



**HAL**  
open science

## Combining genotyping by sequencing and genomic prediction within bi-parental crosses to speed up selection of grapevine cultivars

Timothée Flutre, Amandine Launay, Agnes Doligez, Eric Duchêne, Elisa Marguerit, P.F. Bert, Pierre Mournet, Loic Le Cunff

### ► To cite this version:

Timothée Flutre, Amandine Launay, Agnes Doligez, Eric Duchêne, Elisa Marguerit, et al.. Combining genotyping by sequencing and genomic prediction within bi-parental crosses to speed up selection of grapevine cultivars. 12. International Conference on Grapevine Breeding and Genetics, Jul 2018, Bordeaux, France. hal-01848564

**HAL Id: hal-01848564**

**<https://hal.science/hal-01848564>**

Submitted on 5 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Combining genotyping by sequencing and genomic prediction

## within bi-parental crosses

## to speed up selection of grapevine cultivars

Flutre T.<sup>1</sup>, Launay A.<sup>1</sup>, Doligez A.<sup>1</sup>, Duchêne E.<sup>2</sup>, Marguerit E.<sup>3</sup>, Bert P-F.<sup>3</sup>, Mournet P.<sup>1</sup>, Le Cunff L.<sup>1,4</sup>

<sup>1</sup> AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

<sup>2</sup> SVQV, Université de Strasbourg, INRA, 68000 Colmar, France

<sup>3</sup> UMR EGFV, Bordeaux Sciences Agro, INRA, Université de Bordeaux, F-33882 Villenave d'Ornon, France

<sup>4</sup> IFV, Le Grau-du-Roi, France

timothee.flutre@inra.fr



### 1. Abstract

A major challenge facing viticulture consists in decreasing inputs, especially synthesized fungicides, and adapting to climate change, while maintaining berry quality and differentiated wine styles, as well as reasonable profits to the winegrowers. In this endeavor, breeding new varieties is an important lever. To speed up this process, genomic selection may be advantageous to quickly test in the field candidates precociously selected for various complex traits.

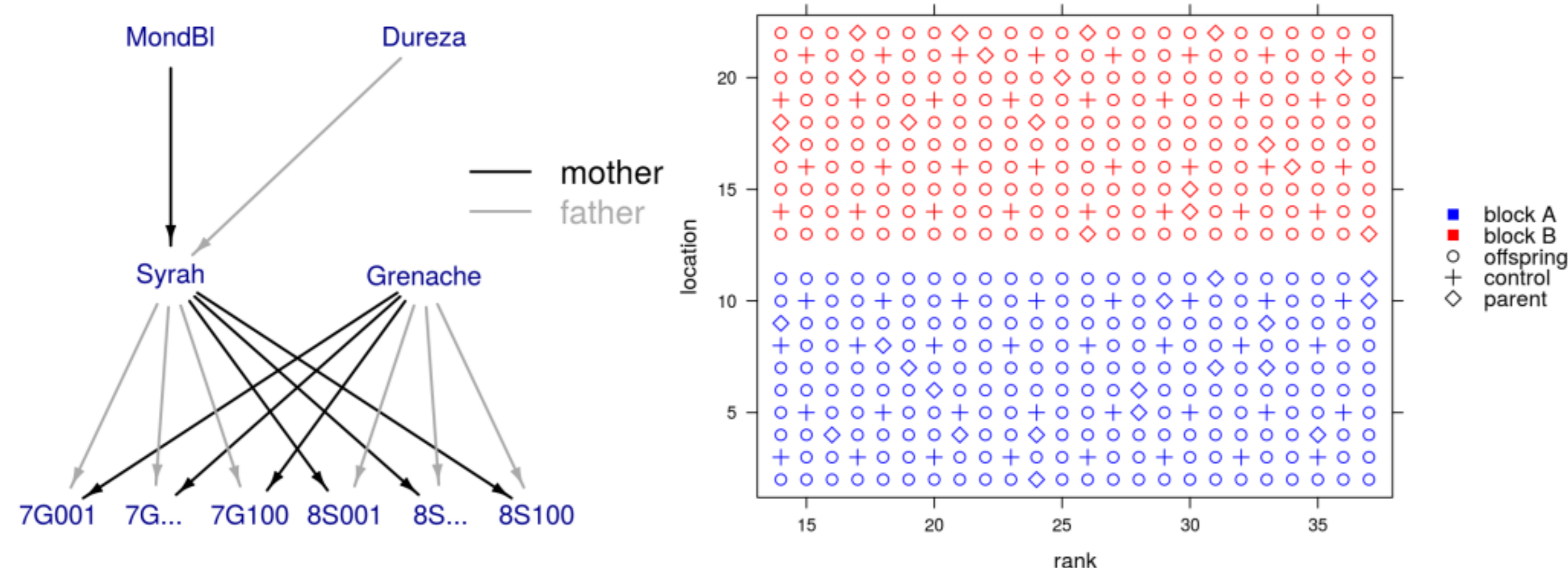
WE assess here the feasibility of this approach based on three bi-parental crosses (Syrah x Grenache, Riesling x Gewurztraminer and Cabernet-Sauvignon x Riparia Gloire de Montpellier), and several types of traits (yield components, phenology, terpenols, transpiration) whose phenotypes were already used for QTL mapping. The plant material was genotyped by sequencing (Keygene patents), providing around 18000 filtered, uniformly distributed SNPs per genotype. Depending on genetic architecture and broad-sense heritability, the average accuracy of genomic prediction obtained by cross-validation can go from 0% (transpiration rate), 50% (budbreak) to 80% (berry weight).

IN conclusion, such an integrated approach aims at exploiting the full extent of available phenotypic, genotypic and genealogical data in order to efficiently keep adapting the plant material. In parallel, the approach is also tested in a specific breeding program for the rosé wines, in order to speed up the selection of genotypes resistant to powdery and downy mildew, and adapted to the rosé vinification. Results will be compared to classical selection operated after offsprings are planted in the vineyard.

### 2. Crosses and experimental designs

cross	#offsprings	design	trait
Syrah x Grenache (SxG)	192	field	berry weight
Riesling x Gewurztraminer (RlxGW)	256	field	budbreak
Cabernet-Sauvignon x Riparia Gloire de Montpellier (CSxRGM)	120	greenhouse	transpiration rate (control condition)

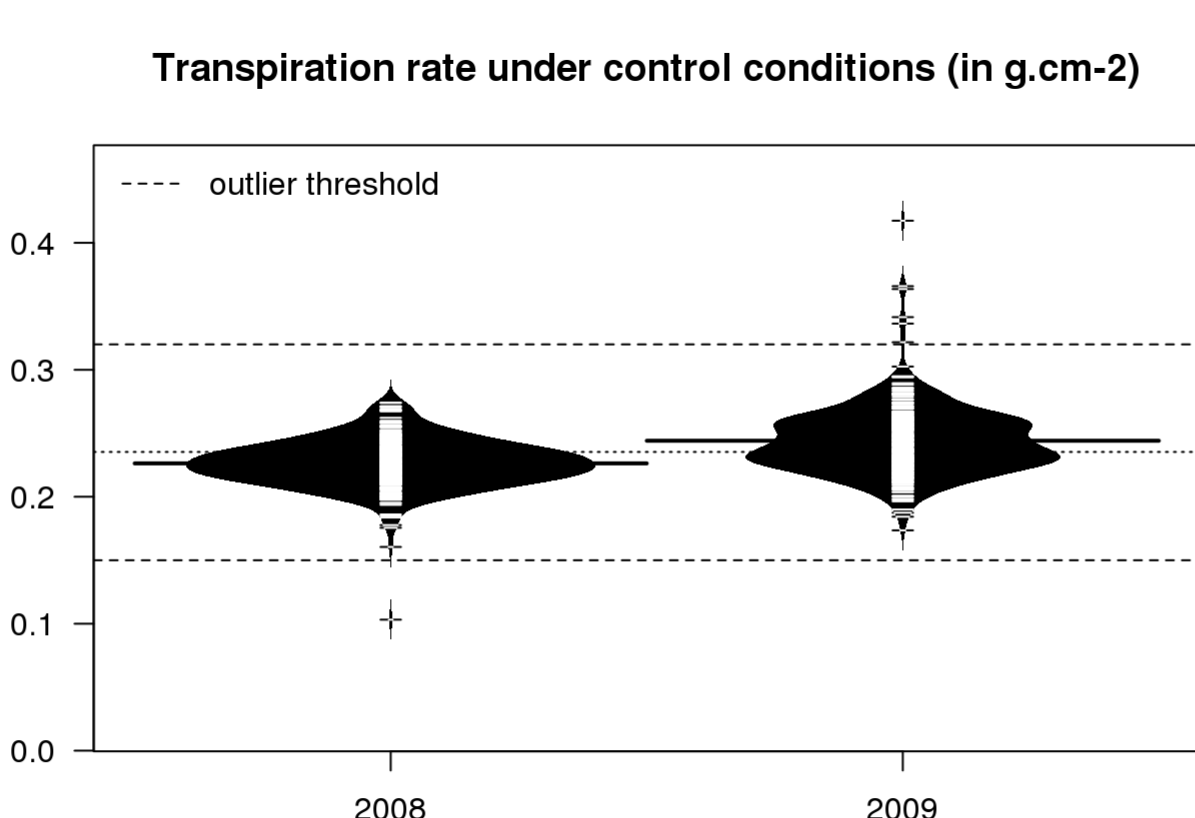
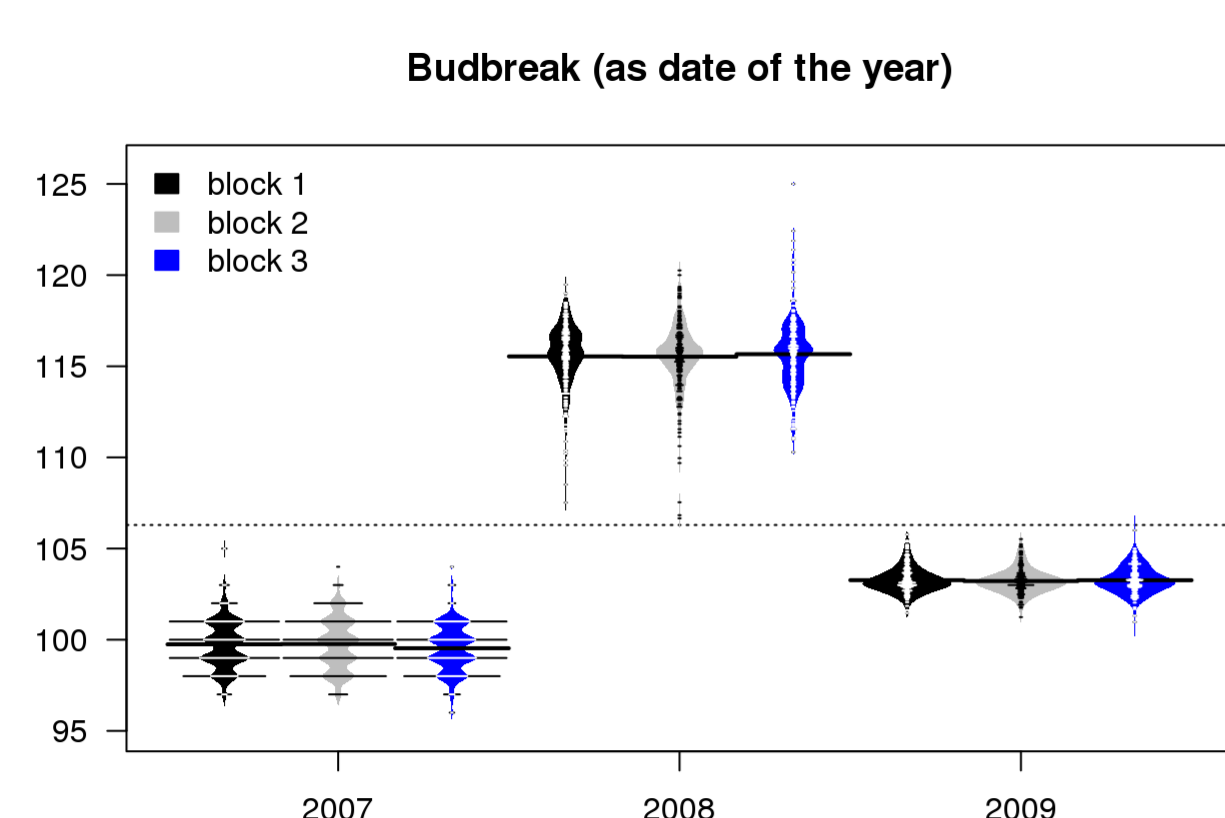
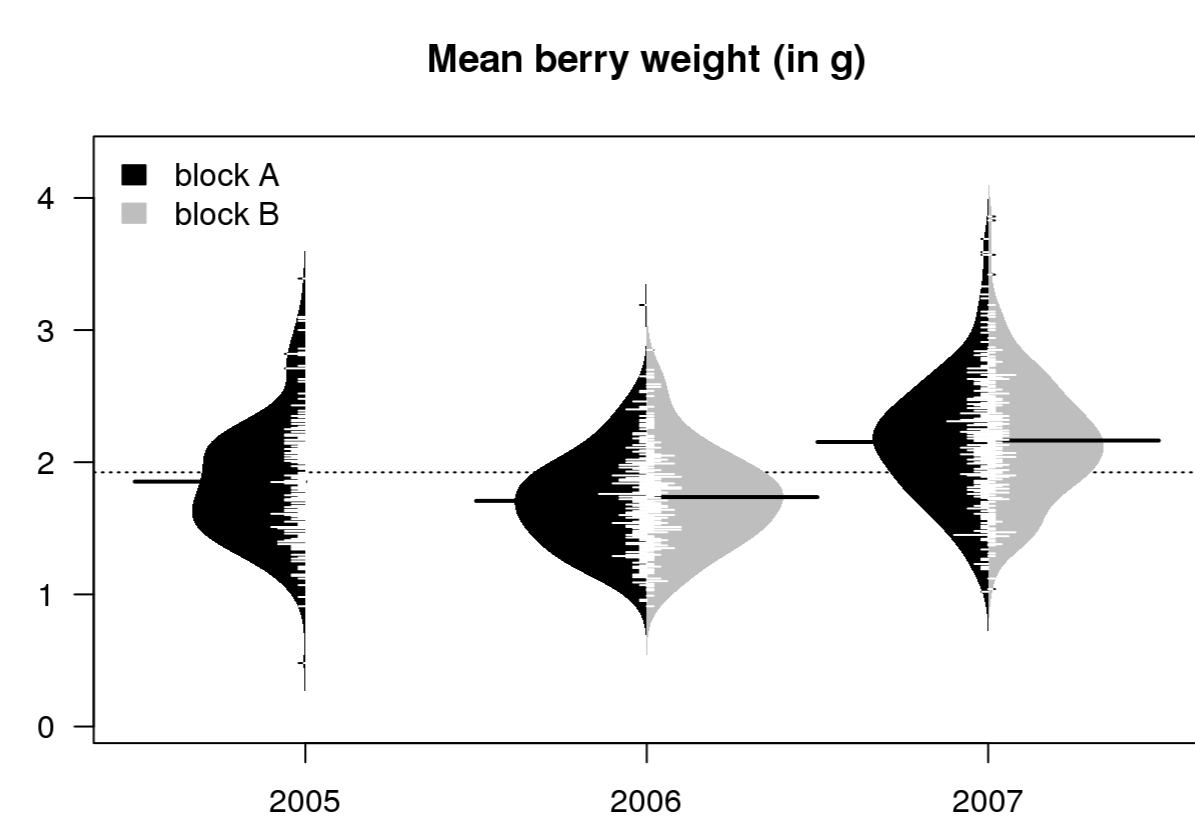
Pedigree and field layout of the reciprocal cross, Syrah x Grenache, in Montpellier :



