

Figure S1 Histograms of the off-diagonal realized relationship coefficients.

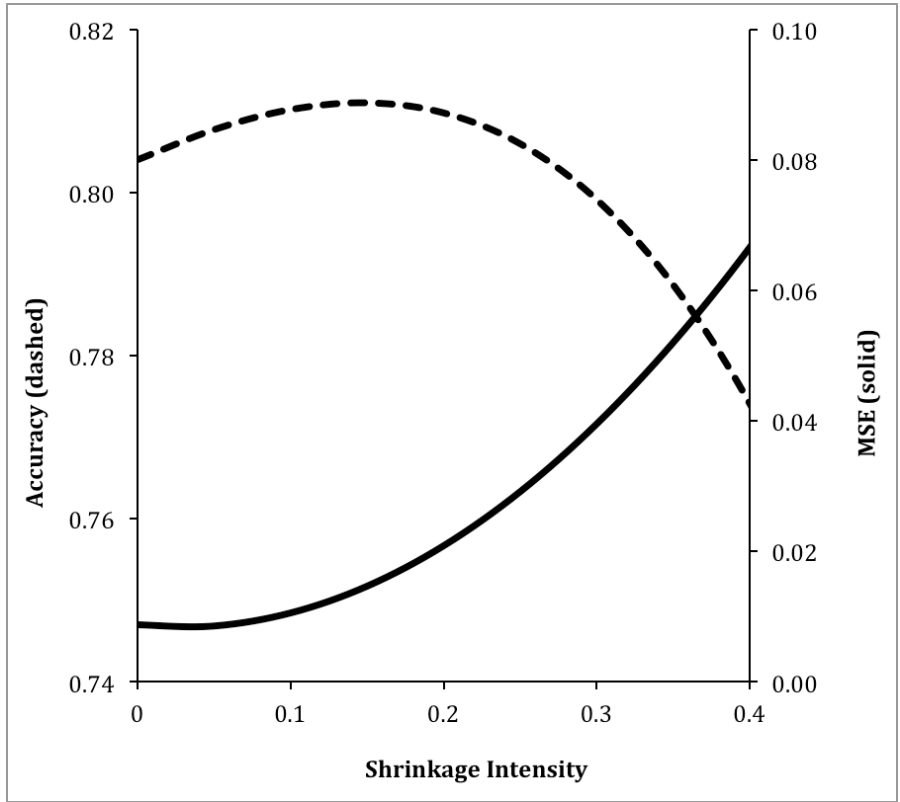


Figure S2 Maximizing accuracy vs. minimizing MSE for the 2+6-row barley population (see Figure 3 caption).

File S1

Starting from the parent-offspring regression (Equation 31):

$$h^2 = \frac{\text{cov}_i[a_i, y_i]}{\text{var}_i[y_i]} = \frac{\mathbf{a}'(\mathbf{y} - \mu\mathbf{1})}{\|\mathbf{y} - \mu\mathbf{1}\|^2}$$

we derive an expression for heritability in terms of the variance components of the mixed model

$\mathbf{y} = \mu\mathbf{1} + \mathbf{a} + \boldsymbol{\varepsilon}$, where $\mathbf{a} \sim N(0, \mathbf{G})$, $\mathbf{G} = \mathbf{A}\sigma^2$, and $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}\sigma_e^2)$. The numerator (N) and

denominator (D) in Equation 31 are univariate random variables that can be written as quadratic forms:

$$N \equiv \mathbf{a}'(\mathbf{y} - \mu\mathbf{1}) = \mathbf{z}'\mathbf{Q}_N\mathbf{z} \tag{S1}$$

$$D \equiv (\mathbf{y} - \mu\mathbf{1})'(\mathbf{y} - \mu\mathbf{1}) = \mathbf{z}'\mathbf{Q}_D\mathbf{z}$$

where the stacked vector \mathbf{z} and matrices $\mathbf{Q}_N, \mathbf{Q}_D$ are

$$\mathbf{z} = \begin{bmatrix} \mathbf{a} \\ \mathbf{y} - \mu\mathbf{1} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{G} \\ \mathbf{G} & \mathbf{G} + \sigma_e^2\mathbf{I} \end{bmatrix}\right)$$

$$\mathbf{Q}_N = \frac{1}{2} \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \tag{S2}$$

$$\mathbf{Q}_D = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

Since $E[\mathbf{z}'\mathbf{Q}\mathbf{z}] = \text{tr}(\mathbf{Q}\mathbf{V})$ for $\mathbf{z} \sim N(\mathbf{0}, \mathbf{V})$ (Searle 1971), it follows that

$$E[N] = E[\mathbf{z}'\mathbf{Q}_N\mathbf{z}] = \frac{1}{2} \text{tr}\left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{G} \\ \mathbf{G} & \mathbf{G} + \sigma_e^2\mathbf{I} \end{bmatrix}\right) = \text{tr}(\mathbf{G})$$

$$E[D] = E[\mathbf{z}'\mathbf{Q}_D\mathbf{z}] = \text{tr}\left(\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{G} \\ \mathbf{G} & \mathbf{G} + \sigma_e^2\mathbf{I} \end{bmatrix}\right) = \text{tr}(\mathbf{G}) + n\sigma_e^2 \tag{S3}$$

