Appendix 1

Expected Prediction Accuracy [$E(r_{MG,\text{r2}})$] with Incomplete Linkage Disequilibrium ($r^2 < 1$) between a Marker and a Quantitative Trait Locus

Assume that $N$ is the number of individuals ($i = 1$ to $N$) and $N_M$ is the number of markers ($j = 1$ to $N_M$). For the $j$th marker-QTL pair, assume $m_j$ is a vector of genotypic indicators for the $j$th marker; $x_j$ is a vector of genotypic indicators for the $j$th QTL; $r_j$ is the correlation between $m_{ij}$ and $x_y$ for any $i$; $b_j$ is the true QTL effect; $\hat{b}_j$ is the estimated QTL effect; $\mu$ is the overall mean; $y$ is a vector of phenotypic values; and $e$ is a vector of residual (error) effects. The $m_j$ vector is corrected for its mean so that $\bar{m}_j = 0$. Each marker-QTL pair is assumed independent of the other pairs. Each QTL is assumed to have an equal variance and we also assume $\text{var}(x) = \text{var}(m)$.

From linear regression,

$$\hat{b}_j = (m_j' m_j)^{-1} m_j' y
= (m_j' m_j)^{-1} m_j' (\mu + x_j b_j + e)$$

Because we have $\bar{m}_j = \sum_{i=1}^N m_{ij} = 0$, we will have $m_j' \mu = 0$ and $m_j' \bar{x}_j b_j = 0$. We can then write $\hat{b}_j$ as:

$$\hat{b}_j = (m_j' m_j)^{-1} m_j' (x_j b_j - \bar{x}_j b_j + e)
= (m_j' m_j)^{-1} m_j' (x_j - \bar{x}_j) b_j + e
= r_j b_j + (m_j' m_j)^{-1} m_j' e$$

We assume the phenotypic variance is equal to 1 (Daetwyler et al., 2008) and we define the following:

$$g_i = \sum_{j=1}^{N_M} x_{ij} b_j$$
\[
\text{var}(g_i) = \sum_{j=1}^{N_M} \text{var}(x) b_j^2 = h^2
\]

\[
\hat{g}_i = \sum_{j=1}^{N_M} m_i \hat{b}_j
\]

\[
m_j m_j = N \text{var}(m)
\]

If we assume \( r_j = r \) for all values of \( j \), then

\[
\text{cov}(g_i, \hat{g}_i) = \sum_{j=1}^{N_M} \text{cov}(x, m_i b_j)
\]

\[
= \sum_{j=1}^{N_M} \text{cov}(x, m_i r_j b_j + m_i (m_j m_j)^{-1} m_j e)
\]

\[
= \sum_{j=1}^{N_M} r_j^2 b_j^2 \text{var}(x) + \sum_{j=1}^{N_M} b_j^2 \text{var}(x) r^2
\]

\[
= h^2 r^2
\]

We let \( \hat{e}_j = (m_j m_j)^{-1} m_j e \), with \( \text{var}(\hat{e}_j) = (N \text{var}(m))^{-1} \sigma^2_e \). With the assumption by Daetwyler et al. (2008) that \( \sigma^2_e = 1 \),

\[
\text{var}(\hat{g}_i) = \text{var}(\sum_{j=1}^{N_M} m_i \hat{b}_j)
\]

\[
= \text{var}(\sum_{j=1}^{N_M} m_i r_j b_j + \hat{e}_j)
\]

\[
= \text{var}(\sum_{j=1}^{N_M} m_i r_j b_j) + \text{var}(\sum_{j=1}^{N_M} m_i \hat{e}_j)
\]

\[
= \sum_{j=1}^{N_M} \text{var}(m_i r_j b_j) + \sum_{j=1}^{N_M} \text{var}(m_i \hat{e}_j)
\]

\[
= \sum_{j=1}^{N_M} r_j^2 b_j^2 \text{var}(m) + \sum_{j=1}^{N_M} \text{var}(m) \text{var}(\hat{e})
\]

\[
= r^2 \sum_{j=1}^{N_M} b_j^2 \text{var}(m) + \sum_{j=1}^{N_M} \text{var}(m) (N \text{var}(m))^{-1}
\]

\[
= h^2 r^2 + N \sigma^2_e / N
\]

Finally,
\[ E(r_{MG, r^2}) = \frac{\text{cov}(g, \hat{g})}{\sqrt{\text{var}(g) \text{var}(\hat{g})}} \]
\[ = \frac{r^2 h^2}{[h^2 (r^2 h^2 + N_r / N)]^{1/2}} \]
\[ = r^2 [Nh^2 / (r^2 Nh^2 + N_r)]^{1/2} \]
Appendix 2

