File S1

Derivation of linkage disequilibrium parameter in progeny for four-way cross and specific case of two-way cross, three-way cross and backcross

Here we derive the linkage disequilibrium parameter of doubled haploid progeny derived from the F_1' generation of a four-way cross (Figure 1 S1), while we give an extension for DH lines generated from higher selfing generations and for recombinant inbred lines in File S2. The crossing scheme for a four-way cross visualizing parental and potential progeny haplotypes is given in Figure 1 S1. Gametes from a four-way cross with four different parents (P1, P2, P3, and P4) correspond to gametes from six biparental crosses (P1xP2, P3xP4, P1xP3, P1xP4, P2xP3, P2xP4).

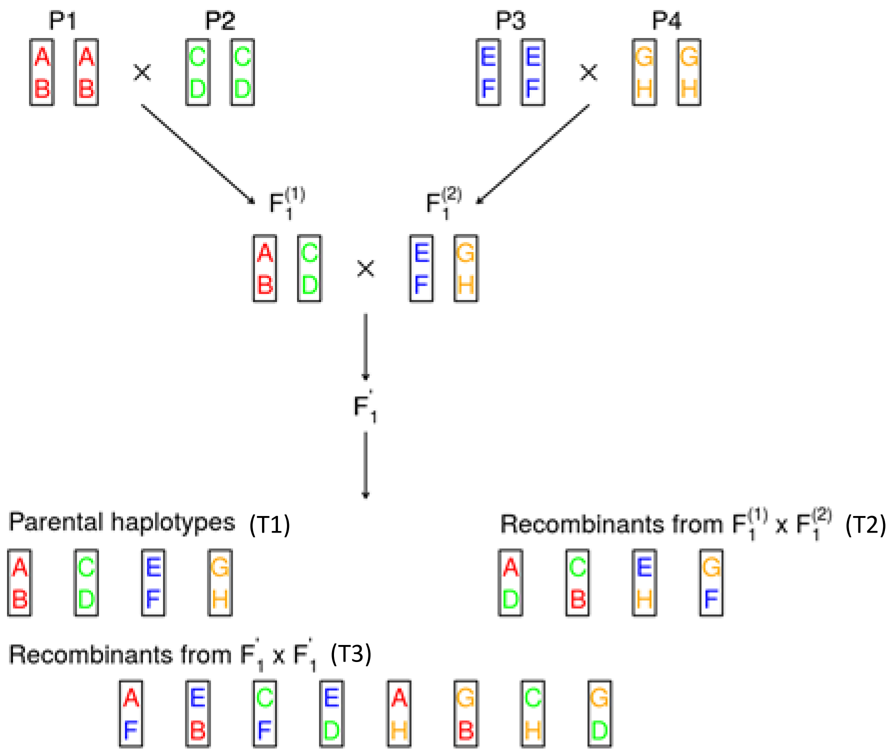

Figure 1 S1 Visualization of crossing scheme and two-locus parental as well as progeny haplotypes of a four-way cross from parents P1, P2, P3, and P4. Potential types of haplotypes are denoted with T1, T2, and T3.

To derive the entries of the Linkage Disequilibrium (LD) matrix $\boldsymbol{D}$ of the progeny of the four-way cross, we derive the frequencies of all different possible haplotypes. For this, three types of haplotypes can be differentiated (namely, T1, T2 and T3).

The first type T1 corresponds to parental haplotypes, for example AB from Figure 1 S1. The frequency of the haplotype AB in the parents is:

$$p_{AB} = \frac{1}{4}$$

The frequency of AB in gametes from the cross $F_1^{(1)} \times F_1^{(2)}$ is:

$$p'_{AB} = \frac{1}{4}\left(1 - c^{(1)}\right),$$

with $c^{(1)}$ the recombination frequency and $(1 - c^{(1)})$ the frequency that no recombination takes place within the cross $F_1^{(1)} \times F_1^{(2)}$.

