

# Quantifying the Drivers of Genetic Change in Plant Breeding

---



**Thiago de Paula Oliveira**

Tolhurst; Pocrnic; Marchal; Heslot; and Gorjanc

`thiago.oliveira@ed.ac.uk`

<https://prof-thiago-oliveira.netlify.com>

The Roslin Institute

20th March 2023



# Introduction and Motivation

---

- **Genetic diversity is indispensable** for plant breeding to improve crops and develop new cultivars with improved agronomics.
- **The importance of germplasm diversity:** **increase the genetic variability** when producing pool-specific **doubled-haploid lines**.
- **Variability and heritability** are fundamental in selecting desirable genotypes.
  - However, there is limited research about **how the germplasm exchange** between breeding programmes **contributes** to each pool's **genetic mean and variance over time**.
  - García-Cortés et al (2008) and Oliveira et al. (2022) proposed a methodology based on pedigree information to partitioning the breeding values into contribution of several paths to **genetic mean and variance**.

# Introduction and Motivation

---

- Genetic diversity is indispensable for plant breeding to improve crops and develop new cultivars with improved agronomics.
- The importance of germplasm diversity: increase the genetic variability when producing pool-specific **doubled-haploid lines**.
- **Variability and heritability** are fundamental in selecting desirable genotypes.
- However, there is limited research about how the germplasm exchange between breeding programmes contributes to each pool's genetic mean and variance over time.
- García-Cortés et al (2008) and Oliveira et al. (2022) proposed a methodology based on pedigree information to partitioning the breeding values into contribution of several paths to genetic mean and variance.

# Introduction and Motivation

---

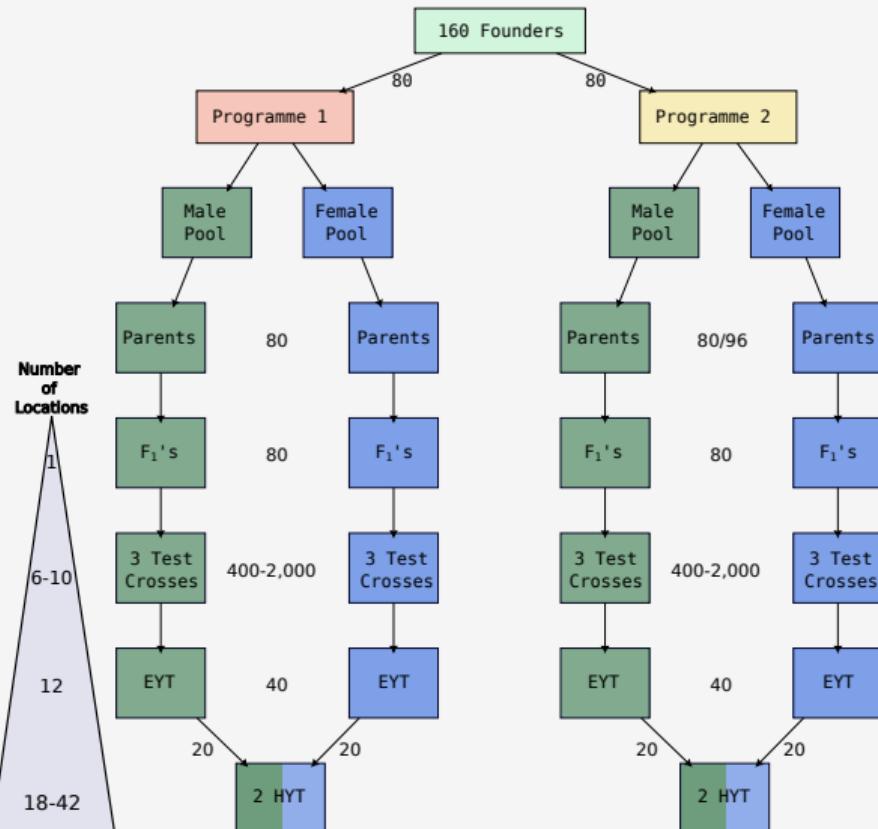
- Genetic diversity is indispensable for plant breeding to improve crops and develop new cultivars with improved agronomics.
- The importance of germplasm diversity: increase the genetic variability when producing pool-specific **doubled-haploid lines**.
- **Variability and heritability** are fundamental in selecting desirable genotypes.
- However, there is limited research about how the germplasm exchange between breeding programmes contributes to each pool's genetic mean and variance over time.
- García-Cortés et al (2008) and Oliveira et al. (2022) proposed a methodology based on pedigree information to partitioning the breeding values into contribution of several paths to genetic mean and variance.

