Horizon 2020 of European Union
This project has received funding from the European Union's Horizon 2020 Programme for Research & Innovation under grant agreement n727312

WorkShop Genomic Evaluation

05/11/2020

Marie PEGARD

Introduction

A brief presentation of myself

- 2015 - 2018 $\rightarrowtail$ PhD : New models for implementation of genome-wide evaluation in black poplar breeding program. - INRAE UMR BioForA
- 2019 $\rightarrowtail$ Post-Doctoral Fellowship : Studying the genetic diversity management in the genomic selection context for two species : the maritime pine and the black poplar. - INRAE UMR Biogeco
- Since January $\rightarrowtail$ PhD on the EUCLEG projet in the WP5 INRE URP3F Lusignan
- My email : marie.pegard@inrae.fr

Workshop Goals

This workshop :

- Takes place in the WP5 of the EUCLEG project.
- Provide knowledge and tools for breeders
- An user-friendly software (Progeno) was developed
- R programming language linked to the Progeno program in command line.

This workshop covers :

- data preparation
- QTL detection
- genomic evaluation
- genomic selection.

⇒ Methods easy to implement and robust

Workshop Organization

- Theoretical part
  - Data Observation and understanding
  - GWAS : Genome Wide Association Studies
  - Genomic selection
  - R
  - Rmarkdown
- Overview of the practical part
  - Presentation of the dataset
  - What are we going to do in this workshop?
  - How the practical part will take place ?
- Pratical part

Data observation and understanding

Obervation and understanding of data

A good understanding of an experimental design and the collected data observation :

- Is part of the experiment.
- Essential pre-requisite for the field data preparation or analysis

I will provide **guidelines** on :

- How to observe and analyse field data
- How to prepare the data
- How to fit a linear model to adjust the phenotypic data before GWAS, GWE or GS analysis.

Why prepare and observe data ?

In genome-wide association studies :

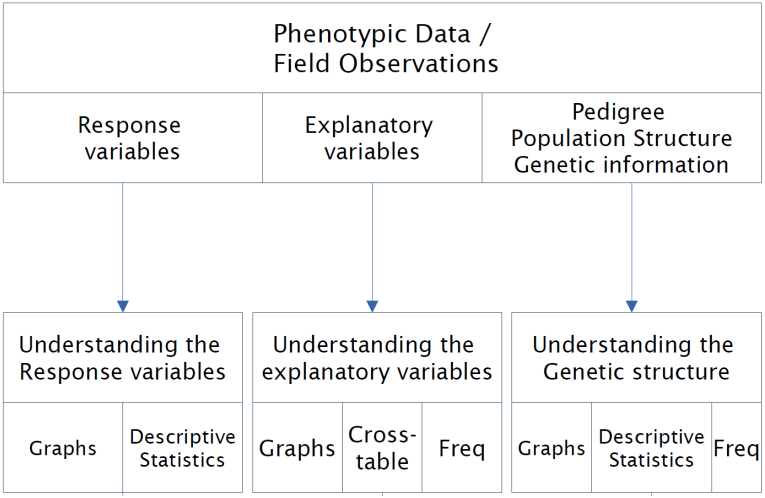
- Determine the **relationships between phenotypes and genotypes**
- Rare phenotypes and their associated genotypes are of great interest

Rare phenotypes $\rightsquigarrow$ the result of technical errors or bad models

False Rare phenotypes $\Rightarrow$ leads to the detection of false QTLs or inaccurate prediction

To avoid this $\Rightarrow$ have the same *rigour* to the analysis of phenotypic and genotypic data.

An overview of the process



Phenotypic Data observation

Raw description of our example:

Let us take the following example: Eucleg's Soy data for 1 trial over 2 years.

There is 15 columns, including :

- Field information : Plot, Row, Colum, Trial, Year
- Genetic information : EntryNo, EUCLEGID, Accession name
- Trait of interest : Seed yield, Protein content
- Explanatory variables : R1 (beginning of flowering), R2 (full flowering), R5 (start of seed filling), R8 (maturity)

For the example we are going to look at the protein content trait only

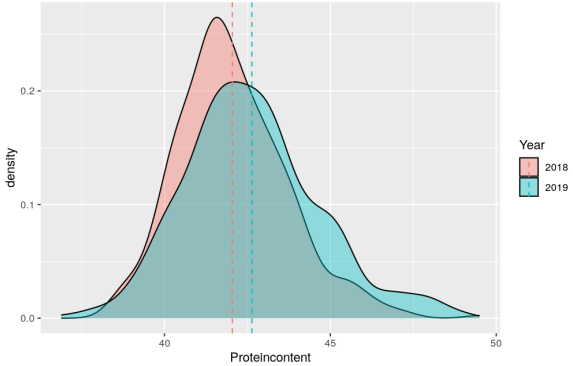
Response variable protein content

Descriptive statistics summarize data to facilitate the exploration :

- Include measures of centrality
  - the mean : `42.34`
  - the median : `42.13`
- Include measures of dispersion
  - the deviation : `1.88`
  - the range `36.9, 49.48`

Response variable protein content

I look at the distribution and the phenotypic mean of the traits depending on the year. The distribution looks normal for the two years



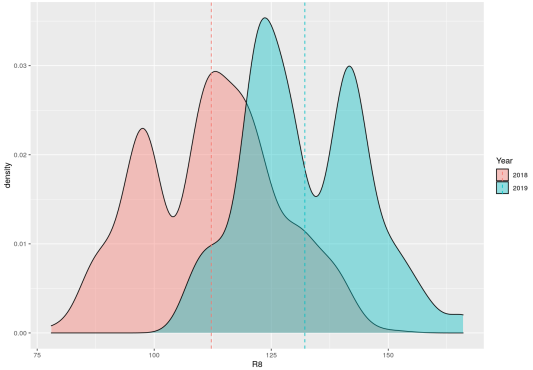
Explanatory variable R8

Descriptive statistics summarize data to facilitate the exploration :

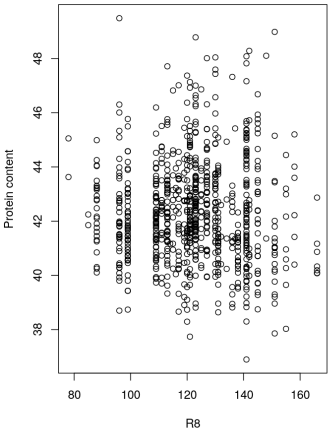
- Include measures of centrality
  - the mean : `122.19`
  - the median : `123`
- Include measures of dispersion
  - the deviation : `16.79`
  - the range `78, 166`

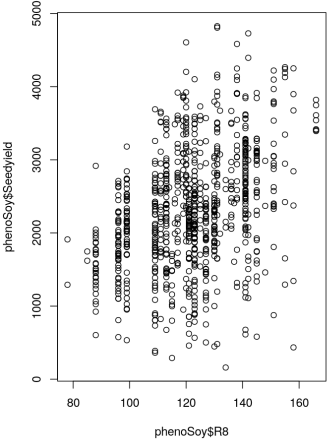
Explanatory variable R8

The distribution is not normal but bimodal.