Similarly, the frequency of AB in gametes from the cross $F'_1 \times F'_1$ is:

$$p''_{AB} = \frac{1}{4} * \left(1 - c^{(1)}\right)^2$$

As there are four different parental haplotypes, the frequency of the type T1 haplotypes is:

$$P(T_1) = p''_{AB} + p''_{CD} + p''_{EF} + p''_{GH} = \left(1 - c^{(1)}\right)^2 \ (1)$$

The second type T2 corresponds to haplotypes formed by recombination in the cross $F_1^{(1)} \times F_1^{(2)}$, for example AD. The frequency of this haplotype in the parents is

$$p_{AD} = 0$$

The frequency of AD in gametes from the cross $F_1^{(1)} \times F_1^{(2)}$ is:

$$p'_{AD} = \frac{1}{2} * \frac{c^{(1)}}{2} = \frac{1}{4} c^{(1)}$$

As $\frac{c^{(1)}}{2}$ is the frequency of recombinants within $F_1^{(1)}$, the frequency in the whole cross is reduced by a factor of 1/2. The frequency of AD in gametes from the cross $F'_1 \times F'_1$ is:

$$p''_{AD} = \frac{1}{4} c^{(1)}\left(1 - c^{(1)}\right),$$

with $(1 - c^{(1)})$ the frequency that no recombination takes place within the cross $F'_1 \times F'_1$.

Overall, the frequency of the type T2 haplotypes is:

$$P(T_2) = p''_{AD} + p''_{CB} + p''_{EH} + p''_{GF} = c^{(1)}\left(1 - c^{(1)}\right) \qquad (2)$$

The third type T3 corresponds to haplotypes formed by recombination in the cross $F_1' \times F_1'$, for example AF. The frequency of these haplotypes in the parents is:

$$p_{AF} = 0$$

The frequency of AF in gametes from the cross $F_1^{(1)} \times F_1^{(2)}$ is:

$$p'_{AF} = 0$$

The frequency of AF in gametes from the cross $F_1' \times F_1'$ can be calculated as:

$$p''_{AF} = \frac{1}{2}\left(1 - c^{(1)}\right) * \frac{1}{2}\left(1 - c^{(1)}\right) * \frac{c^{(1)}}{2} + \frac{c^{(1)}}{2} * \frac{1}{2}\left(1 - c^{(1)}\right) * \frac{c^{(1)}}{2}$$

$$+ \frac{1}{2}\left(1 - c^{(1)}\right) * \frac{c^{(1)}}{2} * \frac{c^{(1)}}{2} + \frac{c^{(1)}}{2} * \frac{c^{(1)}}{2} * \frac{c^{(1)}}{2}$$

$$= \frac{1}{8} c^{(1)}$$

Overall, the frequency of the type T3 haplotypes is:

$$P(T_3) = p''_{AF} + p''_{EB} + p''_{CF} + p''_{ED} + p''_{AH} + p''_{GB} + p''_{CH} + p''_{GD} = c^{(1)} \qquad (3)$$

All the different haplotypes and frequencies are summarized in Table 1 S1.

We define $h_{jl} = \left(h_j, h_l\right)$ a haplotype including loci j and l, with h_j and h_l the alleles of the haplotype at loci j and l, $h_j, h_l \in \{0,1\}$. Using the frequencies of the three types of haplotypes, we derive the LD in the progeny between locus j and l as:

$$D_{jl}^{progeny} = p_{jl} - p_j p_l$$

$$= P\left(h_{jl} = \left(z_j, z_l\right)\right) - P\left(h_j = z_j\right) P\left(h_l = z_l\right)$$

$$= \sum\nolimits^{3}_{k=1} P\big(h_{jl} = (z_j, z_l) \mid T_k\big) P(T_k) - P\big(h_j = z_j\big) P(h_l = z_l), \quad (4)$$

where z_j and z_l denotes realizations of h_j and h_l.respectively.