# Aims

---

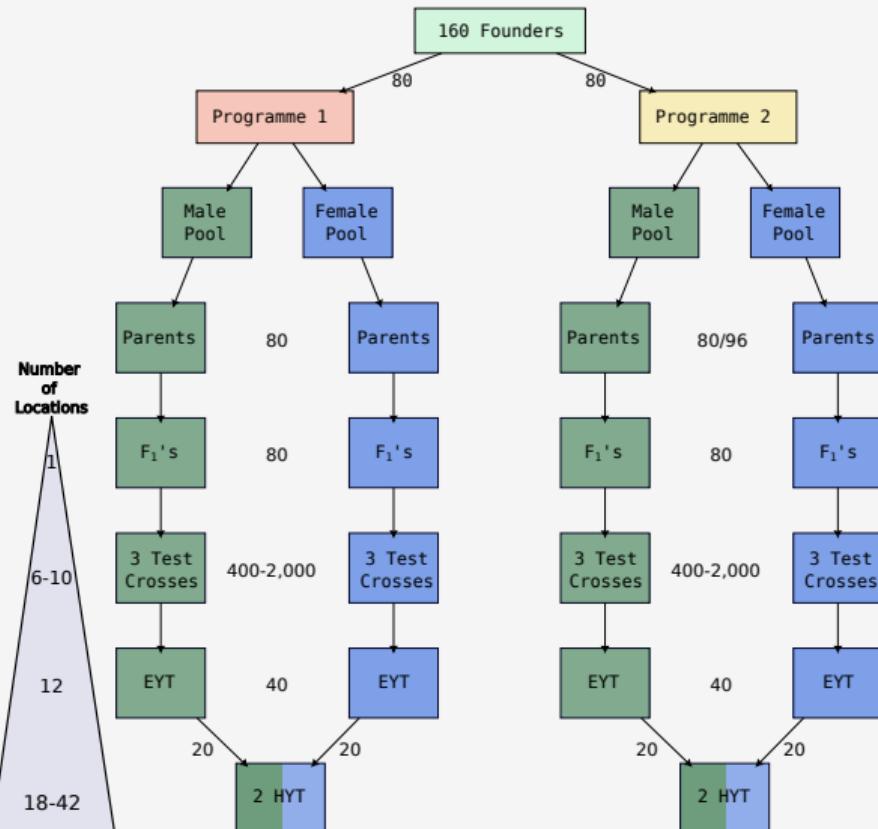
- Simulate **two maize breeding programmes in parallel** with gene flow from one programme to another under **different scenarios of G×L interaction**
- Evaluate the **contribution of germplasm origin** given the heterotic pool into contributions to genetic mean and variance **summarised over the years**
- Evaluate the **impact of G×L** within and across breeding programmes in those contributions

# Hybrid Breeding Scheme - AlphaSimR



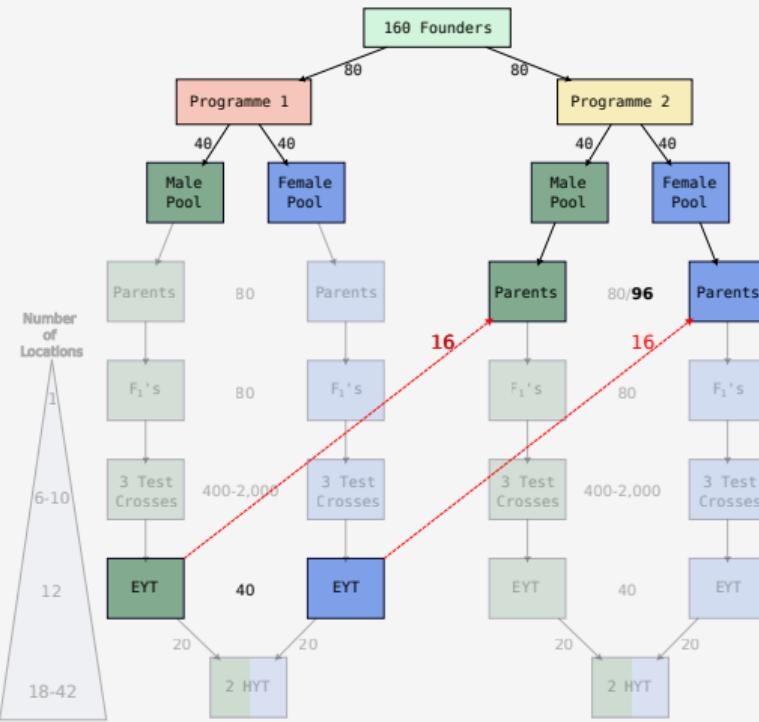
- Trait: Yield per harvested acre in bushels
- Phenotypic Selection
- 20 years of burn-in + 20 future years
- Future breeding:
  - Three scenarios of G×L
- A lag of 10 years between breeding programmes 1 and 2
- Germplasm flow starts in year 21 onwards from programme 1 to 2