Since the trace of \mathbf{G} equals $n(1+f)\sigma^2$, an approximate formula for heritability is

$$E[h^2] = E\left[\frac{N}{D}\right] \approx \frac{E[N]}{E[D]} = \frac{(1+f)\sigma^2}{(1+f)\sigma^2 + \sigma_e^2} \quad (\text{S4})$$

Equation S4 is only approximate because the expectation of the ratio of two random variables does not equal the ratio of expectations. By Taylor series expansion a more accurate formula is (Rice 2006)

$$E\left[\frac{N}{D}\right] = \frac{E[N]}{E[D]} + \frac{(\text{var}[D]E[N] - \text{cov}[N,D]E[D])}{E^3[D]} + \dots \quad (\text{S5})$$

To investigate the magnitude of the correction term, note that for $\mathbf{z} \sim N(\mathbf{0}, \mathbf{V})$,

$$\begin{aligned} \text{var}[\mathbf{z}'\mathbf{Q}_D\mathbf{z}] &= 2\text{tr}\left(\left(\mathbf{Q}_D\mathbf{V}\right)^2\right) \\ \text{cov}[\mathbf{z}'\mathbf{Q}_D\mathbf{z}, \mathbf{z}'\mathbf{Q}_N\mathbf{z}] &= 2\text{tr}\left(\mathbf{Q}_D\mathbf{V}\mathbf{Q}_N\mathbf{V}\right) \end{aligned} \quad (\text{S6})$$

(Searle 1971). Upon substituting S2 into S6 and simplifying, the result is

$$\begin{aligned} \text{var}[D] &= 2n\left[\sigma^4 n \langle A_{ij}^2 \rangle + 2\sigma^2 \sigma_e^2 (1+f) + \sigma_e^4\right] \\ \text{cov}[N, D] &= 2n\left[\sigma^4 n \langle A_{ij}^2 \rangle + \sigma^2 \sigma_e^2 (1+f)\right] \end{aligned} \quad (\text{S7})$$

where

$$\langle A_{ij}^2 \rangle = n^{-2} \sum_{i=1}^n \sum_{j=1}^n A_{ij}^2 \quad (\text{S8})$$

is the mean-squared element of \mathbf{A} . From Equation S7 the correction is

$$\frac{(\text{var}[D]E[N] - \text{cov}[N,D]E[D])}{E^3[D]} = \frac{2\sigma^4 \sigma_e^2 \left[(1+f)^2 - n \langle A_{ij}^2 \rangle \right]}{n \left[\sigma_e^2 + \sigma^2 (1+f) \right]^3} \stackrel{n \rightarrow \infty}{\approx} \frac{-2 \langle A_{ij}^2 \rangle \sigma^4 \sigma_e^2}{\left[\sigma_e^2 + \sigma^2 (1+f) \right]^3} \quad (\text{S9})$$

where the last expression holds for large populations.

Using Equation 13, the mean-squared realized relationship is

$$\begin{aligned}\langle A_{ij}^2 \rangle &= \left(2 \langle \tilde{p}_i \tilde{q}_i \rangle\right)^{-2} n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^{\tilde{m}} \sum_{k'=1}^{\tilde{m}} (\tilde{X}_{ik'} - 2\tilde{p}_{k'}) (\tilde{X}_{ik} - 2\tilde{p}_k) (\tilde{X}_{jk'} - 2\tilde{p}_{k'}) (\tilde{X}_{jk} - 2\tilde{p}_k) \\ &= \left(2 \langle \tilde{p}_i \tilde{q}_i \rangle\right)^{-2} \tilde{m}^{-2} \sum_k \sum_{k'} r_{kk'}^2 \tilde{p}_k \tilde{q}_k \tilde{p}_{k'} \tilde{q}_{k'}\end{aligned}\quad (\text{S10})$$

where

$$r_{kk'} = \left(\tilde{p}_k \tilde{q}_k \tilde{p}_{k'} \tilde{q}_{k'}\right)^{-1/2} n^{-1} \sum_{i=1}^n (\tilde{X}_{ik} - 2\tilde{p}_k) (\tilde{X}_{ik'} - 2\tilde{p}_{k'}) \quad (\text{S11})$$

is the correlation between loci. If the correlation length between causal loci is small compared to

the size of the genome, then $\langle A_{ij}^2 \rangle$ will also be small. For example, if the causal loci are in

linkage equilibrium, then

$$\langle A_{ij}^2 \rangle = \frac{\langle \tilde{p}_k^2 \tilde{q}_k^2 \rangle}{4\tilde{m} \langle \tilde{p}_i \tilde{q}_i \rangle^2} \quad (\text{S12})$$

More generally we can write $\langle A_{ij}^2 \rangle = c/\tilde{m}$ for some finite constant c , which upon substitution in

Equation S9 shows the correction term tends to zero under the infinitesimal model ($\tilde{m} \rightarrow \infty$).

Literature Cited

Rice, J., 2006 *Mathematical Statistics and Data Analysis*, 3rd ed. Duxbury, Pacific Grove.

Searle, S. R., 1971 *Linear Models*. John Wiley & Sons, New York.