

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



Thiago de Paula Oliveira

thiago.oliveira@ed.ac.uk

<https://prof-thiagooliveira.netlify.com>

The Roslin Institute

8th December 2022



@HighlanderLab



Contents

1 Acknowledges and Highlights

2 Introduction and Motivation

3 Methods

4 Simulation

5 Results and Discussion

6 Final Remarks

Acknowledges

Who is involved?



Founders



Highlights



Cold
Spring
Harbor
Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY



New Results



Follow this preprint

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

Thiago de Paula Oliveira, Jana Obsteter, Ivan Pocanic, Nicolas Heslot, Gregor Gorjanc
doi: <https://doi.org/10.1101/2022.01.10.475603>

This article is a preprint and has not been certified by peer review [what does this mean?].

Highlights

- Reproducibility - https://github.com/HighlanderLab/toliveira_alphaPart_variance
- AlphaPart (CRAN) - <https://CRAN.R-project.org/package=AlphaPart>
- AlphaPart (GitHub) - <https://github.com/AlphaGenes/AlphaPart>



Introduction and Motivation

- The observed genetic change is a **sum of the contributions of different groups of individuals**
- Quantifying those sources are essential for identifying the key breeding actions and optimizing breeding programmes
- However, it is **difficult to disentangle the contribution of individual groups** due to the inherent complexity of breeding programmes. Vary due to different
 - Selection intensity
 - Accuracy
 - Genetic variation
 - Generation interval
 - Dissemination of genetic progress

Introduction and Motivation

The breeding value of an individual can be partitioned as

$$a_k = \underbrace{\frac{1}{2}a_{f(k)} + \frac{1}{2}a_{m(k)}}_{\text{Parent Average}} + w_k \quad (1)$$

Garcia Cortés et al. (2008): [change in the genetic mean](#) by partitioning the breeding values into the [contributions](#) of several paths

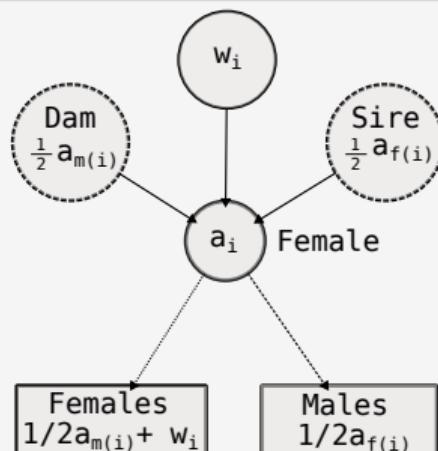
Animal (2008), 2:6, pp 821–824 © The Animal Consortium 2008
doi: 10.1017/S175173110800205X



Partition of the genetic trend to validate multiple selection decisions

L. A. García-Cortés[†], J. C. Martínez-Ávila and M. A. Toro

Introduction and Motivation

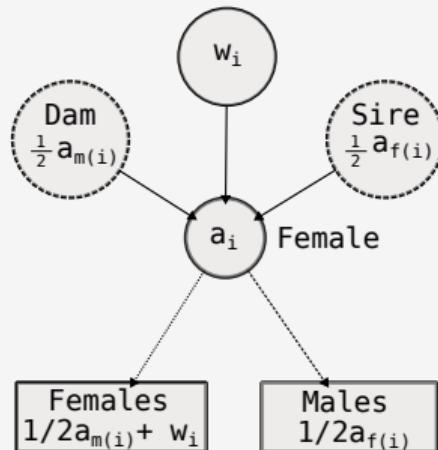


Male and female contribution



Credit: Matthias Zomer

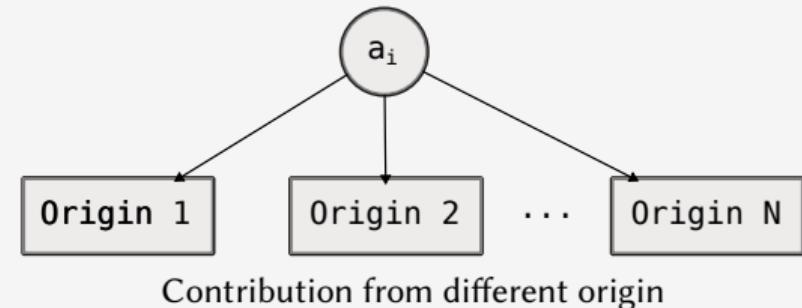
Introduction and Motivation



Male and female contribution

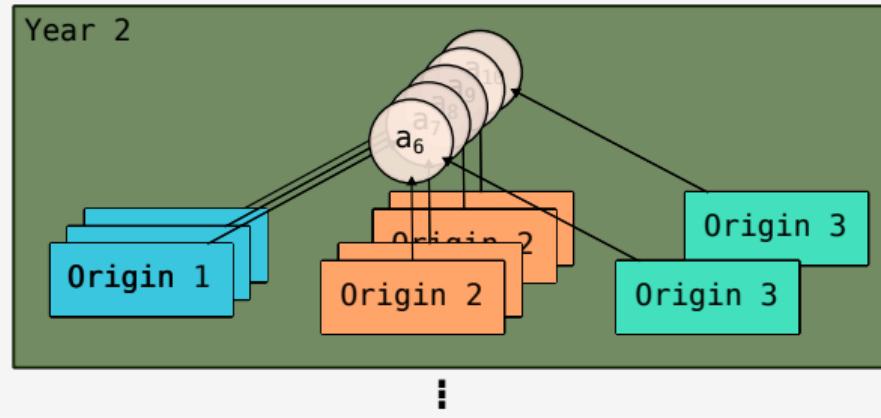
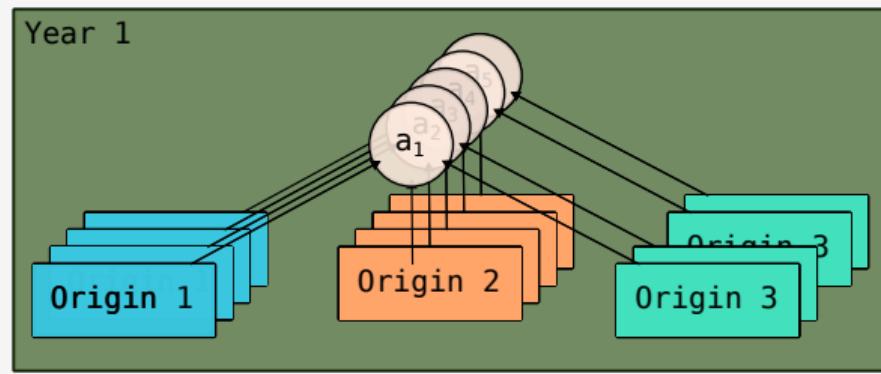


Credit: Matthias Zomer

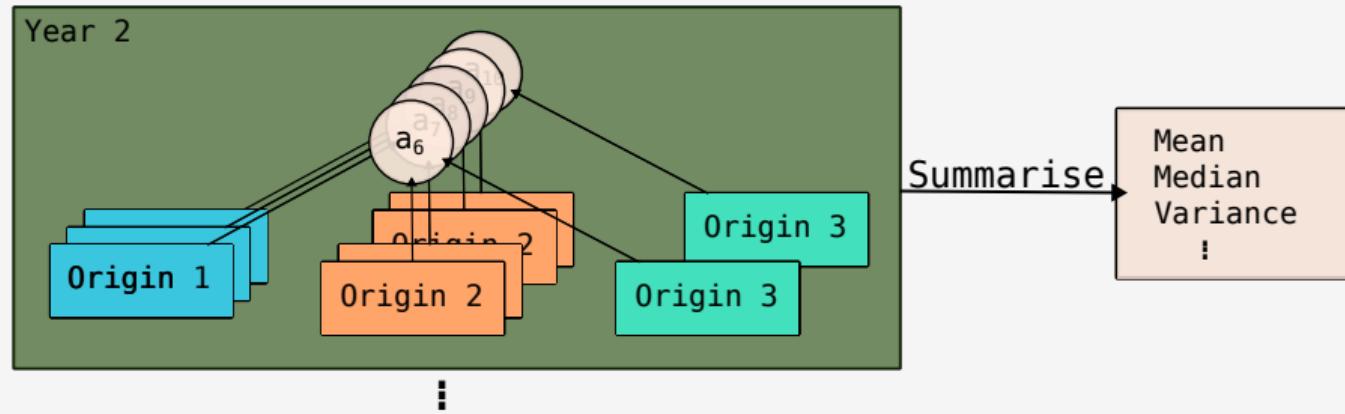
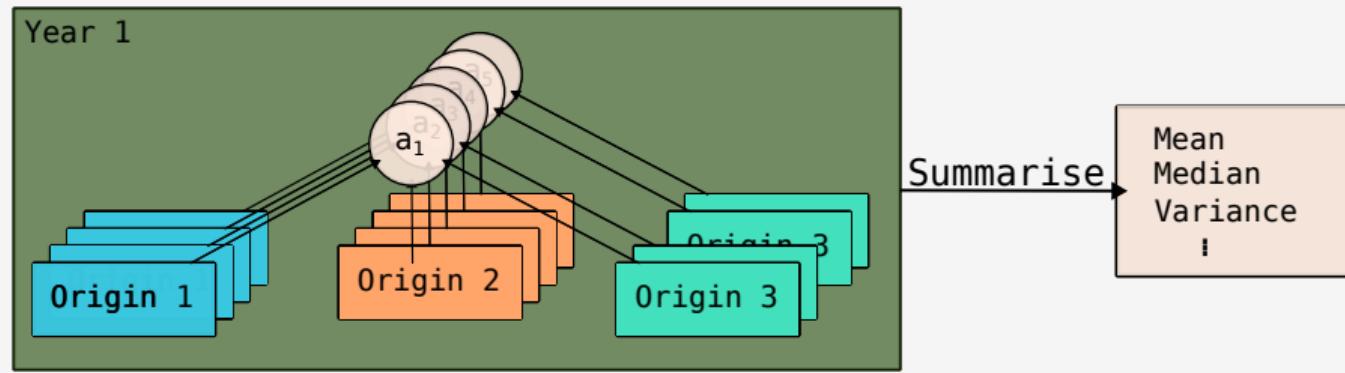


