_{a1a2} & \sigma^2_{a2} \end{bmatrix} \otimes \mathrm{G}$$

- Backsolve SNP-effects (calc_grm)
- Calculate weights as:
- $\sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$

Validation study – bi-variate approach

Data set 2 Training
Gen 207 + 208
4000 individuals

2000 Env1 | 2000 Env2

Data set 2 validation
Gen 209 + 2010
4000 individuals

2000 Env1 | 2000 Env2

- SNP-BLUP (hpblup)

$$y = \mu + Zg + e$$

$$\sigma_g^2 = \sigma_a^2 / 2\Sigma p(1-p)$$

- Apply weights (D) on SNP (co)variance matrix:

- $\sqrt{D_{kk_i}} = \sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$

Assumed heritability and genetic correlation between traits (environments)

	Trait 1	Trait 2
Trait 1	0.23	
Trait 2	0.78	0.30

Results: Estimated heritability and genetic correlation between traits in data set 1

	Trait 1	Trait 2	Trait 1	Trait 2
Trait 1	0.23		0.218	
Trait 2	0.78	0.30	0.762	0.332

Results: Correlation between estimated GBV and TBV within environment

	Homogeneous SNP (co)variance	Heterogeneous SNP (co)variance
Trait 1 – env1	0.566	0.582
Trait 2 – env2	0.605	0.622

Results: Correlation between estimated GBV and TBV across environments

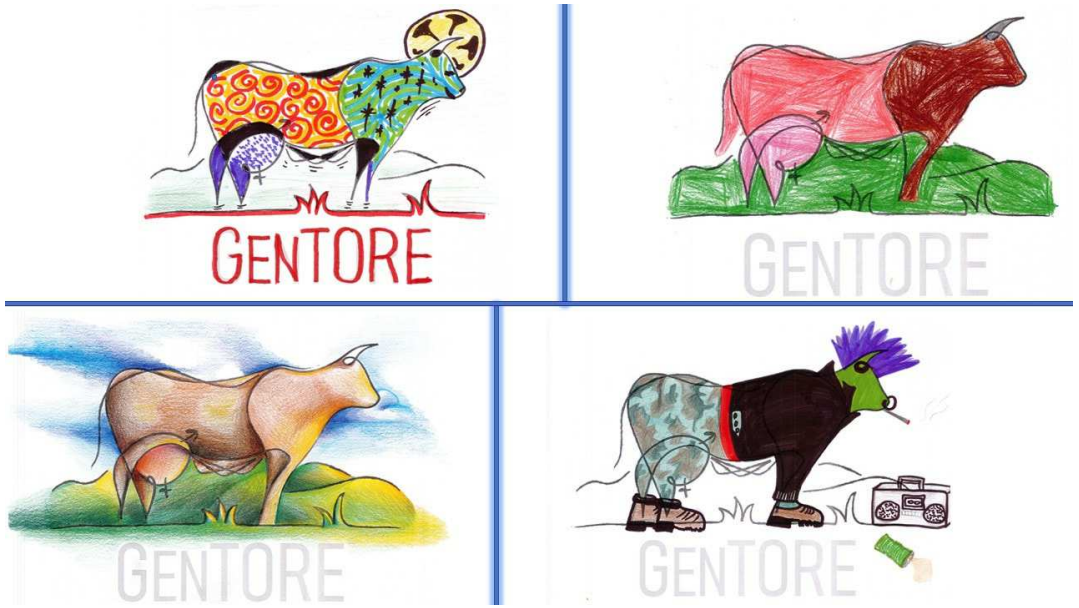
- How well can we predict the breeding value of individuals in environment 1(2) for environment 2(1)?

	Homogeneous SNP (co)variance	Heterogeneous SNP (co)variance
Trait 1 → Trait 2	0.559	0.576
Trait 2 – env2	0.605	0.622
Trait 2 → Trait 1	0.556	0.579
Trait 1 – env1	0.566	0.582

Summary

- Analysis protocol to model heterogeneous SNP variances developed
- Slight increase in accuracy with heterogeneous SNP variances in a multi-variate approach
- Accuracy across and within environment prediction almost equally high
- Test next in reaction norm models

Thank you!



