A method for partitioning trends in genetic mean and variance to understand breeding practices

T. P. Oliveira[1*], J. Obšteter[2], I. Pocrnic[1], G. Gorjanc[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, EH25 9RG, Edinburgh, Scotland, UK;

[2]Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia;

[*] thiago.oliveira@ed.ac.uk

Abstract

Quantifying the sources of genetic change is essential for optimising breeding programmes. However, breeding programmes are often complex because many breeding groups are subject to different breeding actions. Understanding the contribution of these groups to changes in genetic mean and variance is essential to understanding genetic change in breeding programmes. Here we extend the previously developed method for analysing the contribution of groups to changes in genetic mean to analysing changes in genetic variance. We, expectedly, show that the contribution of females and males to change in genetic variance can differ and are not independent, indicating we should not look at the contributions in isolation.

Introduction

Quantifying sources of genetic change is essential for identifying key breeding actions and optimising breeding programmes. Due to the complexity of many breeding programmes, the observed genetic change is a sum of contributions from different groups of individuals (often referred to as selection pathways). These groups vary in their contribution due to different selection intensity, accuracy, genetic variation, generation interval, and dissemination. At the individual level, the breeding value of an individual can be partitioned as $a_i = \frac{1}{2}a_s + \frac{1}{2}a_m + w_i$, where a_i is the breeding values of the individual, a_s and a_m are breeding values of sire and dam, and w_i is the Mendelian sampling term. García-Cortés et al. (2008) provided a methodology to allocate these partitions of breeding values into group-specific contributions to the change in genetic mean. To gain insight into time-trends or other data partitions, they summarised group-specific contributions by a variable such as year, gender, population, etc. While García-Cortés et al. (2008) addressed the contribution of different groups to changes in genetic mean, breeding programmes should also consider the contribution of different groups to changes in genetic variance to fully understand drivers of genetic change in a population (Sorensen et al., 2001; Lara et al., 2021). Here we extend the method of García-Cortés et al. (2008) to analyse changes in genetic variance. We develop the method and demonstrate it with a simulation. We assume throughout we know the true breeding value. Future work will extend the method to estimated breeding values by applying the method on posterior samples of breeding values following the work of Sorensen et al. (1994; 2001) and Lara et al. (2021).

Material and Methods

Statistical Framework. Let $\mathbf{a}$ be a vector of breeding values from a normal distribution with mean $\mathbf{0}$ and covariance $\mathbf{A}\sigma_a^2$. $\mathbf{a}$ can be written as a linear combination of breeding values from ancestors and individual's deviation from the ancestral expectation $\mathbf{a} = \mathbf{T}\mathbf{w}$, where $\mathbf{T}$ is a triangular matrix of expected gene flow between ancestors and individuals and $\mathbf{w} \sim \mathrm{N}(\mathbf{0}, \mathbf{W}\sigma_a^2)$ are Mendelian sampling terms representing deviations, with $\boldsymbol{W}$ being a diagonal matrix of variance coefficients and σ_a^2 the base population additive genetic variance (Henderson, 1976).

Assuming a factor with p groups and for any set $\sum_{j=1}^{p} \mathbf{P}_j = \mathbf{I}$, García-Cortés et al. (2008) partitioned the gene flow into contributions of each group by defining $\mathbf{T}_j = \mathbf{T}\mathbf{P}_j$, and further partitioned the contribution of each group to breeding values *a priori* using the equality $\mathbf{a} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\mathbf{w} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_p$. They further showed that these partitions can be estimated from data collected in breeding programmes (*posteriori*) by first estimating breeding values $\widehat{\boldsymbol{a}} = E(\boldsymbol{a}|\boldsymbol{y})$ from phenotype data $(\boldsymbol{y})$ and then partitioning $\hat{\mathbf{a}} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\widehat{\mathbf{w}} = \hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2 + \cdots + \hat{\mathbf{a}}_p$. By summarising these partitions, they quantified the contribution of each group (for example, males vs. females) to the time-trend in genetic mean, where time-trend represents a change in genetic mean over time. This method has been implemented in the AlphaPart R package (Obšteter et al., 2021), which efficiently calculates (via $\boldsymbol{T}^{-1}$) and summarises such partitions by one or more variables, such as generation, selection status, etc. AlphaPart enables use of any function for the summarisation, that is, $f(\boldsymbol{a}_j)$.