Credit: Daniel Tolhurst

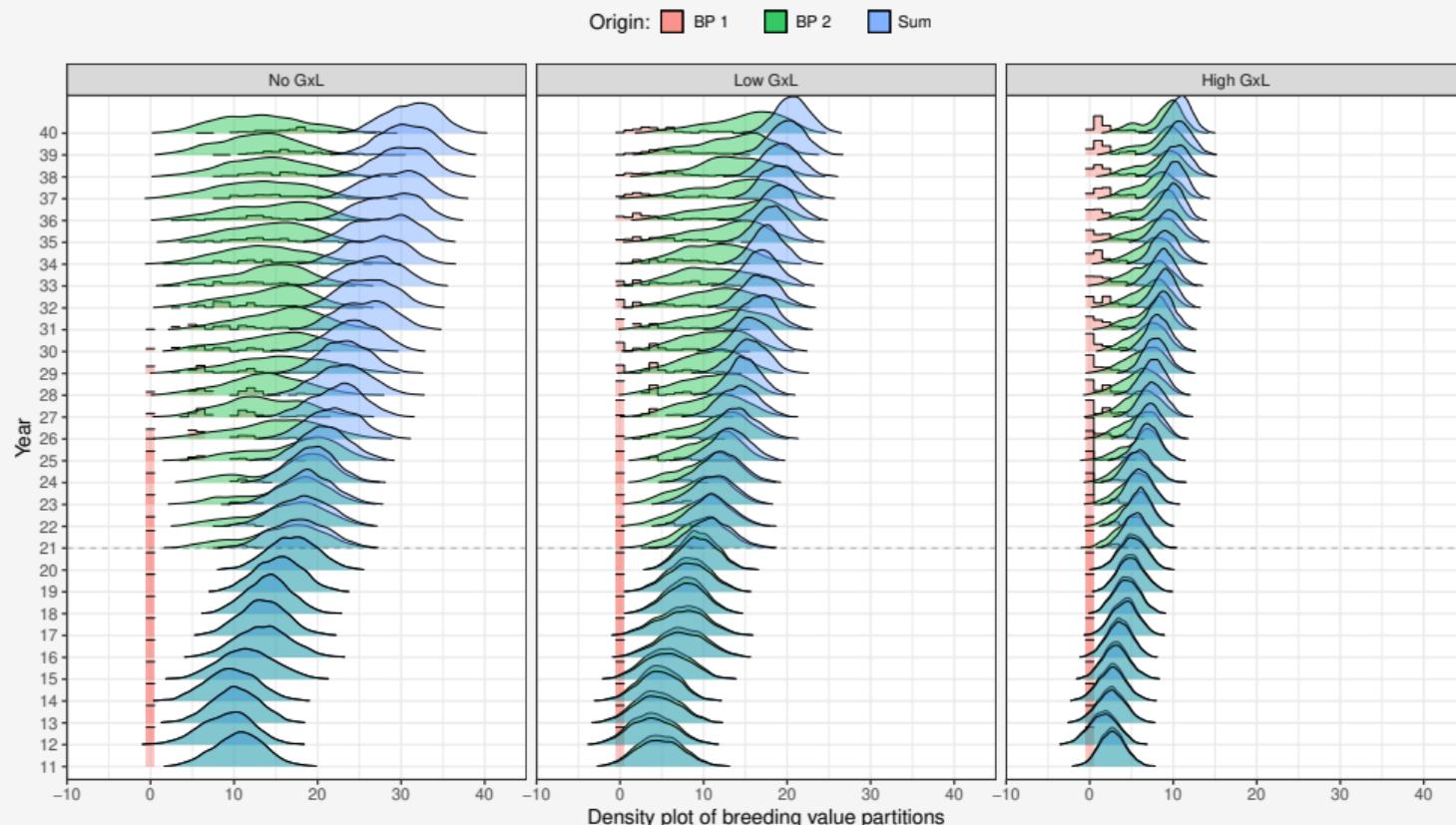
Introduction and Motivation



Introduction and Motivation



Introduction and Motivation: Example of usage



Introduction and Motivation: Example of usage

INTERBULL BULLETIN NO. 44. Stavanger, Norway, August 26 – 29, 2011

Partitioning of International Genetic Trends by Origin in Brown Swiss Bulls

G. Gorjanc¹, K. Potočnik¹, L. A. García-Cortés², J. Jakobsen³ and J. Dürr³

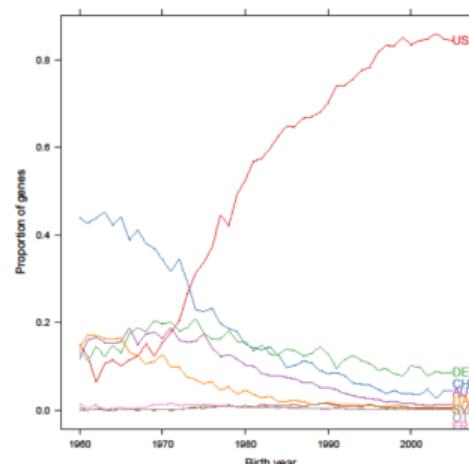


Figure 1. Proportion of genes by origin from 1960 onwards.

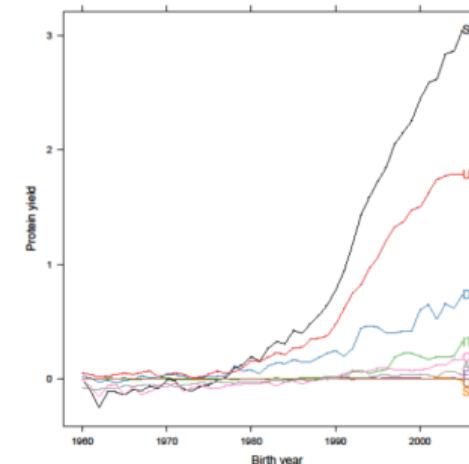


Figure 3. Partitioning of global genetic trend by origin for PRO.

Introduction and Motivation: Example of usage

Abdollahi-Arpanahi et al.
Genetics Selection Evolution (2021) 53:89
<https://doi.org/10.1186/s12711-021-00683-6>



RESEARCH ARTICLE

Open Access



Dissecting genetic trends to understand breeding practices in livestock: a maternal pig line example

Rostam Abdollahi-Arpanahi^{1*} Daniela Lourenco¹, Andres Legarra² and Ignacy Misztal¹

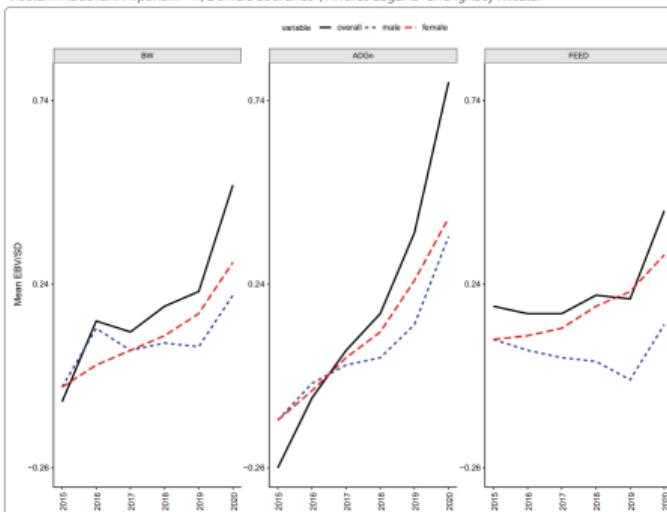


Fig. 4 Decomposition of overall genetic trends into males and females for birth weight (BW), average daily gain through the end of the nursery (ADG), and feed intake (FEED) for Landrace pigs. Average breeding values are presented on the additive genetic standard deviation scale

Objective



- Partition the genetic mean and **variance**
- Implement the method in AlphaPart software
- Apply the partitioning method to estimated breeding values using posterior sampling following the work of Sorensen et al.(2001) and Lara et al. (2022).

Obsteter et al. *Genet Sel Evol* (2021) 53:30
<https://doi.org/10.1186/s12711-021-00600-x>



SOFTWARE

Open Access

AlphaPart—R implementation
of the method for partitioning genetic trends



Jana Obsteter^{1*} , Justin Holl², John M. Hickey³ and Gregor Gorjanc³

Partitioning genetic trend

Post-processing method:

- We already have fitted the model and obtained the EBV or GEBV, or we know a priori TBV
- We assume the best fit line between EBVs and TBVs has no deviation from the equality line

General idea

- Genetic values are assumed to be multivariate and normally distributed $\mathbf{a} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{A})$
- Decomposition: $\mathbf{A}\sigma_a^2 = \mathbf{T}\mathbf{W}\mathbf{T}^T\sigma_a^2$
- We can then transform the additive genetic value as:

$$\mathbf{a} = \mathbf{T}\mathbf{w}, \text{ with } \mathbf{w} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{W}) \quad (2)$$

$$Var(\mathbf{a}) = Var(\mathbf{T}\mathbf{w}) = \mathbf{T}Var(\mathbf{w})\mathbf{T}^T = \mathbf{T}\mathbf{W}\mathbf{T}^T\sigma_a^2$$

Partitioning genetic trend

General idea

- Partitioning the response to selection into partial responses originated by several groups of Mendelian sampling effects will consist of dividing \mathbf{T} accordingly.
- Suppose any set $\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \dots + \mathbf{P}_m = \mathbf{I}$, thus:

$$\begin{aligned}\mathbf{a} &= \mathbf{T}(\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \dots + \mathbf{P}_m) \mathbf{w} \\ &= (\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_p) \mathbf{w}\end{aligned}\tag{3}$$

- As $\hat{\mathbf{a}} = \mathbf{T}\hat{\mathbf{w}}$, then $\hat{\mathbf{w}} = \mathbf{T}^{-1}\hat{\mathbf{a}}$. Consequently,

$$\begin{aligned}\hat{\mathbf{a}} &= \mathbf{T}_1\mathbf{T}^{-1}\hat{\mathbf{a}} + \mathbf{T}_2\mathbf{T}^{-1}\hat{\mathbf{a}} + \mathbf{T}_3\mathbf{T}^{-1}\hat{\mathbf{a}} + \dots + \mathbf{T}_p\mathbf{T}^{-1}\hat{\mathbf{a}} \\ &= \hat{\mathbf{a}}_1 + \hat{\mathbf{a}}_2 + \hat{\mathbf{a}}_3 + \dots + \hat{\mathbf{a}}_p\end{aligned}\tag{4}$$

where $\mathbf{a}_j, j = 1, 2, \dots, p$, is a part of the breeding value that can be assigned to the group i .

