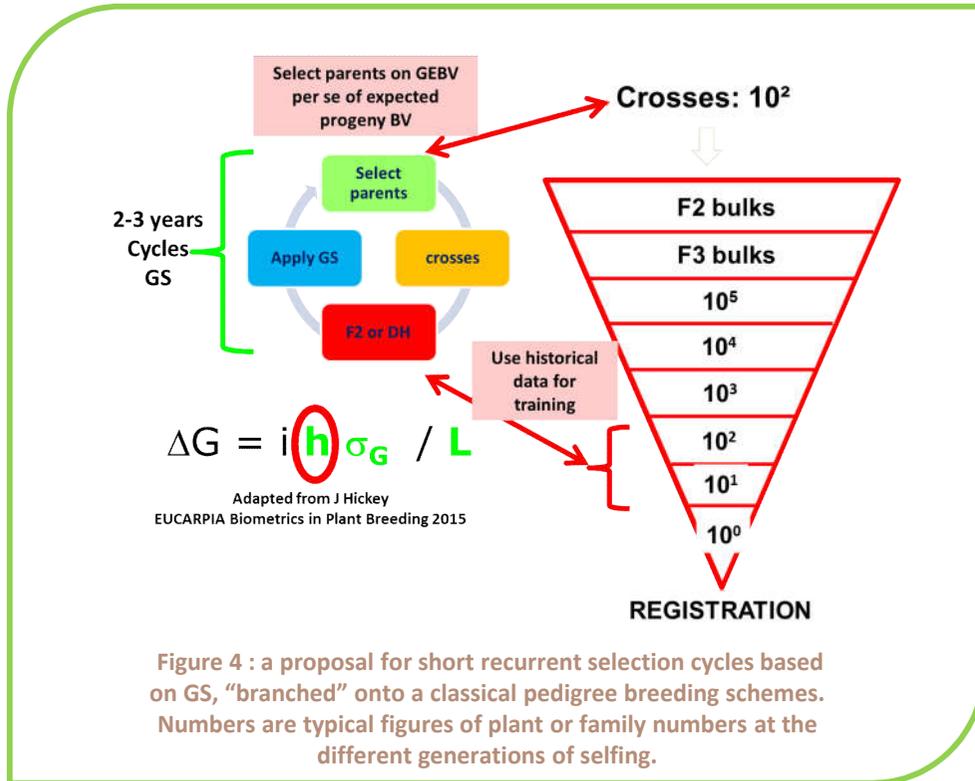


There are questions that remain to be addressed, such as:

- Where applying GS in a breeding scheme: should we incorporate GS into an existing pedigree selection without any other change, or should we develop new, accelerated breeding cycles based solely on GS, or a mix between these two extremes (Figure 4)?
- Wheat improvement must consider a lot of traits, some are true quantitative (yield, protein content), other are binary (resistant/susceptible) or ordinal (e.g. quality classes with fuzzy limits). Multi-traits GS prediction should then consider traits with various genetic architecture, which is less trivial than classical index selection.



More information

www.breedwheat.fr

Coordinator : **Dr. Jacques Le Gouis**, UMR GDEC - jacques.legouis@clermont.inra.fr

Project manager : **Emmanuelle Legendijk**, INRA Transfert - emmanuelle.legendijk@paris.inra.fr

Communication : **Bernard BEJAR**, Céréales Vallée - bernard.bejar@cereales-vallee.org

This project receives funding from the French Government managed by the Research National Agency (ANR) in the framework of the Investments for the Future (ANR-10-BTBR-03), France Agrimer and the French Fund to support Plant Breeding (FSOV).