For the conditional haplotype probabilities it holds:

$$P\big(h_{jl} = (z_j, z_l) \mid T_k\big) = \frac{1}{|T_k|} \sum_{v_{jl} \in T_k} \mathbf{1}_{v_j == z_j} \times \mathbf{1}_{v_l == z_l},$$

with $|T_k|$ the number of haplotypes of type k, $v_{jl} = \big(v_j, v_l\big)$ a haplotype of type k, $\mathbf{1}_{v_j == z_l}$ ($\mathbf{1}_{v_j == z_l}$) an indicator equal to 1 if $v_j = z_j$ ($v_l = z_l$) and 0 otherwise.

For the allele frequencies it holds:

$$P\big(h_j = z_j\big) = \frac{1}{4}\ (\ \mathbf{1}_{A == z_j} + \mathbf{1}_{C == z_j} + \mathbf{1}_{E == z_j} + \mathbf{1}_{G == z_j})$$

$$P(h_l = z_l) = \frac{1}{4}\ (\ \mathbf{1}_{B == z_l} + \mathbf{1}_{D == z_l} + \mathbf{1}_{F == z_l} + \mathbf{1}_{H == z_l})$$

Table 1 S1 Different haplotype types, their frequency in the parents (G0), after the first cross (G1), after the second cross (G2) and the Linkage Disequilibrium (LD) in G2.

Type	G0	G1	G2	LD
T1[a]	$\frac{1}{4}$	$\frac{1}{4}\big(1 - c^{(1)}\big)$	$\frac{1}{4} * \big(1 - c^{(1)}\big)^2$	$\frac{1}{4} * \big(1 - c^{(1)}\big)^2 - \frac{1}{16}$
T2[b]	0	$\frac{1}{4}c^{(1)}$	$\frac{1}{4}c^{(1)} * \big(1 - c^{(1)}\big)$	$\frac{1}{4}c^{(1)} * \big(1 - c^{(1)}\big) - \frac{1}{16}$
T3[c]	0	0	$\frac{1}{8}c^{(1)}$	$\frac{1}{8}c^{(1)} - \frac{1}{16}$

[a] Haplotypes: AB, CD, EF, GH (parental haplotypes)
[b] Haplotypes: AD, BC, EH, FG (recombinant from $F_1^{(1)} \times F_1^{(2)}$)
[c] Haplotypes: AF, AH, CF, CH, EB, ED, GB, GD (recombinant from $F_1' \times F_1'$)

Further, we use the linkage disequilibrium among two parents between loci j and l, which is exemplified for parent 1 and 2:

$$D_{jl}^{12} = p_{jl}^{12} - p_j^{12} p_l^{12}$$

$$= \frac{1}{2}\Big(\mathbf{1}_{A == z_j} \times \mathbf{1}_{B == z_l} + \mathbf{1}_{C == z_j} \times \mathbf{1}_{D == z_l}\Big) - \frac{1}{4}\Big(\mathbf{1}_{A == z_j} + \mathbf{1}_{C == z_j}\Big)\big(\mathbf{1}_{B == z_l} + \mathbf{1}_{D == z_l}\big)$$

$$= \frac{1}{4}\left(\mathbf{1}_{A==z_j} \times \mathbf{1}_{B==z_l} + \mathbf{1}_{C==z_j} \times \mathbf{1}_{D==z_l} - \mathbf{1}_{A==z_j} \times \mathbf{1}_{D==z_l} - \mathbf{1}_{C==z_j} \times \mathbf{1}_{B==z_l}\right).$$