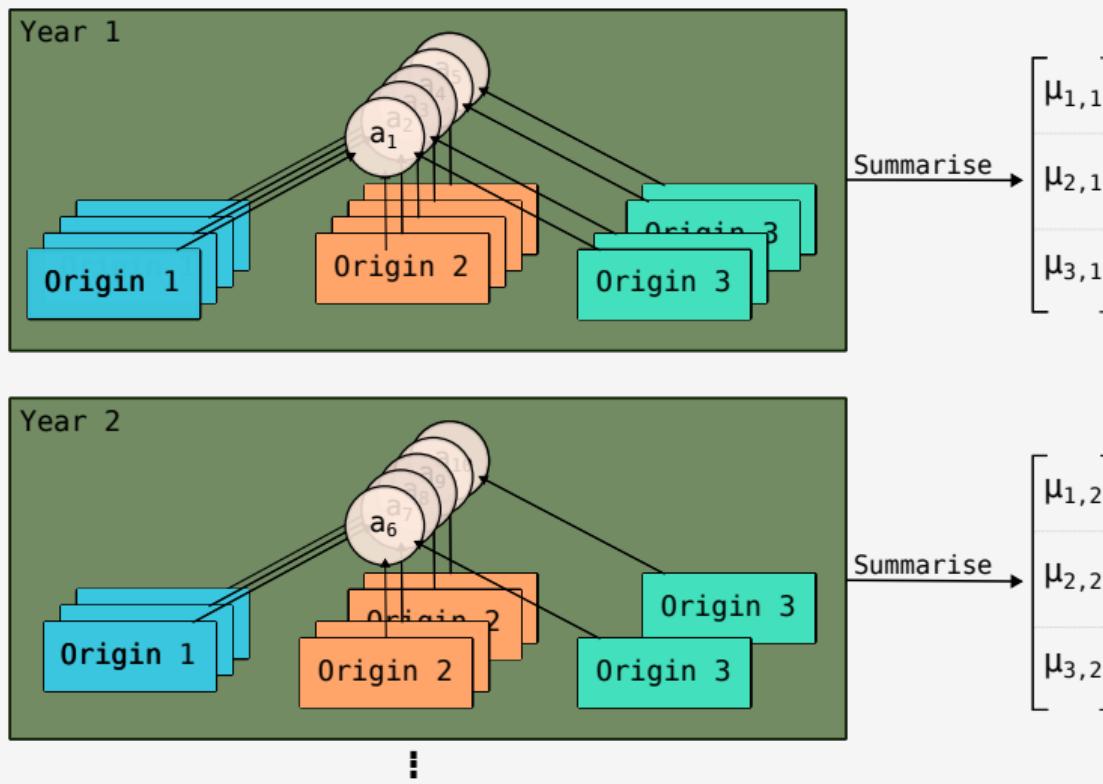
Partitioning genetic trend - variance

Here we extend the partitioning method to analyse the contribution of groups to genetic variance.

$$\begin{aligned}
 \text{Var}(\mathbf{a}) &= \text{Var} [(\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_p) \mathbf{w}], \\
 &= \sum_{j=1}^p \text{Var}(\mathbf{T}_j \mathbf{w}) + 2 \sum_{j=1}^{p-1} \sum_{j'=j+1}^p \text{Cov}(\mathbf{T}_j \mathbf{w}, \mathbf{T}_{j'} \mathbf{w}), \\
 &= \sum_{j=1}^p \mathbf{T}_j \mathbf{W} \mathbf{T}_j^T \sigma_a^2 + 2 \sum_{j=1}^{p-1} \sum_{j'=j+1}^p \mathbf{T}_j \mathbf{W} \mathbf{T}_{j'}^T \sigma_a^2, \\
 &= \left(\sum_{j=1}^p \mathbf{A}_j + 2 \sum_{j=1}^{p-1} \sum_{j'=j+1}^p \mathbf{A}_{j,j'} \right) \sigma_a^2,
 \end{aligned} \tag{5}$$

- \mathbf{A}_j and $\mathbf{A}_{j,j'}$ → dense matrices while \mathbf{T}^{-1} → sparse matrix
- We can use \mathbf{T}^{-1} to efficiently calculate the partitions $\mathbf{a}_1 + \mathbf{a}_2 + \dots + \mathbf{a}_p$
- Do not ignore parental inbreeding coefficients in setting up the \mathbf{A}^{-1}

Summarising partitions - mean



Summarising partitions - mean

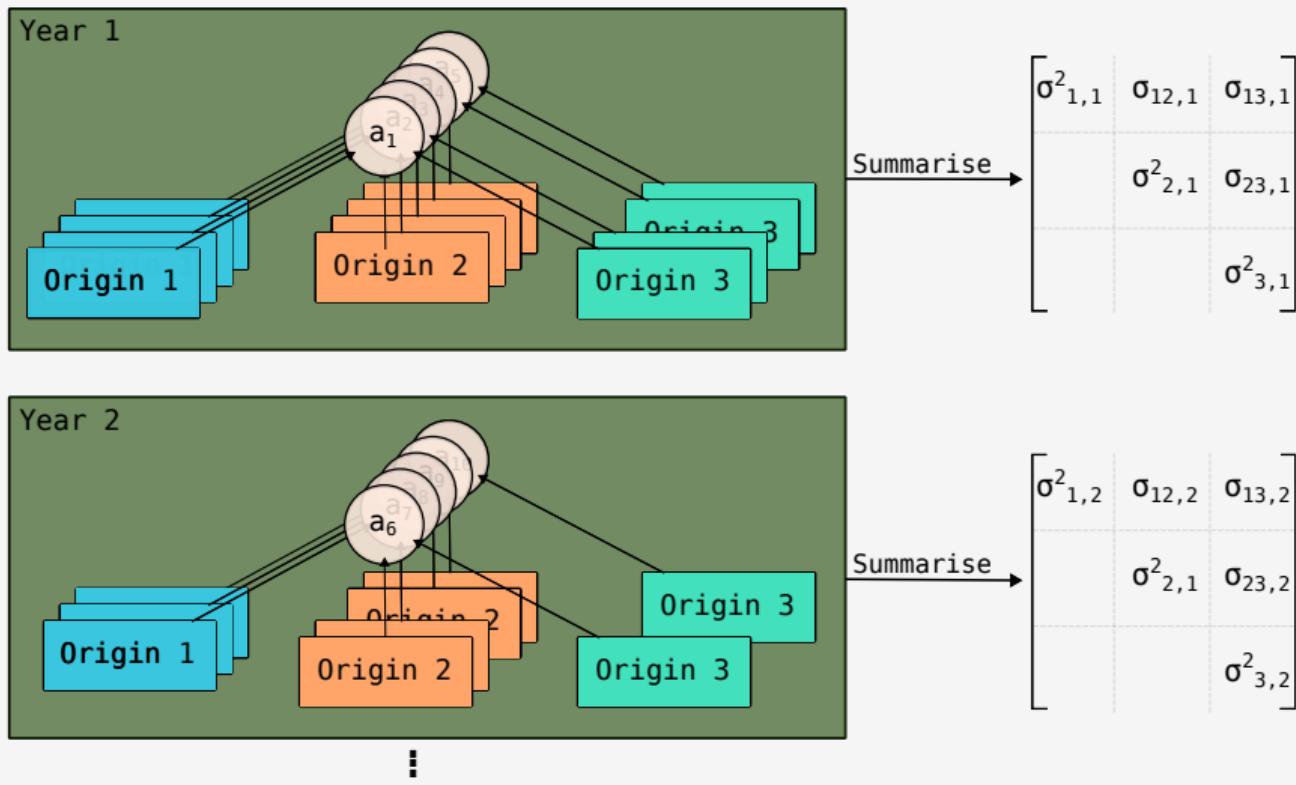
- Let $\mathbf{a}_j, j = 1, 2, \dots, p$, be part of the breeding value that can be assigned to the group j .
- Thus, we can set a linear regression of \mathbf{a}_j as a function of the explanatory variables (\mathbf{X}) we are interested in summarise \mathbf{a} .

$$\mathbf{a}_j = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_j \text{ with } \mathbf{e}_j \sim N\left(\mathbf{0}, \sigma_{e_j}^2\right) \quad (6)$$

The expected value (mean) of the partition \mathbf{a}_j is then given by

$$E[\mathbf{a}_j] = E[\mathbf{X}\boldsymbol{\beta} + \mathbf{e}_j] = \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\mu}_j \quad (7)$$

Summarising partitions - variance



Summarising partitions - variance

Assuming x_t , $t = 1, 2, \dots, m$, has m distinct categories, we can define genetic variance for the category t that has $n_k \leq nI$ individuals, $k^* = 1, 2, \dots, n_k$, as:

$$\begin{aligned} Var(\mathbf{a}_{j_t}) &= E(\mathbf{a}_{j_t}^2) - E^2(\mathbf{a}_{j_t}), \\ &= \frac{1}{n_k} \sum_{k^*=1}^{n_k} (a_{j_t, k^*} - \bar{a}_{j_t})^2, \end{aligned} \tag{8}$$

- n_t is the number of observations in the data that belongs to the category t
- $\bar{a}_{j_t} = \frac{1}{n_k} \sum_{k^*=1}^{n_k} a_{j_t, k^*}$.