# Hybrid Breeding Scheme - AlphaSimR

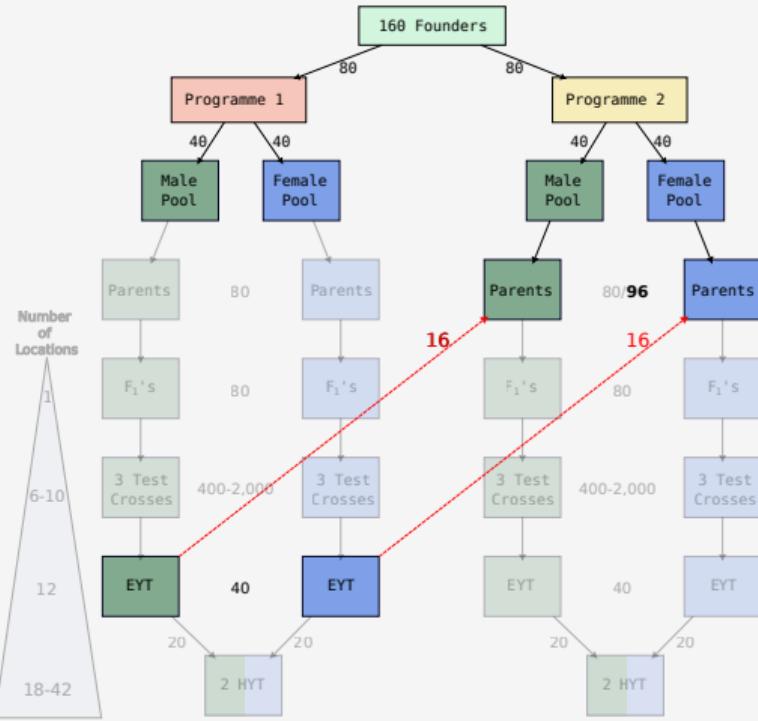


- Trait: Yield per harvested acre in bushels
- Phenotypic Selection
- 20 years of burn-in + 20 future years
- Future breeding:
  - Three scenarios of G×L
- A lag of 10 years between breeding programmes 1 and 2
- Germplasm flow starts in year 21 onwards from programme 1 to 2

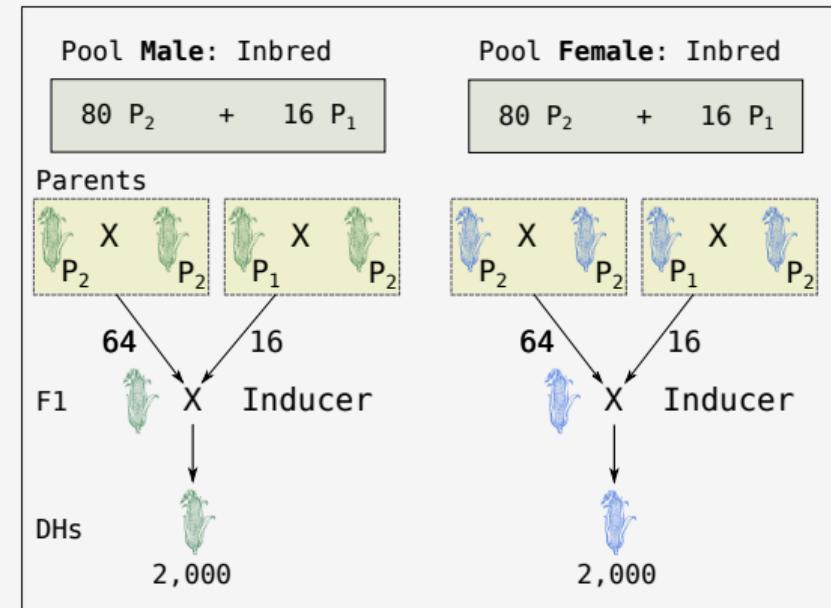
# Germplasm Exchange



# Germplasm Exchange



## Breeding Programme 2 + Importing from 1



# Correlation matrices for additive and dominance

- Let  $\mathbf{a}_i = [a_{i1}, a_{i2}, \dots, a_{in_L}]^T$  be the additive effects for  $i$ th genotype considering  $n_L$  locations

$$\mathbf{a}_i \sim N_{n_L} (\mathbf{0}, \sigma_A^2 \mathbf{R}_A)$$

- Dominance degree:  $\delta_i \sim N_{n_L} (\boldsymbol{\mu}_\delta, \sigma_\delta^2 \mathbf{R}_D) \rightarrow$  dominance effect  $\mathbf{d}_i = \delta_i \odot |\mathbf{a}_i|$

- Scenarios:

- No additive and dominance by location interaction (No G×L)

$$\mathbf{R}_A = \mathbf{R}_D = \mathbf{J}, \text{ with } \mathbf{J} \text{ being a matrix of ones} \quad (1)$$

- Low/High additive and dominance by location interaction

$$\mathbf{R}_A = \begin{bmatrix} \mathbf{R}_A^{(P_1)} & \mathbf{R}_A^{(P_1, P_2)} \\ \mathbf{R}_A^{(P_1, P_2)} & \mathbf{R}_A^{(P_2)} \end{bmatrix} \text{ and } \mathbf{R}_D = \begin{bmatrix} \mathbf{R}_D^{(P_1)} & \mathbf{R}_D^{(P_1, P_2)} \\ \mathbf{R}_D^{(P_1, P_2)} & \mathbf{R}_D^{(P_2)} \end{bmatrix} \quad (2)$$

# Correlation matrices for additive and dominance

- Let  $\mathbf{a}_i = [a_{i1}, a_{i2}, \dots, a_{in_L}]^T$  be the additive effects for  $i$ th genotype considering  $n_L$  locations

$$\mathbf{a}_i \sim N_{n_L} (\mathbf{0}, \sigma_A^2 \mathbf{R}_A)$$

- Dominance degree:  $\delta_i \sim N_{n_L} (\mu_\delta, \sigma_\delta^2 \mathbf{R}_D) \rightarrow$  dominance effect  $\mathbf{d}_i = \delta_i \odot |\mathbf{a}_i|$

- Scenarios:
  - No additive and dominance by location interaction (No G×L)

$$\mathbf{R}_A = \mathbf{R}_D = \mathbf{J}, \text{ with } \mathbf{J} \text{ being a matrix of ones} \quad (1)$$

- Low/High additive and dominance by location interaction

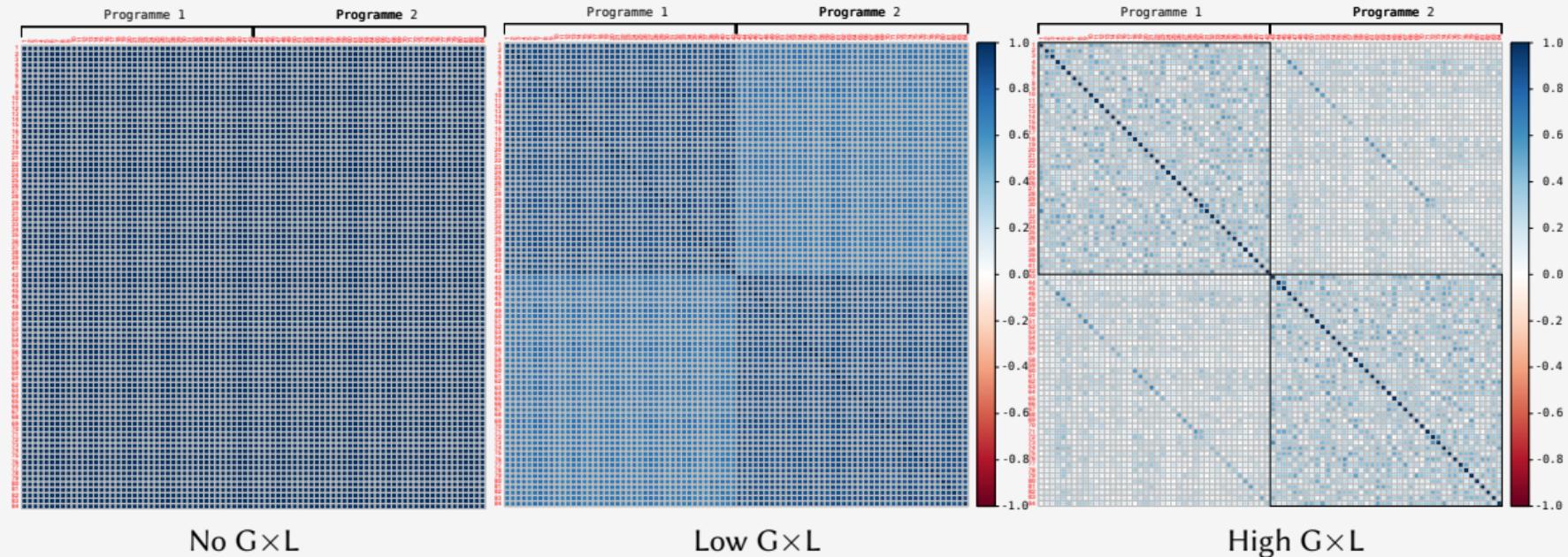
$$\mathbf{R}_A = \begin{bmatrix} \mathbf{R}_A^{(P_1)} & \mathbf{R}_A^{(P_1, P_2)} \\ \mathbf{R}_A^{(P_1, P_2)} & \mathbf{R}_A^{(P_2)} \end{bmatrix} \text{ and } \mathbf{R}_D = \begin{bmatrix} \mathbf{R}_D^{(P_1)} & \mathbf{R}_D^{(P_1, P_2)} \\ \mathbf{R}_D^{(P_1, P_2)} & \mathbf{R}_D^{(P_2)} \end{bmatrix} \quad (2)$$

# G×L Scenarios - Additive Genetic Correlation

1.00	1.00
1.00	1.00

0.84 - 0.93	0.65 - 0.81
0.65 - 0.81	0.84 - 0.93

0.02 - 0.62	0.00 - 0.71
0.00 - 0.71	0.02 - 0.62



# Partitioning genetic trend - AlphaPart

## General idea

- $\mathbf{a} = \mathbf{T}\mathbf{w}$ , with  $\mathbf{w} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{W})$
- Suppose any set  $\mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3 + \dots + \mathbf{P}_m = \mathbf{I}$ , thus:

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

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

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

where  $\mathbf{a}_k$ ,  $k = 1, 2, \dots, m$ , is a part of the breeding value that can be assigned to the group  $k$ .