This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 727213

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Genomic GxE approaches modelling heterogeneous SNP variances: applied to simulated data

Birgit Gredler-Grandl and Mario Calus (WUR)

Virtual Interbull Meeting 2021

GENTORE

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Genomic GxE models

- Multi-trait approach and reaction norm model can be implemented relying on:
  - Genomic relationships (GREML)
  - Random regression on SNP genotypes (RR-REML)

- GREML and RR-REML are equivalent
  - Homogeneous (co)variance assumed for all SNPs

- Certain regions in genome may harbour QTL → assumption of equal (co)variances is violated

Genomic GxE models

Can we model heterogeneous SNP (co)variances and do those models improve accuracy of genomic prediction?

Model heterogeneous SNP variances

- Make SNP (co)variances heterogeneous by **weighing**
  - (1) Weights derived from estimated SNP effects
  - (2) Re-compute SNP-effects using those weights

- Issue: computing (1) & (2) from the same data may inflate large SNP-effects

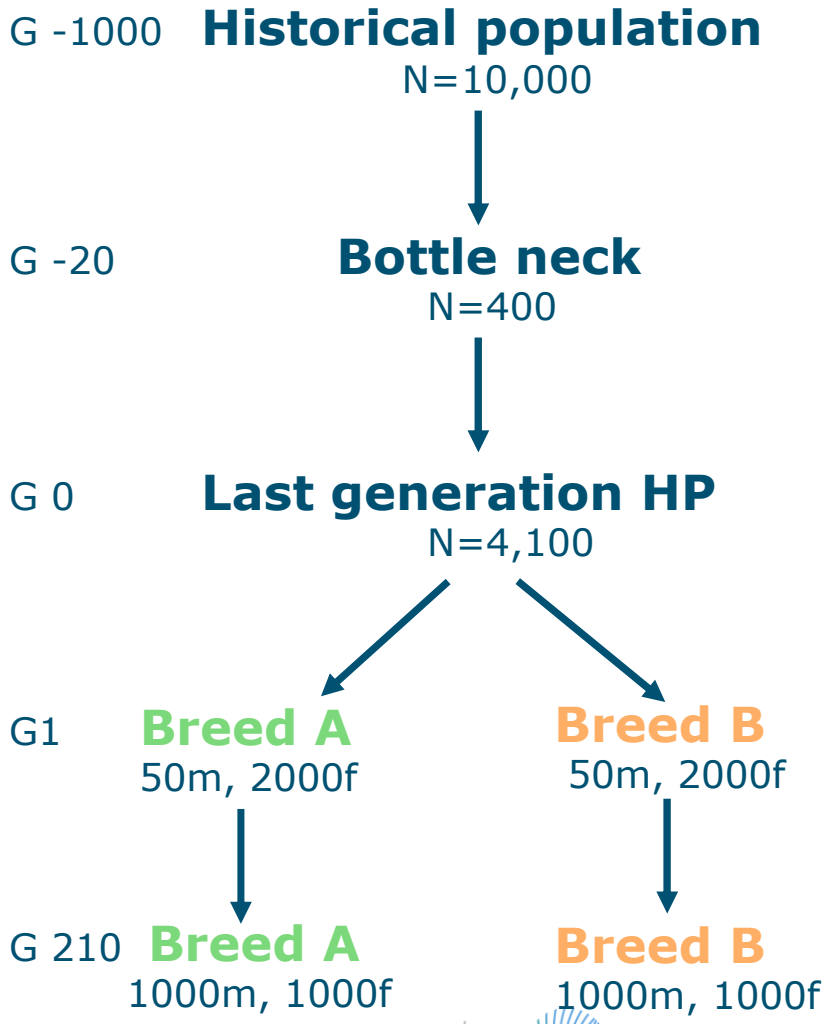
Proposed solution: split data in two

- Estimate SNP-effects assuming equal (co)variances for all SNP
- Calculate SNP specific weights within environment

- Estimate GEBV using the 2nd subset, applying weights on SNP (co)variance matrix within environment

picture source: freepik.com

Simulation (1) (QMSim, Sargolzaei and Schenkel, 2009)