Correlation between Response and Explanatory variable





Genotyping data - missing values

An important aspect in genotyping data is missing data:

- too much missing data $\rightsquigarrow$ to remove markers or individuals
- reasonable number of missing data $\rightsquigarrow$ several solutions are possible
  - imputation software (BEAGLE, FImpute, AlphaImpute) can be used, or
  - imputation with the mean or allelic frequency

Genotyping data - missing values

There is no fixed threshold for determining whether the number of missing data is reasonable or not, it depends on the population of interest:

- Access to pedigree and full genotyping in the parents $\rightsquigarrow$ imputation of 80% or even 90%
- Complete independent individuals $\rightsquigarrow$ beyond 1% or 2% of the population, the bias brought about by imputation can be critical.

Genotyping data

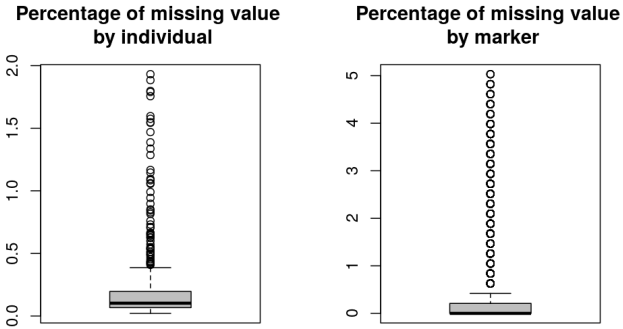
In the following example there is :

- 360 phenotyped individuals
- 357 genotyped individuals
- 226798 SNPs

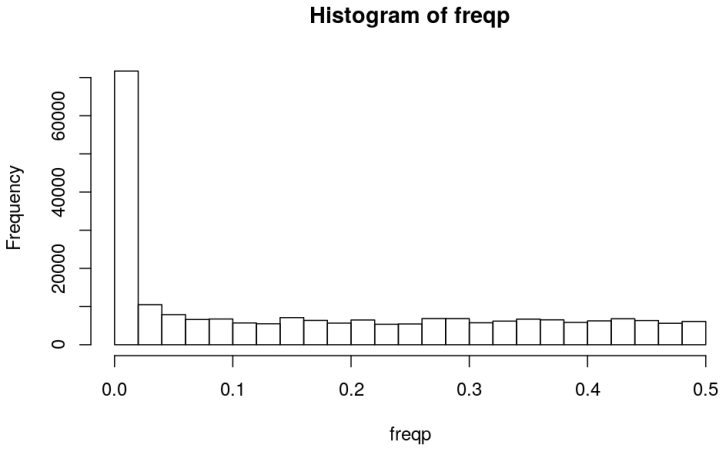
Visualisation of missing data

In the present situation, the number of missing value is really low (`0.211%`)

⇒ I replaced the missing value by the minor allelic frequency



Visualisation of the Minor allelic frequency



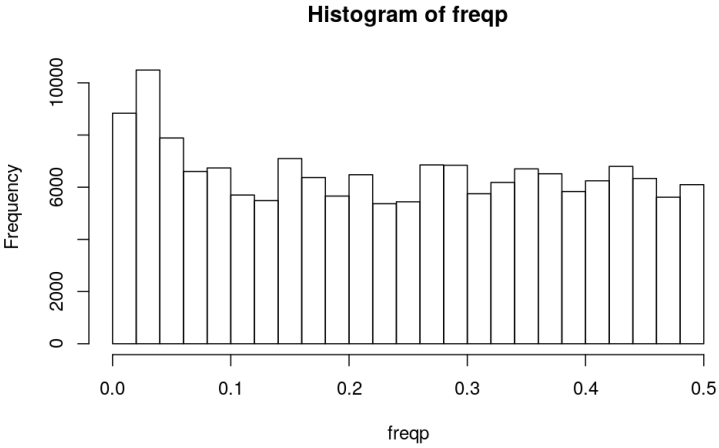
SNPs trimming

We remove SNPs with a MAF of less than 0.01 (62,863 SNPs are concerned)

At this frequency a very large number of individuals are needed to be able to estimate their effect correctly.

To let them bias the analysis would be a mistake.

SNP trimming



Visualization Marker density along the genome

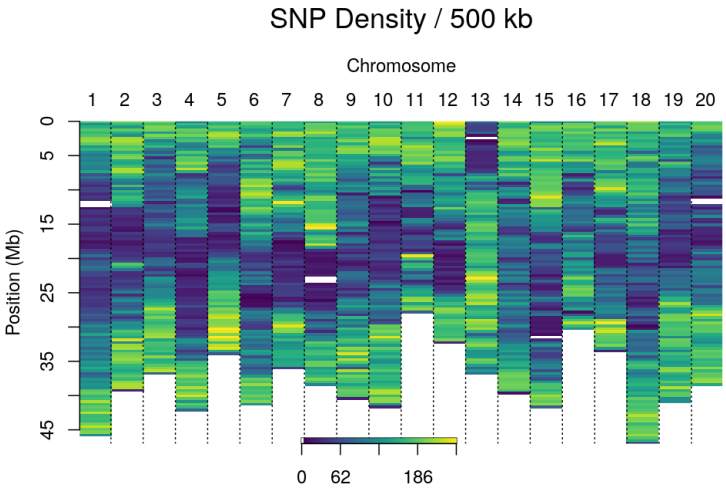
Observation of the distribution of markers along the chromosomes.

There is 99.2041376 % of the markers on the chromosomes, the rest on the scaffolds.

When the physical position of the markers is available, it can be interesting to look at their distribution.

Are there areas of the genome that are absent or less well covered by genotyping?

Visualization Marker density along the genome



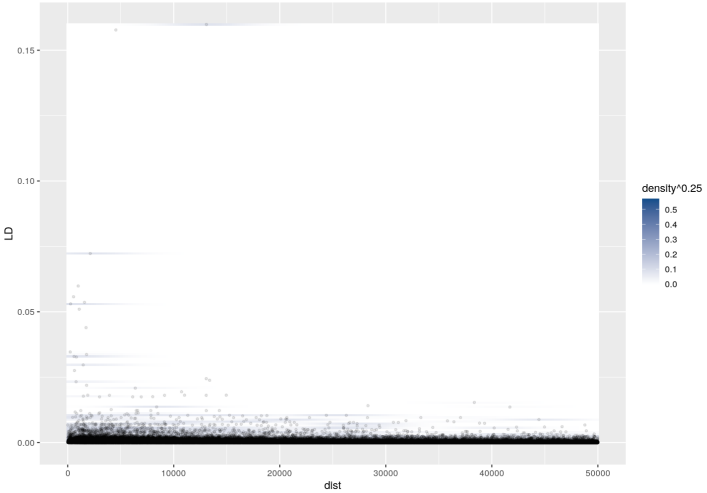
Linkage disequilibrium

The Linkage disequilibrium is the Non random association between alleles from different (linked) loci.

A quick way to do this is to calculate a partial correlation squared.

When we have the physical (or genetic map) we can represent it according to the distance as here with chromosome 1

Linkage disequilibrium - visualisation



Genetic structure

The genetic structure can be asses with several softwares and methods :

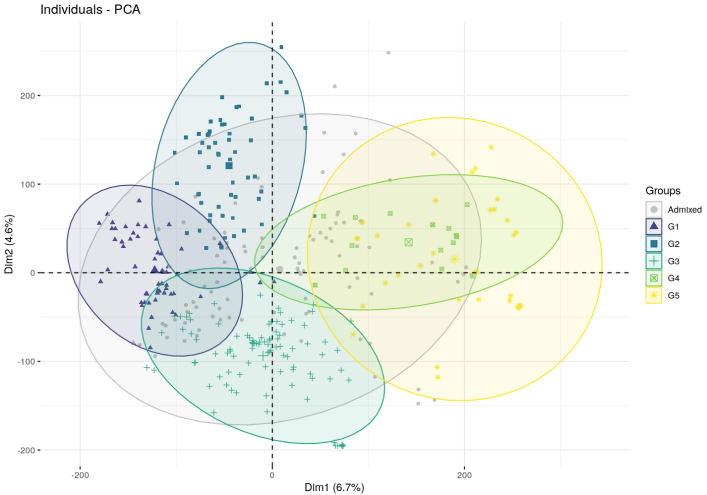
- STRUCTURE, ADMIXTURE,
- The R package Adegenet (Discriminant Analysis of Principal Components (DAPC))

If a structure in the population is observed. It is always interesting to represent it graphically.

The structure can be used later for :

- QTL detection
- data adjustment
- to optimize the creation of training and validation populations.

Visualisation - Genetic structure



Visualisation - Pedigree

In some case, the pedigree information is available but not the genotypes

The pedigree is a kind of genetic structure

```
##          Name         P1         P2
## 1: EUC_GM_009 EUC_GM_239       <NA>
## 2: EUC_GM_050 EUC_GM_179       <NA>
## 3: EUC_GM_055 EUC_GM_204       <NA>
## 4: EUC_GM_056 EUC_GM_328       <NA>
## 5: EUC_GM_072 EUC_GM_328       <NA>
## 6: EUC_GM_097 EUC_GM_604 EUC_GM_615
```

Understanding the data

We carried out the first steps of observing the available data

Understanding the Response variables		Understanding the explanatory variables			Understanding the Genetic structure		
Graphs	Descriptive Statistics	Graphs	Cross-table	Freq	Graphs	Descriptive Statistics	Freq

Understanding the data

What did we learn from our observations

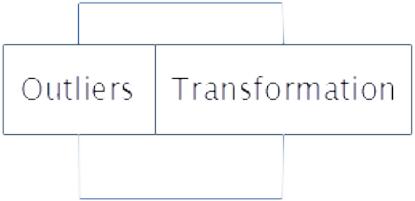
1. The response variables have a distribution close to normal
2. The explanatory variable R8 is not correlated with the response variable Proteincontent but has it with the variable Seedyield
3. The genotyping data are of good qulity
4. We have replaced the NA and removed the low frequency markers
5. A genetic structure was detected in the data
6. Not all phenotyped individuals are genotyped
7. ...

Understanding the data

The next step concerns the outliers detection and the data transformation

There is no need to transform our response variable in our example.

In all case, It is necessary to be careful about data transformation

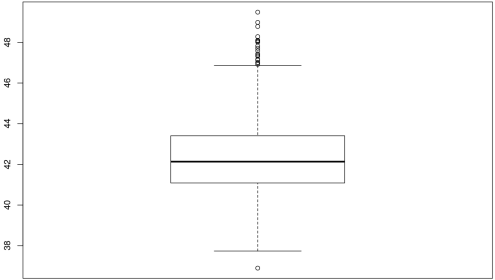


Outliers

Outliers can be detected with a boxplot.

The point after the brackets are outliers.

In the true worlds when there is a continuum like here, it is not outliers.



Linear models

Linear models are a good way of extracting the components of the phenotype and remove all undesired effects

Fit a linear model	
ANOVA	Model diagnostics

Linear models and Phenotypes

Let us specify the writing of the linear model for any individual i ($i = 1, \ldots, n$) of a sample of size n :

$$Y_i = \sum_{j=1}^{p} \beta_j X_i^j + U_i$$

- Y_i is the true response random variable
- U_i is the true error random variable, assumed to be $\mathcal{N}(0, \sigma^2)$ and independent
- β_j are coefficients, unknown parameters, to be estimated : generally random effects
- X_i^j are the values of the explanatory variables : generaly fixed effects

Linear models and Phenotypes

Matrix notation

$$Y = X\beta + U$$

- Y and U are random vectors
- X is a matrix $n \times p$
- β is the vector p parameters.

In our example

The model :

$$Y_{jk} = \mu + s_k + g_j + e_{jk}$$

- Y_{jk} : phenotype of the jth genotype in the kth spatial location (row and column)
- μ : overall mean
- s_k : the effect of the kth spatial location
- g_j : the genetic effect of the jth genotype
- e_{jk} : the residual

In our example

The model :

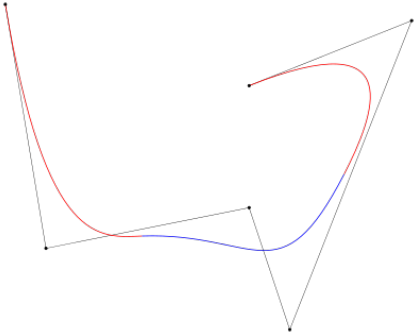
$$Y_{jk} = \mu + s_k + g_j + e_{jk}$$

We are going to integrate in our model:

- the intercept
- a genetic effect: via the EUCLEDID code, the genetic structure or molecular markers
- a spatial effect: plot (xy coordinates), calculated with a Bi-spline model
- the residual

Bi-splines

- A B-spline is a linear combination of positive splines
- B-splines are the generalisation of the Bézier curves
- The shape of the basic functions is determined by the position of the nodes.
- The curve is within the convex envelope of the control points.



Linear model and Phenotypic adjustment

For GWAS or GS cases, the only parts of the phenotype we are interested in are :

- the genetic part
- the residual part.

We want the phenotyes to be cleaned of ground and environmental effects

Linear model and Phenotypic adjustment

It is possible to do this year by year or to combine years.

Let's take the example of the year 2018 for the protein content.

I will use an "EuclegID" effect to represent the genetic effect

I use the R package breedR to do it.

Linear model and Phenotypic adjustment

BreedR function

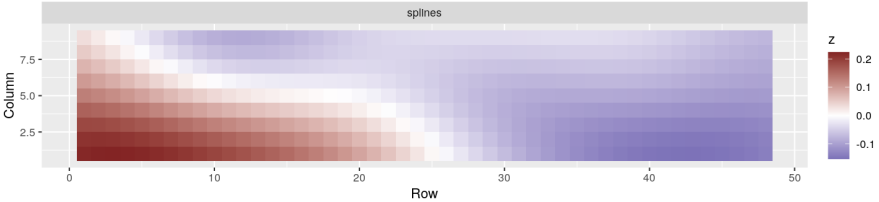
```
resA1 <- remlf90(
  fixed  =  Proteincontent  ~ 1  ,
  random = ~ EUCLEGID, #,
  spatial = list(model = 'splines',
                 coord = phenoSoy[phenoSoy$Year == 2018,
                                  c('Row', 'Column')]),
  data = phenoSoy[phenoSoy$Year == 2018,],
  method = 'em'
)
```

Linear model and Phenotypic adjustment

```
## Formula: Proteincontent ~ 0 + Intercept + EUCLEGID + spatial
##    Data: phenoSoy[phenoSoy$Year == 2018, ]
##   AIC     BIC logLik
##  1484 unknown   -739
##
## Parameters of special components:
## spatial: n.knots: 12 9
##
## Variance components:
##          Estimated variances
## EUCLEGID            2.47300
## spatial             0.03121
## Residual            0.23300
##
```

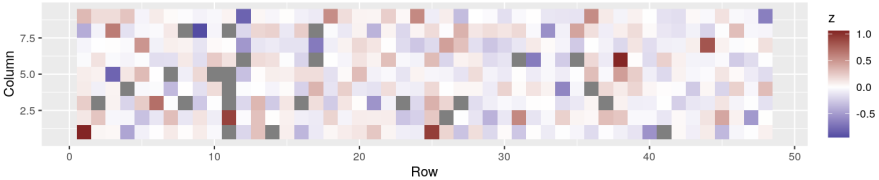
Linear model and Phenotypic adjustment

Spatial effet :



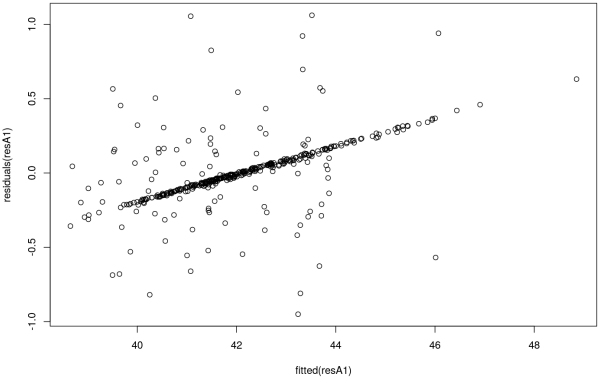
Linear model and Phenotypic adjustment

Residuals



Linear model and Phenotypic adjustment

Fitted values vs Residuals



Linear model and Phenotypic adjustment

Heritability

```
## [1] 0.9034747
```

Linear model and Phenotypic adjustment - 2019

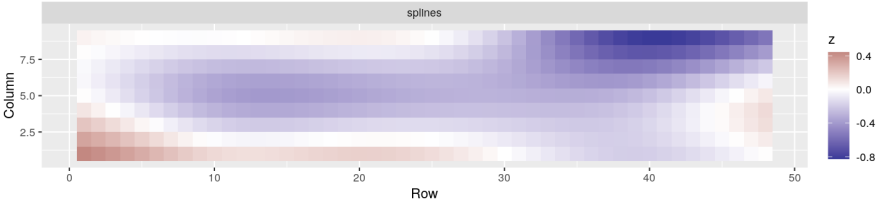
We are doing the same for the second year.

Heritability

```
## [1] 0.8198598
```

Linear model and Phenotypic adjustment - 2019

It is clear that the field effect is different between the two year.



Combined

Now We make a model that combines the two years. This will allow a better estimate of the genetic value of individuals because several years of phenotyping will be cumulated.

The model :

$$Y_{ijk} = \mu + s_{kl} + g_j + a_i + e_{ijkl}$$

- Y_{ijk} : phenotype of the jth genotype in the ith year of the kth spatial location (row and column)
- μ : overall mean
- s_k : the effect of the kth spatial location
- g_j : the genetic effect of the jth genotype
- a_i : the year effect
- e_{ijk} : the residual

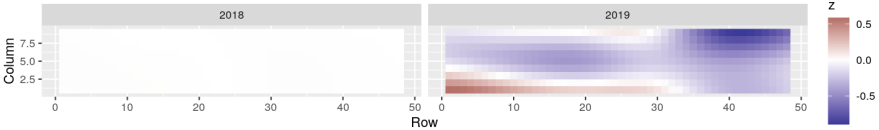
Combined

```
Glob_res <- remlf90(
  fixed  =  Proteincontent ~ 1 + Year,
  random = ~  EUCLEGID,
  generic = list(sp18 = list(inc.sp18,
                             breedR:::get_structure(sp18)),
                 sp19 = list(inc.sp19,
                             breedR:::get_structure(sp19))
  ),
  data = phenoSoy,
  method = 'em'
)
```

Combined

```
## Formula: Proteincontent ~ 0 + Intercept + Year + EUCLEGID
##    Data: phenoSoy
##   AIC     BIC logLik
##  3023 unknown  -1508
##
## Parameters of special components:
##
##
## Variance components:
##          Estimated variances
## EUCLEGID            2.415000
## sp18                0.002286
## sp19                0.338500
## Residual            0.890300
```

Combined



Combined

What if we integrate genomic information

This is equivalent to making a genomic evaluation.

A basic formula is :

$y = \mu + Xg + \epsilon$

- g is a vector $(p \times 1)$ of GEBV,
- X a design matrix linking observations to individuals.

GEBVs follow a normal distribution, with a *covariance* between effects modelled through an *additive relationship matrix* that is derived from markers for GEBV or from pedigree in classical pedigree-based BLUPs.

I will give the details when we discuss GWAS and genomic selection.

What if we integrate genomic information

In breedR

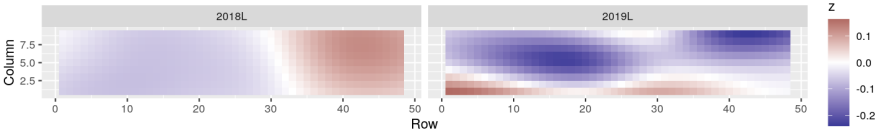
```
Glob_res_lopez <- remlf90(
  fixed  =  Proteincontent ~ 1 + Year,
  # random = ~  1, # If you need random effects
  generic = list(genetic = list(inc.H,HX),
                 sp18 = list(inc.sp18,
                             breedR:::get_structure(sp18)),
                 sp19 = list(inc.sp19,
                             breedR:::get_structure(sp19))
  ),
  data = phenoSoy,
  method = 'em'
)
```

What if we integrate genomic information

```
## Formula: Proteincontent ~ 0 + Intercept + Year
##    Data: phenoSoy
##   AIC     BIC logLik
##  2876 unknown  -1434
##
## Parameters of special components:
##
##
## Variance components:
##                 Estimated variances
## generic_genetic             2.46900
## sp18                        0.01584
## sp19                        0.06144
## Residual                    0.95260
```

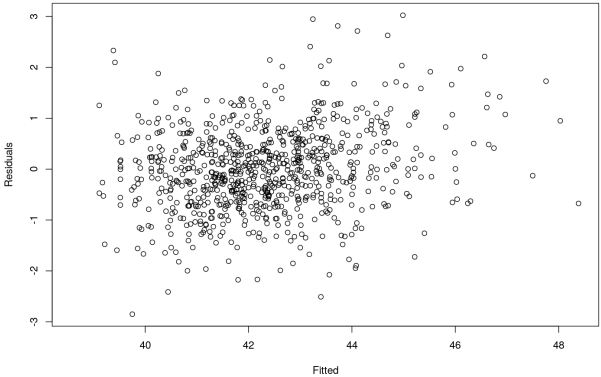
What if we integrate genomic information

Spatial effect



What if we integrate genomic information

```
## [1] 0.7056544
```



Summary

We have just seen that the model combining the data from the two rating years gives better results.

For the following steps, two solutions are possible:

- Either we use the combined model without the modular data.
- Either we use the combined model with the molecular data.

In both cases, the part due to the year and the spatial effect will be removed from the phenotype, leaving only the genetic and residual part.

We are ready to perform GWAS and/or GS

GWAS

Introduction

The GWAS can be used in several situations.

Firstly, when seeking to determine the genetic architecture of a trait of ecological or agronomic interest.

It is also useful in the context of Marker Assisted Breeding and Genomic Selection.

Introduction

GWAS aims at determine the association between the phenotype and genotype of individuals in a population by the way of linear regression.

Some parameters influence the GWAS results :

- LD : Non random association between alleles from different (linked) loci
- Population structure : Allele frequencies vary across sub-populations and cause long-range disequilibrium, i.e. between unlinked loci

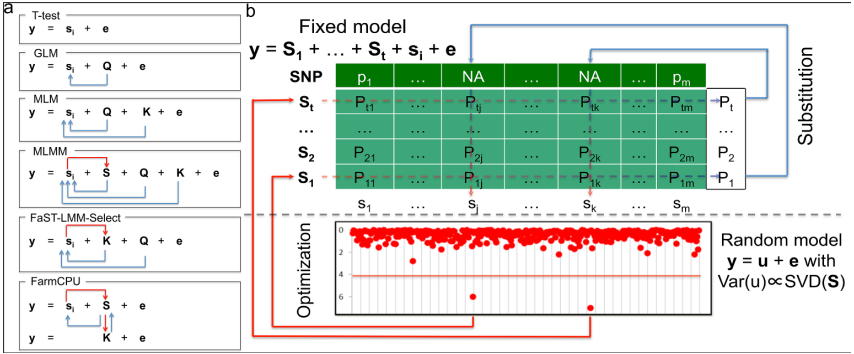
Introduction

Several methods have been developed over the years to perform GWAS :

- GRAMMAR (Aulchenko et al., 2007)
- P3D (Zhang et al., 2010) & EMMAX (Kang et al., 2010)
- FaST-LMM (Lippert et al., 2011)
- GEMMA (Zhou & Stephens, 2012)
- MixABEL (Aulchenko, et al.)
- CMLM (Zhang et a., 2010)
- ECMLM (Li et al., 2014)
- MLMM (Segura et al., 2012)
- FarmCPU (Liu et al, 2016)
- ...

Introduction

A very good illustration in the paper of Liu et al, 2016 shows the differences between the methods.



MLMM

We are going to test the MLMM with the genomic relationship matrix.

As the genomic relationship matrix includes information on both population structure and relatedness, it is in general not useful to consider admixture information as fixed effects covariates (Astle and Balding 2009).

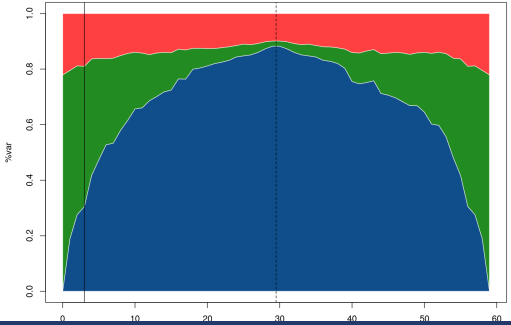
This method is a stepwise model regression with a forward inclusion and a backward elimination.

This limits the space of the model to be explored and makes the method computationally efficient.

MLMM - Proteincontent

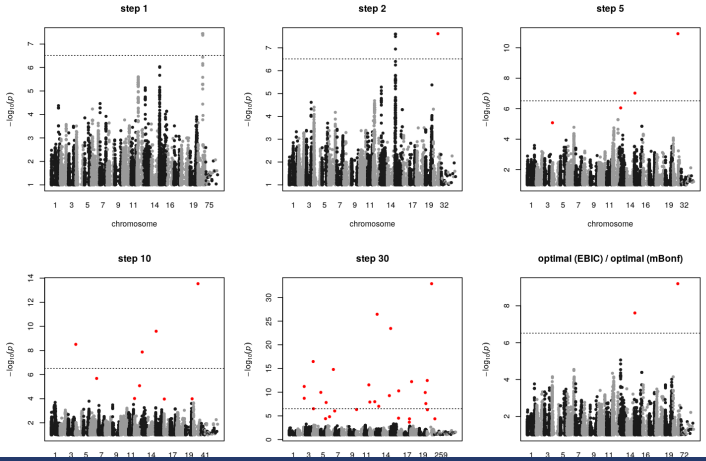
Graph showing the variance explained by the SNPs at each stage :

- In red the residual variance
- In green the genetic variance
- In blue the variance explained by the QTLs detected.



MLMM - Proteincontent

Representation of the QTLs detected with different step of the MLMM.

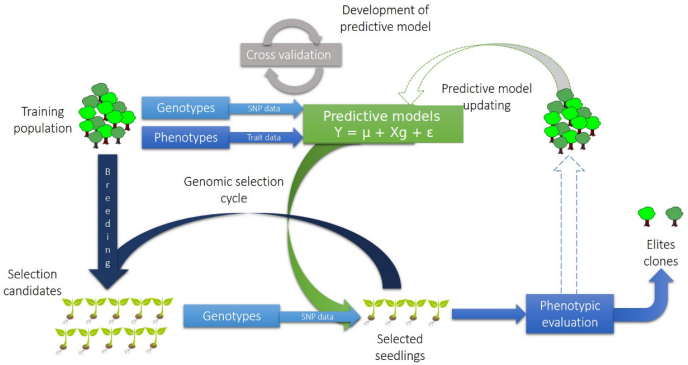


Genomic Selection

Introduction

Genomic selection (GS) is the direct descendant of marker-assisted selection (SAM)

The genomic selection



Accuracy of genomic selection

The accuracy of genomic prediction : the correlation between the true breeding value (TBV) and its estimate (GEBV : Genomic Estimated Breeding Value).

If the observed values are phenotypes, the accuracy of prediction : $\frac{predictive-ability}{\sqrt{h2}}$

($r_{GEBV,TBV} = r_{GEBV,Phenotypes}/\sqrt{h^2}$) (Falconer, 1981).

Accuracy of genomic selection

Imposed by the study population

Adjustment variables

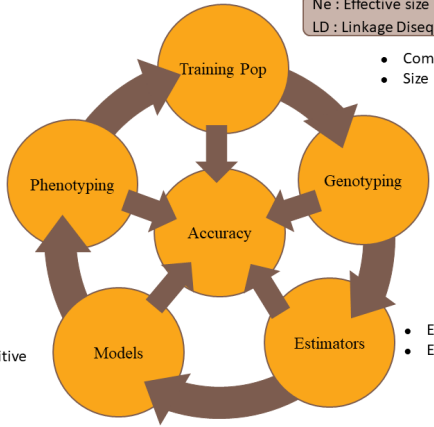
Ne : Effective size

LD : Linkage Disequilibrium

- Composition
- Size

Trait architecture heritability

- Phenotypes Quality



- Univariate or Multivariate
- Consideration of non-additive effects

- Estimation of genetic values
- Estimation of marker effects

Other quality criteria

The predicting ability (or accuracy) is the most common criteria

Other complementary criteria for prediction quality are :

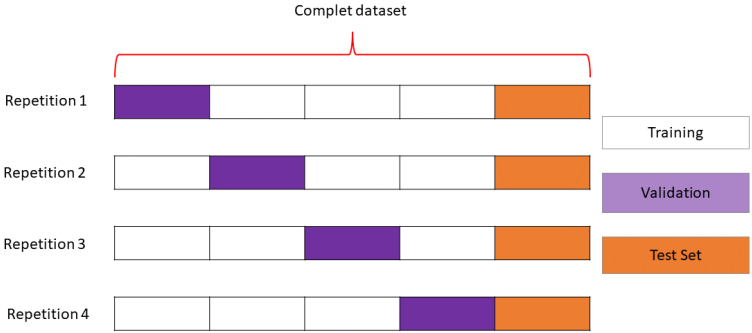
- the slope : should be close to 1
- the intercept : close to zero (but it is not a critical biais)

These two parameter allow to observe model's deficiencies :

- wrong variance partition
- incomplete model,
- nonrandom choice of individuals for training and validation population

Cross-validation

Cross validation & Test Set



GBLUP and other models

Two main groups can be distinguished :

- approaches that estimate an additive effect associated with each marker
- approaches that directly give the additive value of individuals.

In recent years, several methods have been developed :

- derivatives of BLUP methods (Henderson, 1975),
- Bayesian methods and non-parametric methods (Gianola and van Kaam, 2008; Neves et al., 2012 . . .).

Roughly, the different strategies are thought to accommodate implicitly different underlying genetic a architectures of quantitative traits:

- from simpler infinitesimal-like architectures
- to those more complex with a variety of heterogeneous effects across genes.

GEBV

The formula to estimate GEBV : $y = \mu + Xg + \epsilon$

- g is a vector (p $\times$ 1) of GEBV,
- X a design matrix linking observations to individuals.

GEBVs follow a normal distribution, with a **covariance** between effects modelled through an **additive relationship matrix** that is derived from :

```
* markers for GEBV (GRM)
* from pedigree in classical pedigree-based BLUPs (NRM)
```

GRM matrice

There are several methods for calculating this matrix G (or GRM) :

- the first methods uses a similarity index calculated for each locus : This method is based on the assumption that all identical alleles (IBS) are all identical by descent (IBD).
- This formula can be corrected by taking into account the probability of an allele being IBS for a locus by using the founder's genotypes.
- Another method is to correct the matrix G by twice the allele frequency of the minor allele as in equation (VanRaden 2007,Habier 2008).

GS example

We will see an example of ganomic predictions.

Here I varied the size of the training population.

I divided the data as follows:

- 20% in test
- 80% in cross-validation
  - Training: 20%, 50% or 70%.
  - Validation: 80%, 50%, 70%.

Just as a reminder : The phenotypes of the individuals in the validation and test population are masked and the model must predict them.

GS example

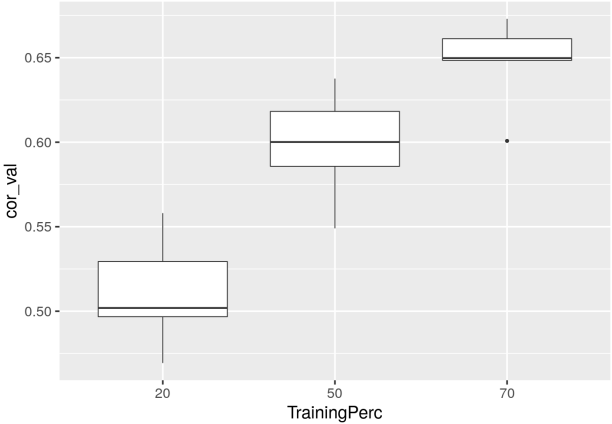
We will look at the quality of the prediction with the following parameters:

- Predictive ability
- Accuracy
- the Ranking
- the slope

10 repetitions have been performed

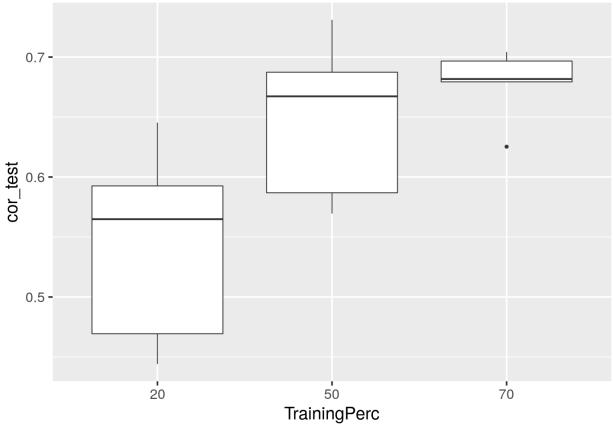
GS example

The predicting ability in validation



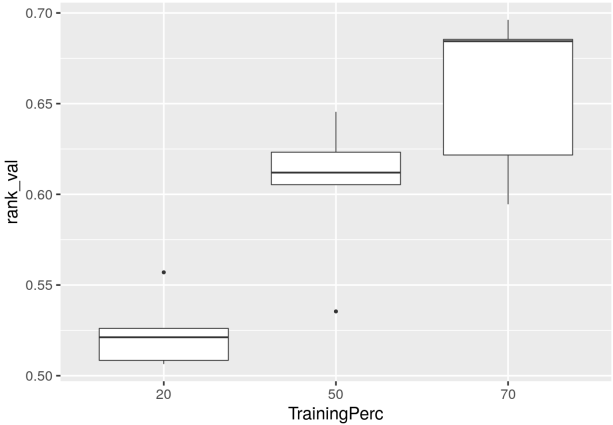
GS example

In the test Set



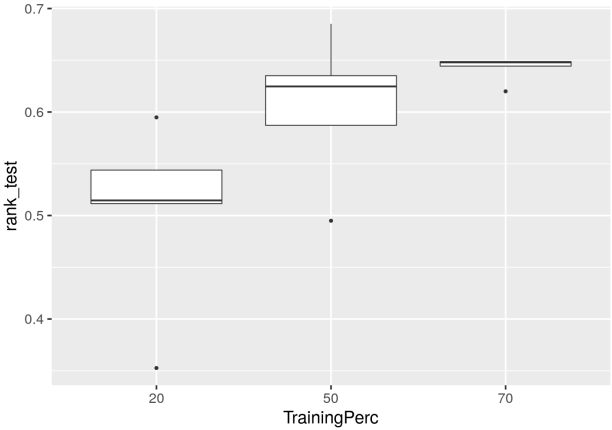
GS example

The ranking in validation



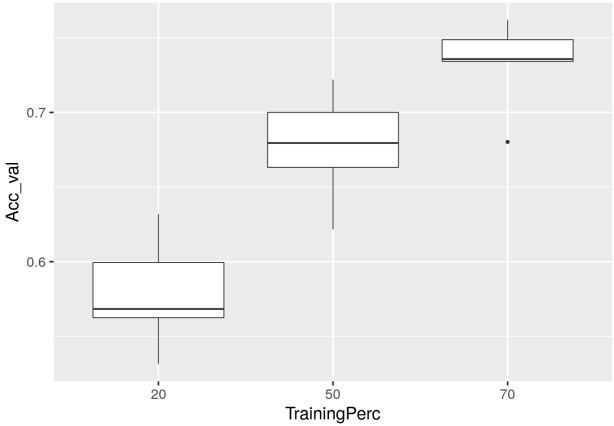
GS example

Then in Test Set



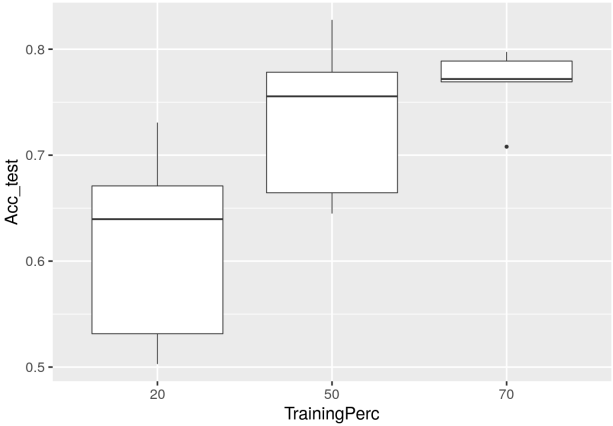
GS example

The accuracy in validation



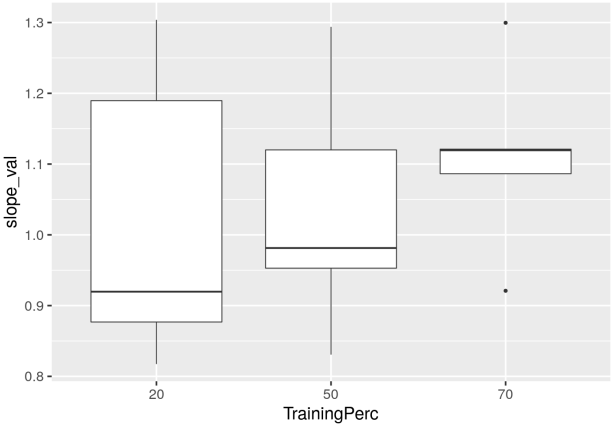
GS example

Then in Test Set



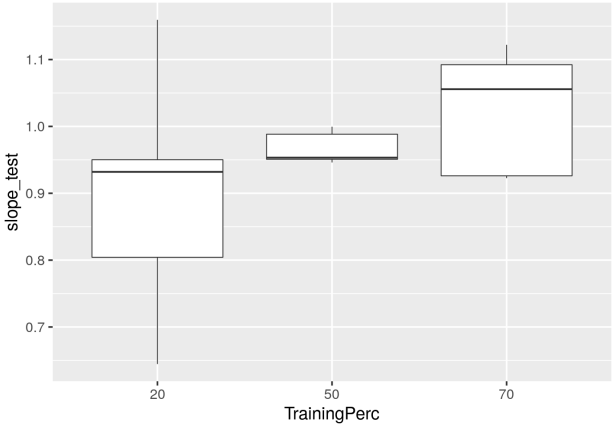
GS example

The slope in validation



GS example

Then in Test Set



To go further

This example is very simple, to go further in your analysis you can :

- Test several different marker sets
  - selected according to the LD
  - Around the QTLs
  - According to the recombination rate
- Optimising the choice of individuals of the training population
- Using non infinitesimal models
- Using GWAS data in your predictions
- Multiple-trait
- ...

R, Rstudio and RMarkdown

Quick introduction

In recent years R has become the statistical programming language :

- statisticians
- the most widely used generic environment for analysis of high-throughput genomic data.

R's core strength are :

- the literally thousands of packages freely available
- There is a good chance that for any given task there already is a package that can do the job at hand

Quick introduction

More specifically for genome-wide association studies (GWAS) or genomic Selection (GS), there are hundreds of packages available for the various analytical steps.

There are packages for :

- importing a wide range of data formats,
- preprocessing data,
- to perform quality control tests,
- to run the analysis per se,
- A large number of new algorithms and methods are published and at the same time released as an R package,

Quick introduction

R is **free** (released under the GNU public license).

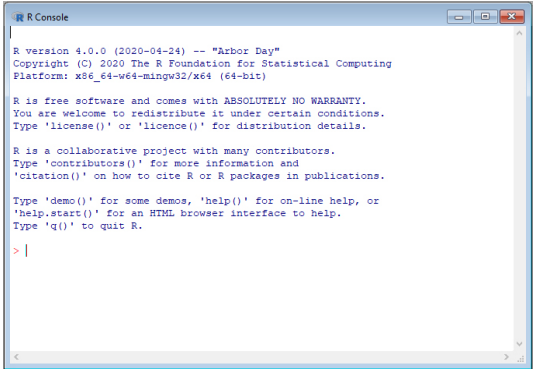
It is free to *use*, free to *modify*, and an open source.

R is also **platform independent**.

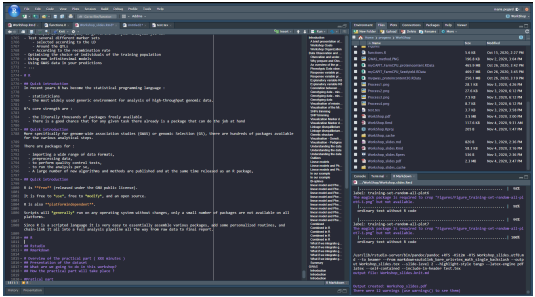
Scripts will *generally* run on any operating system without changes, only a small number of packages are not available on all platforms.

Since R is a scripted language it is very easy to essentially assemble various packages, add some personalized routines, and chain-link it all into a full analysis pipeline all the way from raw data to final report.

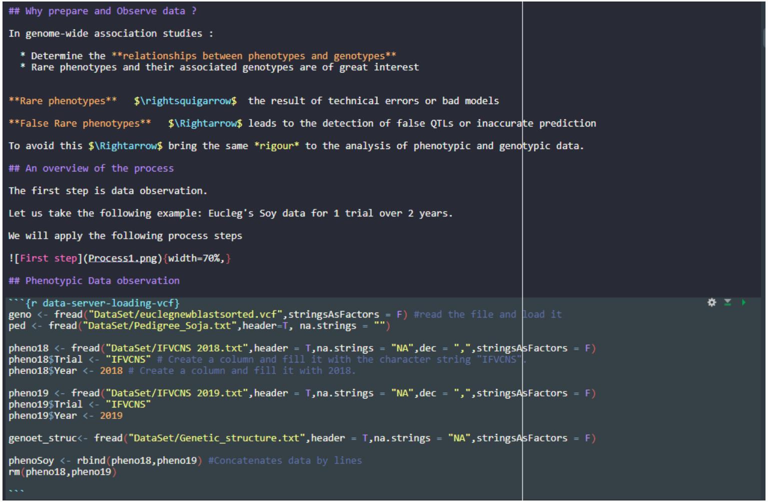
R


R Console
R version 4.0.0 (2020-04-24) -- "Arbor Day"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>

Rstudio



Rmarkdown



~~~
## Why prepare and Observe data ?

In genome-wide association studies :

  * Determine the **relationships between phenotypes and genotypes**
  * Rare phenotypes and their associated genotypes are of great interest


**Rare phenotypes**   $\rightsquigarrow$  the result of technical errors or bad models

**False Rare phenotypes**   $\Rightarrow$ leads to the detection of false QTLs or inaccurate prediction

To avoid this $\Rightarrow$ bring the same *rigour* to the analysis of phenotypic and genotypic data.

## An overview of the process

The first step is data observation.

Let us take the following example: Eucleg's Soy data for 1 trial over 2 years.

We will apply the following process steps

![First step](Process1.png){width=70%,}

## Phenotypic Data observation