A new way to use molecular markers to facilitate breeding: the Genomic Selection

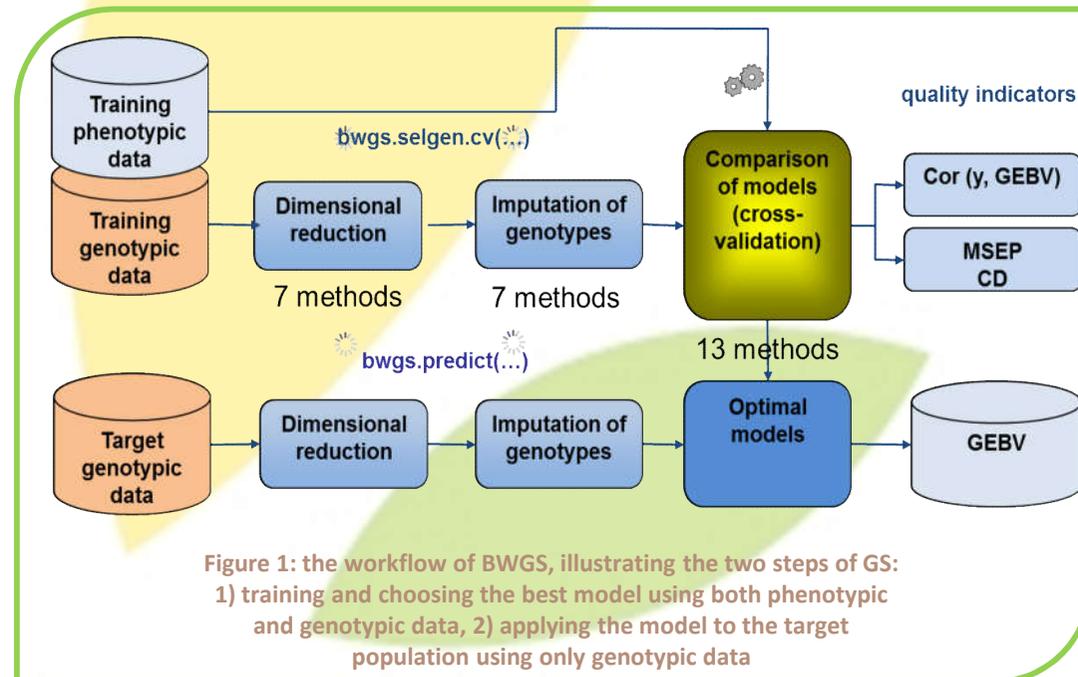
May 2016

Genomic selection has been proposed in 2001 as an alternative to "marker assisted selection" (MAS), when available molecular markers are enough to densely cover the genome of the animal or plant species.

In its principle, it no longer requires to perform a two-step process of 1) detecting markers associated to chromosomal regions influencing a trait of interest and 2) building a prediction model for this trait using those markers. In Genomic Selection (GS), all markers are fitted together in a single step prediction model. Since the number of predictor variables (markers) usually exceeds the number of observations (a trait measured on different genotypes), the usual linear model framework is not applicable and specific statistical models adapted to the "big data" problem must be used.

The objectives in BreedWheat (WP4) were to 1) assemble various statistical tools in a friendly-to-use R package for breeders (BWGS), 2) apply GS in a real breeding programme conducted by INRA-Agri-Obtentions (AO) partners, and 3) develop simulation tools to forecast the interest of GS in an applied wheat breeding programme from both technical and economic points of view.

How does Genomic Selection work?



To “train” the prediction model, we need a “training” or reference population, with both genotyping (markers) and phenotyping (observations) data (Figure 1). When the number of observations is limited, it may be useful to reduce the number of markers and the risk of “over parametrization”. Then missing data must be “imputed” and finally several statistical models can be fitted and compared using cross-validation techniques. Once an “optimal model” is selected, it can be applied to a target population of candidates with only genotype data. By replacing a long (and possibly costly) phenotypic evaluation of complex traits by their genomic prediction, if genotyping is cheaper, one should be able to 1) screen more candidates, thus increasing selection intensity, enlarge diversity and 2) select at early stages in the breeding schemes, thus accelerating selection cycles and therefore improving time-related efficiency.