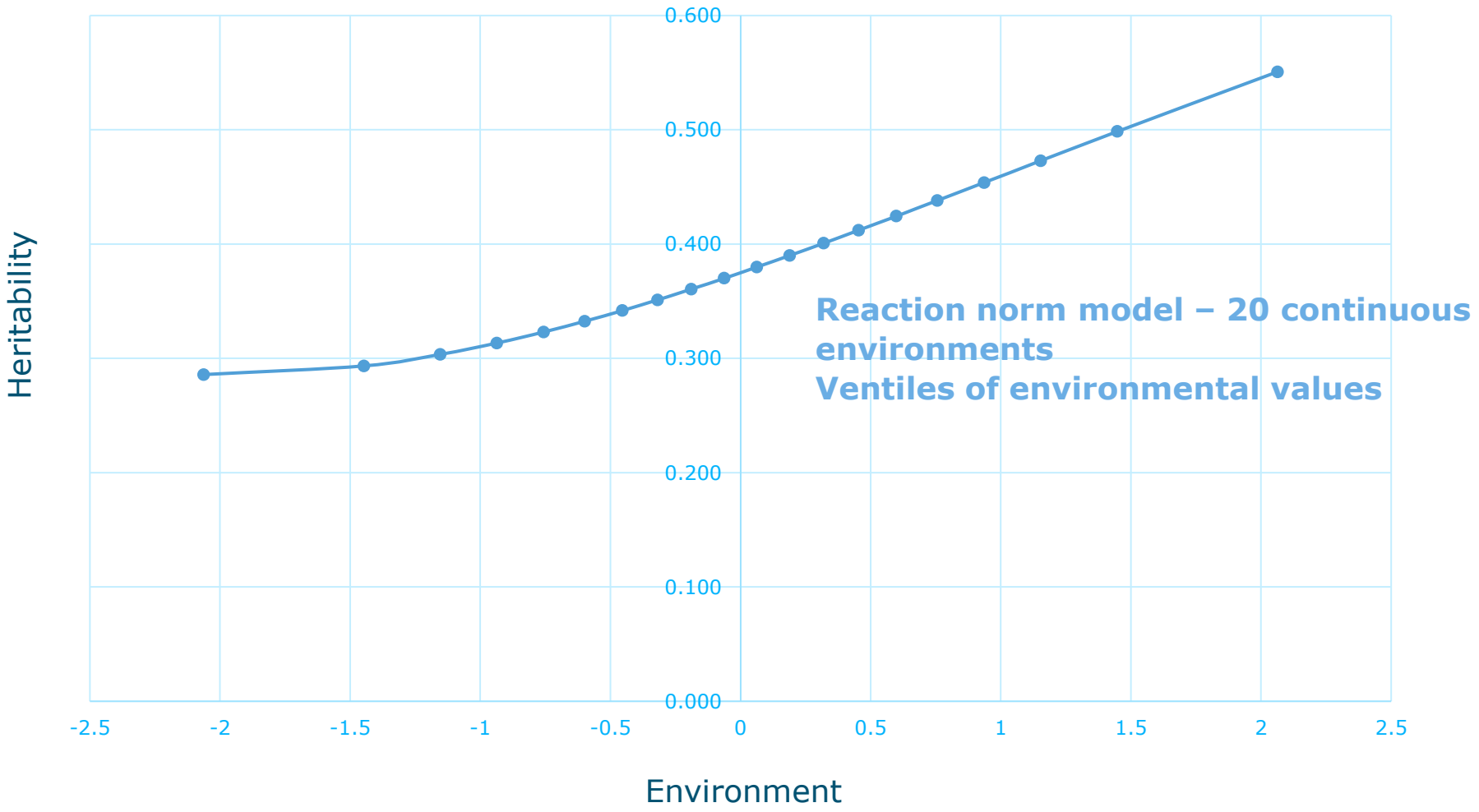
- Random mating and selection
- Genome 30 Chr
- 100 cM length
- 1700 markers per Chr
- 150 QTL per Chr
- ~ 51,000 markers
- ~ 4,500 QTL
- 5 replicates

Simulation of phenotypes

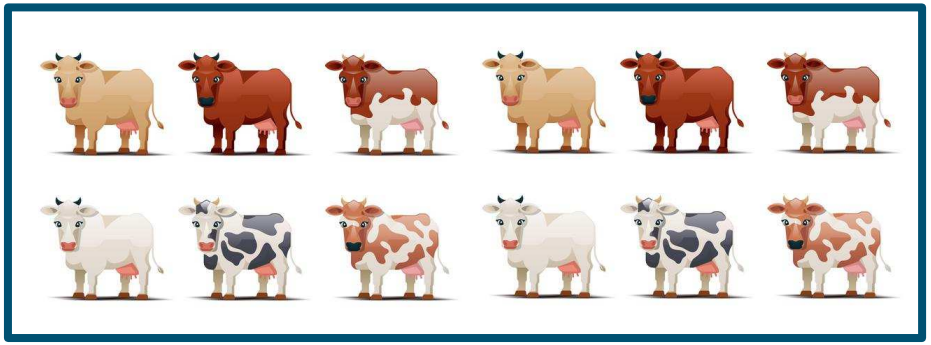
- Phenotypes follow a reaction norm model
- Input: environmental values, genetic & residual (co)variances
- QTL-effects are simulated for QTLs simulated in QMSim
- Phenotype: environmental value * TBV + residual error

Gen cov matrix Reaction norm model		
	b_0	b_1
b_0	0.3	
b_1	0.05	0.025
Environmental variance 0.5		

Heritability across environments



Validation study



Data set 1
Gen 205 + 206
4000 individuals

Data set 2
Training
Gen 207 + 208
4000 individuals

Data set 2
validation
Gen 209 + 210
4000 individuals

Individuals in each generation are randomly assigned to environments

Model Data set 1

- Reaction norm model (mtg2)

$$y = \mu + \beta_0 + \beta_1 * x + e$$

- Backsolve SNP-effects (calc_grm)
- Calculate weights as:
- $\sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$

Model Data set 2

- SNP-BLUP (MiXBLUP)

$$y = \mu + Z\beta_0 + ZQ\beta_1 + e$$

- Apply weights (D) on SNP (co)variance matrix:

- $\sqrt{D_{kk_i}} = \sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$

- $Var([\beta_0, \beta_1]') = \sqrt{D_{kk_i}}$ G $\sqrt{D_{kk_i}}$

Results: Estimated genetic covariance matrix for b_0 and b_1 in data set 1

	b_0	b_1	b_0	b_1
b_0	0.3		0.35	
b_1	0.05	0.025	0.04	0.031

Results: Correlation between estimated GBV and TBV for b_0 and b_1

	Homogeneous SNP (co)variance	Heterogeneous SNP (co)variance
b_0	0.521	0.551
b_1	0.588	0.601

Application in Irish beef crossbred data set

- Trait: age at slaughter (Berry et al., 2017)
- 14,668 genotyped bulls, steers, heifers
- HD imputed genotypes (662,011 SNPs)
- Yield deviation as phenotypes
- CG-effects as continuous descriptor of environment