**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY



New Results



[Follow this preprint](#)

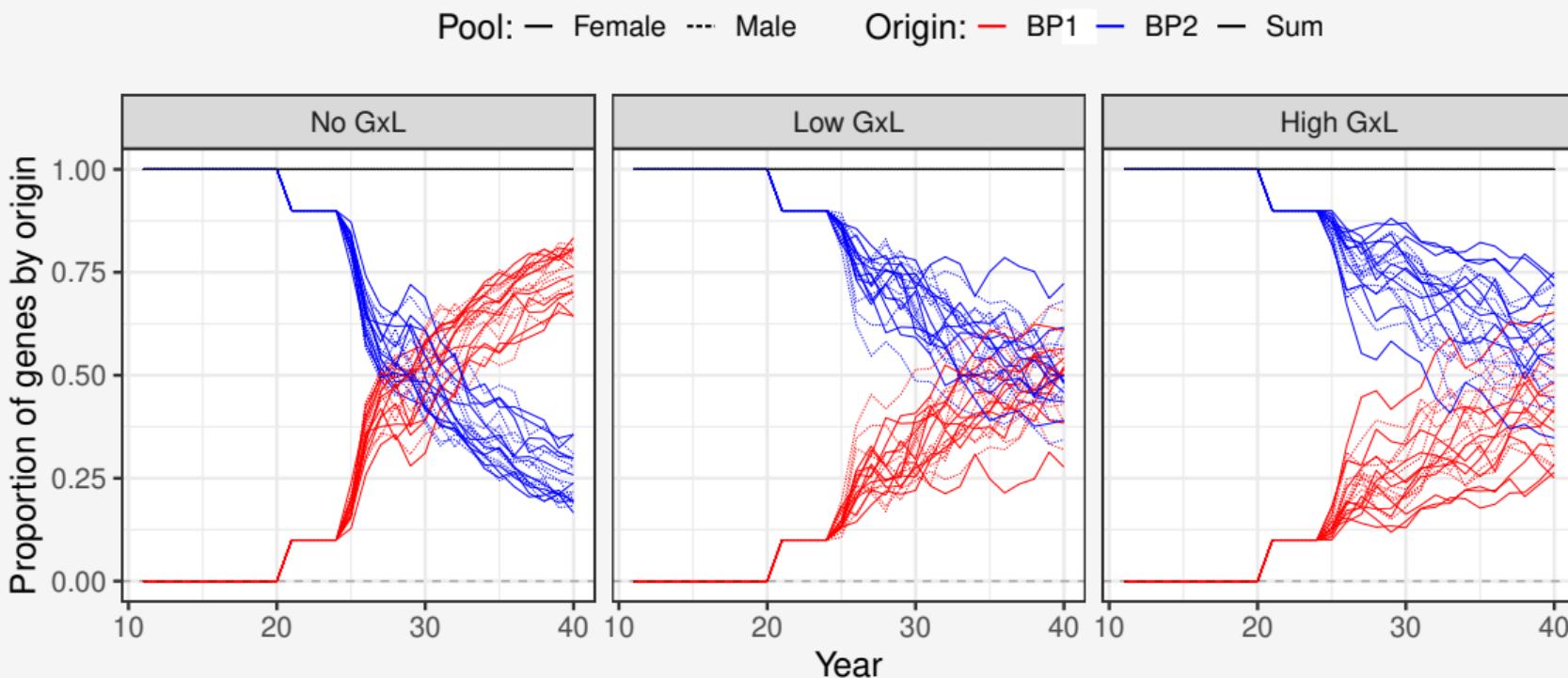
## A method for partitioning trends in genetic mean and variance to understand breeding practices

Thiago de Paula Oliveira, Jana Obsteter, Ivan Pocrnic, 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?].

