



# GenTORE

## *Genomic management Tools to Optimise Resilience and Efficiency*

Grant agreement n°: 727213

**H2020 - Research and Innovation Action**

### **D5.3**

## ***Report on the potential and accuracy of genomic mating plans with an associated software prototype for at least one population***

**Due date:** M48 (May 2021)

**Actual submission date:** M48 (May 2021)

**Project start date:** 1<sup>st</sup> June 2017      **Duration:** 60 months

**Workpackage concerned:** WP5

**Concerned workpackage leader:** Donagh Berry

**Lead Beneficiary:** TEAGASC

**Dissemination level:**

- PU:** Public (must be available on the website)
- CO:** Confidential, only for members of the consortium (including the Commission Services)
- CI:** Classified, as referred to in Commission Decision 2001/844/EC

GenTORE – GA n° 727213

D5.3 - Potential and accuracy of genomic mating plans with an associated software prototype for at least one population



## Table of content

1. Summary.....	3
2. Introduction .....	3
3. Approach.....	3
4. Results.....	5
5. Conclussions.....	<b>Erreur ! Signet non défini.</b>

## 1. Summary

ALLICE in collaboration with INRAe, has developed and evaluated a simulation tool to depict the expected extent of variability in offspring from potential matings. This tool is based on genotypes and marker effect computed in the official national routine evaluation. The initial objective was to optimize mating at the breeding company level to create bulls suiting the best breeder's goal. The mating simulation procedure is now routinely available for breeding company since early 2020 in dairy cattle and is currently tested at a R&D level for beef breed. In parallel with the simulation, a tool to easily compute gametic variance i.e. variability of offsprings has been developed and evaluated for beef and dairy bulls in both Ireland and France.

## 2. Introduction

Genomic selection in France and Ireland was implemented in 2010 and 2008, respectively, and more and more females are genotyped annually. Hence, breeders now have access to more precise estimated genomic breeding values and knowledge of the variants carried for individuals genes of interest for both sire and dam. This has shifted breeding schemes from traditional progeny testing to embryo donor station but have also questioned the status quo on mating plans. At herd level, it was already proposed that genomic relationships should be considered rather than the traditional pedigree relationship when allocating mates in the pursuit of reducing the accumulation of inbreeding (Bérodier et al., 2020; Carthy et al., 2019). Formula were also proposed to account for inbreeding and the probability of carrying specific gene of interest when making mating plan (Bérodier et al, 2020; Pryce et al., 2012).

In the present task, focus was given to the mating plan at the level of the herd (Ireland) but also at the level of the breeding company (France) when planning the mating to conceive the future elite population. Focusing on elite population maximized the chances that both the candidate parents were genotyped making it easier to simulate but is also interesting because this population is less studied despite their importance for genetic progress.

The general idea here was to simulate offspring from potential matings enabling the investigation of the linkage disequilibrium between QTL within traits observed through the heterogeneity of progeny, the linkage disequilibrium between QTL across traits studied through genomic correlations, and the linkage disequilibrium between QTL and loci affecting gene of interest. The results are highly practical for dairy and beef farmers in the pursuit of homogenous progeny but also for breeders in the pursuit of heterogenous progeny hoping that some will excel and be useful as a parent of the next generation.

## 3. Approach

### Simulation:

The input used in the simulation software are French routine evaluation files. This allows us to directly use the most-up-to date information in our simulation and was thought to be efficient in a perspective of routine use.

The input for mating simulation are output of routine genomic evaluation :

- Imputed (if necessary) and phased genotypes

GenTORE – GA n° 727213

D5.3 - Potential and accuracy of genomic mating plans with an associated software prototype for at least one population

- Effect of SNP and haplotypes for all routinely evaluated traits \* breeds
- Haplotypes associated with alleles of genes of interest or genetic defects
- Formulas for computation of composite index (i.e. net merit)

Starting from a list of genotyped males and females, this tool simulates the genotypes of hundreds of offspring for each mating. Then, GEBVs of all simulated offspring are estimated using marker effects calculated during the national routine evaluation. Status for genes of interest (e.g. polled, caseins but also genetic defects) are also taken into account. A first tool was developed in 2018 showed interesting results but lacked efficiency and was too computationally demanding (time and memory) for routine. An alternative software developed by INRAe in 2019 gained efficiency by avoiding the export of genotypes. It allowed the dividing of the total computing time by 2 and real time by 10 thanks to parallelisation while drastically reducing memory required. It is now possible to generate DGV plus gene of interest for ten million offspring issued from 20 000 mating in less than 24 hours.

