

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

1. Introduction

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

Understanding the contributions of genetic and environmental factors to phenotypic variance has been a core pursuit in quantitative genetics. The inclusion of social interactions has been pivotal in advancing breeding programs, particularly in plants and animals with significant social structures.

We will consider a theoretical situation that could mimick the situation of a farmer's selection within heterogeneous populations.

The selection can be made either by selecting the best performing phenotypes or by getting measurements of individual plant phenotypes within their neighborhood knowing their genotypes, i.e., several genotypes are present within the plot and every plant is identified for its genotype. We consider a simple situation where genotypes are completely repeatable, e.g. pure lines, or clones or F1 hybrids). You will explore the consequences of social interactions when they are determined genetically.

2. Key theoretical Concepts

2.1 Decomposition of the Phenotypic Value

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

$$P = G + E$$

Where:

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

Let's rigorously derive the total phenotypic variance V_P where the phenotypes may not be independent, either because of genetic relatedness or structured environment such as neighborhoods.

2.2 Independent data

Consider the phenotypes P_i and $P_{i'}$ of two individuals i and i' . In the general case, $V(P) = E(Var(P_i)) + 2cov(P_i, P_{i'})$ and if genotypes are independent and if there is a random distribution of individuals in a homogeneous environment then $cov(P_i, P_{i'}) = 0$ and $V(P) = V(G) + V(E)$.

2.3 Phenotypic variance in a complex neighbourhood : Genetic effects of interacting plants

The general formula $V(P) = E(Var(P_i)) + 2cov(P_i, P_{i'})$ has to be reconsidered since in our workshop situation, we 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 |

Direct and indirect effects

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 on the phenotype of i ,
- and,
- $E_{DGE_{Focal_i}}$: environmental effect on Direct Genetic Effect on the focal plant i ,
- $E_{IGE_{Neighbour_j}}$: environmental effect on Indirect Genetic Effect of the neighbour j ,

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

Of course, in an heterogeneous populations, any individual i has a DGE_i and also has a IGE_i . These two values can be correlated.

Genetic correlation between DGE and IGE

A correlation between DGE_i and IGE_i will affects 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, i.e., a negative IGE_i value.

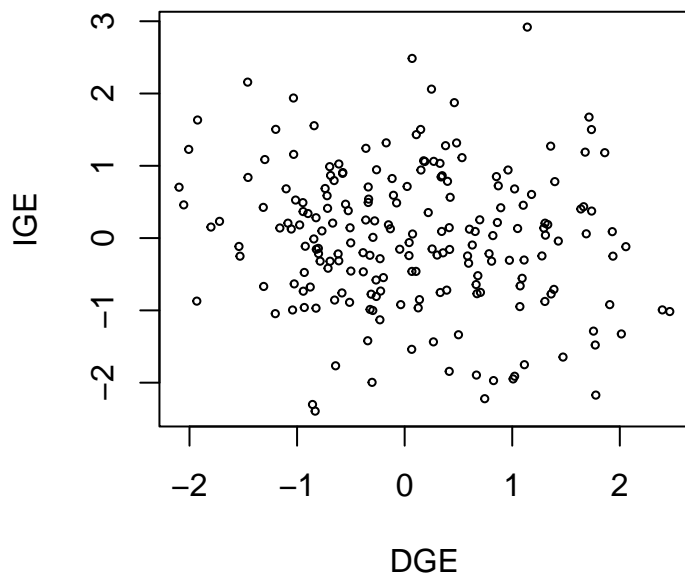
According to the trait, the correlation between the DGE_i and their IGE_i can theoretically vary from -1 to 1. This correlation will thus translate in the phenotypes into a covariance between the DGE_i and IGE_i values, denoted $cov(DGE : IGE)$ with:

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

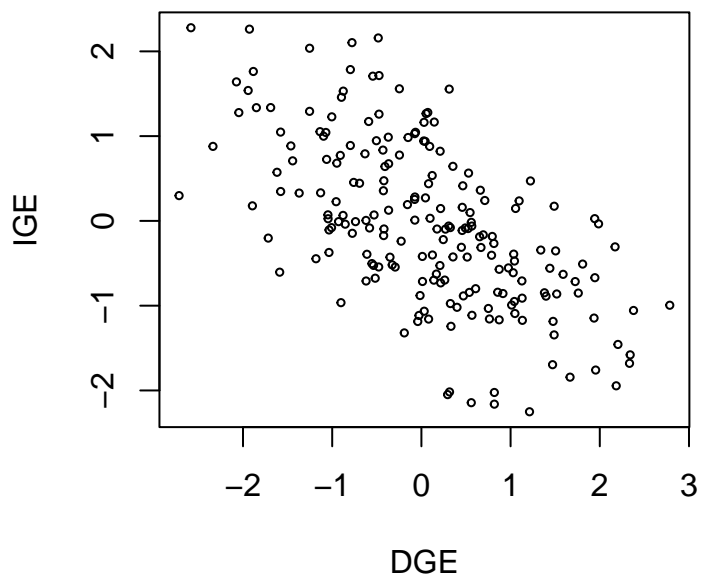
Illustration

Comment the different following situations :