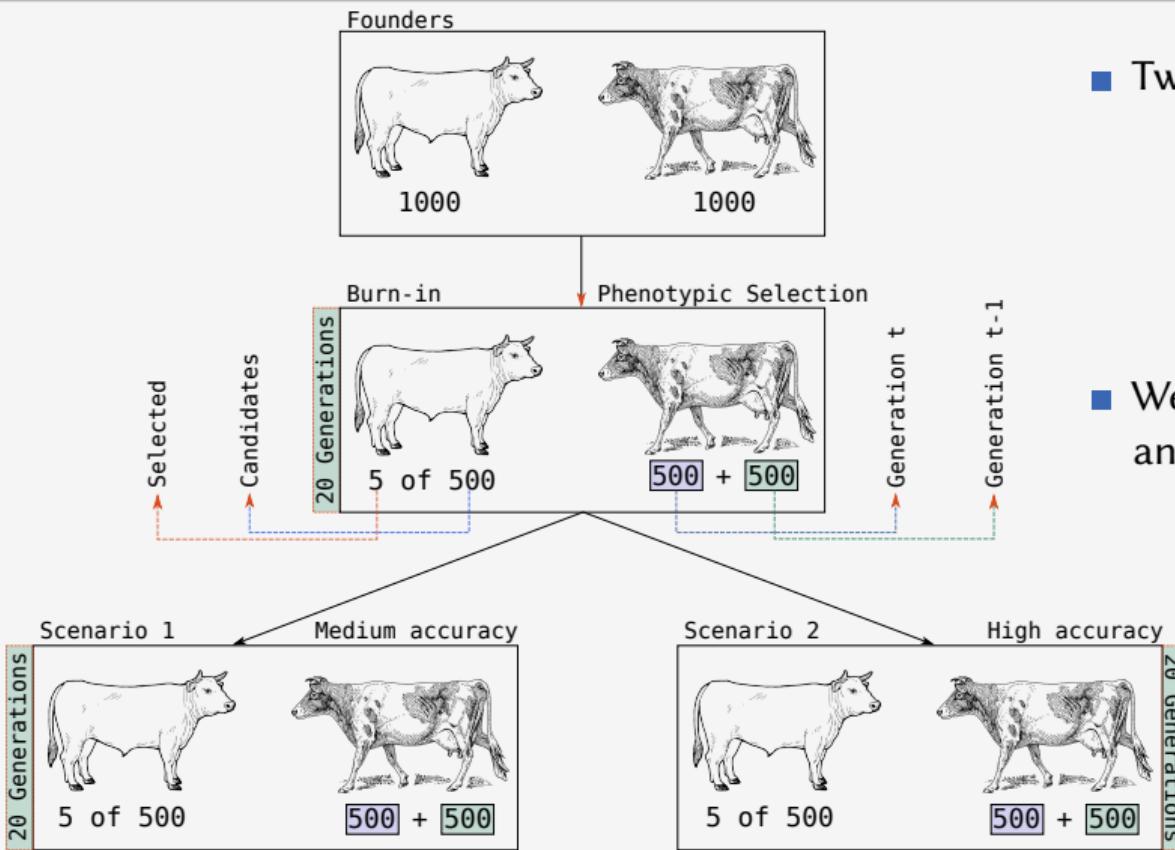
Summarising partitions - variance

We can also define covariance between paths for the category t as:

$$\begin{aligned} \text{Cov}(\mathbf{a}_{j_t}, \mathbf{a}_{j'_t}) &= E(\mathbf{a}_{j_t} \mathbf{a}_{j'_t}) - E(\mathbf{a}_{j_t}) E(\mathbf{a}_{j'_t}), \\ &= \frac{1}{n_k} \sum_{k^*=1}^{n_k} (a_{j_t, k^*} - \bar{a}_{j_t}) (a_{j'_t, k^*} - \bar{a}_{j'_t}). \end{aligned} \tag{9}$$

Simulation

Cattle Breeding Programme - summary



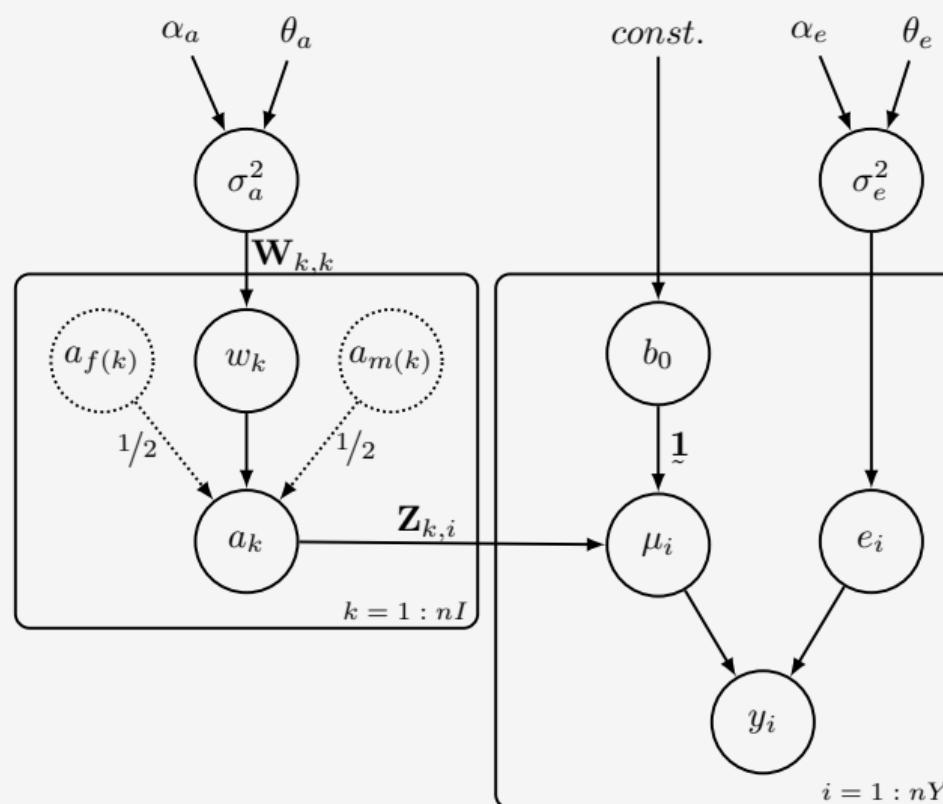
- Two scenarios:
 - Phenotypic (**Medium accuracy**, $r = 0.3$)
 - TBV (**High accuracy**, $r = 1.0$) selection
- We split T by using $P_m + P_f = I$ and $P_m^s + P_m^n = P_m$

Estimated breeding value - Statistical Model

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e}, \\ \mathbf{a} &\sim N(\mathbf{0}, \sigma_a^2 \mathbf{A}), \text{ and } \mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}) \end{aligned} \quad (10)$$

- Matrix \mathbf{A} can be decomposed as $\mathbf{A} = \mathbf{T}\mathbf{W}\mathbf{T}^T$ using LDL decomposition
- \mathbf{W} can be computed according to specific scenarios
 - $\mathbf{W}_{k,k} = \frac{1}{2} - \frac{1}{4}(F_{f(k)} + F_{m(k)})$ when both parents are known
 - $\mathbf{W}_{k,k} = \frac{3}{4} - \frac{1}{4}F_{m(k)}$ or $\mathbf{W}_{k,k} = \frac{3}{4} - \frac{1}{4}F_{f(k)}$ when one parent are known
 - $\mathbf{W}_{k,k} = 1$ when both parents are unknown
- $Var(a_k | \mathbf{W}_{k,k}) = (1 + F_k) \sigma_a^2$ and under no inbreeding $\rightarrow Var(a_k | \mathbf{W}_{k,k}) = \sigma_a^2$
- Verify the impacts of accounting or not for inbreeding when calculating the partitioning

DAG



Full Bayesian approach

Posterior

- We used Markov Chain Monte Carlo (MCMC) to generate samples from the posterior distribution using Gibbs sampler algorithm

$$p(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_e^2) p(\mathbf{b}) p(\mathbf{a} | \mathbf{A}, \sigma_a^2) p(\sigma_a^2) p(\sigma_e^2).$$

In this case we are interested in the posterior distribution $p(\mathbf{a} | \mathbf{A}, \sigma_a^2, \mathbf{y})$.

Full Bayesian approach

Posterior

- We used Markov Chain Monte Carlo (MCMC) to generate samples from the posterior distribution using Gibbs sampler algorithm

$$p(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2 | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{b}, \mathbf{a}, \sigma_e^2) p(\mathbf{b}) p(\mathbf{a} | \mathbf{A}, \sigma_a^2) p(\sigma_a^2) p(\sigma_e^2).$$

In this case we are interested in the posterior distribution $p(\mathbf{a} | \mathbf{A}, \sigma_a^2, \mathbf{y})$.

MCMC samples

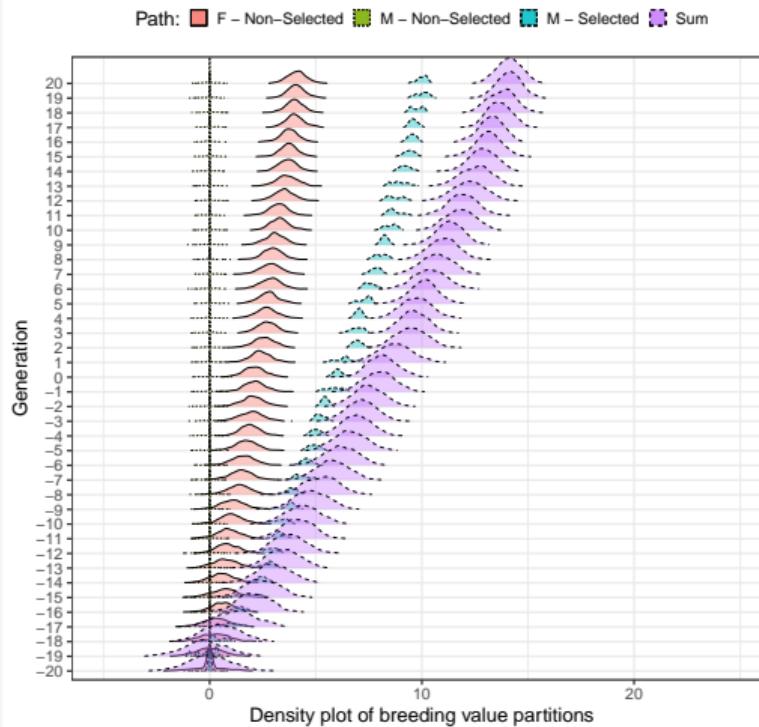
80,000 samples from where

- 20,000 iterations are burn-in
- 60,000 iterations were stored using a thinning of length 40
- 1,500 samples of EBV's were produced to be used in AlphaPart

