

# 2024 ELLS summer school Darwinian agriculture

## Workshop Social evolution under mass selection or under true breeding values

Jacques David, Nicolas Salas, Peter Bourke

2024-06-20

## Introduction

Welcome to this practical session on the decomposition of phenotypic value and the effects of selection. In this session, we will explore how the phenotypic value of a plant is influenced by its genotype and by the genotypes of its neighbors, and how selection on these traits affects the next generation.

## Key theoretical Concepts

find the ref Bijma , Muir etc

## Decomposition of the Phenotypic Value

Classically, in one single macro-environment, the phenotypic value ( $P$ ) of an individual can be decomposed into multiple components:

$$P = G + E$$

Where:

- $G$  represents the genetic contribution to the phenotype,
- $E$  represents the micro-environmental contribution to the phenotype.

In the case of heterogeneous populations, such as Bulks or landraces or farmers varieties, individuals of different genotypes are neighbouring each others. So, if we introduce the fact that the phenotype of a *focal* individual,  $i$ ,  $P_{Focal_i}$ , can be determined by its own genotype and by the genotype of its neighbour,  $j$ ,  $P_{Focal_i, Neighbour_j}$  can now be decomposed into a Direct Genetic Effect ( $DGE$ ) due to  $i$  and the Indirect Genetic Effect ( $IGE$ ) from its neighbour  $j$ , the model becomes:

$$P_{Focal_i, Neighbour_j} = DGE_{Focal_i} + IGE_{Neighbour_j} + E_{DGE_{Focal_i}} + E_{IGE_{Neighbour_j}}$$

Where:

- $DGE_{Focal_i}$  is a Direct Genetic Effect of  $i$ , i.e., the direct contribution of the individual's own genotype to its own phenotype,
  - $IGE_{Neighbour_j}$  is the Indirect genetic contribution of the genotype of the neighbour  $j$
- and,
- $E_{DGE_{Focal_i}}$  : environment effect on Direct Genetic Effect) on the focal plant
  - $E_{IGE_{Neighbour_j}}$  : environment effect on Indirect Genetic Effect of the neighbour

These components have their associated variances :  $V_{DGE}, V_{IGE}, V_{E_{DGE}}, V_{E_{IGE}}$

Of course, any individual in an heterogeneous populations, if a individual  $i$  has a  $DGE_i$  it also have a  $IGE_i$  than can be correlated.

## Correlation between DGE and IGE

A correlation between  $DGE_i$  and  $IGE_i$  will affect the overall distribution of the phenotypic values and hence the response to selection. For a social trait, a strong competitor,  $i$ , can have a strong positive  $DGE_i$  value while reducing the trait value of its neighbours with a negative  $IGE_i$  value.

According to the trait, the correlation between the  $DGE_i$  and their  $IGE_i$  counterpart can theoretically vary from -1 to 1. This correlation will thus translate in the data into a covariance between the  $DGE_i$  and  $IGE_i$  values, denoted  $cov(DGE : IGE)$  with

$$r_{(DGE:IGE)} = \frac{cov(DGE : IGE)}{\sqrt{V_{DGE}V_{IGE}}}$$

## Phenotypic variance in a complex neighbourhood

In our workshop situation, we will consider that plants are grown on a grid and any of them have 8 neighbours with which it is interacting.

Neighbour1	Neighbour2	Neighbour3
Neighbour4	Focal	Neighbour5
Neighbour6	Neighbour7	Neighbour8

$$P_{i,(n1..n8)} = DGE_i + IGE_{n1} + .. + IGE_{n8} + E_{DGE_i} + E_{IGE_{n1}} + .. + E_{IGE_{n8}}$$

Sinc

If a mass selection is performed based on phenotypic value, the selection acts on both the individual's genotype and the genotypes of its neighbors, a concept known as the social phenotype.

**True Breeding Value (TBV)** The true breeding value (TBV) integrates both DGE and IGE. It is an important metric for predicting the genetic value of the next generation, particularly when social interactions are considered. The TBV is influenced by the correlation between DGE and IGE and can be calculated as:

$$TBV = DGE + \sum IGE$$

where  $\sum IGE$  represents the sum of the indirect genetic effects from all relevant neighbors.

**Environmental Effects** Environmental effects ( $E$ ) play a significant role in shaping the phenotypic value. These effects can be unique to each individual (specific environment) or shared among individuals (common environment). In our model, we consider environmental variances for both DGE and IGE.

**Bibliographic Context** Understanding the contributions of genetic and environmental factors to phenotypic variance has been a core pursuit in quantitative genetics. The inclusion of social interactions, as captured by IGE, has been pivotal in advancing breeding programs, particularly in plants and animals with significant social structures. Studies (ref) have laid the foundation for integrating IGEs into genetic models, highlighting their impact on selection response and breeding strategies.

**Simulation Algorithm** The simulation algorithm used in this practical session involves the following steps:

1. **Matrix Preparation:** Genetic (G) and environmental (E) variance-covariance matrices are prepared based on user inputs.
2. **Genotype and Neighborhood Generation:** Genotypes and environmental effects are generated for a grid of plants, and neighbors are assigned randomly.
3. **Phenotypic Value Calculation:** The phenotypic value for each plant is calculated as the sum of its DGE, the IGEs from its neighbors, and environmental effects.
4. **Selection:** A proportion of the top-performing plants is selected based on their phenotypic values.
5. **Next Generation Prediction:** The mean genetic values for the next generation are predicted based on the selected plants, taking into account the variances and covariance between DGE and IGE.