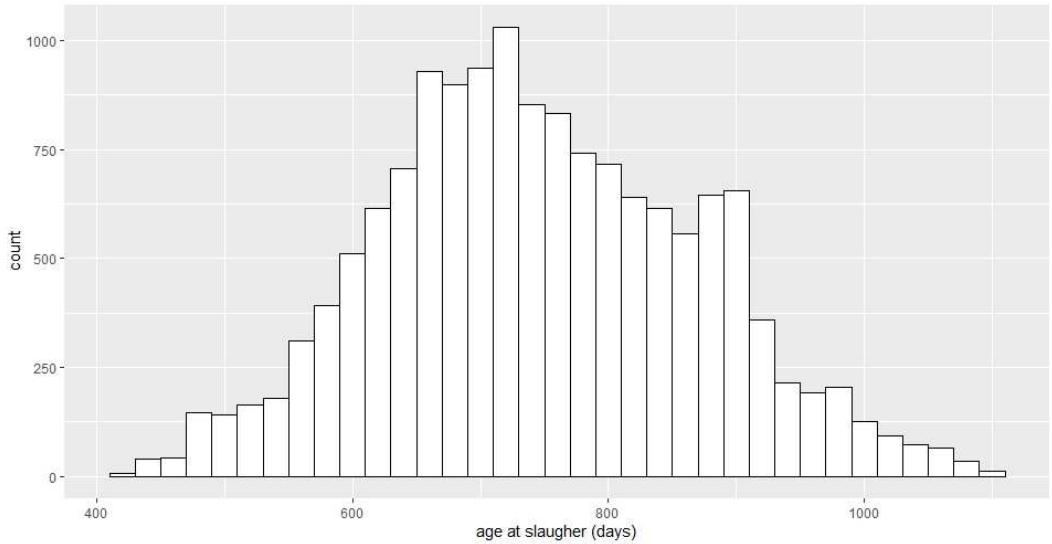
Age at slaughter in days

mean = 746.7

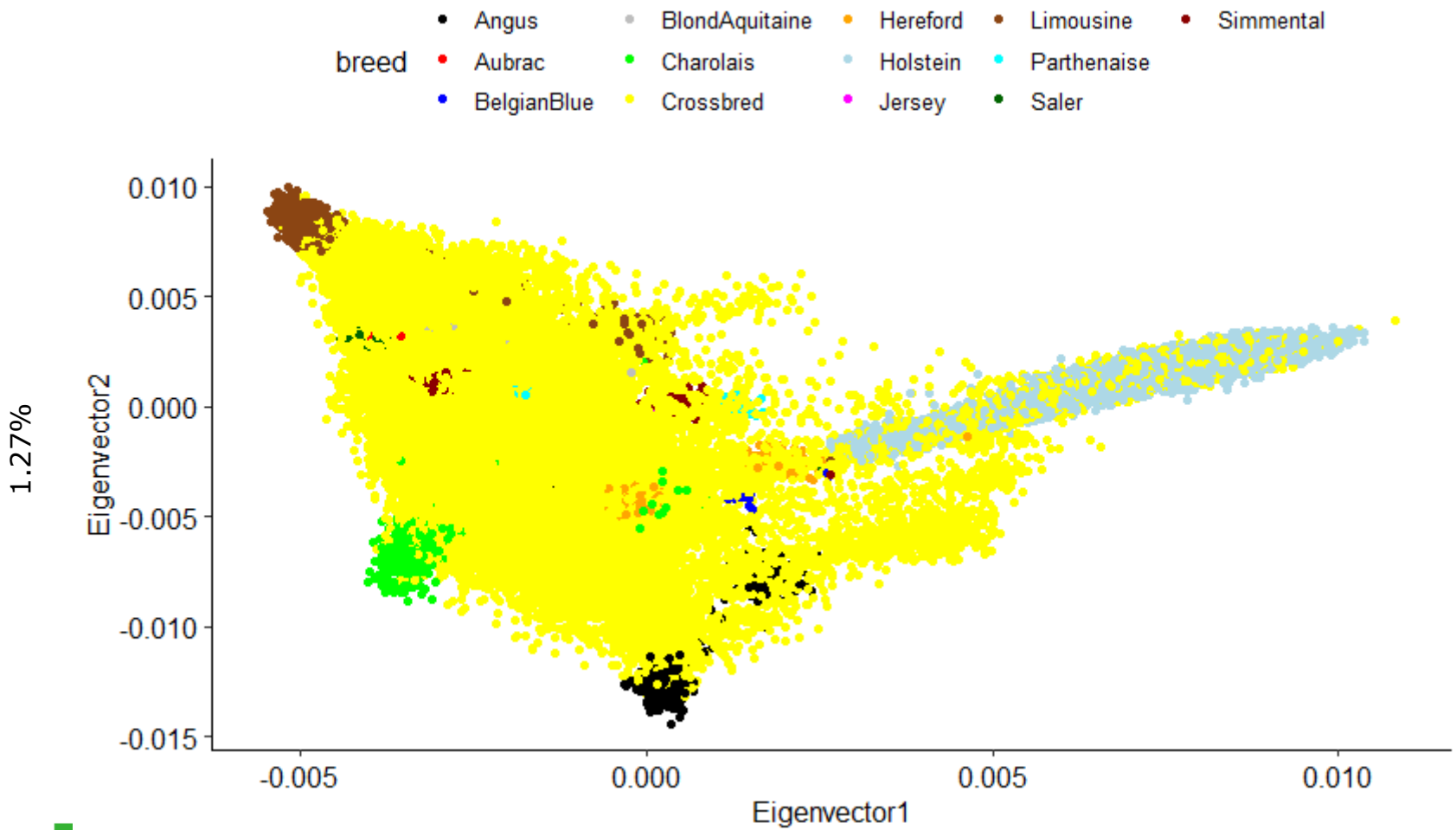
sd = 123.5

min = 427.0

max = 1094.0



Breeds: PCA G-Matrix purebred and crossbred animals

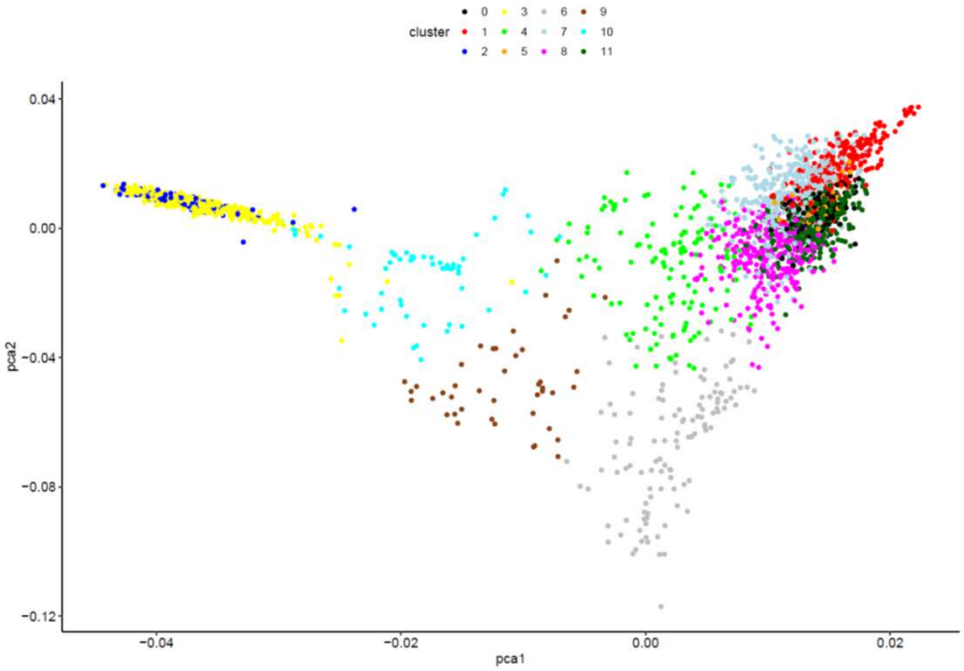


How to define sets for analysis?

- K-means clustering approach (similar Saatchi et al., 2011)
- Distance matrix between individuals computed as follows:

$$d_{ij} = 1 - \frac{a_{ij}}{\sqrt{a_{ii} \cdot a_{jj}}}$$