# Proportion of genes by origin (Programme 2)

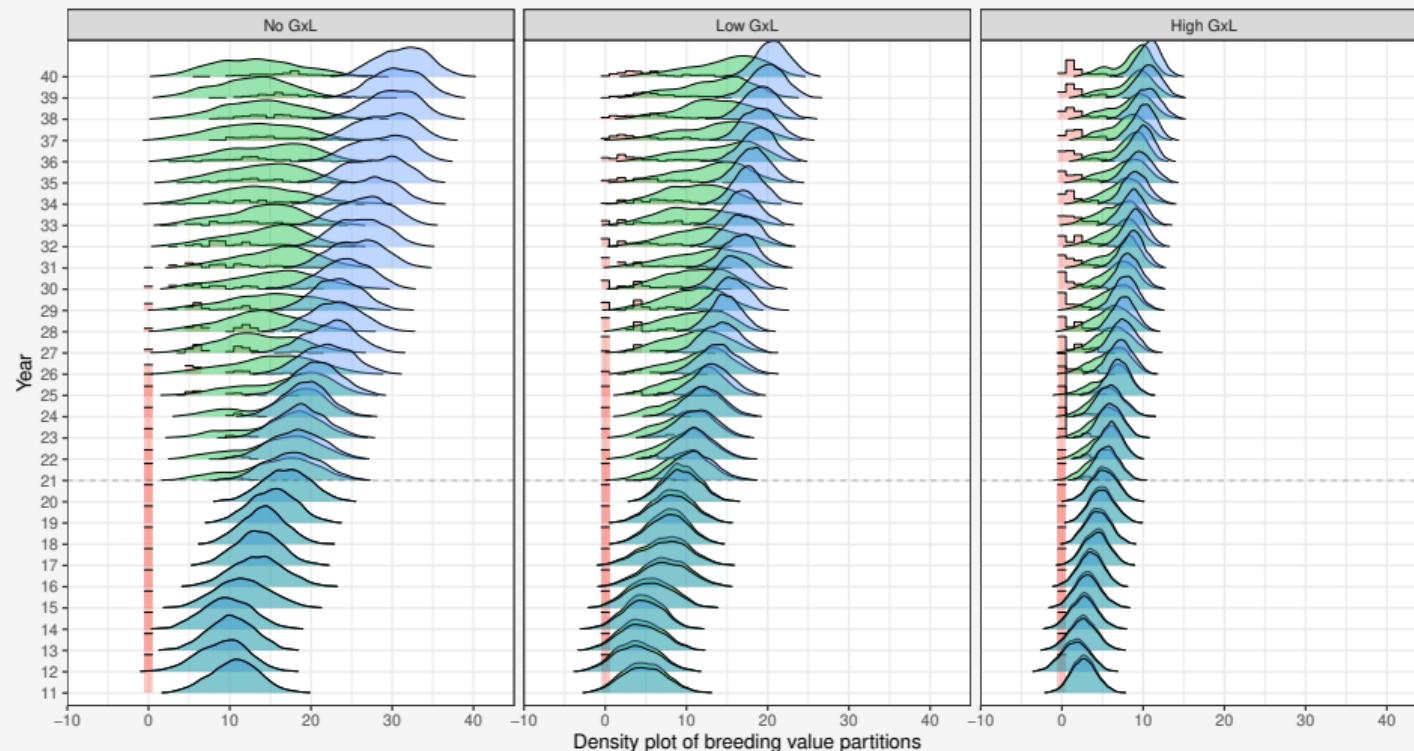
10 rep. & DH stage



# Additive genetic values by origin (Programme 2)

1 rep. & DH stage

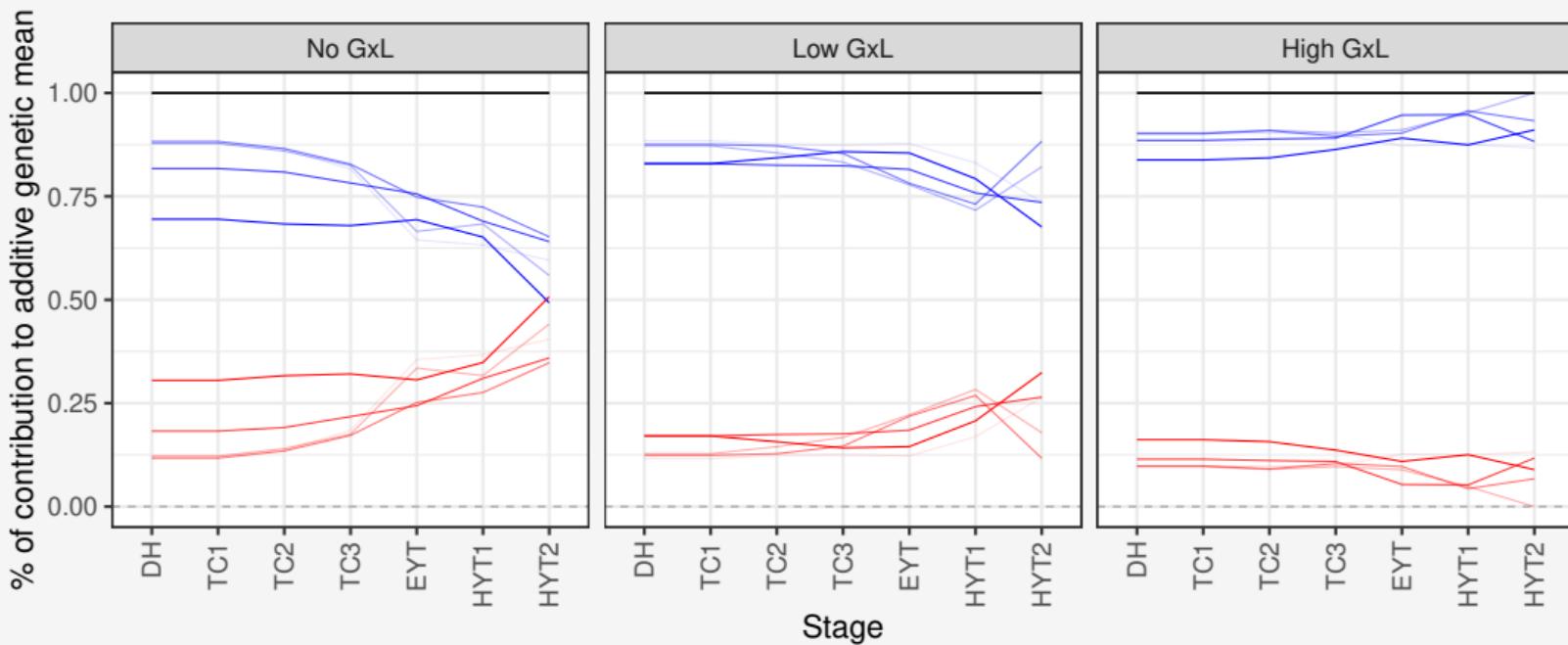
Origin: ■ BP 1 ■ BP 2 ■ Sum



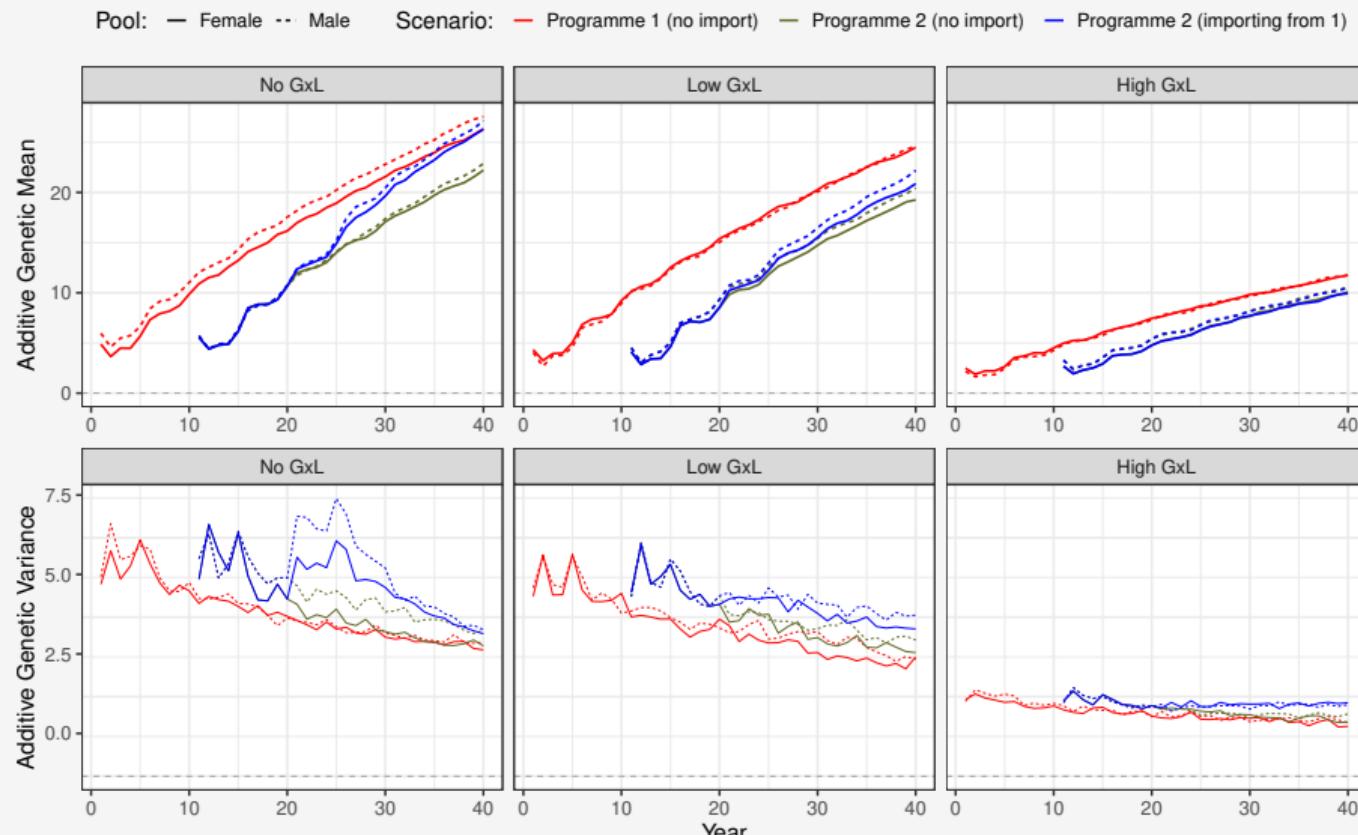
# Contribution to additive genetic mean per stage

1 rep.

BC: — 21 — 22 — 23 — 24 — 25      Origin: — BP1 — BP2 — Sum



# Total additive genetic mean and variance (All programmes) - DH stage



# Final Remarks

- First application of the partitioning method in plant breeding, enabling a clear analysis of the impact of import.
- Import expectedly increased genetic mean and variance.
- The benefit of import is highly dependent on the degree of G×L.
- Import increased mean similar to programme 1 in the last year of the simulation under a non G×L scenario and a lag of 10 years between breeding programmes.
- Next step is to extend the partitioning method with genomic data and quantify its impact.

# Final Remarks

---

- First application of the partitioning method in plant breeding, enabling a clear analysis of the impact of import.
- Import expectedly increased genetic mean and variance.
- The benefit of import is highly dependent on the degree of G×L.
- Import increased mean similar to programme 1 in the last year of the simulation under a non G×L scenario and a lag of 10 years between breeding programmes.
- Next step is to extend the partitioning method with genomic data and quantify its impact.