Derivation of $r^2_{MM/2}$, the Squared Correlation between a Marker and QTL when the QTL Is Assumed at the Midpoint of Two Markers

If there are two markers, A and C, with a QTL (B) at the midpoint of the two markers, the recombination rate between AB and BC should be equal. Assuming no interference, the recombination rate between AC can be calculated as (Haldane, 1919):

$$c_{AC} = c_{AB} + c_{BC} - 2c_{AB}c_{BC} = 2c_{AB} - 2c_{AB}^2.$$ Therefore, $c_{AB} = \left[1 - (1 - 2c_{AC})^{1/2}\right]/2$.

Consider the gametes produced by F1 or F2 plants. Assume $x_1$ is the allele at marker 1 and $x_2$ is the allele at marker 2 in a gamete; $p_1$ is the frequency of parent 1 allele at marker 1; and $p_2$ is the frequency of parent 1 allele at marker 2; $q_1$ is the frequency of the parent 2 allele at marker 1; and $q_2$ is the frequency of the parent 2 allele at marker 2. Alleles from parent 1 are coded as 1 and alleles from parent 2 are coded as 0. The $r^2$ is the squared correlation between marker haplotypes; $c$ is the recombination fraction between two markers.

Given that $E(x_1) = p_1 = 0.5$, $E(x_2) = p_2 = 0.5$, and $E(x_1x_2)=c/2$:

$$r^2 = \frac{\text{cov}^2(x_1, x_2)}{\text{var}(x_1)\text{var}(x_2)}$$

$$= E^2[(x_1 - p_1)(x_2 - p_2)]/(p_1q_1p_2q_2)$$

$$= E^2[x_1x_2 - p_1x_2 - p_2x_1 + p_1p_2]/(p_1q_1p_2q_2)$$

$$= (c/2 - 0.25)^2/0.5^4$$

$$= (2c - 1)^2$$
For BC populations and DH populations derived from the F\(_1\), the squared correlation between marker genotypes is the same as the squared correlation between haplotype genotypes of the gametes produced by F\(_1\) plants. For DH populations derived from the F\(_2\), the squared correlation between marker genotypes is the same as the squared correlation between haplotype genotypes of the gametes produced by F\(_2\) plants.

For F\(_2\) populations, assume: \(g_1\) is the genotype at marker 1; \(g_2\) is the genotype at marker 2; \(x_1\) and \(x_2\) are the haplotypes from one chromosome; \(x_1'\) and \(x_2'\) are the haplotypes from the homologous chromosome. For an F\(_2\) population:

\[
\begin{align*}
\text{cov}(g_1, g_2) &= \text{cov}(x_1 + x_1', x_2 + x_2') \\
&= \text{cov}(x_1, x_2) + \text{cov}(x_1, x_2') + \text{cov}(x_1', x_2) + \text{cov}(x_1', x_2') \\
&= 2\text{cov}(x_1, x_2) \\
\text{var}(g_1) &= \text{var}(x_1 + x_1') \\
&= 2\text{var}(x_1) \\
\text{var}(g_2) &= \text{var}(x_2 + x_2') \\
&= 2\text{var}(x_2) \\
\text{r}^2(g_1, g_2) &= \frac{\text{cov}^2(g_1, g_2)}{\text{var}(g_1)\text{var}(g_2)} \\
&= \frac{\text{cov}^2(x_1, x_2)}{\text{var}(x_1)\text{var}(x_2)}
\end{align*}
\]

So, for an F\(_2\) population, the squared correlation between marker genotypes is the same as the squared correlation between haplotypes of the gametes produced by F\(_1\) plants. Such relationship does not apply to a BC\(_1\)F\(_2\) population.
Based on the above analysis, if we know that the squared correlation between marker genotypes is $r_{\text{MM}}^2$, then for F$_2$, BC, DH populations from F$_1$ and DH populations derived from from the F$_2$, the following relationships exist: $r_{\text{MM}}^2 = (1 - 2c_{\text{MM}})^2$ and $c_{\text{MM}/2} = [1 - (1 - 2c_{\text{MM}})^{1/2}] / 2 = [1 - (r_{\text{MM}}^2)^{1/4}] / 2$. We can therefore calculate the correlation squared between the marker and the midpoint QTL as $\sigma_e^2 = 1$.

References:
