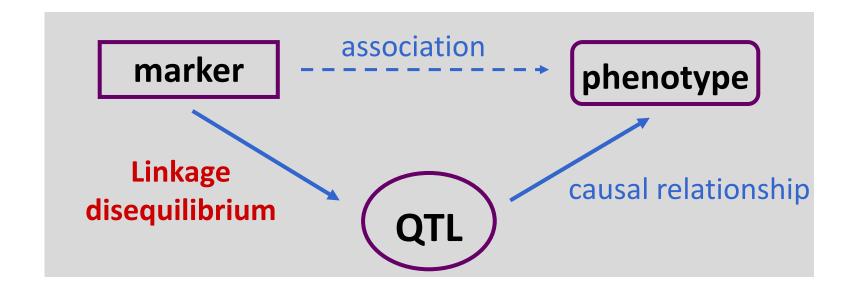




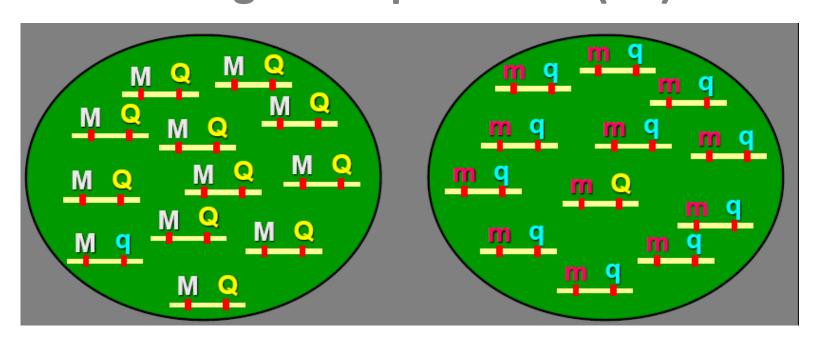
# Genomic Prediction: Applications in R

FA21 ANSCI 610

Genomic selection is a form of marker-assisted selection in which genetic markers covering the whole genome are used so that all quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker



### Linkage Disequilibrium (LD)



M is more often associated with Q than m is associated with Q

The marker genotype is associated with the phenotype if Q/q has an effect on the phenotype

The *r*<sup>2</sup> is generally the preferred measure of LD for biallelic markers, which is the squared correlation of alleles at 2 *loci*.

$$r^{2} = \frac{D^{2}}{f(A_{1})f(A_{2})f(B_{1})f(B_{2})}$$

Where  $D=f(A_1B_1)-f(A_1)f(B_1)$ , and  $f(A_1B_1)$ ,  $f(A_1),f(B_1),f(A_2)f(B_2)$  are the observed frequencies of haplotype  $A_1B_1$  and of alleles  $A_1$ ,  $A_2$ ,  $B_1$ , and  $B_2$ , respectively.

### Marker A

Marker B B1

B1 B2

Frequency 0.5

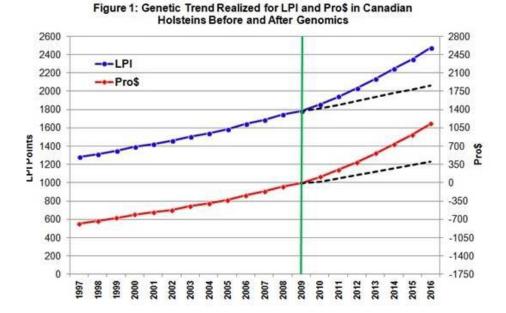
| A1  | A2  | Frequency |
|-----|-----|-----------|
| 0.4 | 0.1 | 0.5       |
| 0.1 | 0.4 | 0.5       |
| 0.5 | 0.5 |           |

$$D = 0.15$$

$$r^2 = \frac{0.15^2}{0.5 \times 0.5 \times 0.5 \times 0.5}$$

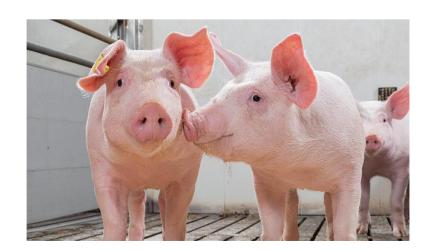
$$r^2 = 0.36$$

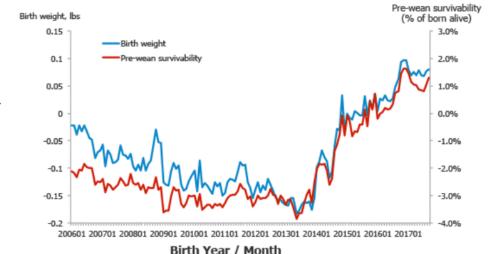




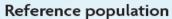
#### PIC IMPROVES BIRTH WEIGHT & PRE-WEAN SURVIVABILITY

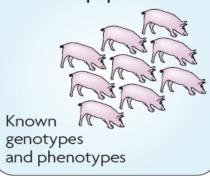
Trend: genetic improvement in birth weight and pre-wean survivability (PIC Genetic Nucleus)



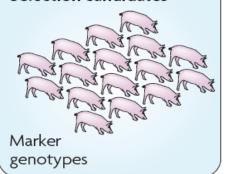


#### **Box 2 | Genomic selection**





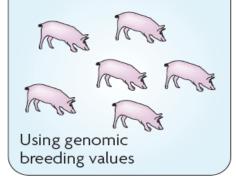
#### Selection candidates



#### Prediction equation

Genomic breeding value =  $W_1X_1 + W_2X_2 + W_3X_3.....$ 

#### Selected breeders



The general model for the prediction equation can be written as:

$$y_i = \mu + \sum_{j=1}^p x_{ij} + e_i$$

Where:  $y_i$  is the (pseudo) phenotype of individual i  $\mu$  is the overall mean  $x_{ij}$  is the genotype of individual i at locus j p is the number of markers loci  $e_i$  is the residual term

### Challenges:

- Imputation
- ❖ Low → High resolution chip (Cost Reduction)
- ☐ Curse of dimensionality (n << p)
- Possible solutions:

Variable Selection

PCA

Bayesian Inference (e.g. Bayesian Alphabet)

Linear Mixed Models (e.g. GBLUP and single step GBLUP)

### Other relevant questions:

- ☐ Which (Pseudo) phenotype to use?
- Raw phenotype
- Adjusted phenotype
- **EBV**
- Deregressed EBV
- Daughter Yield Deviation (DYD) or Progeny Yield Deviation (PYD)
- ☐ How to validate the prediction ability?
- Cross-validation (e.g. k-fold CV, LOOCV, k-fold with random repetitions)
- Forward validation

### Other relevant questions:

- ☐ How to measure the prediction ability?
- Pearson Correlation
- Mean squared error
- Bias metrics
- AUC and other classification metrics
- ☐ Which prediction method to use?
- Bayesian regression models (e.g. Bayesian LASSO, BayesA, Bayes C, etc.)
- Genomic relationship based models (e.g. GBLUP, ss-GBLUP)
- Semi-parametric methods (e.g. Reproducing Kernel Hilbert Space regressions)
- Machine Learning, Deep Learning

### **SOFTWARES**

- ☐ Fortran 90/95 programs
- \* ASREML https://www.vsni.co.uk/software/asreml/
- BLUPF90 <a href="http://nce.ads.uga.edu/wiki/doku.php">http://nce.ads.uga.edu/wiki/doku.php</a>
- ♦ DMU <a href="https://mbg.au.dk/en/">https://mbg.au.dk/en/</a>
- **GS3** <a href="http://genoweb.toulouse.inra.fr/~alegarra/gs3\_folder/">http://genoweb.toulouse.inra.fr/~alegarra/gs3\_folder/</a>
- **WOMBAT** http://didgeridoo.une.edu.au/km/wombat.php
- □ R packages
- ❖ BGLR
- Gaston
- MCMCglmm
- rrBLUP
- Sommer



Genetics. 2014 Oct; 198(2): 483-495.

Published online 2014 Jul 9. doi: 10.1534/genetics.114.164442

PMCID: PMC4196607 PMID: 25009151

### Genome-Wide Regression and Prediction with the BGLR Statistical Package

Paulino Pérez\*,1 and Gustavo de los Campos†

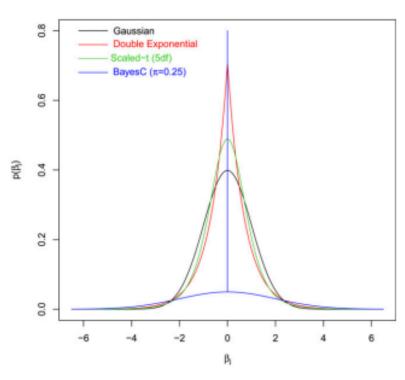


Figure 1 Prior densities of regression coefficients implemented in BGLR (all densities in the figure have null mean and unit variance).

Table 1 Prior densities available for regression coefficients in the BGLR package

| Model (prior density)        | Hyperparameters  | Treatment in BGLR <sup>a</sup>   |
|------------------------------|--|--|
| Flat (FIXED)                 | Mean $(\mu_{\beta})$<br>Variance $(\sigma_{\beta}^2)$    | $\mu_{\beta} = 0$ $\sigma_{\beta}^2 = 1 \times 1^{10}$   |
| Gaussian (BRR)               | Mean $(\mu_{\beta})$<br>Variance $(\sigma_{\beta}^2)$    | $\mu_{\beta} = 0$ $\sigma_{\beta}^{2} \sim \chi^{-2}$  |
| Scaled-t (BayesA)            | Degrees of freedom $(df_{\beta})$<br>Scale $(S_{\beta})$ | User specified<br>(default value, 5)<br>$S_B \sim Gamma$   |
| Double<br>exponential (BL)   | $\lambda^2$  | $\lambda$ fixed, user<br>specified, or<br>$\lambda^2 \sim Gamma$ , or<br>$\lambda/max \sim Beta^b$ |
| Gaussian mixture<br>(BayesC) | π (prop. of<br>nonnull effects)                          | $\pi \sim$ Beta  |
|                              | $df_{oldsymbol{eta}}$                                    | User specified<br>(default value, 5)   |
|                              | $S_{\beta}$  | $S_{\beta} \sim Gamma$   |
| Scaled-t mixture<br>(BayesB) | π (prop. of nonnull effects)                             | $\pi \sim$ Beta  |
|                              | $df_{oldsymbol{eta}}$                                    | User specified<br>(default value, 5)   |
|                              | Sβ   | $S_{\beta} \sim Gamma$   |

<sup>\*</sup>Further details are given in the supporting information (Section A of File S1).

<sup>&</sup>lt;sup>b</sup> This approach is further discussed in de los Campos et al. (2009b).

Change Language