For sake of clarity, we abbreviate in the following $\mathbf{1}_{A==z_j}$ with $\mathbf{1}_A$, $\mathbf{1}_{B==z_l}$ with $\mathbf{1}_B$ and accordingly for the rest (C, D, E, F, G, H). Then we can reform the LD in the progeny as a function of the recombination frequency $c_{jl}^{(1)}$ and the LD among two parents between loci j and l:

$$D_{jl}^{progeny} = \sum_{k=1}^{3} P\left(h_{jl} = \left(z_j, z_l\right) \mid T_k\right) P(T_k) - P\left(h_j = z_j\right) P(h_l = z_l)$$

$$\begin{aligned}
&= \frac{1}{4}\left(\mathbf{1}_A\mathbf{1}_B + \mathbf{1}_C\mathbf{1}_D + \mathbf{1}_E\mathbf{1}_F + \mathbf{1}_G\mathbf{1}_H\right)\left(1 - c_{jl}^{(1)}\right)^2 \\
&\qquad + \frac{1}{4}\left(\mathbf{1}_A\mathbf{1}_D + \mathbf{1}_C\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_H + \mathbf{1}_G\mathbf{1}_F\right) c_{jl}^{(1)}\left(1 - c_{jl}^{(1)}\right) \\
&\qquad + \frac{1}{8}\left(\mathbf{1}_A\mathbf{1}_F + \mathbf{1}_A\mathbf{1}_H + \mathbf{1}_C\mathbf{1}_F + \mathbf{1}_C\mathbf{1}_H + \mathbf{1}_E\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_D + \mathbf{1}_G\mathbf{1}_B + \mathbf{1}_G\mathbf{1}_D\right) c_{jl}^{(1)} \\
&\qquad - \frac{1}{16}\left(\mathbf{1}_A + \mathbf{1}_C + \mathbf{1}_E + \mathbf{1}_G\right)\left(\mathbf{1}_B + \mathbf{1}_D + \mathbf{1}_F + \mathbf{1}_H\right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4}\left[\left(\left(1 - c_{jl}^{(1)}\right)^2 - \frac{1}{4}\right)\left(\mathbf{1}_A\mathbf{1}_B + \mathbf{1}_C\mathbf{1}_D + \mathbf{1}_E\mathbf{1}_F + \mathbf{1}_G\mathbf{1}_H\right)\right. \\
&\qquad + \left(c_{jl}^{(1)}\left(1 - c_{jl}^{(1)}\right) - \frac{1}{4}\right)\left(\mathbf{1}_A\mathbf{1}_D + \mathbf{1}_C\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_H + \mathbf{1}_G\mathbf{1}_F\right) \\
&\qquad \left. + \left(\frac{c_{jl}^{(1)}}{2} - \frac{1}{4}\right)\left(\mathbf{1}_A\mathbf{1}_F + \mathbf{1}_A\mathbf{1}_H + \mathbf{1}_C\mathbf{1}_F + \mathbf{1}_C\mathbf{1}_H + \mathbf{1}_E\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_D + \mathbf{1}_G\mathbf{1}_B + \mathbf{1}_G\mathbf{1}_D\right)\right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4}\left[\left(1 - c_{jl}^{(1)}\right)^2\left(\mathbf{1}_A\mathbf{1}_B + \mathbf{1}_C\mathbf{1}_D + \mathbf{1}_E\mathbf{1}_F + \mathbf{1}_G\mathbf{1}_H\right)\right. \\
&\qquad + c_{jl}^{(1)}\left(1 - c_{jl}^{(1)}\right)\left(\mathbf{1}_A\mathbf{1}_D + \mathbf{1}_C\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_H + \mathbf{1}_G\mathbf{1}_F\right) \\
&\qquad - \frac{1}{4}\left(\mathbf{1}_A\mathbf{1}_B + \mathbf{1}_C\mathbf{1}_D + \mathbf{1}_E\mathbf{1}_F + \mathbf{1}_G\mathbf{1}_H\right) - \frac{1}{4}\left(\mathbf{1}_A\mathbf{1}_D + \mathbf{1}_C\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_H + \mathbf{1}_G\mathbf{1}_F\right) \\
&\qquad - \frac{1}{4}\left(1 - 2c_{jl}^{(1)}\right)\left(\mathbf{1}_A\mathbf{1}_F + \mathbf{1}_A\mathbf{1}_H + \mathbf{1}_C\mathbf{1}_F + \mathbf{1}_C\mathbf{1}_H + \mathbf{1}_E\mathbf{1}_B + \mathbf{1}_E\mathbf{1}_D + \mathbf{1}_G\mathbf{1}_B\right. \\
&\qquad \left.\left. + \mathbf{1}_G\mathbf{1}_D\right)\right]
\end{aligned}$$