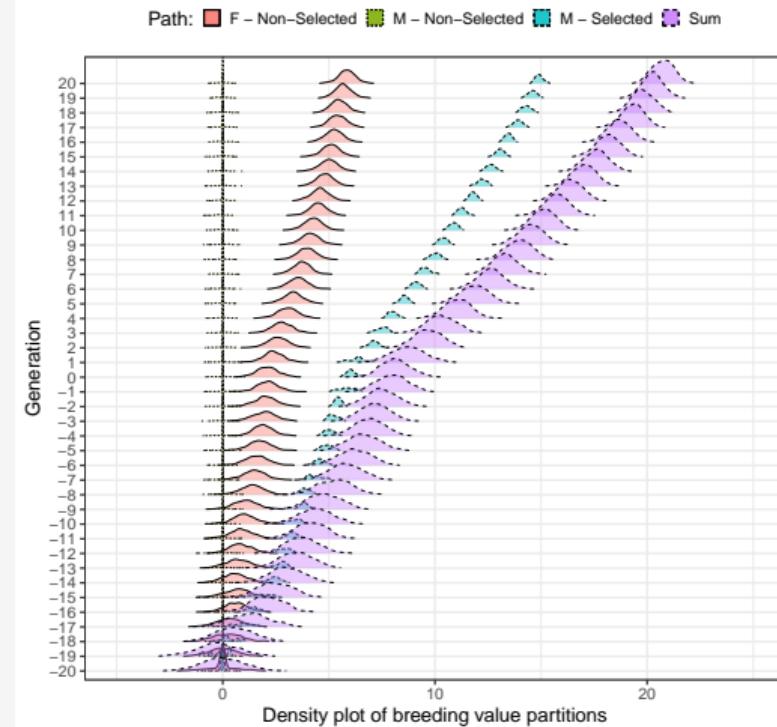
Results and Discussion

Paths distribution

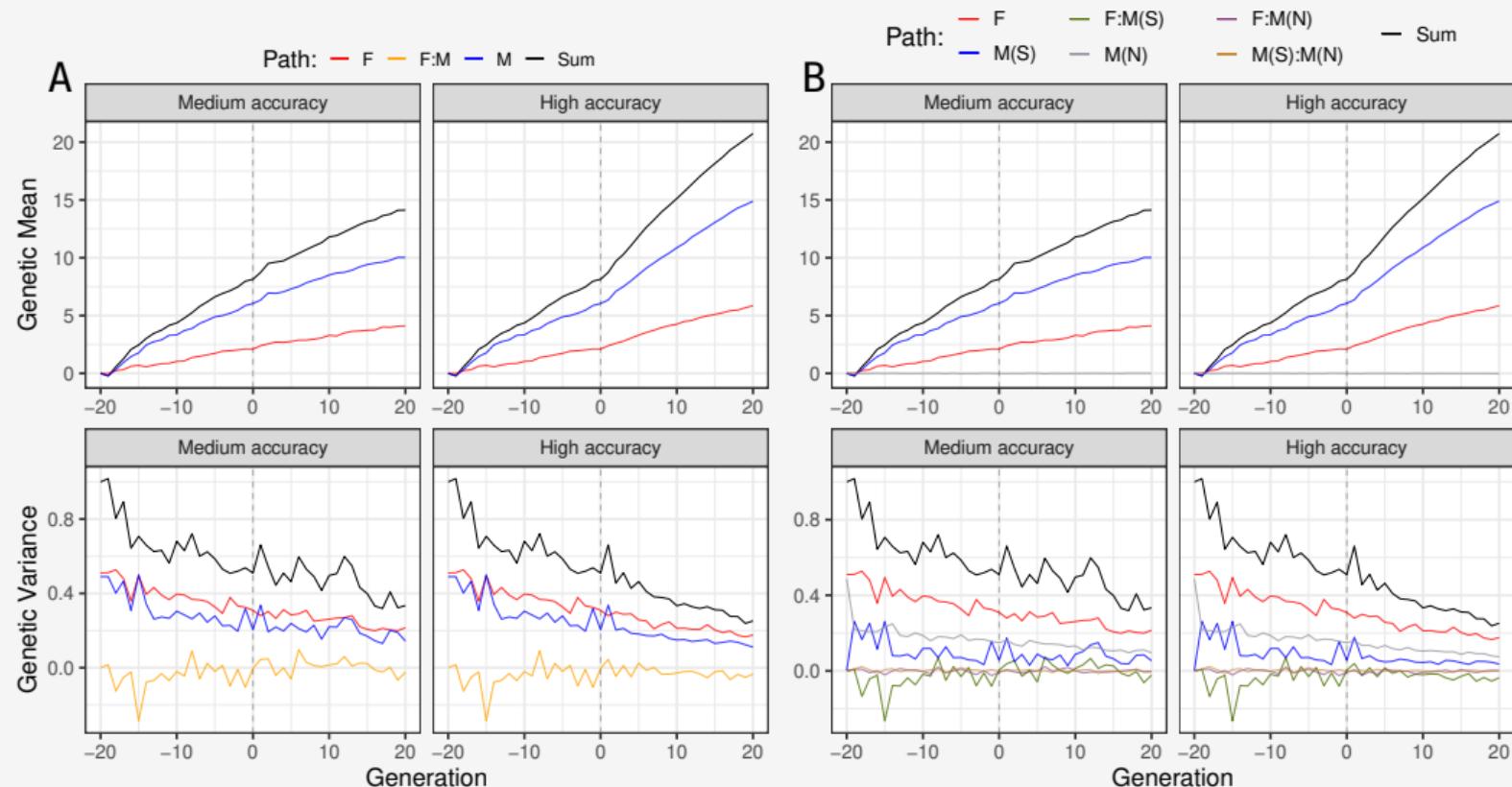
Medium accuracy

Medium Accuracy - $r = 0.3$

High accuracy

High Accuracy - $r = 1.0$

Partitioning genetic mean and variance - TBVs

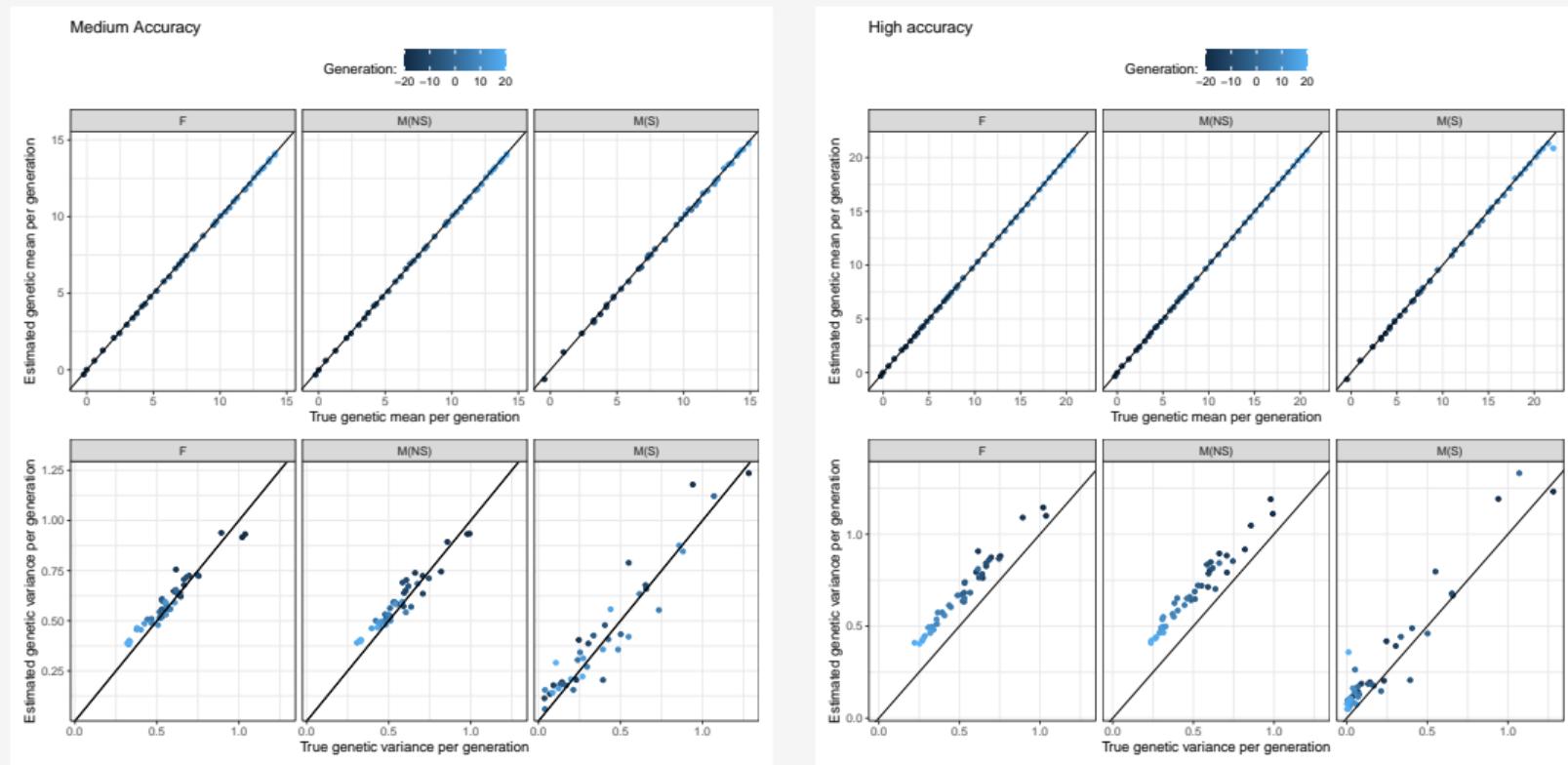


Model results

Table: Variance components (VC) true values, point estimates, and their 95% highest posterior density (HPD) interval