### 3. Statistical analysis of the phenotypes

$$y = X\beta + Zg + \epsilon \text{ where } g \sim \mathcal{N}(0, \sigma_g^2 \text{Id}) \text{ and } \epsilon \sim \mathcal{N}(0, \sigma_e^2 \text{Id}) : \text{fitted with lme4.}$$

1. fit by maximum likelihood (ML) several linear mixed models with different fixed effects ( $\beta$ )
2. select the best based on the Akaike Information Criterion (AIC) and check its assumptions visually (e.g. residuals  $\hat{\epsilon}$ )
3. re-fit the best model by restricted maximum likelihood (ReML) to obtain the broad-sense heritability ( $H^2$ ) and the Best Linear Unbiased Predictors (BLUPs) of the total genotypic values ( $g$ )



### 4. Genotyping

After classical DNA extraction, samples were digested with the ApeKI restriction enzyme. Libraries multiplexing 96 samples were sequenced on Illumina HiSeq 3000 (paired-end : 2x75). As this resulted in substantial non-uniformity, a second sequencing phase was performed (single-end : 1x150). N.B. : Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies.

Raw reads were demultiplexed, cleaned with CutAdapt and aligned with BWA. SNP and genotypes were called with GATK, followed by filters on depth, quality and mendelian errors.

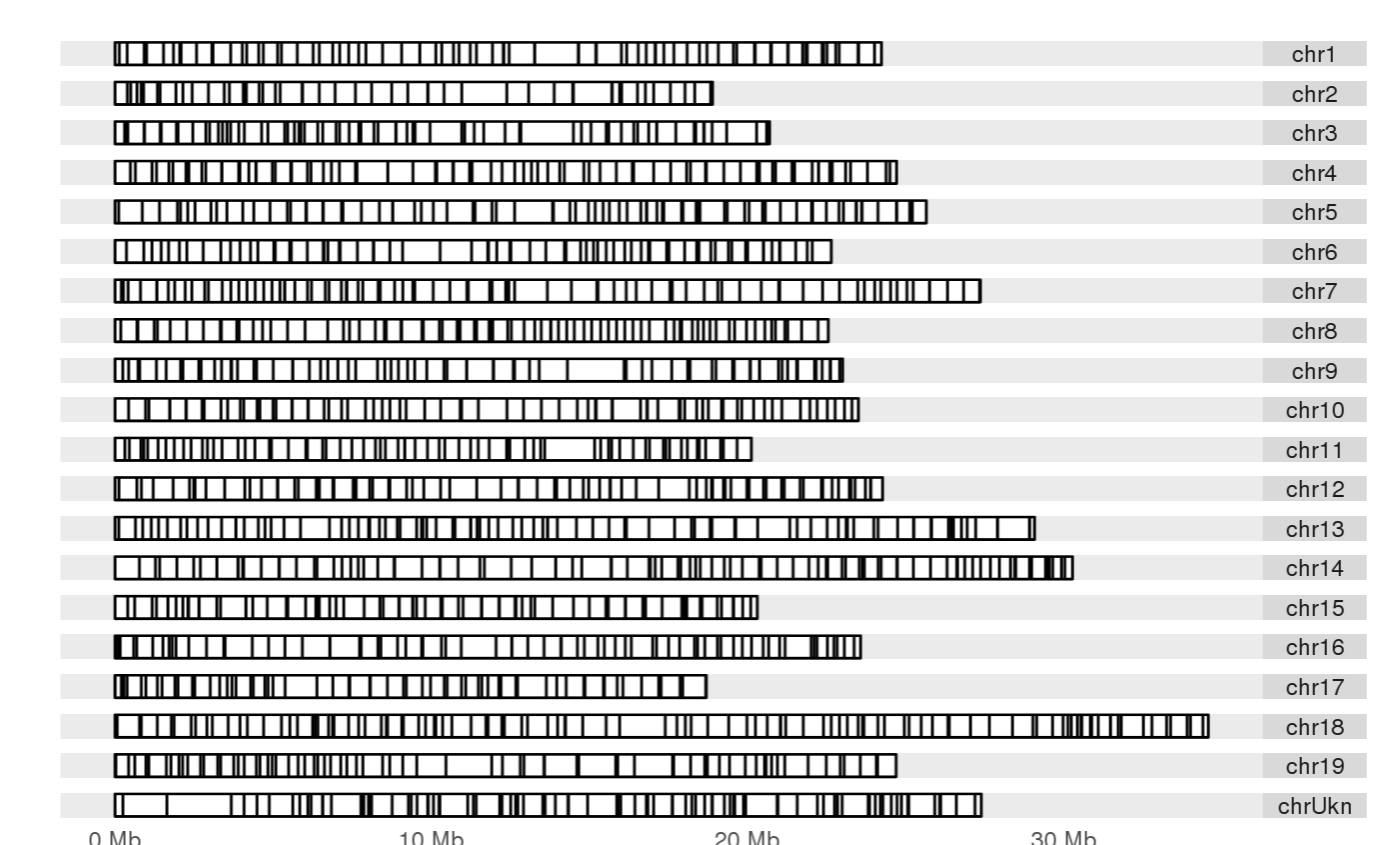
cross	med(breadth)	med(depth)	#SNPs
SxG	4.9%	22.1	28624
RlxGW	4.3%	19.7	113353
CSxRGM	3.7%	22.5	45861