Output from the simulation software are therefore very similar to what is computed routinely for newly genotyped animals

- DGV for routinely evaluated traits (30 to 40 depending on the breeds)
- DGV for composite index (production, morphology, health, net merit)
- Status for gene of interest (3 to 10 depending on the breed)

Simulations studied here are based on real data and therefore on real linkage disequilibrium between genes of interest and QTL. We offer the possibility to compare progeny of bulls on specific criteria considering large number of traits and therefore taking into account correlations between traits. It became possible to compute statistics for several groups within progeny such as carrier of genetic defect or not and then take into account for linkage disequilibrium between QTL and genes of interest.

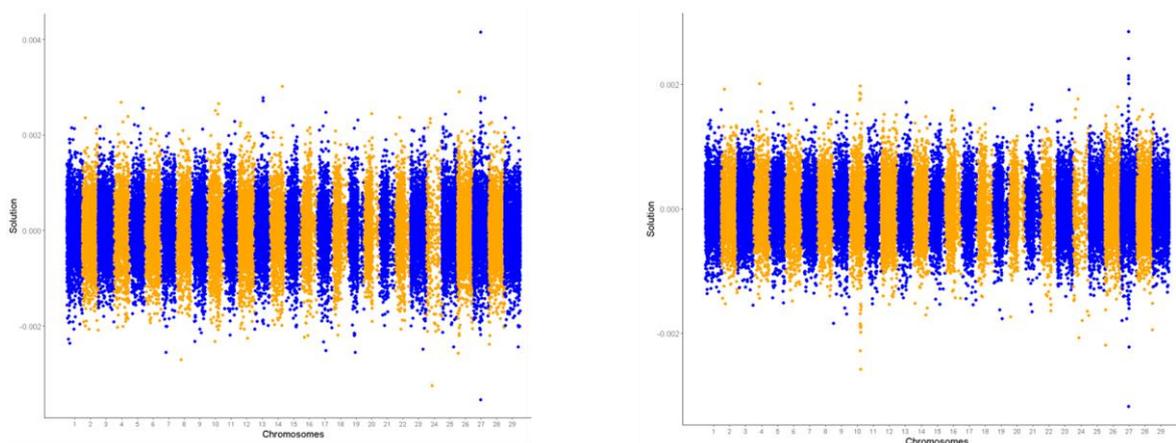
### Deterministic approach

The deterministic equation used to quantify the gametic variance per animal was:

$$\sigma_{\text{gamete}}^2 = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}' \begin{bmatrix} 0.25 & \dots & al_{N,1} \left( -\frac{cM_{N,1}}{200} + 0.25 \right) \\ \vdots & \ddots & \vdots \\ al_{N,1} \left( -\frac{cM_{N,1}}{200} + 0.25 \right) & \dots & 0.25 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$$

Where alpha represents the SNP effects from the national genomic evaluations,  $al_{jk}$  is a phase indicator for loci  $j$  and  $k$ , with value 1 when both loci have the reference allele on the same chromosome and  $-1$  otherwise;  $cM_{jk}$  is the genetic distance between the 2 loci (in centimorgans). Any 2 loci with genetic distance  $>50$  cM on the same chromosome, or on different chromosomes, are assumed to be independent and thus have zero values.

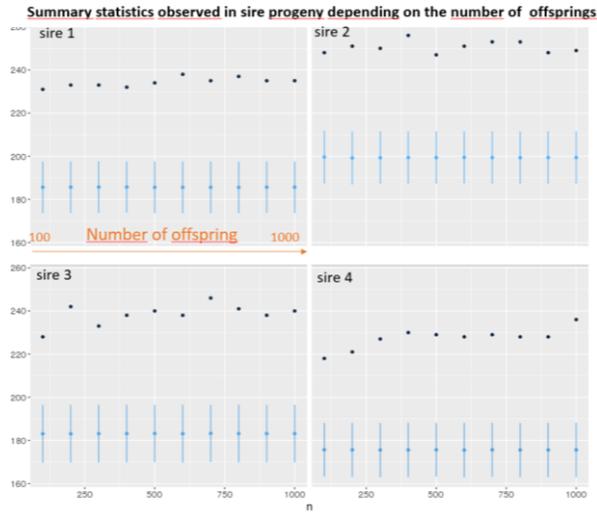
Calving difficulty records from the years 2005 to 2020 were extracted from Irish dairy herds; these were from dairy or beef bull matings on dairy females. Only herd-years that recorded some variability in calving difficulty were retained. Genetic evaluations for calving difficulty in Ireland is undertaken for primiparous and multiparous cows separately; both were treated separately in this analysis. The residuals from a model which adjusted calving difficulty (in primiparae and multiparae separately) for contemporary group, parity\*calf gender, heterosis, recombination loss and dam genetic merit for calving difficulty were calculated and the standard deviation of such estimated per sire. This was compared to the gametic variance per sire calculated using the deterministic equation and the SNP effects from the Irish national genomic evaluation below