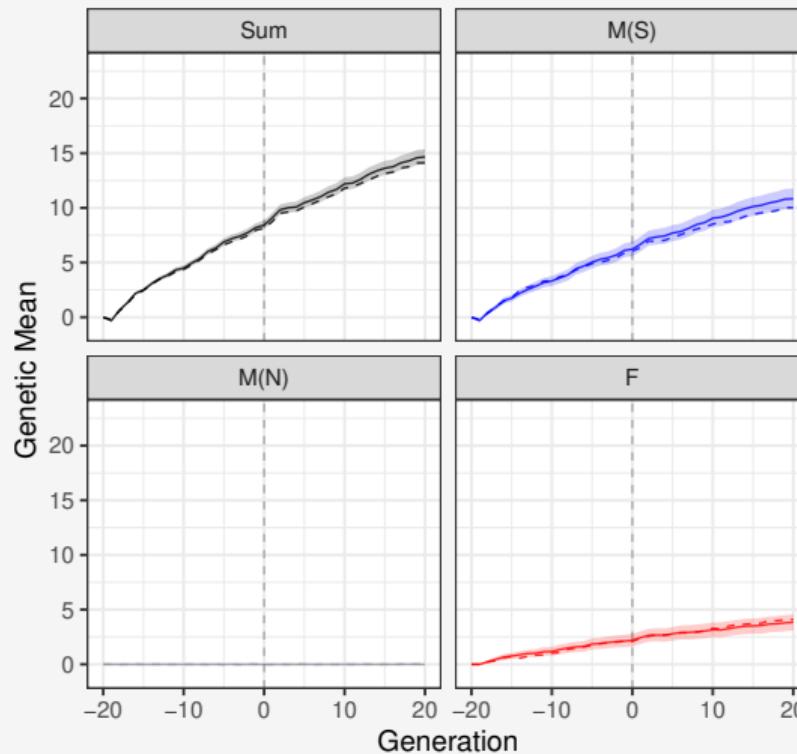
Scenario	VC	True	Estimate	95% HPD	
				Lower	Upper
Medium accuracy	σ_a^2	0.3	0.27	0.25	0.30
	σ_e^2	0.7	0.69	0.68	0.71
High accuracy	σ_a^2	0.3	0.35	0.33	0.38
	σ_e^2	0.7	0.66	0.64	0.67

Model results

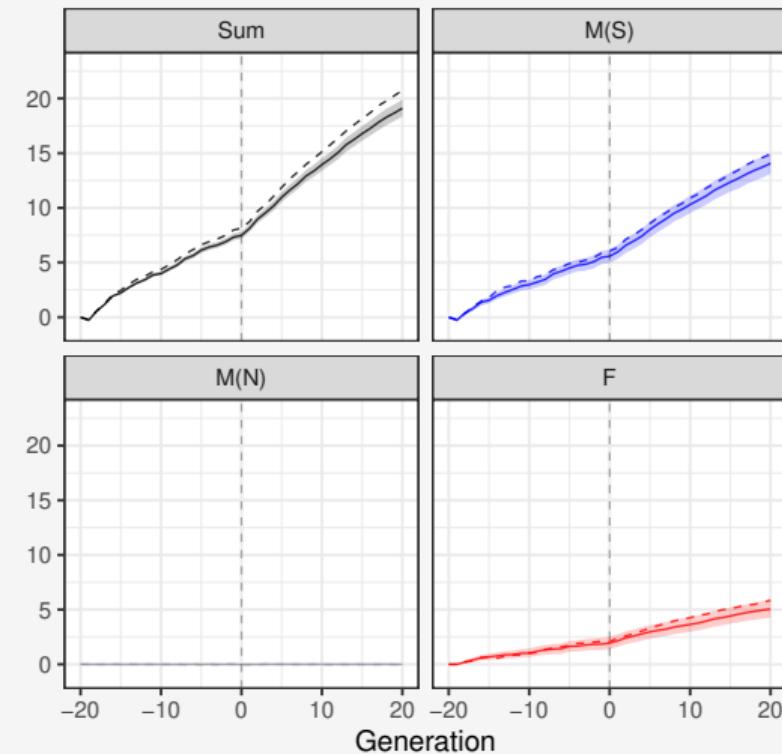


Partitioning genetic mean - EBVs

A Medium Accuracy



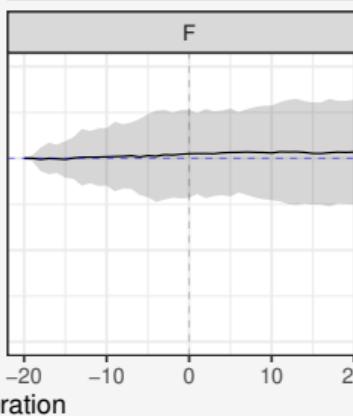
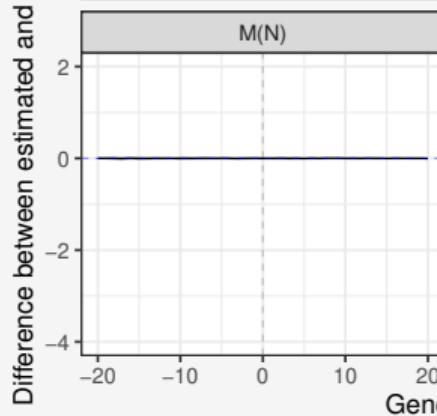
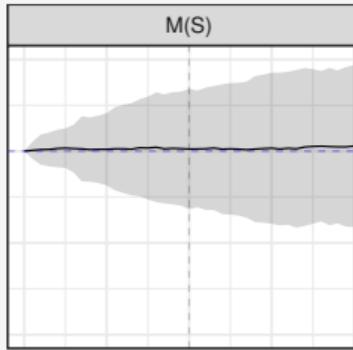
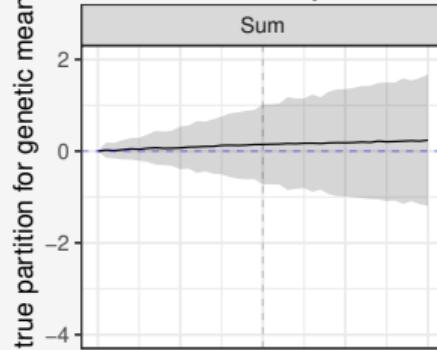
B High accuracy



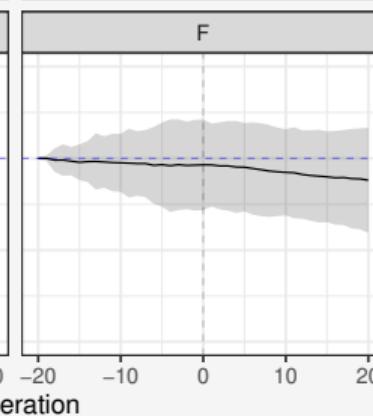
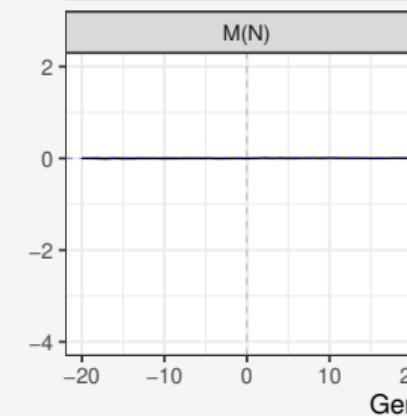
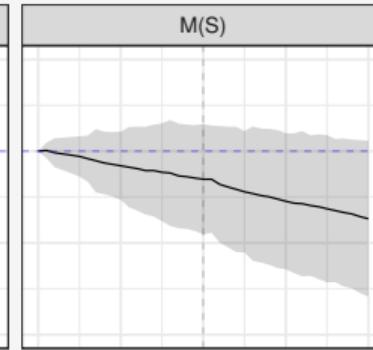
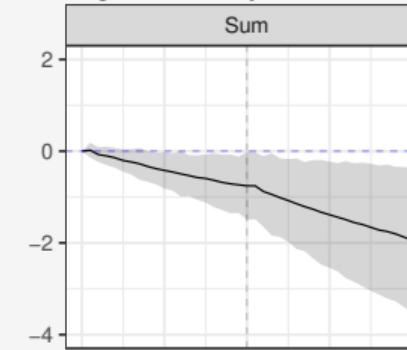
Partitioning genetic mean - EBVs (30 replicates)

