

Additional File 2

We applied the Usefulness Criterion Parental Contributions approach (UCPC) proposed by Allier *et al.* (2019a) and further extended in Allier *et al.* (2019b) to evaluate the interest of a set of two-way crosses regarding the performance and the diversity in the best fraction of the progeny of each cross.

Prediction of the mean expected breeding value and parental contributions in the selected fraction of progeny

Considering two inbred lines P_1 and P_2 and the cross $P_1 \times P_2$ and $(x_1, x_2)'$ denotes their $(2 \times m)$ -dimensional genotyping matrix at the $m = 2,000$ SNP markers. x_p denotes the $(m \times 1)$ -dimensional genotype vector of parent $P_{p \in \{1,2\}}$ with the j^{th} element coded as 1 or -1 for the genotypes AA or aa at QTL j . Following Lehermeier *et al.* (2017), the DH progeny mean and progeny variance of the breeding values in the progeny before selection can be computed as:

$$\hat{\mu}_T = 0.5 (x'_1 \hat{\beta} + x'_2 \hat{\beta}), \text{ (Eq. 1a)}$$

$$\hat{\sigma}_T^2 = \hat{\beta}' \Sigma \hat{\beta}, \text{ (Eq. 1b)}$$

where $\hat{\beta}$ is $(m \times 1)$ -dimensional vector of estimated marker effects and Σ is the $(m \times m)$ -dimensional variance covariance matrix of marker genotypes in DH progeny defined in Lehermeier *et al.* (2017). We define the $(m \times 1)$ -dimensional vector β_{C1} to follow P_1 genome contribution to progeny as $\beta_{C1} = \frac{x_1 - x_2}{(x_1 - x_2)'(x_1 - x_2)}$. The mean and variance of P_1 contribution in the progeny before selection are computed as:

$$\mu_{C1} = 0.5 (x'_1 \beta_{C1} + x'_2 \beta_{C1} + 1), \text{ (Eq. 2a)}$$

$$\sigma_{C1}^2 = \beta_{C1}' \Sigma \beta_{C1}. \text{ (Eq. 2b)}$$

The progeny mean for P_2 contribution is then $\mu_{C2} = 1 - \mu_{C1}$.

Following Allier *et al.* (2019a), the covariance between the breeding values and P_1 contribution in progeny is:

$$\hat{\sigma}_{T,C1} = \hat{\beta}' \Sigma \beta_{C1}. \text{ (Eq. 3)}$$

The expected mean breeding value of the selected fraction of progeny, i.e. usefulness criterion (Schnell and Utz 1975), of the cross $P_1 \times P_2$ is:

$$\widehat{UC}^{(i,h)} = \hat{\mu}_T + ih\hat{\sigma}_T, \text{ (Eq. 4)}$$

where i is the within family selection intensity and h the within family selection accuracy. The correlated responses to selection on P_1 and P_2 contributions to the selected fraction of progeny are (Falconer and Mackay 1996):

$$\hat{c}_1^{(i,h)} = \mu_{C1} + ih \frac{\hat{\sigma}_{T,C1}}{\hat{\sigma}_T} \text{ and } \hat{c}_2^{(i,h)} = 1 - \hat{c}_1^{(i,h)}. \text{ (Eq. 5)}$$

Optimal cross selection accounting for within family variance

Considering N homozygote candidate parents, $N(N-1)/2$ two-way crosses are possible. We define a crossing plan $\mathbf{nc}$ as a set of $|\mathbf{nc}|$ crosses out of possible two-way crosses, giving the index of selected crosses, i.e. with the i^{th} element $nc(i) \in [1, N(N-1)/2]$. The $(N \times 1)$ -dimensional vector of candidate parents estimated contributions in the selected fraction of progeny of each cross $\hat{\mathbf{c}}^{(i,h)}$ is:

$$\hat{\mathbf{c}}^{(i,h)} = \frac{1}{|\mathbf{nc}|} (\mathbf{Z}_1 \hat{\mathbf{c}}_1^{(i,h)} + \mathbf{Z}_2 \hat{\mathbf{c}}_2^{(i,h)}), \text{ (Eq. 6)}$$

where $\mathbf{Z}_1$ (respectively $\mathbf{Z}_2$) is a $(N \times |\mathbf{nc}|)$ -dimensional design matrix that links each N candidate parent to the first (respectively second) parent in the set of crosses $\mathbf{nc}$, $\hat{\mathbf{c}}_1^{(i,h)}$ (respectively $\hat{\mathbf{c}}_2^{(i,h)}$) is a $(|\mathbf{nc}| \times 1)$ -dimensional vector containing the estimated contributions of the first (respectively second) parent to the selected fraction of the progeny of the crosses in $\mathbf{nc}$.

The expected performance $V(\mathbf{nc})$ for this set of two-way crosses is defined as the expected mean performance of the selected DH progeny, i.e. usefulness criterion:

$$\hat{V}^{(i,h)}(\mathbf{nc}) = \frac{1}{|\mathbf{nc}|} \sum_{j \in \mathbf{nc}} \widehat{UC}^{(i,h)}(j). \text{ (Eq. 7)}$$

The constraint on diversity $\widehat{D}^{(i,h)}(\mathbf{nc})$ in the selected progeny is:

$$\widehat{D}^{(i,h)}(\mathbf{nc}) = 1 - \hat{\mathbf{c}}^{(i,h)'} \mathbf{K} \hat{\mathbf{c}}^{(i,h)}, \text{ (Eq. 8)}$$

where $\mathbf{K}$ is the $(N \times N)$ -dimensional identity by state (IBS) coancestry matrix at markers between the N candidates. Allier *et al.* (2019b) showed that $\widehat{D}^{(i,h)}(\mathbf{nc})$ is a good proxy of the genomewide diversity

in the selected fraction of progeny $He^{(i,h)} = \frac{1}{m} \sum_{j=1}^m 2p_j^{(i,h)}(1 - p_j^{(i,h)})$ where $p_j^{(i,h)}$ is the frequency of the genotypes AA at marker j in the selected fraction of progeny.

Literature cited

- Allier A., L. Moreau, A. Charcosset, S. Teyssèdre, and C. Lehermeier, 2019a Usefulness Criterion and Post-selection Parental Contributions in Multi-parental Crosses: Application to Polygenic Trait Introgression. *G3 Genes Genomes Genet.* 9: 1469–1479.
- Allier A., C. Lehermeier, A. Charcosset, L. Moreau, and S. Teyssèdre, 2019b Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. *Front. Genet.* 10.
- Falconer D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics. 4th ed.* Pearson, Harlow, England.
- Lehermeier C., S. Teyssèdre, and C.-C. Schön, 2017 Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. *Genetics* 207: 1651–1661.
- Schnell F., and H. Utz, 1975 F1-Leistung und Elternwahl in der Züchtung von Selbstbefruchtern., pp. 243–248 in *Bericht über die Arbeitstagung der Vereinigung österreichischer Pflanzenzüchter.*, BAL Gumpenstein, Austria.