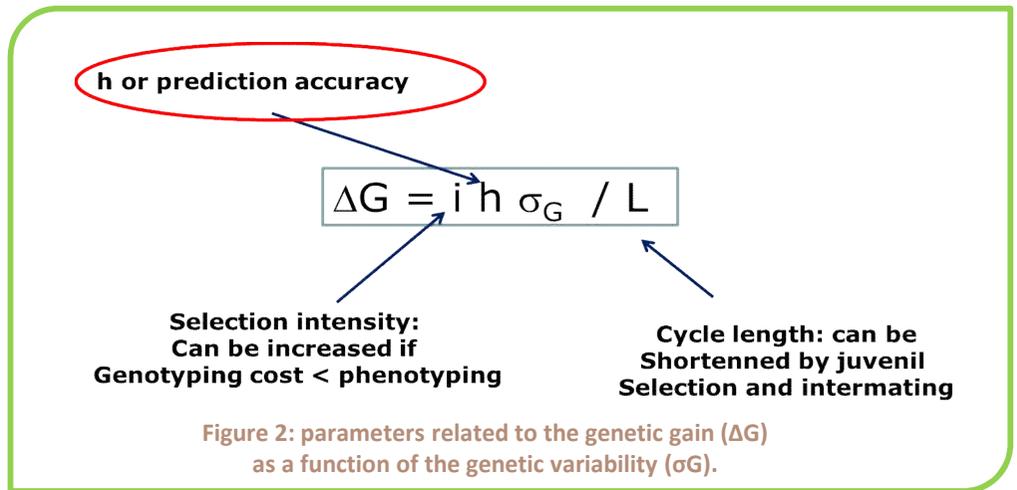
Does Genomic Selection work?

From historical trials carried out by INRA and Agri-Obtentions (AO) organisations, we constructed a reference population of 760 breeding lines that were genotyped with TaBW420K SNP chip (see BW booklet #2). Phenotypic traits were pre-processed by appropriate statistical models to correct for fixed effects (e.g. year, location...). BWGS was then used to evaluate prediction accuracy of adjusted phenotypes with 12 statistical models, using different numbers and types of SNP molecular markers. Results show that 5 000 to 10 000 randomly sampled markers are enough to achieve good accuracy with all methods (Table 1).

TRAIT	$h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{GE}^2 + \sigma_e^2)$	N (training)	$r = \text{cor}(\text{GEBV}, y)$
Yield and protein %:			
Yield	0,307	760	0,558
Protein	0,513	760	0,557
Alveograph:			
dough strength W	0,705	357	0,536
tenacity P	0,757	357	0,622
extensibility L	0,564	357	0,574
P / L	0,062	357	0,301
Bread making			
dough score	0,392	370	0,404
crumb score	0,371	370	0,448
bread score	0,275	370	0,405
total score	0,433	370	0,452
loaf volume	0,44	370	0,427
Other:			
hagberg FN	0,505	315	0,427
dietary fibre (visco)	0,908	267	0,68
Fusarium HB score	0,563	672	0,63
	very high accuracy		
	high accuracy		
	moderate accuracy		
	Insufficient accuracy		

Table 1: heritability, training population size and accuracy (correlation between Genomic Estimate of Breeding Value GEBV and adjusted phenotype) for different traits important for wheat improvement.

These accuracies are encouraging for an efficient use of GS in real breeding programmes. Moreover, accuracy is only one parameter in the expression of genetic gain (Figure 2).



GS could enable breeders to screen larger candidate populations (thus increasing selection intensity while saving more diversity) and/or make selection at earlier stages (e.g. young generation), and finally accelerating breeding cycles.

And now ?

The efficiency of genomic (GS) vs phenotypic (PS) selection can also be addressed through simulations. In BreedWheat, we developed simulation tools in R to compare different strategies for selecting parental lines and designing crosses. Figure 3 illustrates an example of 5 cycles of virtual selection.