$$= \frac{1}{4}\Big[\left(1-c_{jl}^{(1)}\right)(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H)$$

$$-c_{jl}^{(1)}\left(1-c_{jl}^{(1)}\right)(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H-\mathbf{1}_A\mathbf{1}_D-\mathbf{1}_C\mathbf{1}_B-\mathbf{1}_E\mathbf{1}_H$$

$$-\mathbf{1}_G\mathbf{1}_F)-\frac{1}{4}(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H+\mathbf{1}_A\mathbf{1}_D+\mathbf{1}_C\mathbf{1}_B+\mathbf{1}_E\mathbf{1}_H+\mathbf{1}_G\mathbf{1}_F)$$

$$-\frac{1}{4}\left(1-2c_{jl}^{(1)}\right)(\mathbf{1}_A\mathbf{1}_F+\mathbf{1}_A\mathbf{1}_H+\mathbf{1}_C\mathbf{1}_F+\mathbf{1}_C\mathbf{1}_H+\mathbf{1}_E\mathbf{1}_B+\mathbf{1}_E\mathbf{1}_D+\mathbf{1}_G\mathbf{1}_B$$

$$+\mathbf{1}_G\mathbf{1}_D)\Big]$$

$$= \frac{1}{4}\Big[\left(1-c_{jl}^{(1)}\right)(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H)-4c_{jl}^{(1)}\left(1-c_{jl}^{(1)}\right)\left(D_{jl}^{12}+D_{jl}^{34}\right)$$

$$-\frac{1}{4}(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H+\mathbf{1}_A\mathbf{1}_D+\mathbf{1}_C\mathbf{1}_B+\mathbf{1}_E\mathbf{1}_H+\mathbf{1}_G\mathbf{1}_F)$$

$$+\frac{1}{4}\left(1-2c_{jl}^{(1)}\right)\left(4D_{jl}^{13}+4D_{jl}^{14}+4D_{jl}^{23}+4D_{jl}^{24}-2\mathbf{1}_A\mathbf{1}_B-2\mathbf{1}_C\mathbf{1}_D-2\mathbf{1}_E\mathbf{1}_F\right.$$

$$\left.-2\mathbf{1}_G\mathbf{1}_H\right)\Big]$$

$$= \frac{1}{4}\left[\left(1-c_{jl}^{(1)}-\frac{1}{4}-\frac{2}{4}\left(1-2c_{jl}^{(1)}\right)\right)(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H)\right.$$

$$-4c_{jl}^{(1)}\left(1-c_{jl}^{(1)}\right)\left(D_{jl}^{12}+D_{jl}^{34}\right)-\frac{1}{4}(\mathbf{1}_A\mathbf{1}_D+\mathbf{1}_C\mathbf{1}_B+\mathbf{1}_E\mathbf{1}_H+\mathbf{1}_G\mathbf{1}_F)$$

$$\left.+\frac{1}{4}\left(1-2c_{jl}^{(1)}\right)\left(4D_{jl}^{13}+4D_{jl}^{14}+4D_{jl}^{23}+4D_{jl}^{24}\right)\right]$$

$$= \frac{1}{4}\Big[\frac{1}{4}(\mathbf{1}_A\mathbf{1}_B+\mathbf{1}_C\mathbf{1}_D+\mathbf{1}_E\mathbf{1}_F+\mathbf{1}_G\mathbf{1}_H-\mathbf{1}_A\mathbf{1}_D-\mathbf{1}_C\mathbf{1}_B-\mathbf{1}_E\mathbf{1}_H-\mathbf{1}_G\mathbf{1}_F)$$