## Detailed Steps in the Simulation

### 1. Matrix Preparation:

- Calculate the covariance between DGE and IGE based on the user-provided genetic correlation.
- Form the genetic variance-covariance matrix  $G$  and the environmental variance-covariance matrix  $E$ .

```
covdgE_IGE <- input$r * (input$varG11 * input$varG22)^0.5
G <- matrix(c(input$varG11, covdgE_IGE, covdgE_IGE, input$varG22), nrow = 2, byrow = TRUE)
E <- matrix(c(input$varE11, 0, 0, input$varE22), nrow = 2, byrow = TRUE)
```

### 2. Genotype and Neighborhood Generation:

- Generate genotype values using a multivariate normal distribution for a grid of  $N \times N$  plants.
- Generate environmental effects similarly.
- Assign neighbors randomly on a grid using a neighborhood matrix.

```
genotype <- mvrnorm(n = N^2, mu = c(0, 0), Sigma = G)
enviro <- mvrnorm(n = N^2, mu = c(0, 0), Sigma = E)
voisinage <- matrix(sample(1:N^2, 4*N^2, replace = TRUE), nrow = 2*N, ncol = 2*N)
```

### 3. Phenotypic Value Calculation:

- For each plant, calculate its DGE and sum the IGEs from its neighbors.
- Add environmental effects to obtain the total phenotypic value.

```
Valeur_G <- function(x, y, genotype, voisinage) {
  geno <- c()
  for (i in (x-1):(x+1)) {
    for (j in (y-1):(y+1)) {
      if (!(i == x && j == y)) {
        geno <- c(geno, voisinage[i, j])
      }
    }
  }
  DGE <- genotype[voisinage[x, y], 1] + enviro[voisinage[x, y], 1]
  IGE <- sum(genotype[geno, 2] + enviro[geno, 2])
  return(DGE + IGE)
}
```

```

Pheno <- c()
tri_geno <- c()
cpt <- 0

for (x in 5:(N+5)) {
  for (y in 5:(N+5)) {
    cpt <- cpt + 1
    Pheno[cpt] <- Valeur_G(x, y, genotype, voisinage)
    tri_geno[cpt] <- voisinage[x, y]
  }
}

```

#### 4. Selection:

- Rank individuals by their phenotypic values.
- Select the top proportion ( $p$ ) of individuals.

```

N_sel <- round(cpt * input$p)
geno_sel <- genotype[tri_geno[order(Pheno, decreasing = TRUE)][1:N_sel], ]

```

#### 5. Next Generation Prediction:

- Calculate the mean genetic values for the next generation based on the selected individuals.
- Adjust the mean values considering the response to selection, including the correlation between DGE and IGE.

```

S <- c()
S[1] <- mean(genotype[tri_geno[order(Pheno, decreasing = TRUE)][1:N_sel], 1])
S[2] <- mean(genotype[tri_geno[order(Pheno, decreasing = TRUE)][1:N_sel], 2])

R <- G %%% solve(P) %%% S
mus <- mu
mus[1] <- mu[1] + R[1]
mus[2] <- mu[2] + R[2]

new_geno <- mvrnorm(n = N^2, mu = mus, Sigma = G)
new_voisinage <- matrix(sample(1:N^2, 4*N^2, replace = TRUE), nrow = 2*N, ncol = 2*N)

new_Pheno <- c()
new_tri_geno <- c()
cpt <- 0

for (x in 5:(N+5)) {
  for (y in 5:(N+5)) {
    cpt <- cpt + 1
    new_Pheno[cpt] <- Valeur_G(x, y, new_geno, new_voisinage)
    new_tri_geno[cpt] <- new_voisinage[x, y]
  }
}

```

By understanding these steps and using the Shiny application, you will gain insights into the dynamics of genetic selection and the impact of social interactions on breeding outcomes.

**Shiny Application Instructions** To use the Shiny application for this practical session, follow these steps:

1. Open the Shiny application in your web browser.
2. Set the number of individuals (N) using the slider.
3. Adjust the genetic variances for DGE and IGE, the genetic correlation, and the environmental variances using the provided inputs.
4. Set the selection pressure to determine the proportion of top-performing plants to be selected.
5. If desired, set a seed for reproducibility by checking the box and entering a seed value.
6. Click the “Run Simulation” button to start the simulation.
7. View the results in the “Graph” and “Summary” tabs, which show the genotype distribution, phenotypic distributions before and after selection, and a summary of the calculations.

**Suggested Parameters** To help you understand the effects of selection, try using the following parameter values:

- **Number of individuals (N):** 20
- **Genetic variance DGE ( $V_{DGE}$ ):** 0.75
- **Genetic variance IGE ( $V_{IGE}$ ):** 0.07
- **Genetic correlation DGE : IGE ( $r$ ):** 0.5
- **\*\*Environmental variance DGE**

**( $V_{DGE}$ ): 0.7 - Environmental variance IGE ( $V_{IGE}$ ): 0.07 - Selection pressure ( $p$ ):\*\*:** 0.3

Observe how changes in these parameters affect the selection outcomes and the predicted genetic values of the next generation.