Purebred and crossbred genomic evaluation and mate allocation strategies to exploit dominance in pig crossbreeding schemes

David González-Diéguez\*1, Llibertat Tusell\*, Alban Bouquet†,‡ , Andres Legarra\*, Zulma G. Vitezica\*

\* GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France.

† IFIP Institut du Porc, BP35104, 35651 Le Rheu, France.

‡ France Génétique Porc, BP35104, 35651 Le Rheu, France.

Running title: Mate allocation in crossbreeding schemes

Keywords: Purebred and crossbred evaluation, non-additive genetic effects, heterosis, purebred-crossbred genetic correlation.

1Corresponding authors:

David González-Diéguez

INRAE Toulouse, UMR 1388 GenPhySE
24 Chemin de Borde Rouge
31326 CASTANET-TOLOSAN Cedex,

France

Tel : +33 561285153

david-omar.gonzalez-dieguez@inrae.fr

# SUPPLEMENTAL MATERIAL

## S1. Simulation of QTL effects for two parental breeds and their crossbred

At each locus, three (P1, P2 for the two parental breeds and C for the CB progeny) QTL additive and three (for P1, P2 and C) QTL dominance effects were sampled from two normal distributions with 0 mean and covariance matrices

$$Var\left(\begin{matrix}a^{P1}\\a^{P2}\\a^{c}\end{matrix}\right)=\left(\begin{matrix}σ\_{a\_{QTL}^{P1}}^{2}&σ\_{a\_{QTL}^{P1,P2}}&σ\_{a\_{QTL}^{P1,c}}\\σ\_{a\_{QTL}^{P2,P1}}&σ\_{a\_{QTL}^{P2}}^{2}&σ\_{a\_{QTL}^{P2,c}}\\σ\_{a\_{QTL}^{c,P1}}&σ\_{a\_{QTL}^{c,P2}}&σ\_{a\_{QTL}^{c}}^{2}\end{matrix}\right)$$

$$Var\left(\begin{matrix}d^{\*P1}\\d^{\*P2}\\d^{\*c}\end{matrix}\right)=\left(\begin{matrix}σ\_{d\_{QTL}^{\*P1}}^{2}&σ\_{d\_{QTL}^{\*P1,P2}}&σ\_{d\_{QTL}^{\*P1,c}}\\σ\_{d\_{QTL}^{\*P2,P1}}&σ\_{d\_{QTL}^{\*P2}}^{2}&σ\_{d\_{QTL}^{\*P2,c}}\\σ\_{d\_{QTL}^{\*c,P1}}&σ\_{d\_{QTL}^{\*c,P2}}&σ\_{d\_{QTL}^{\*c}}^{2}\end{matrix}\right)$$

Variances in the diagonal of these matrices were obtained backsolving from estimates in the animal scale in Xiang *et al.* (2016). For instance, additive $(σ\_{a\_{QTL}^{P1}}^{2})$ and dominance $(σ\_{d\_{QTL}^{\*P1}}^{2})$ QTL variances in the P1 breed with allele frequencies $p\_{QTL\_{j}}^{P1}$ and $q\_{QTL\_{j}}^{P1}$ at the $j^{th}$ QTL, are equal to:

$$σ\_{d\_{QTL}^{\*P1}}^{2}=σ\_{v^{P1}}^{2}/\sum\_{j=1}^{nQTL}\left(2p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1}\right)^{2}$$

$$σ\_{a\_{QTL}^{P1}}^{2}=(σ\_{u^{P1}}^{2}-\sum\_{j=1}^{nQTL}(2p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1}\left(q\_{QTL\_{j}}^{P1}-p\_{QTL\_{j}}^{P1}\right)^{2})σ\_{d\_{QTL}^{\*P1}}^{2})/\sum\_{j=1}^{nQTL}(2p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1})$$

where $σ\_{u^{P1}}^{2}$ and $σ\_{v^{P1}}^{2}$ are the additive and dominance genetic variances in P1. QTL variances were computed in the same manner for P2. In the CB population, additive $(σ\_{a\_{QTL}^{c}}^{2})$ and dominance $(σ\_{d\_{QTL}^{\*c}}^{2})$ QTL variances were obtained from Eq. (5), (6) and (7) in Vitezica *et al.* (2016), as:

$$σ\_{d\_{QTL}^{\*c}}^{2}=σ\_{v^{c}}^{2}/\sum\_{j=1}^{nQTL}(4p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1}p\_{QTL\_{j}}^{P2}q\_{QTL\_{j}}^{P2})$$