```{r data-server-loading-vcf}
geno <- fread("DataSet/euclegnewblastsorted.vcf",stringsAsFactors = F) #read the file and load it
ped <- fread("DataSet/Pedigree_Soja.txt",header=T, na.strings = "")

pheno18 <- fread("DataSet/IFVCNS 2018.txt",header = T,na.strings = "NA",dec = ",",stringsAsFactors = F)
pheno18$Trial <- "IFVCNS" # Create a column and fill it with the character string "IFVCNS".
pheno18$Year <- 2018 # Create a column and fill it with 2018.

pheno19 <- fread("DataSet/IFVCNS 2019.txt",header = T,na.strings = "NA",dec = ",",stringsAsFactors = F)
pheno19$Trial <- "IFVCNS"
pheno19$Year <- 2019

genoet_struc<- fread("DataSet/Genetic_structure.txt",header = T,na.strings = "NA",stringsAsFactors = F)

phenoSoy <- rbind(pheno18,pheno19) #Concatenates data by lines
rm(pheno18,pheno19)

```
~~~

Overview of the practical part

Presentation of the dataset

Soybean Dataset field experiment of IFVCNS in 2018 and 2019

There is 15 columns, including :

- Field information : Plot, Raow, Colum, Trial, Year
- Genetic information : EntryNo, EUCLEGID, Accessionname
- Trait of interest : Seed yield, Protein content
- Explanatory variables : R1 (beginning of flowering), R2 (full flowering), R5 (start of seed filling), R8 (maturity)

The genomic data :

- 360 phenotyped individuals
- 357 genotyped individuals
- 226798 SNPs

What are we going to do in this workshop?

- Obervation and understanding of data
  - Dataset Visualization and Preparation
- Relationship matrice construction
  - NRM matrice
  - GRM matrice
  - Hybrid matrice
- Heritability estimation
- Estimation of genetic correlation with a multiple-trait model
- Phenotypic adjustment
- GWAS
- Genomic Selection
  - GBLUP
  - Q-GBLUP

How the practical part will take place ?

The differents steps :

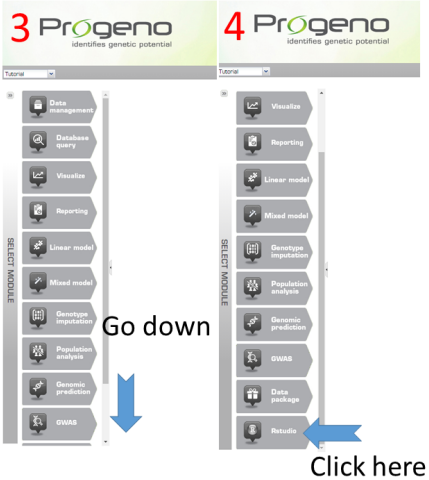
1. download the following files from the platform:
   - complete dataset file
   - the slides
   - the workshop file
2. Go to the Rstudio de progeno (procedure in the following slides)
3. Create a working folder
4. Download the data from your computer to progeno or use the database to extract and import them into Rstudio.
5. Start the tutorial
6. Try on your own data or use the sample data

Pratical part

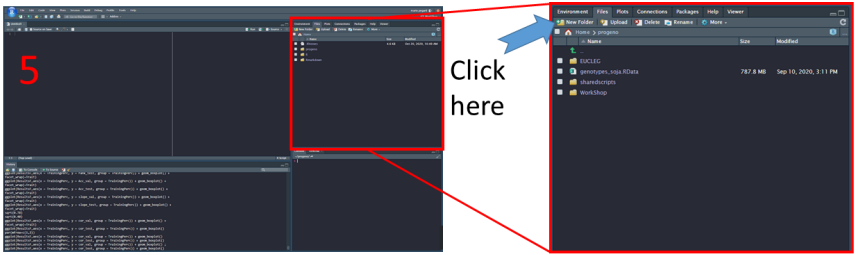
step 2

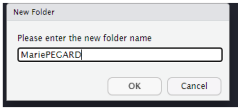

1
Progeno
identifies genetic potential
LOGIN
Username
Password
Database
Tutorial
login

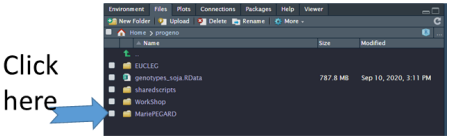

2
Progeno
identifies genetic potential
Tutorial
Breading activities
Analytics
Decision support
Help
Click here

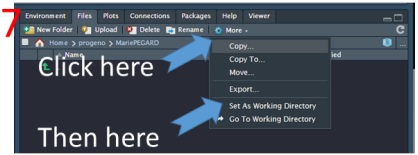

3
Progeno
identifies genetic potential
Tutorial
SELECT MODULE
Data management
Database query
Visualize
Reporting
Linear model
Mixed model
Genotype imputation
Population analysis
Genomic prediction
GWAS
Go down
4
Progeno
identifies genetic potential
Tutorial
SELECT MODULE
Visualize
Reporting
Linear model
Mixed model
Genotype imputation
Population analysis
Genomic prediction
GWAS
Data package
Retudio
Click here

Step 3

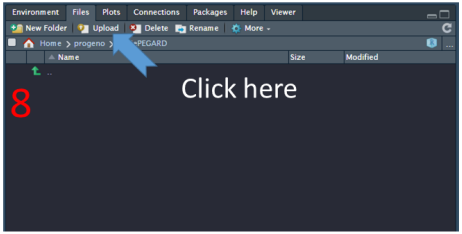

5
Click
here
Environment
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home > progeno
Name
Size
Modified
EUCLEG
genotypes_soja.RData
787.8 MB
Sep 10, 2020, 3:11 PM
sharedscripts
WorkShop

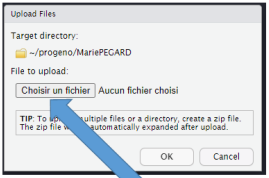

6
New Folder
Please enter the new folder name
MariePEGARD
OK
Cancel


Click
here
Environment
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home > progeno
Name
Size
Modified
EUCLEG
genotypes_soja.RData
787.8 MB
Sep 10, 2020, 3:11 PM
sharedscripts
WorkShop
MariePEGARD


7
Environment
Files
Plots
Connections
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home > progeno > MariePEGARD
Name
Copy...
Copy To...
Move...
Export...
Set As Working Directory
Go To Working Directory
Click here
Then here

Step 4





Click here and chose the download files

Acknowledgements

Acknowledgements

- Bernadette Julier and Philippe Barre
- Hilde Muylle, Jonas Asper and Aamir Saleem
- Steven Maenhout
- Cloe Paul-Victor

Let's GO !

just few last point

If you needs help please email me : marie.pegard@inrae.fr

I may have forgot tu write some explaination to understand how works a function use ?function_name (example ?summary)

Take your time to test and look the results. Even if you don'ont have the time to do eveything !

At the end I will give you the scripts that I used to generate the pdf and the slides.

I will also provide a script with "all-inclusive" fucntion to performe genomic prediction.