$$-4c_{jl}^{(1)}\left(1-c_{jl}^{(1)}\right)\left(D_{jl}^{12}+D_{jl}^{34}\right)+\left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{13}+D_{jl}^{14}+D_{jl}^{23}+D_{jl}^{24}\right)\Big]$$

$$= \frac{1}{4}\Big[\frac{1}{4}\left(4D_{jl}^{12}+4D_{jl}^{34}\right)-4c_{jl}^{(1)}\left(1-c_{jl}^{(1)}\right)\left(D_{jl}^{12}+D_{jl}^{34}\right)$$

$$+\left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{13}+D_{jl}^{14}+D_{jl}^{23}+D_{jl}^{24}\right)\Big]$$

$$= \frac{1}{4}\left[\left(D_{jl}^{12}+D_{jl}^{34}\right)\left(\frac{1}{4}-4c_{jl}^{(1)}(1-c_{jl}^{(1)})\right)+\left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{13}+D_{jl}^{14}+D_{jl}^{23}+D_{jl}^{24}\right)\right]$$

$$= \frac{1}{4}\left[\left(1-2c_{jl}^{(1)}\right)^2 \left(D_{jl}^{12} + D_{jl}^{34}\right) + \left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{13} + D_{jl}^{14} + D_{jl}^{23} + D_{jl}^{24}\right)\right]$$

$$= \frac{1}{4}\left(1-2c_{jl}^{(1)}\right)\left[\left(D_{jl}^{13} + D_{jl}^{14} + D_{jl}^{23} + D_{jl}^{24}\right) + \left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{12} + D_{jl}^{34}\right)\right]$$

$$= \frac{1}{4}\left(1-2c_{jl}^{(1)}\right)\left[\Phi_{2\ jl} + \left(1-2c_{jl}^{(1)}\right)\Phi_{1\ jl}\right] \qquad (5)$$

with $\Phi_{1\ jl} = D_{jl}^{12} + D_{jl}^{34}$ summing the LD values among parents that can be considered to be involved as biparental crosses in $F_1^{(1)} \times F_1^{(2)}$ and with $\Phi_{2\ jl} = D_{jl}^{13} + D_{jl}^{14} + D_{jl}^{23} + D_{jl}^{24}$ summing the LD values among parents that can be considered to be involved as biparental crosses in $F_1' \times F_1'$.

The linkage disequilibrium parameter $\Phi_1$and Φ_2 and equation (5) can be simplified in the case of two-way, three-way and backcrosses (Table 2 S1). For two-way crosses we arrive at the same variance covariance matrix elements Σ_{jl} as given by Lehermeier et al. (2017).

Table 2 S1 Linkage disequilibrium parameter between QTLs j and l in pairs of parental lines depending on the mating design.

	$\Phi_{1\ jl}$	$\Phi_{2\ jl}$	Σ_{jl}
Four-way	$D_{jl}^{12} + D_{jl}^{34}$	$D_{jl}^{13} + D_{jl}^{14} + D_{jl}^{23} + D_{jl}^{24}$	$\left(1-2c_{jl}^{(1)}\right)\left(\left(D_{jl}^{13} + D_{jl}^{14} + D_{jl}^{23} + D_{jl}^{24}\right) + \left(1-2c_{jl}^{(1)}\right)\left(D_{jl}^{12} + D_{jl}^{34}\right)\right)$
Three-way	D_{jl}^{12}	$2\left(D_{jl}^{14} + D_{jl}^{24}\right)$	$\left(1-2c_{jl}^{(1)}\right)\left(2\left(D_{jl}^{14} + D_{jl}^{24}\right) + \left(1-2c_{jl}^{(1)}\right)D_{jl}^{12}\right)$
Backcross	D_{jl}^{14}	$2\,D_{jl}^{14}$	$\left(1-2c_{jl}^{(1)}\right)\left(3-2c_{jl}^{(1)}\right)D_{jl}^{14}$
Two-way	0	$4\,D_{jl}^{14}$	$4\left(1-2c_{jl}^{(1)}\right)D_{jl}^{14}$

Literature cited

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.