 And

$$σ\_{a\_{QTL}^{c}}^{2}=\frac{\left[σ\_{u^{c}}^{2}-\left[∑\left(p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1}\left(q\_{QTL\_{j}}^{P2}-p\_{QTL\_{j}}^{P2}\right)^{2}\right)+∑\left(p\_{QTL\_{j}}^{P2}q\_{QTL\_{j}}^{P2}\left(q\_{QTL\_{j}}^{P1}-p\_{QTL\_{j}}^{P1}\right)^{2}\right)\right]σ\_{d\_{QTL}^{\*c}}^{2}\right]}{\left[∑\left(p\_{QTL\_{j}}^{P1}q\_{QTL\_{j}}^{P1}\right)+∑\left(p\_{QTL\_{j}}^{P2}q\_{QTL\_{j}}^{P2}\right)\right]}$$

where $p\_{QTL\_{j}}^{P2}$ and $q\_{QTL\_{j}}^{P2}$ refers to the frequencies in the second parental breed (P2). Note that QTL allelic frequencies differ between parental breeds.

The covariance matrix for residual effects was as follows:

$$Var\left(\begin{matrix}e\_{1}\\e\_{2}\\e\_{c}\end{matrix}\right)=\left(\begin{matrix}σ\_{e1}^{2} &0&0\\0&σ\_{e2}^{2}&0\\0&0&σ\_{ec}^{2}\end{matrix}\right)⊗I$$

## S2. Genomic evaluation models

Genomic evaluations were performed based on two models, depending on each scenario. The first model was a univariate GBLUP model included only additive genetic effects plus genomic inbreeding, and was written as follows:

$$y=1μ+fb+Zu+e,$$

where $y$ is a vector of phenotypes of PB females; $μ$ is the general mean; $f$ is the vector of genomic inbreeding coefficients, calculated as the average homozygosity per individual; and $b$is the inbreeding depression parameter. Further, $u$is the vector of additive genetic effects and $e$ is a vector of random residual effects $(e\~N\left(0, Iσ\_{e}^{2}\right))$. Breeding values were distributed as $u\~N\left(0,Gσ\_{u}^{2}\right)$, where $G$is the additive genomic relationship matrix, and $Z$ is the incidence matrix that related phenotypes to breeding values. Terms $σ\_{u}^{2}$ and $σ\_{e}^{2}$ refer to the additive genetic and residual variances, respectively, which were assumed known.

The additive genomic relationship matrix was calculated according to VanRaden (2008), as follows:

$$G=\frac{MM´}{2\sum\_{k=1}^{m}p\_{k}q\_{k}},$$

where $M$ is a matrix with dimensions equal to the number of animals $(n)$ by number of SNPs $(m)$, with elements equal to $\left(2-2p\_{k}\right)$, $\left(1-2p\_{k}\right)$ and $-2p\_{k}$, for genotypes $AA, Aa,$ and $aa$ respectively; $p\_{k}$ is the frequency for allele $A$ of the $k^{th}$ SNP (for $k=1,…,m$), and $q\_{k}=1-p\_{k}$. This model was implemented with *Blupf90* software (Misztal *et al.* 2002). The $EBV\_{P}$ for candidates were obtained directly from the solutions of the mixed model equations.

The second model was a trivariate SNP-BLUP, with correlated SNP effects across the three populations. This model can be written according to Xiang *et al.* (2016) as:

$$y^{P1}=1μ^{P1}+f^{P1}b^{P1}+Z^{P1}a^{P1}+W^{P1}d^{\*^{P1}}+e^{P1}$$

$$y^{P2}=1μ^{P2}+f^{P2}b^{P2}+Z^{P2}a^{P2}+W^{P2}d^{\*^{P2}}+e^{P2}$$

$$y^{c}=1μ^{c}+f^{c}b^{c}+Z^{c}a^{c}+W^{c}d^{\*^{c}}+e^{c}$$

where $y^{P1}$, $y^{P2}$, and $y^{c}$ are the vector containing the phenotype of females for P1 and P2 breeds, and CB population, respectively; $μ^{P1}$, $μ^{P2}$, and $μ^{c}$ are the respective general mean; $a^{P1}$, $a^{P2}$, $a^{c}$ and $d^{\*^{P1}}$, $d^{\*^{P2}}$, $d^{\*^{c}}$ are the functional additive and dominance SNP effects for each population, respectively. Note that estimated dominance SNP effects $d^{\*}$ are centered ($d^{\*}=d-μ\_{d}$). Then $f^{P1}$, $f^{P2}$ and $f^{c}$ are the genomic inbreeding coefficient as explained above; and $b^{P1}$, $b^{P2}$ and$ b^{c}$ are the inbreeding depression parameters for P1, P2, and CB, respectively, which are transformed into directional depression estimates $μ\_{d}=-\frac{b}{m}$ so that the final estimate of the dominance effects is $\hat{d}=  \hat{d^{\*}}+1\hat{μ\_{d}}$. $Z$ and $W$ areincidence matrices that related the data to additive and dominance SNP effects, respectively. SNP genotypes were coded as 1, 0, and -1 for additive effects and 0, 1 and 0 for dominance effects, for genotypes $AA, Aa,$ and $aa$, respectively. The covariance structures for the additive and dominance SNP effects across populations are:

$$Var\left(\begin{matrix}a^{P1}\\a^{P2}\\a^{c}\end{matrix}\right)=\left(\begin{matrix}σ\_{a\_{SNP}^{P1}}^{2}&σ\_{a\_{SNP}^{P1,P2}}&σ\_{a\_{SNP}^{P1,c}}\\σ\_{a\_{SNP}^{P2,P1}}&σ\_{a\_{SNP}^{P2}}^{2}&σ\_{a\_{SNP}^{P2,c}}\\σ\_{a\_{SNP}^{c,P1}}&σ\_{a\_{SNP}^{c,P2}}&σ\_{a\_{SNP}^{c}}^{2}\end{matrix}\right)⨂I$$

$$Var\left(\begin{matrix}d^{\*P1}\\d^{\*P2}\\d^{\*c}\end{matrix}\right)=\left(\begin{matrix}σ\_{d\_{SNP}^{\*P1}}^{2}&σ\_{d\_{SNP}^{\*P1,P2}}&σ\_{d\_{SNP}^{\*P1,c}}\\σ\_{d\_{SNP}^{\*P2,P1}}&σ\_{d\_{SNP}^{\*P2}}^{2}&σ\_{d\_{SNP}^{\*P2,c}}\\σ\_{d\_{SNP}^{\*c,P1}}&σ\_{d\_{SNP}^{\*c,P2}}&σ\_{d\_{SNP}^{\*c}}^{2}\end{matrix}\right)⨂I$$

The additive and dominance SNP (co)variances were assumed known (backsolving from estimates in the animal scale in Xiang *et al.* (2016)). This SNP-BLUP was implemented using own Fortran program. The $EBV\_{C}$ was obtained from the equations used to simulate the $TBV\_{C}$, but instead of using QTL genotypes and effects, the SNP frequencies from P1 and P2 and the estimated SNP effects of the CB progeny were used. The SNP effects were estimated each generation by using the current allele frequencies.

Genomic evaluations were performed each generation, and animals in previous generations were included in the analyses. For instance, for S1 and S2, the number of PB animals evaluated by GBLUP in the last generation was equal to 27,144. However, when the trivariate SNP-BLUP model was implemented, a maximum of 5 generations was included in the evaluations, due to the high computing demanding resources and time. For example, in S3 and S4, the total number of animals evaluated in the last generation was 36,720 (12,240 animals from each of the three populations). Note that, as CB animals were created after selecting PB animals, CB information used in the genetic evaluation came from the previous generations. For instance, for the first genomic evaluation in generation 1, we included PB information (genotypes) from generation 1 and founders (genotypes and records of females), but the CB information (genotypes) comes only from the CB reference population derived from founder animals (see Figure 2 in the manuscript).

# LITERATURE CITED

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet *et al.*, 2002 BLUPF90 and related programs (BGF90), pp. 28 in *Proceedings of the 7th World Congress Applied to Livestock Production.*, INRA, Montpellier, France.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423.

Vitezica, Z. G., L. Varona, J.-M. Elsen, I. Misztal, W. Herring *et al.*, 2016 Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. Genet. Sel. Evol. 48: 6.

Xiang, T., O. F. Christensen, Z. G. Vitezica, and A. Legarra, 2016 Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. Genet. Sel. Evol. 48: 92.