After segregation filtering, imputation was performed with FImpute accounting for pedigree and linkage disequilibrium.

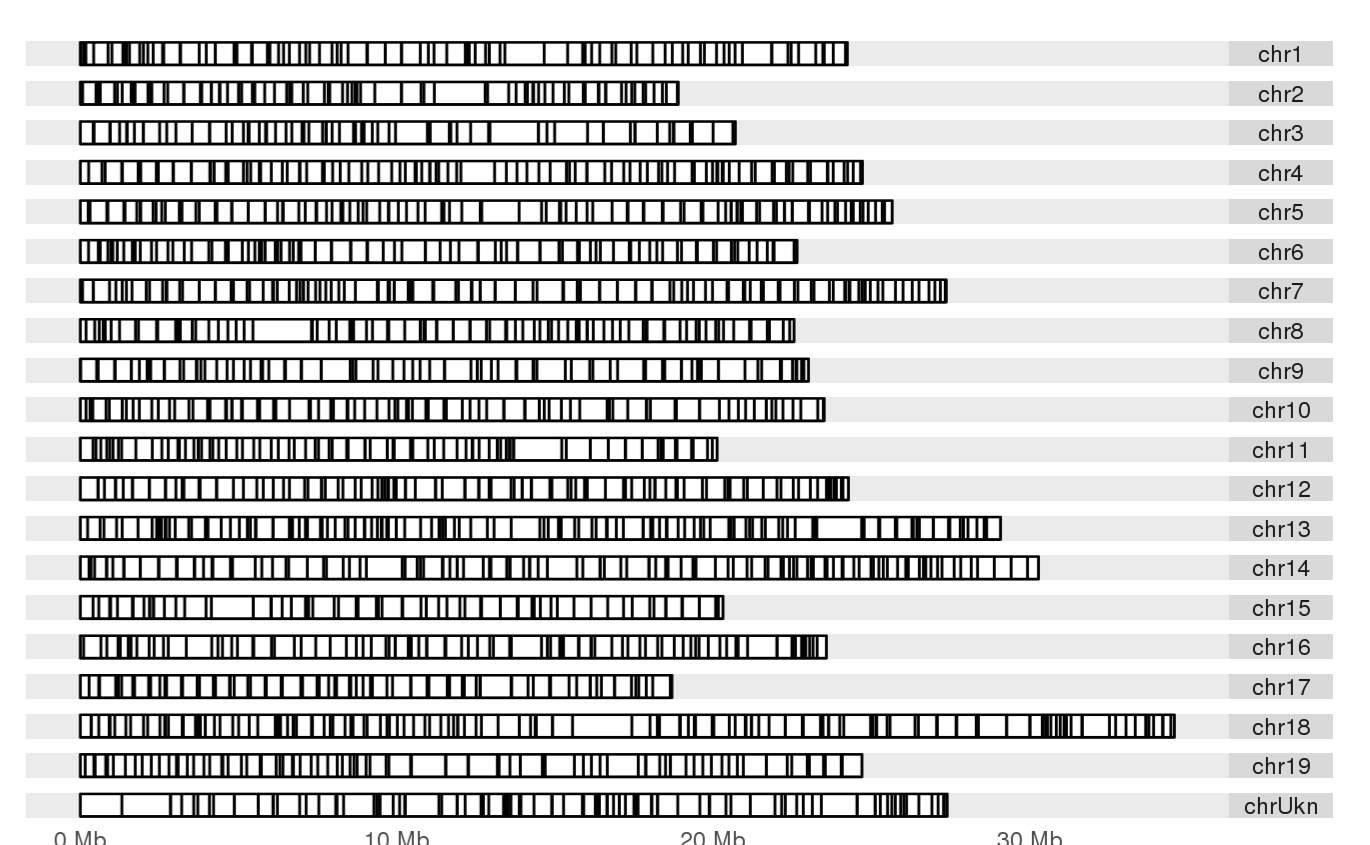
SxG : 1728 SNPs with LD < 0.4



RlxGW : 1551 SNPs with LD < 0.4



CSxRGM : 2016 SNPs with LD < 0.4



### 5. Genomic prediction

$$\text{BLUP}(g) = 1\mu + M\alpha + \epsilon \text{ where } \alpha \sim \mathcal{N}(0, \sigma_\alpha^2 \text{Id}) \text{ and } \epsilon \sim \mathcal{N}(0, \sigma_e^2 \text{Id}) : \text{fitted with GS3.}$$

The proportion of variance explained (PVE) shows that genotyped SNPs capture most of the genetic variance. Prediction accuracy is assessed by correlation ( $\rho$ ) via cross-validation (50 random replicates).

cross	trait	#genos	#SNPs	PVE	$\hat{\rho}$	$\hat{H}^2$
SxG	mbw	190	1728	0.90 [0.80,0.98]	0.73 ± 0.06	0.92 [0.89,0.94]
RlxGW	bud	122	1552	0.70 [0.43,0.88]	0.47 ± 0.15	0.75 [0.66,0.82]
CSxRGM	transpC	132	2016	0.41 [0.02,0.99]	-0.09 ± 0.19	0.12 [0.00,0.40]

It is helpful to interpret the differences in prediction accuracy ( $\rho$ ) along with the broad-sense heritability ( $H^2$ , computed here as if the data sets were balanced) which, as a measure of repeatability, quantify how well  $g$  is approximated by its BLUP. This depends on the quality of the model but also of the data : e.g., for the transpiration rate,  $H^2$  is too low because of a very small genetic variance ( $\sigma_g^2 \approx 10^{-4}$ ), leading to a very uncertain PVE and a  $\hat{\rho}$  of zero.

### 6. Genomic selection

The IFV started in 2015 a pilot project to breed new varieties for rosé wine and experiment with a strategy to speed-up the whole process. The goal is to obtain several varieties, not only with two resistance genes for each of downy and powdery mildew, but also adapted to the peculiarities of rosé vinification.