## 4. Results

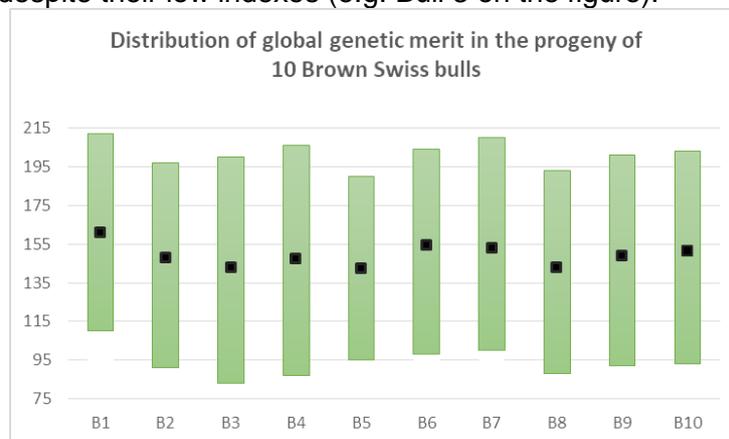
### Simulation

The initial research was to determine the optimal number of descendants to be simulated. The mating of 4 sires with 100 dams randomly selected from the population were studied. For each pair, 100 to 1000 offsprings were simulated. The mean and standard deviation of the progeny per sire was calculated. The figure below shows the mean net merit (blue dot) and maximum net merit (black dot) and standard deviation (blue error bars) observed in the progeny of every sire.



The conclusion was that 100 simulated offspring was enough to predict genetic level of the mating. Interestingly, the variability in offspring differed between bull with sire3 in the figure above having more diverse progeny.

The distribution of possible offspring for a given bull, a given cow and for a given genes of interest was then derived. We observed that for traits governed by large QTL such as production traits, the variability of offspring is three times larger than for more polygenic traits such as fertility. The average level of offspring for net merit index depending on the bull is in the Figure below. Interestingly, these simulations highlighted some bulls that can produce top ranking progeny, despite their low indexes (e.g. Bull 3 on the figure).



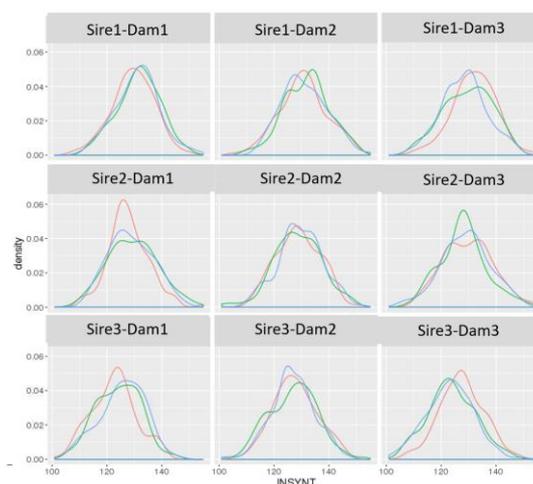
Considering the whole set of traits at an offspring level, and therefore accounting for correlation between traits within mating, also enables the maximization of the probability to produce a bull fitting with marketing segment.

This can be of particular interest to deal with negative genetic correlation between functional and production traits, where bulls excelling in genetic merit for fertility with limited impact on milk production is important. In the Table below is an example for 9 mating to have an offspring filling the criteria to be selected as “production”, “functional” or “conformation” bull. We also observed here that the probability of producing a bull matching the marketing is segment computed out of the simulation procedure has a correlation of 0.58 with the genetic level of ISU.