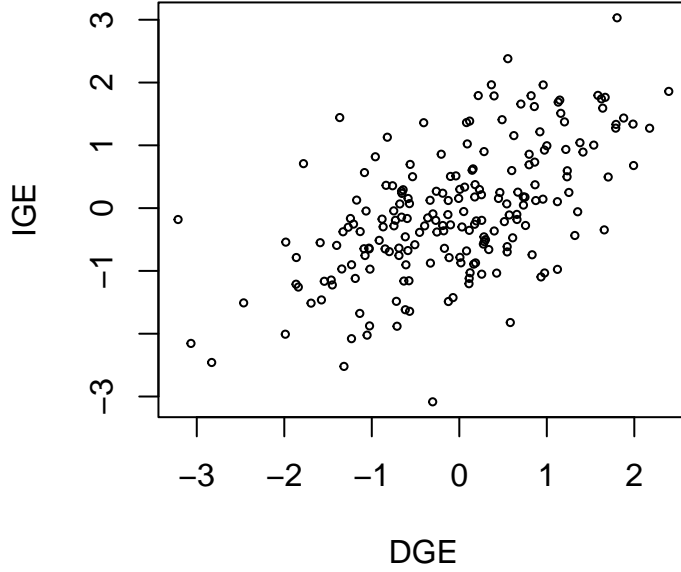
Case A



Case B



Case C



2.4 Phenotypic values and variance with neighborhoods

P_i has to be written according to its 8 neighbors.

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

So the $V(P)$ becomes :

$$V(P) = V(DGE) + 8V(IGE) + V(E_{DGE}) + 8V(E_{IGE}) + 2cov(P_i, P_j)$$

where $2cov(P_i, P_j)$ is 0 when i and j are in different neighborhoods

BUT

non null when they are in the same neighbourhood because:

- they have common neighbours, e.g., the focal plant shares the plant N1 with plants N2 and N4 as a common neighbour and thus, they share IGE_{N1} in common ($cov(IGE_{N1}, IGE_{N1}) = V(IGE)$) and,
- that the phenotypic value of j contains IGE_i that is correlated to DGE_i , so $cov(DGE_i, IGE_i) = cov(DGE : IGE)$. Here this is true for the 8 neighbors of any plant i .

You are going to explore this situation with the simulations to develop your understanding.

If you detect important deviation to the classical case, then it may be of importance to get estimate for each genotype of its DGE_i and IGE_i values.

We won't go into all the theoretical developments here. For more details, see Bourke's course and the papers cited at the end of these documents.

2.5 True Breeding Value (TBV)

Definition

The true breeding value of a genotype, TBV integrates both its DGE and IGE. In mixture of two genotypes the IGE are applying on only one neighbor on average thus :

$$TBV_i = DGE_i + IGE_i$$

But in our case, each plant has 8 neighbours so

$$TBV_i = DGE_i + 8IGE_i$$

This leads to the variance of TBV :

$$\begin{aligned} V(TBV) &= V(DGE_i + 8IGE_i) \\ &= V(DGE) + 8 * 8 V(IGE) + 2 * 8 cov(DGE : IGE) \\ &= V(DGE) + 64 V(IGE) + 16 cov(DGE : IGE) \end{aligned}$$

An alternative heritability measure T^2 is defined as the proportion of the total phenotypic variance attributable to the variance of the total breeding value:

$$T^2 = \frac{V(TBV)}{V_P}$$

with $V(P) = V(TBV) + V(E)$

T^2 is an important metric for predicting the genetic value of the next generation, particularly when social interactions are considered. Since the $cov(DGE : IGE)$ can be negative, T^2 can diminish compared to the classical H^2 .

BLUPs of DGE's and IGE's

We propose here to estimate the values using the dat with the *SOMMER* package. We can thus get Best Linear Unbiased Prediction (BLUP) of DGE and IGE of each genotype present and repeated in the simulated population.

For motivated students, here's the call to the `mmer` function in the *SOMMER* package. `Pheno` is the phenotype, `Focal` is the occurrence matrix containing the focal genotype, `Zv` is the neighborhood matrix.

```
Mod <- mmer(fixed = Pheno ~ 1, random = ~vsr(Focal) + vsr(Zv), rcov = ~units, data = DATA, nIters = 4)
```

Please explore the app.R code for more details.

Here the possibility of declaring a covariance between IGE and DGE is not considered but *SOMMER* and even better *ASREML* can take it into account and provide an estimation.

The `Mod` object contains the BLUPs for DGE and IGE which will be used for building an index of selection. The Shiny app will provide graphs for you to visualise the precision of the estimates compared to the true values.

