File S2

Validation of four-way cross formulas for DH-k and RIL-k and evolution of RIL variance depending on selfing generations

1   In File S1, we considered DH lines generated from F1' (DH-1), i.e., only two meioses took place.

2   Progeny variance for DH-1 is expressed in terms of parental expected recombination frequency $c^{(1)}$

3   (Table 2 S1). For recombinant inbred lines (RILs) or when DH lines are generated from higher selfing

4   generations, the expected frequency of recombinants increases depending on the number of selfing

5   generations. In the following k denotes the generation from which progeny are derived (Figure 1). The

6   expected frequency of recombinants in generation k can be derived from the genotype probabilities

7   given in Broman (2012) as done in File S1 of Lehermeier et al. (2017). Hence, for DH lines after k

8   generations, $c^{(1)}$ in Table 2 S1 should then be replaced by $c^{(k)}$, leading to the general four-way DH-k

9   formula as shown in Table 1:

10
$$c^{(k)} = \frac{2c^{(1)}}{1 + 2c^{(1)}} \left(1 - 0.5^k \left(1 - 2c^{(1)}\right)^k\right), \forall\, k \in \mathbb{N}^*$$

11   In case of RILs, no doubling of gametes takes place and the covariance for RILs after generation k is

12   obtained by updating $c^{(k)}$ by $c^{(k)} + 0.5 \left[0.5\left(1 - 2c^{(1)}\right)\right]^k, \forall\, k \in \mathbb{N}^*$ (Table 1). Note that the variance-

13   covariance of DH-k and RIL-k converge with increasing k.

14   Formulas for DH-k and RIL-k in the general case of four-way crosses have been validated by

15   simulations for $k \in [\![1,6]\!]$ (Table 1 S2 and Table 2 S2). The observed high positive correlations (Table 1

16   S2) and low mean squared differences (Table 2 S2) between predicted (derivation) and empirical (*in*

17   *silico*) values validate the presented formulas. Lower squared correlations between predicted and

18   empirical values were observed for $\boldsymbol{\mu}_C^{(sel)}$ and $\boldsymbol{\mu}_{C(+)}^{(sel)}$ compared to the variances and covariances. This

19   can be explained by sampling bias in *in silico* simulations (50,000 progenies) where the P_1 parental

20   genome contribution before selection slightly differed from the expected value of 0.25 for four way

21   crosses (ranging from 0.249 to 0.251).

Table 1 S2 Squared correlations (R²) between empirical values (*in silico*) and predictions (derivation) per generation and type of progeny.

Generation	σ_T^2	σ_C^2	$\sigma_{C(+)}^2$	σ_{T-C}	$\sigma_{T-C(+)}$	UC_T	$\mu_C^{(sel)}$	$\mu_{C(+)}^{(sel)}$
DH1	0.999	0.960	0.995	0.999	0.999	1.000	0.900	0.946
DH2	0.999	0.964	0.995	0.998	0.998	1.000	0.909	0.952
DH3	0.999	0.966	0.995	0.999	0.999	1.000	0.914	0.955
DH4	0.999	0.969	0.995	0.999	0.999	1.000	0.912	0.955
DH5	0.999	0.961	0.994	0.998	0.998	1.000	0.914	0.955
DH6	0.999	0.963	0.994	0.998	0.998	1.000	0.913	0.955
RIL1	0.999	0.957	0.994	0.999	0.999	1.000	0.938	0.967
RIL2	0.999	0.957	0.994	0.999	0.999	1.000	0.917	0.957
RIL3	0.999	0.960	0.994	0.998	0.998	1.000	0.918	0.958
RIL4	0.999	0.962	0.994	0.998	0.998	1.000	0.915	0.956
RIL5	0.999	0.962	0.994	0.998	0.998	1.000	0.912	0.955
RIL6	0.999	0.962	0.994	0.999	0.998	1.000	0.911	0.954

Table 2 S2 Mean squared difference between empirical values (*in silico*) and predictions (derivation) per generation and type of progeny.