SIRE	DAM	%_prod	%_funct	%_conf
S1	D1	0.26	0.01	0.482

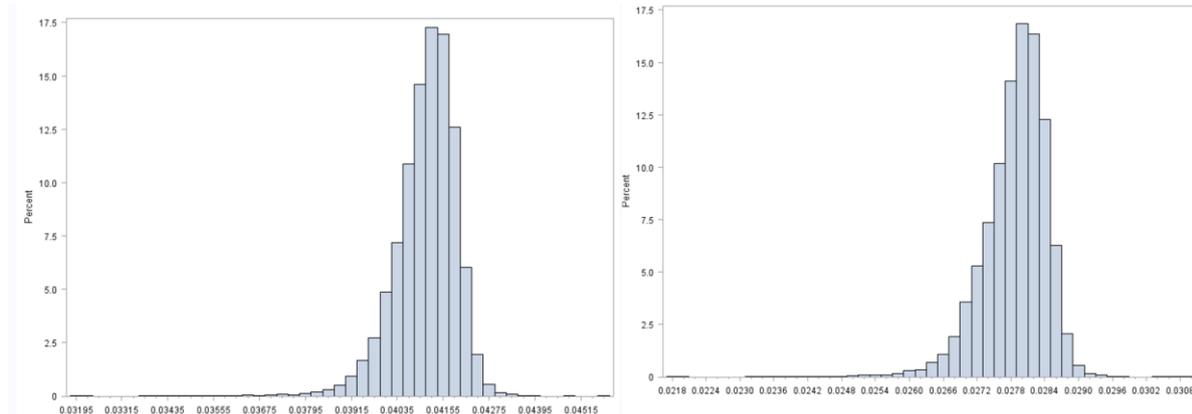
S1	D2	0.132	0.122	0.48
S1	D3	0.16	0.006	0.358
S2	D1	0.18	0	0.234
S2	D2	0.308	0.002	0.418
S2	D3	0.112	0	0.15
S3	D1	0.16	0	0.224
S3	D2	0.136	0.026	0.252
S3	D3	0.066	0	0.128

The expected genotype for a gene of interest is also computed during offspring simulation using haplotypes in high linkage disequilibrium with the mutation. We were then able to observe the genetic merit of the offspring depending on the status for gene of interest. The figure below, is an example of the distribution of net-merit index depending of the status of a genetic defect. We observed here that, depending on the status the mean level of offspring may differ meaning that you have to take this information into account when making your mating decisions.



Also observed during the simulation process was that one critical factor was the heterogeneity of offspring which depends on gametic variance. This can be estimated by simulation of offspring but the procedure is quite computationally demanding and cannot be applied at herd level. Following that, we developed a tool to be able to directly compute the gametic variance reflecting the variability of offspring level. Again, the computation is based on routine evaluation files such as phased genotypes and marker effects from national genomic evaluations. Then, we computed the contrasts of effects between the alleles carried by paternal and maternal haplotype and extended it to the whole genome. Indicators were compared to empirical variability obtained by simulation and the correlation ranged from 0.90 to 0.95 between the predicted and the empirical variability in France. It means that such an indicator could be interesting to predict the level of extreme progeny. It may be useful at a breeding company level to predict the probability to have an elite bull. Bijma et al, 2020 suggested to breed bulls based on a combination of genetic level and gametic variance to increase genetic progress. Breeders may also be interested in such indicators for traits governed by major genes and therefore where large differences in variability of offsprings can be observed. It was tested in Ireland to investigate the gametic variability of calving traits for beef breeds. Unfortunately they observed that a low variability within individuals due to a very polygenic traits, but further work on production traits seems promising.

The results from the Irish population were not as convincing when applied to calving difficulty without any major gene. The gametic variance for sires for calving difficulty in heifers (left) and cows (right) are demonstrated in the figure below. After adjusting for the mean genetic level, the correlation between the estimated gametic variance and residual variance was just 0.05 and 0.08 for primiparae and multiparae, respectively.



## 5. Conclusions

In conclusion, we produced a tool based on simulation and deterministic equations of offspring to optimize the mating when breeding animals. We observed that the probability of having a calf selected to be marketed is only partly linked to the level of genetic merit of parents. It will offer an opportunity to account for gametic variance when selecting animal which may lead to a more diverse choice of bulls and an increase in genetic diversity. Finally, accounting for gene of interest at the mating procedure may also lead to a better handling of genetic defects. Little association existed between the deterministically calculated gametic variance and the variance in calving difficulty of individual bulls.