Selection index

A selection index is based on a weighted average between the BLUPs of DGE and IGE. Selection can thus be carried out on sorted values of such an index.

$$I = b(DGE_{BLUP}) + (1 - b)(IGE_{BLUP})$$

where b is the relative weight of DGE on IGE.

You can expect that a selection based on this index will result in a different choice than that obtained on best phenotype or the best genotypic average.

You will thus be able to compare mass selection with a selection in which you have BLUPs of DGE and IGE of each genotype. You could choose which term to maximise using the value of

3. Simulated populations and selection

3.1 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. **BLUPS** Prediction of DGE and IGE are obtained from phenotypic values and the neighbourhood position using the SOMMER package
5. **Selection:** A proportion of the top-performing plants is selected based on their phenotypic values or their index value based on their BLUPS
6. **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.
7. **New phenotypic values** The new genotypes replications (plants) are distributed on a grid and the phenotypes are computed accordingly. Results are provided.

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.

3.2 Shiny Application Instructions

Build the population

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 genotype (N) using the slider,
3. Set the number of repeats of each genotype (rep) using the slider and consider that the product N.rep not going above 1000 individuals (you can but it will be quite long) .
4. Adjust the genetic variances for DGE and IGE, their genetic correlation, and the environmental variances using the provided inputs.

Then **click the red button** 1.Run simulation .

If you want to reproduce exactly the same results another time, tick “Set seed for reproducibility”. If this box is not ticked, results will differ for each run.

In the first result panel, graphs will appear permitting to discover the realised properties the population according to the parameters values you have chosen.

Breeding

5. Set the selection pressure to determine the proportion of top-performing plants to be selected.
6. Choose the value of the selection index, the relative weight of DGE to the IGE
7. **click the Blue button** 2.Breed. You can modify the parameters of selection without rebuilding a population.
8. View the results in the “Selection” and “Summary” tabs, which show the genotype distribution, phenotypic distributions before and after selection, and a summary of the calculations. They will renew each time you click the Breed button.

You get two main graphs, one using Mass selection on direct phenotypic values and the other when using the knowledge on the DGE and IGE BLUPS estimated from the data.

The summary tab provide a number of stats allowing you to compare the efficiency on your selection processes.

4.Explore different situations

Once you have understood the functioning of the Shiny app, group by 3 students and start discussing.

4.1 Classical modeling

What would be the values of the different parameters that are representative of a classical situation : no social interaction

4.2 Understanding the impact of social genetics

To explore the impact of social genetic interactions on selection outcomes and the predicted genetic values of the next generation, follow these detailed steps:

1. Set Up Basic Parameters:

- **Initial Population:** Define the initial population size. Ensure that the total population size multiplied by the number of replicates does not exceed 1000 for faster calculations. Distribute your plots between the number of genotypes and the number of repeats
- **Genetic Variances:**
 - **Direct Genetic Effect (DGE):** Adjust the genetic variance for direct effects
 - **Indirect Genetic Effect (IGE):** Adjust the genetic variance for indirect effects.
 - **Genetic Correlation:** explore the effect of the genetic correlation between DGE and IGE.
- **Environmental Variances:** Define the environmental variances for both direct and indirect effects.

2. Explore Different Scenarios:

- **Parameter Grid:** Create a grid of parameter values to explore different scenarios. This involves varying the values of DGE, IGE, their correlation, and environmental variances systematically.
- **Systematic Variation:** By changing one parameter at a time while keeping others constant, you can isolate the effects of each parameter. Alternatively, you can change multiple parameters simultaneously to see combined effects.

3. Run Simulations:

- **Initial Run:** Conduct an initial run with a baseline set of parameters to establish a reference point.
- **Subsequent sel:** For each set of parameters in your grid, run the simulation and record the outcomes.

4. Analyze Results:

- **Selection Outcomes:** Observe how the changes in parameters affect the selection outcomes. Look for patterns in the performance of selected individuals.
- **Predicted Genetic Values:** Examine the predicted genetic values of the next generation under different scenarios.
- **Comparative Analysis:** Compare results across different runs to identify which parameters have the most significant impact on selection outcomes and genetic values.

5. Report Findings:

- **Document Observations:** Keep detailed notes on the results of each simulation, highlighting any notable trends or unexpected results.
- **Graphs and Tables:** Use graphs and tables to visualize the data and make comparisons easier to understand.
- **Summary:** Summarize your findings, discussing how different parameters influence the selection process and the genetic improvement of the population.

By following these steps, you will gain a deeper understanding of how social genetic interactions affect breeding outcomes and be better equipped to make informed decisions in practical breeding programs.

5. References

Piter Bijma, William M Muir, Esther D Ellen, Jason B Wolf, Johan A M Van Arendonk, Multilevel Selection 2: Estimating the Genetic Parameters Determining Inheritance and Response to Selection, Genetics, Volume 175, Issue 1, 1 January 2007, Pages 289–299, <https://doi.org/10.1534/genetics.106.062729>