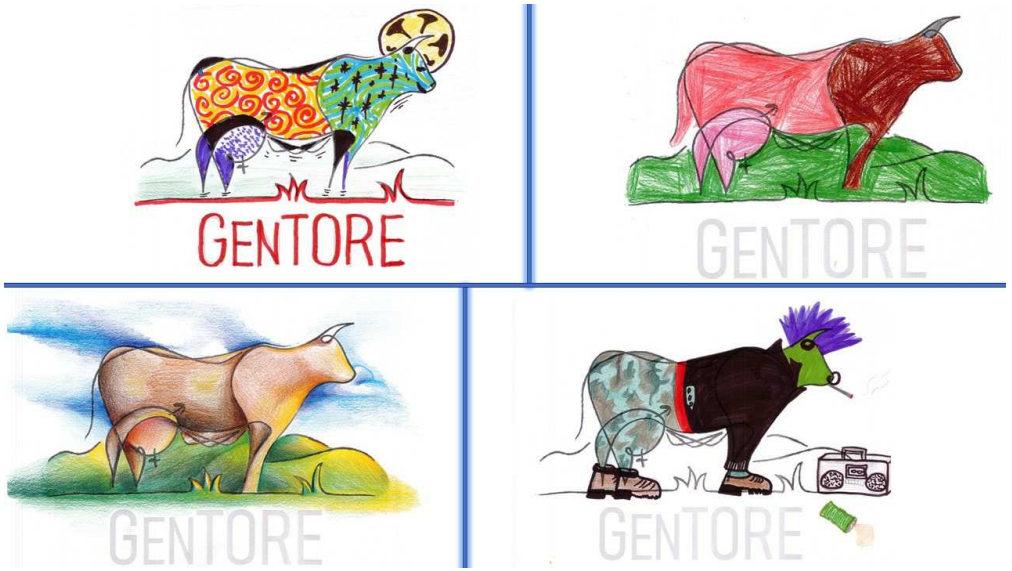
- Apply on herds
- Set up GRM for herds
- Define sets according to cluster results



Summary

- Analysis protocol to model heterogeneous SNP variances developed
- Slight increase in accuracy with heterogeneous SNP variances in reaction norm models in simulated data
- Currently investigating added value in real data

Thank you!



Alan Twomey and Donagh Berry

This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 727213