Here we extend the partitioning method to analyse the contribution of groups to genetic variance. Variance of breeding values is, *a priori*, $Var(\boldsymbol{a}) = Var(\boldsymbol{T}\boldsymbol{w}) = \boldsymbol{T}\boldsymbol{W}\boldsymbol{T}^T\sigma_a^2$. Using the equality $\mathbf{a} = (\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\mathbf{w} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_p$, we can partition genetic variance as $Var(\boldsymbol{a}) = Var\left((\mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_p)\boldsymbol{w}\right) = Var(\mathbf{T}_1\boldsymbol{w}) + Var(\mathbf{T}_2\boldsymbol{w}) + \cdots + Var(\mathbf{T}_p\boldsymbol{w}) + 2Cov(\mathbf{T}_1\boldsymbol{w}, \mathbf{T}_2\boldsymbol{w}) + \cdots + 2Cov(\mathbf{T}_p\boldsymbol{w}, \mathbf{T}_{p-1}\boldsymbol{w}) = \sum_{j=1}^{p} \mathbf{T}_j\boldsymbol{W}\boldsymbol{T}_j^T\sigma_a^2 + 2\sum_{j=1}^{p-1}\sum_{j'=j+1}^{p} \mathbf{T}_j\boldsymbol{W}\boldsymbol{T}_{j'}^T\sigma_a^2 = \sum_{j=1}^{p}\sigma_{a_j}^2 + 2\sum_{j=1}^{p-1}\sum_{j'=j+1}^{p}\sigma_{a_j,a_{j'}}$. While this "theoretical" partitioning involves matrix products, we can also summarise partitions $\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_p$ (calculated via $\boldsymbol{T}^{-1}$) by calculating variance of each partition $f(\boldsymbol{a}_j) = Var(\boldsymbol{a}_j)$ and covariance of each pair of partitions $f(\boldsymbol{a}_j, \boldsymbol{a}_{j'}) = Cov(\boldsymbol{a}_j, \boldsymbol{a}_{j'})$. The partitions can be summarised in many ways, for example, to quantify the contribution of males and females to change in genetic variance over time. We implemented this method in the AlphaPart R package (Obšteter et al., 2021) with the summary method: `summary(part, by = 'time', FUN = var)`, where `part` is an object with group-partitioned breeding values. This partitioning is for *a priori* or true breeding values. To estimate partitions of genetic variance from data collected in a breeding programme (*posteriori*) we would have to use methods from Sorensen et al. (2001) and Lara et al. (2021). While this is straightforward, we left this to future work.

Simulation. To illustrate the method, we have simulated a breeding programme with 1,000 individuals over 40 generations. For the burning phase, generations -20 to 0, we selected the best 5 males (out of 500) as sires based on their phenotype and mated them with all 500 females from the previous generation and all 500 females from the current generation. These matings produced 1,000 selection candidates for the next generation. After the burning phase, we tested two selection scenarios over additional 20 generations: we selected 5 best males from 500 male

candidates based on their phenotypes (scenario "phenotype") or true breeding values (scenario "TBV"). The simulation was done with AlphaSimR version 1.0 (Gaynor et al., 2020) using cattle genome, where we assumed 30,000 quantitative trait loci (QTL) with additive effects sampled from a normal distribution, and phenotype with heritability 0.3. With this simulation, we aimed to show how trends in genetic mean and variance come about due to the selection of males. We split $\mathbf{T}$ by using $\mathbf{P}_m + \mathbf{P}_f = \mathbf{I}$ and $\mathbf{P}_m^s + \mathbf{P}_m^n = \mathbf{P}_m$, where $\mathbf{P}_m$ is a diagonal matrix with ones in rows for males and zeros otherwise; $\mathbf{P}_f = \mathbf{I} - \mathbf{P}_m$; $\mathbf{P}_m^s$ is a diagonal matrix with ones in rows for selected males, and $\mathbf{P}_m^n = \mathbf{P}_m - \mathbf{P}_m^s$is a diagonal matrix with ones in rows for non-selected males.

Results

Figure 1A shows time-trends in the genetic mean and variance and its partitioning by males and females, while Figure 1B represents trends partitions by selected males, non-selected males, and females. We scaled the genetic mean and variance of the base population respectively to zero and one to facilitate interpretation.

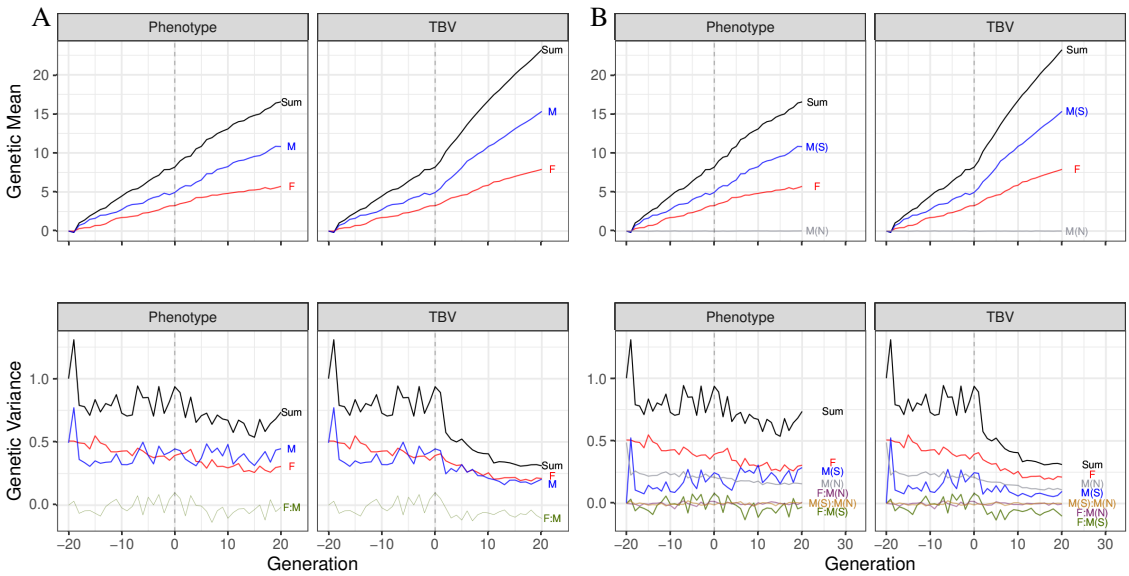

Figure 1 Partitioning of the total genetic mean and variance (Sum) over generation by (A) gender (male (M) and female (F)) and (B) gender and status (selected males (S) and non-selected males (N)). Each sub-figure shows the scenario with selection on phenotype (Phenotype, left) and selection on true breeding value (TBV, right).

Figure 1A shows that selection of males contributed almost twice as much as females, which contributed only via dissemination of past male selection. However, if only sex is used to investigate the contribution of the males and females to the genetic variance, we can see that the contributions of males and females are very similar for both selection scenarios. In fact, males seem to be contributing a bit more to genetic variance than females in the phenotypic selection scenario. This observation raises a question, how can males contribute similar proportions to the genetic variance as females if we are selecting among males? The answer to this question is demonstrated in Figure 1B. Although non-selected sires do not contribute to the change in the genetic mean, they still contribute to the genetic variance. Selection on true breeding values clearly showed that the contribution of groups to genetic variance is a function of the intensity of selection. Selected males contributed least to genetic variance, followed by

non-selected males and females. Partitioning in the scenarios with phenotype selection showed a similar result, but surprisingly selected males had a higher or comparable contribution to genetic variance as non-selected males.

Splitting the male contribution into selected and non-selected also shows that the negative covariance between male and female contributions in Figure 1A is driven by the covariance between female and selected male contributions (F:M(S)). The latter was consistently negative from generation 2 to 20 in the scenario with selection on true breeding values, resulting in a mean correlation of -0.52 (±0.12). This negative correlation means that the contribution of females and selected males to the genetic variance are not independent. That is, the total genetic variance is smaller than the sum of individual contributions.

Discussion

We presented a simple method for partitioning the trends in genetic mean and variance into contributions from different groups as an extension of the work of García-Cortés et al. (2008). We implement this method in the AlphaPart R package available from CRAN. This method can be a powerful and valuable tool to understand how different breeding groups contribute to genetic mean and variance change. We demonstrated that the choice of groups is essential and that contributions are not independent. Hence, they should not be analysed in isolation from each other. Future work will apply the method to estimate trends in the genetic mean and variance following the work of Sorensen et al. (2001) and Lara et al. (2021).

References

García-Cortés, L.A., Martínez-Ávila, J.C., Toro, M.A. (2008) Partition of the genetic trend to validate multiple selection decisions. Animal 2:821–824.

Gaynor, R.C., Gorjanc, G., Hickey, J.M. (2020) AlphaSimR: An R-package for Breeding Program Simulations. bioRxiv 21. https://doi.org/10.1101/2020.08.10.245167

Gorjanc, G., Obšteter, J., Oliveira, T.P. (2021) AlphaPart: Partition/Decomposition of Breeding Values by Paths of Information. https://CRAN.R-project.org/package=AlphaPart

Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83.

Lara, L.A.C., Pocrnic, I., Oliveira, T. P., Gaynor, C., Gorjanc, G. (2021) Temporal and genomic analysis of additive genetic variance in breeding programmes. Heredity. https://doi.org/10.1038/s41437-021-00485-y

Obšteter, J., Holl, J., Hickey, J.M., Gorjanc, G. (2021) AlphaPart—R implementation of the method for partitioning genetic trends. Genet. Sel. Evol. 53(30). https://doi.org/10.1186/s12711-021-00600-x

Sorensen, D., Wang, C., Jensen, J., Gianola, D. (1994) Bayesian analysis of genetic change due to selection using Gibbs sampling. Genet. Sel. Evol. 26(333). https://doi.org/10.1186/1297-9686-26-4-333

Sorensen, D., Fernando, R., Gianola, D. (2001) Inferring the trajectory of genetic variance in the course of artificial selection. Genet. Res. 77:83–94. https://doi.org/10.1017/S0016672300004845