A

Medium accuracy

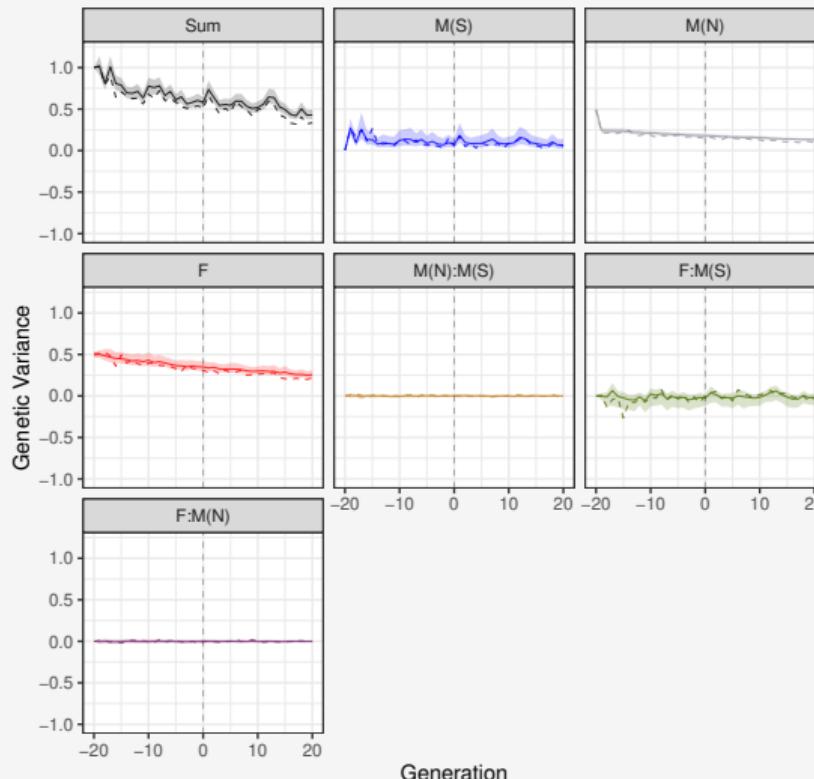
**B**

High accuracy

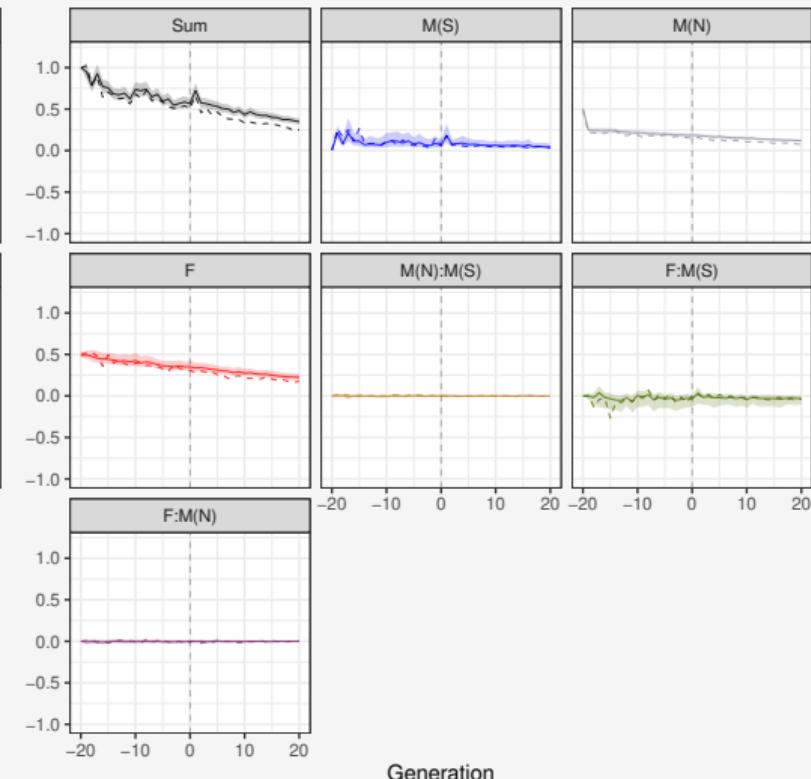


Partitioning genetic variance - EBVs

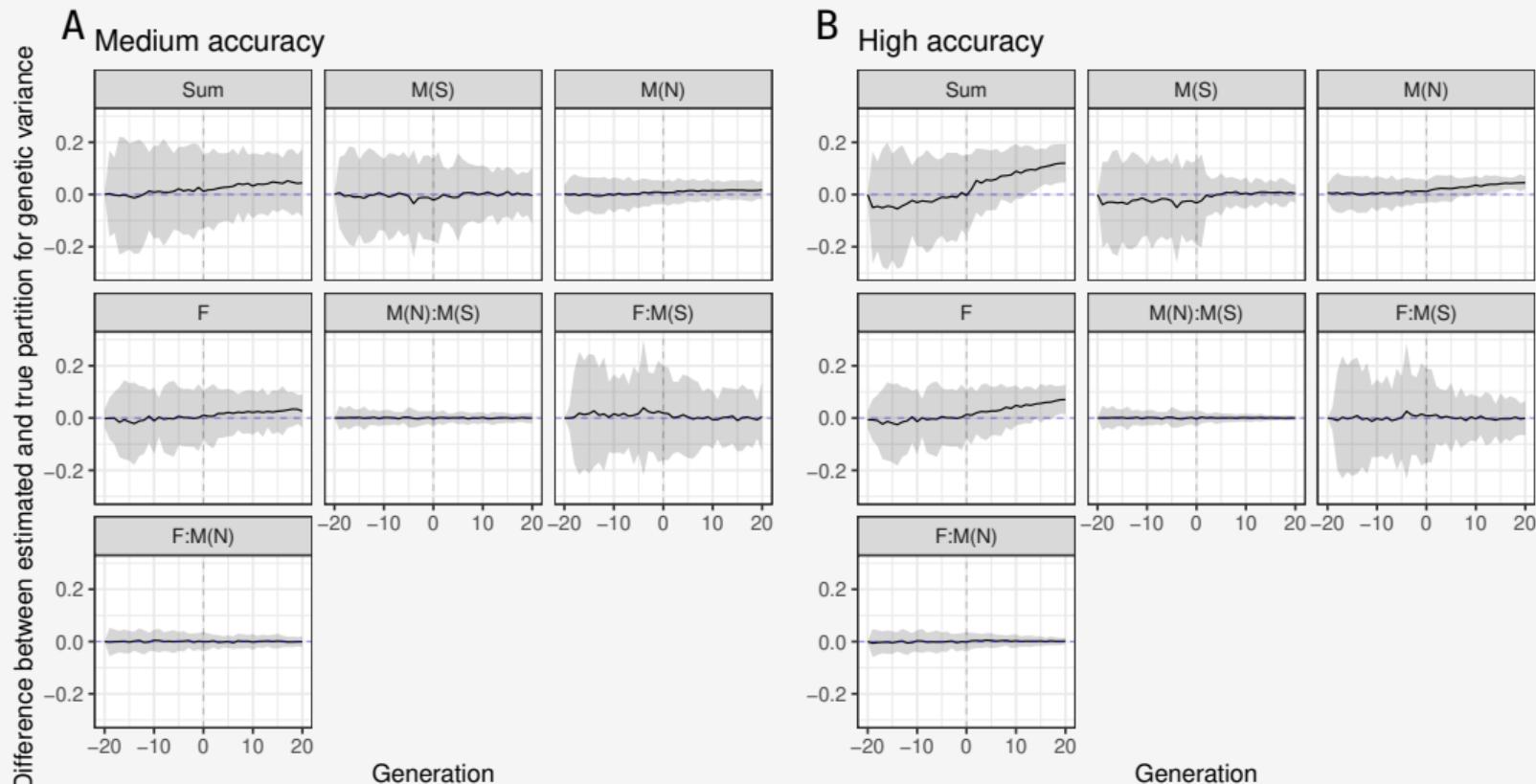
A Medium Accuracy



B High accuracy

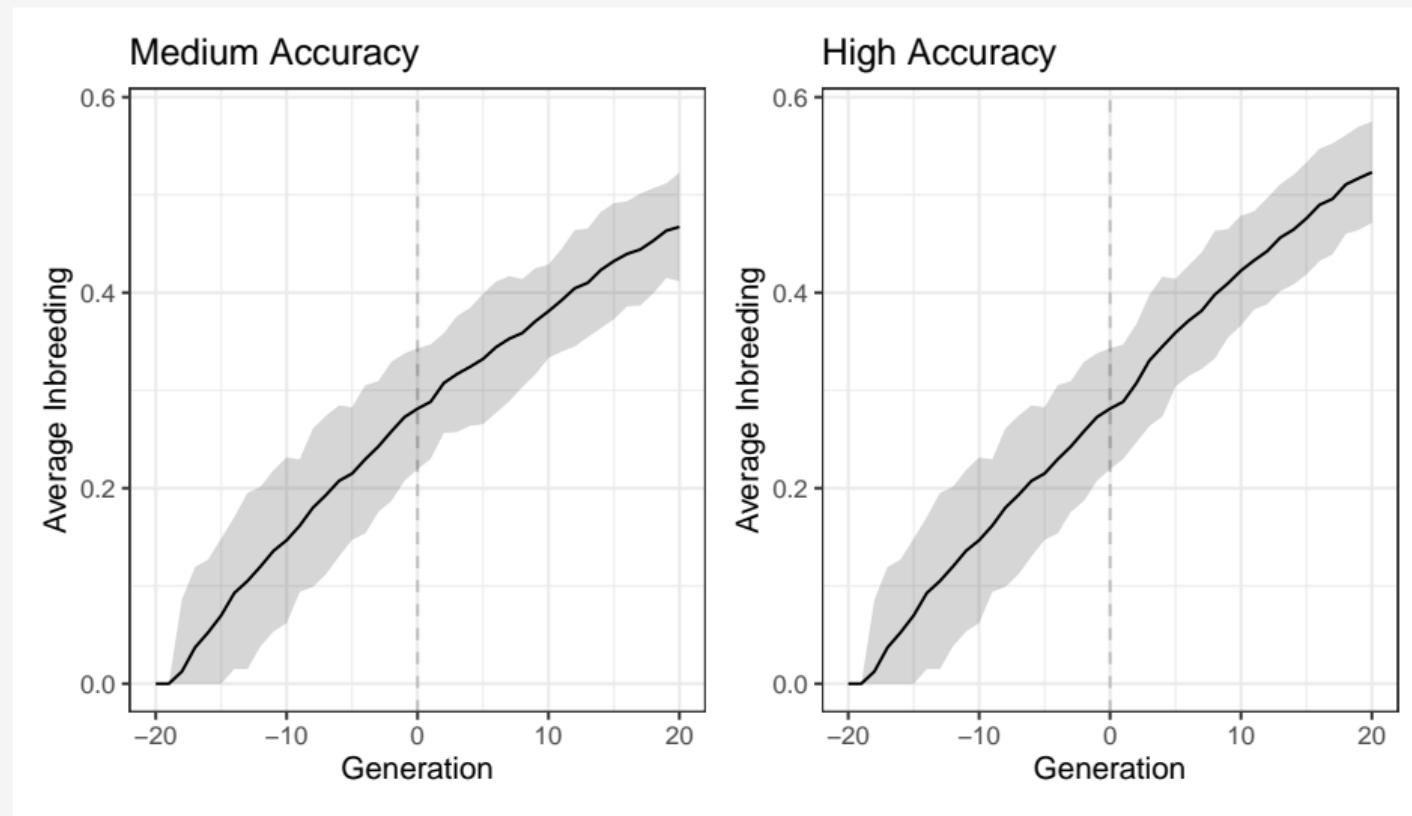


Partitioning genetic variance - EBVs (30 replicates)

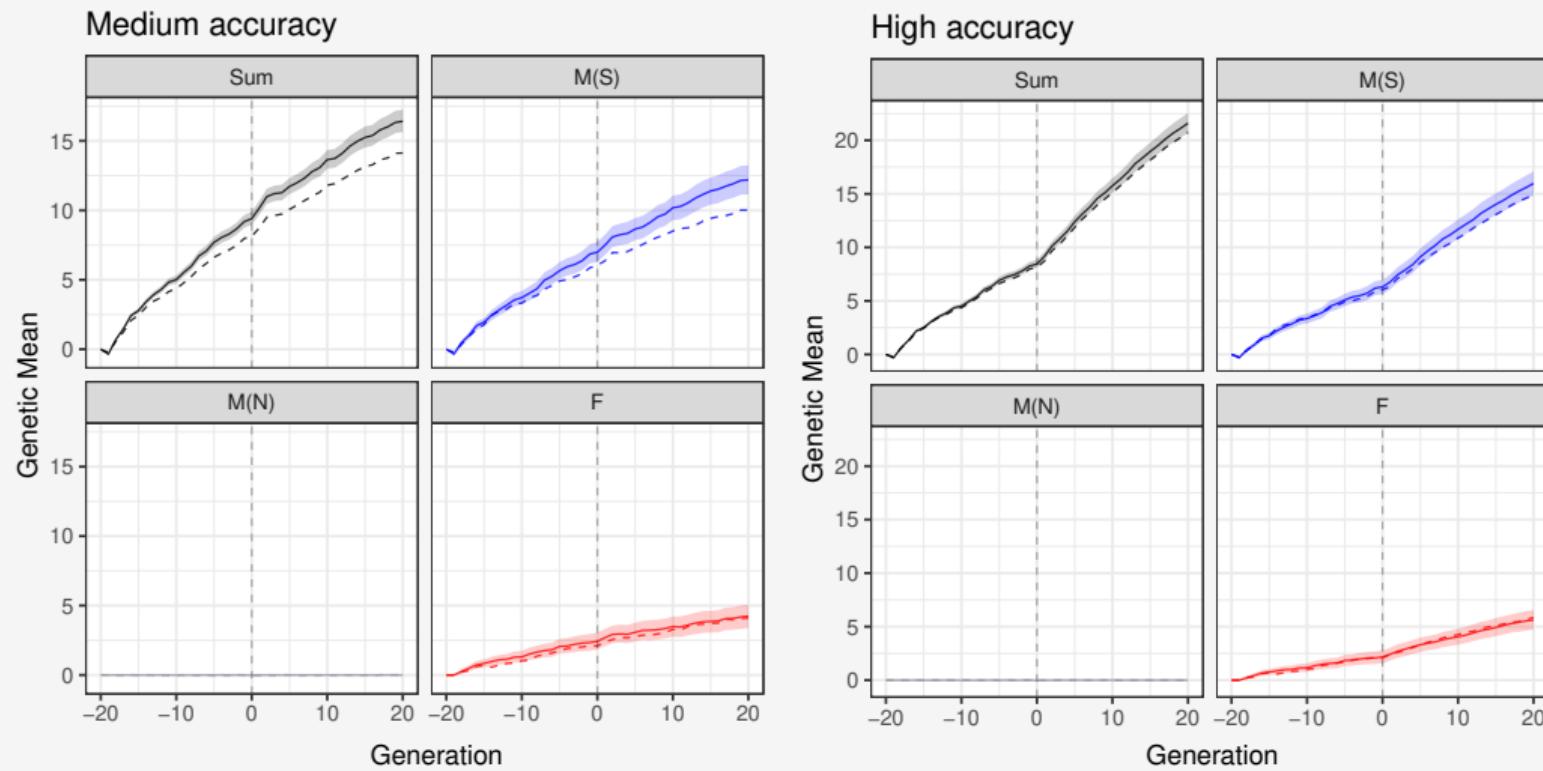


What happens if we do not consider inbreeding to build W ?

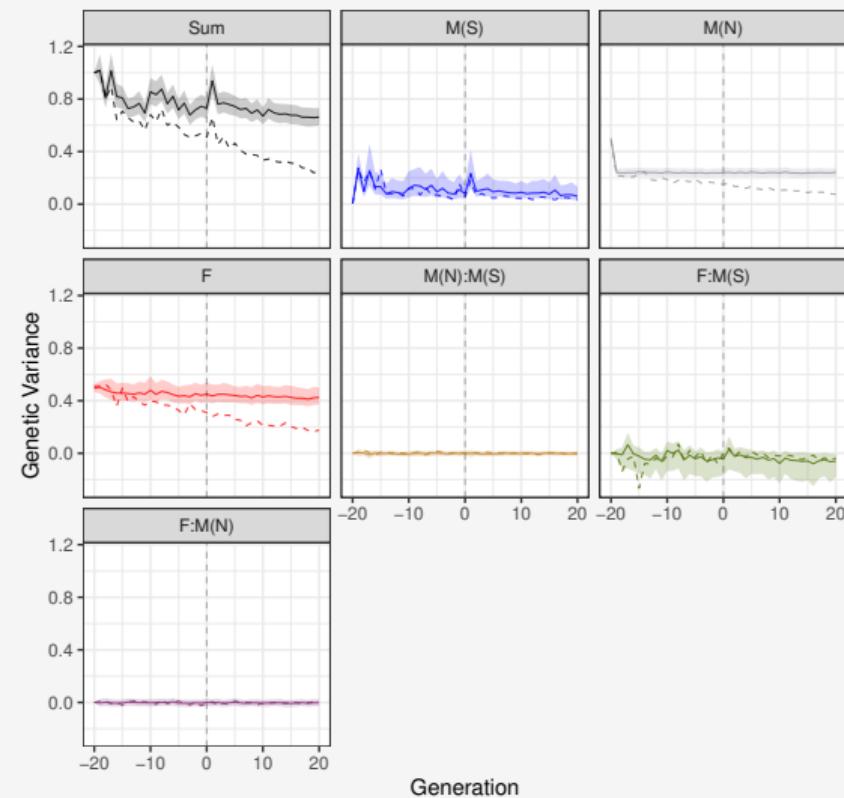
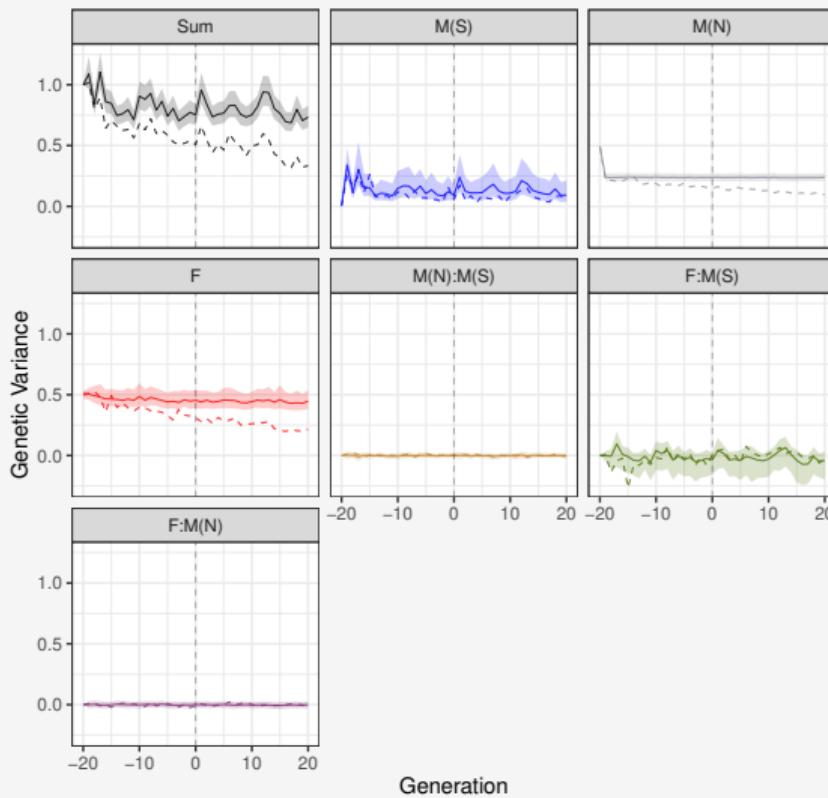
What happens if we do not consider inbreeding to build W ?



What happens if we do not consider inbreeding to build W ?



What happens if we do not consider inbreeding to build W ?



Final Remarks

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.
- It can **help breeders and researchers** to better **understand the genetic gain and variance dynamics** in specific breeding programmes.

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.
- It can **help breeders and researchers** to better **understand the genetic gain and variance dynamics** in specific breeding programmes.
- We demonstrate that the **method is not limited to working with TBVs** for quantifying sources of genetic mean and variance when applying the partitioning method

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.
- It can **help breeders and researchers** to better **understand the genetic gain and variance dynamics** in specific breeding programmes.
- We demonstrate that the **method is not limited to working with TBVs** for quantifying sources of genetic mean and variance when applying the partitioning method
- **Covariance** between paths can make a **substantial contributions** to the **genetic variance**

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.
- It can **help breeders and researchers** to better **understand the genetic gain and variance dynamics** in specific breeding programmes.
- We demonstrate that the **method is not limited to working with TBVs** for quantifying sources of genetic mean and variance when applying the partitioning method
- **Covariance** between paths can make a **substantial contributions** to the **genetic variance**
- We **should not consider the contribution of different groups in isolation** (they are not necessarily independent)

Final Remarks

- We propose a method to quantify the **sources of genetic variance** in breeding programmes by partitioning path variables into specific contributions.
- It can **help breeders and researchers** to better **understand the genetic gain and variance dynamics** in specific breeding programmes.
- We demonstrate that the **method is not limited to working with TBVs** for quantifying sources of genetic mean and variance when applying the partitioning method
- **Covariance** between paths can make a **substantial contributions** to the **genetic variance**
- We **should not consider the contribution of different groups in isolation** (they are not necessarily independent)
- How to **assess the uncertainty of the contributions** by drawing samples from the **posterior distribution**

References

- Abdollahi-Arpanahi, R., Lourenco, D., Legarra, A., Misztal, I.: Dissecting genetic trends to understand breeding practices in livestock: a maternal pig line example. **Genetics Selection Evolution** 53(1), 89 (2021).
- García-Cortés, L.A., Martínez- Ávila, J.C., Toro, M.A.: Partition of the genetic trend to validate multiple selection decisions. **Animal** 2(6), 821–824 (2008).
- Gorjanc, G., Hely, F., Amer, P.: Partitioning international genetic trends by origin in Holstein bulls (2012)
- Lara, L.A.d.C., Pocrnic, I., Oliveira, T.d.P., Gaynor, R.C., Gorjanc, G.: Temporal and genomic analysis of additive genetic variance in breeding programmes. **Heredity** (2021)
- Obsteter, J., Holl, J., Hickey, J.M., Gorjanc, G.: AlphaPart-R implementation of the method for partitioning genetic trends. **Genetics Selection Evolution** 53(1), 30 (2021).
- Sorensen, D., Fernando, R., Gianola, D.: Inferring the trajectory of genetic variance in the course of artificial selection. **Genetical Research** 77(1), 83–94 (2001).



THE UNIVERSITY
of EDINBURGH



Biotechnology and
Biological Sciences
Research Council