Crosses between iconic varieties for rosé and resistant progenitors from IFV are performed over several years. In phase 1 of the breeding program, classical marker-assisted selection on resistance genes is performed, and the selected genotypes can go on into phase 2.

In addition, the whole progeny is genotyped by sequencing. Moreover, among the offsprings produced in the first year of the project, 200 are trained in such a way that they will have berries in two years to phenotype traits related to the ideotype defined by experts, viticulturists and winemakers, such as juice oxidability, berry acidity, etc. Prediction models can then be trained on this subset, and used to predict genotypic values of non-phenotyped offsprings produced in the subsequent years. At this stage, the idea is to select a subset of resistant genotypes with sufficiently high performance on the other traits so that they can bypass phase 2 and be directly assessed in several field sites (phase 3).

### 7. Conclusion and perspectives

Genomic prediction looks promising enough to be experimented in grapevine breeding programs, yet its usefulness depends on several key factors. Genotyping is now flexibly achieved via restriction-assisted DNA sequencing, and statistical models and software implementations are available. The crucial component remains phenotyping protocols and tools, which should be cheap enough to be performed on large progenies. The phenotyped traits should then also be closely related to the ideotype of interest.

### 8. Funding and acknowledgments

INRA (FruitSelGen project), CASDAR (EDGARR project), South Green and URGI platforms.