# Final Remarks

---

- First application of the partitioning method in plant breeding, enabling a clear analysis of the impact of import.
- Import expectedly increased genetic mean and variance.
- The benefit of import is highly dependent on the degree of G×L.
- Import increased mean similar to programme 1 in the last year of the simulation under a non G×L scenario and a lag of 10 years between breeding programmes.
- Next step is to extend the partitioning method with genomic data and quantify its impact.

# Final Remarks

---

- First application of the partitioning method in plant breeding, enabling a clear analysis of the impact of import.
- Import expectedly increased genetic mean and variance.
- The benefit of import is highly dependent on the degree of G×L.
- Import increased mean similar to programme 1 in the last year of the simulation under a non G×L scenario and a **lag of 10 years** between breeding programmes.
- Next step is to extend the partitioning method with genomic data and quantify its impact.

# Final Remarks

---

- First application of the partitioning method in plant breeding, enabling a clear analysis of the impact of import.
- Import expectedly increased genetic mean and variance.
- The benefit of import is highly dependent on the degree of G×L.
- Import increased mean similar to programme 1 in the last year of the simulation under a non G×L scenario and a **lag of 10 years** between breeding programmes.
- Next step is to extend the partitioning method with genomic data and quantify its impact.



```
for (Year in 1:10) {  
  Variety = selectInd(EYT, nInd = 1)  
  EYT = selectInd(AYT, nInd = 10)  
  AYT = selectInd(PYT, nInd = 50)  
  PYT = selectInd(HDRW, nInd = 500)  
  HDRW = makeDH(F1, nDH = 100)  
  Parents = c(EYT, AYT)  
  F1 = randCross(Parents, nCrosses = 100)  
}
```



Free short online course

# Breeding Programme Modelling with AlphaSimR



THE UNIVERSITY  
of EDINBURGH



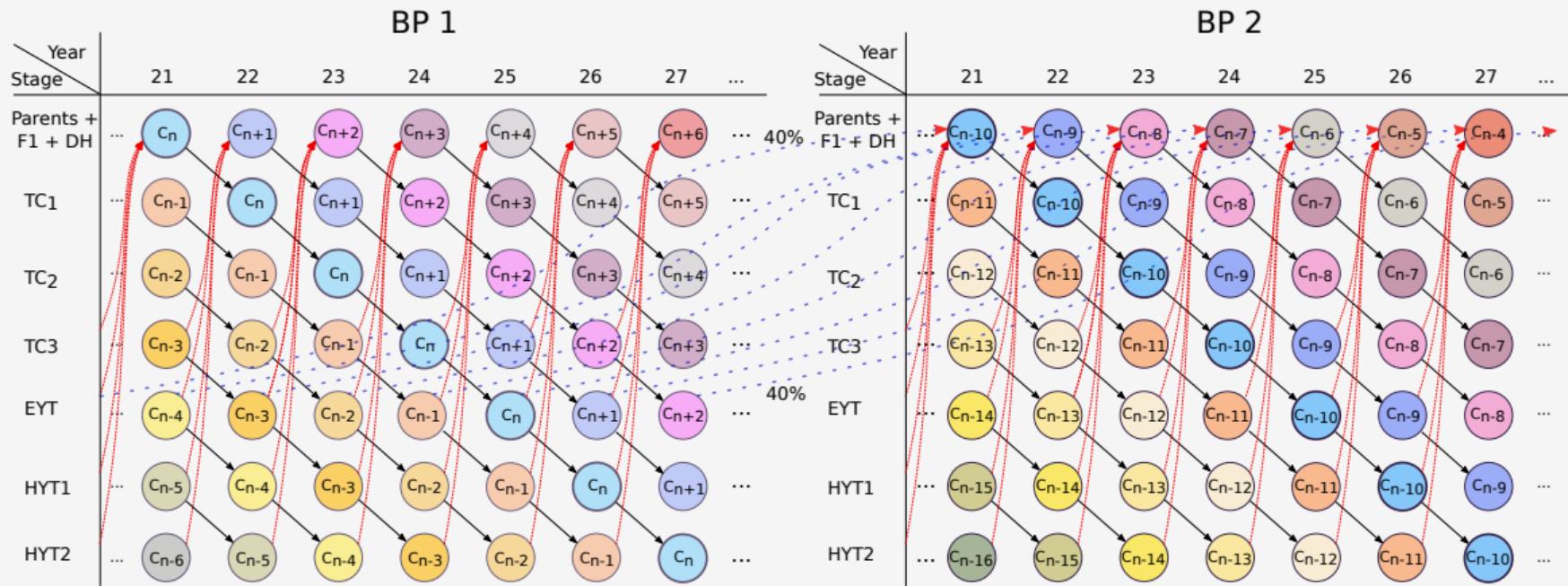
THE  
DATA LAB  
value from data



DDI Data-Driven  
Innovation

Part of the Edinburgh & South East Scotland City Region Deal

## Breeding programme (BP) in parallel



# Important values

---

- runMacs and AlphaSimR
- 10 chromosome pairs
- genetic length of 2 Morgans
- physical length of  $2 \times 10^8$  base pairs
- recombination rate set to  $1 \times 10^{-8}$  base per pair
- mutation rate set to  $1 \times 10^{-8}$  per base pair

# Accuracy of Selection