Generation	σ_T^2	σ_C^2	$\sigma_{C(+)}^2$	σ_{T-C}	$\sigma_{T-C(+)}$	UC_T	$\mu_C^{(sel)}$	$\mu_{C(+)}^{(sel)}$
DH1	5.20E-06	3.28E-09	3.52E-10	5.99E-08	2.07E-08	8.44E-04	4.92E-05	1.42E-05
DH2	5.09E-06	2.81E-09	3.16E-10	6.65E-08	2.24E-08	7.02E-04	3.83E-05	1.12E-05
DH3	5.36E-06	2.56E-09	2.97E-10	4.74E-08	1.51E-08	6.49E-04	3.50E-05	1.03E-05
DH4	4.56E-06	2.30E-09	2.87E-10	5.16E-08	1.66E-08	6.85E-04	3.55E-05	1.05E-05
DH5	4.83E-06	2.88E-09	3.32E-10	5.95E-08	1.99E-08	6.40E-04	3.47E-05	1.03E-05
DH6	4.76E-06	2.74E-09	3.14E-10	6.08E-08	1.96E-08	6.77E-04	3.47E-05	1.04E-05
RIL1	2.25E-06	1.56E-09	1.81E-10	2.96E-08	9.80E-09	4.30E-04	2.51E-05	7.54E-06
RIL2	3.26E-06	2.29E-09	2.69E-10	4.09E-08	1.37E-08	5.73E-04	3.40E-05	1.00E-05
RIL3	3.93E-06	2.58E-09	3.05E-10	5.28E-08	1.72E-08	6.22E-04	3.34E-05	9.84E-06
RIL4	4.49E-06	2.59E-09	3.02E-10	5.64E-08	1.81E-08	6.59E-04	3.43E-05	1.01E-05
RIL5	4.91E-06	2.69E-09	3.10E-10	5.59E-08	1.83E-08	6.65E-04	3.53E-05	1.04E-05
RIL6	4.91E-06	2.71E-09	3.13E-10	5.54E-08	1.83E-08	6.63E-04	3.59E-05	1.06E-05

23   Predicted RIL progeny variance for the simulated agronomic trait increased with the number of selfing

24   generations considered (k) and converged toward DH progeny variance after five generations of selfing

25   ($k = 5$) (Figure 1 S2). We observed that some crosses profited more from an increase in selfing

26   generations by generating more variance compared to others. An example with two crosses is shown

27   in Figure 2 S2. While the cross visualized in blue showed a higher variance in generation RIL-1 than the

28   cross visualized in orange, it reached a plateau faster and showed a lower variance than the orange

29   cross with $k \geq 3$. Differences in the speed to release variance between crosses is likely due to

30   differences in the recombination frequency between segregating QTLs in parental lines. This

31   underlines the interest of predicting RIL progeny variance using proposed algebraic formula.
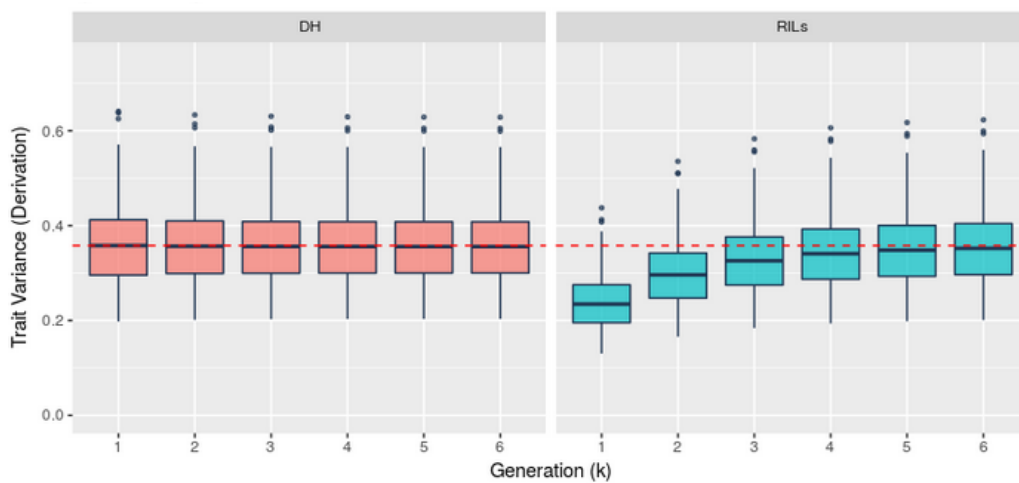

Figure 1 S2 Evolution of predicted progeny trait variance depending on progeny type (DH, left or RIL, right) and generation (k). The red dotted line presents the median DH progeny variance over 100 crosses.
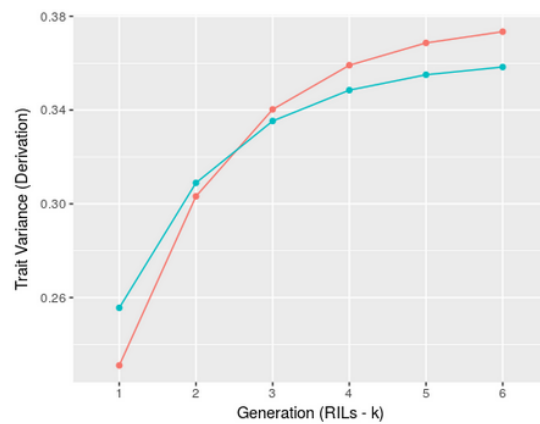
Figure 2 S2 Example of two crosses showing different evolutions of predicted RIL progeny variance depending on the selfing generation (k).

Literature cited

Broman K. W., 2012 Genotype probabilities at intermediate generations in the construction of

recombinant inbred lines. Genetics 190: 403–412.

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.