File S1: Tables S1–S2, Figures S1–S3, and Supporting Methods

Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato

Endelman *et al.*

Table S1 Number of new vs. returning clones by year in the training population.

Year	No. new clones	No. returning clones
2012	78	0
2013	59	45
2014	82	58
2015	43	56
2016	128	52
2017	181	67

Table S2 Genetic variance estimates (and SE) for the G+GG+D model, as a proportion of total genetic variance.

	Yield	Specific Gravity	Fry Color
V_a	0.45 (0.13)	0.20 (0.10)	0.45 (0.17)
V_{aa}	0.03 (0.24)	0.51 (0.22)	0.44 (0.23)
V_d	0.07 (0.10)	0.00 (NA[a])	0.00 (NA)
V_r	0.45 (0.20)	0.29 (0.18)	0.12 (0.17)

[a] REML solution on the boundary

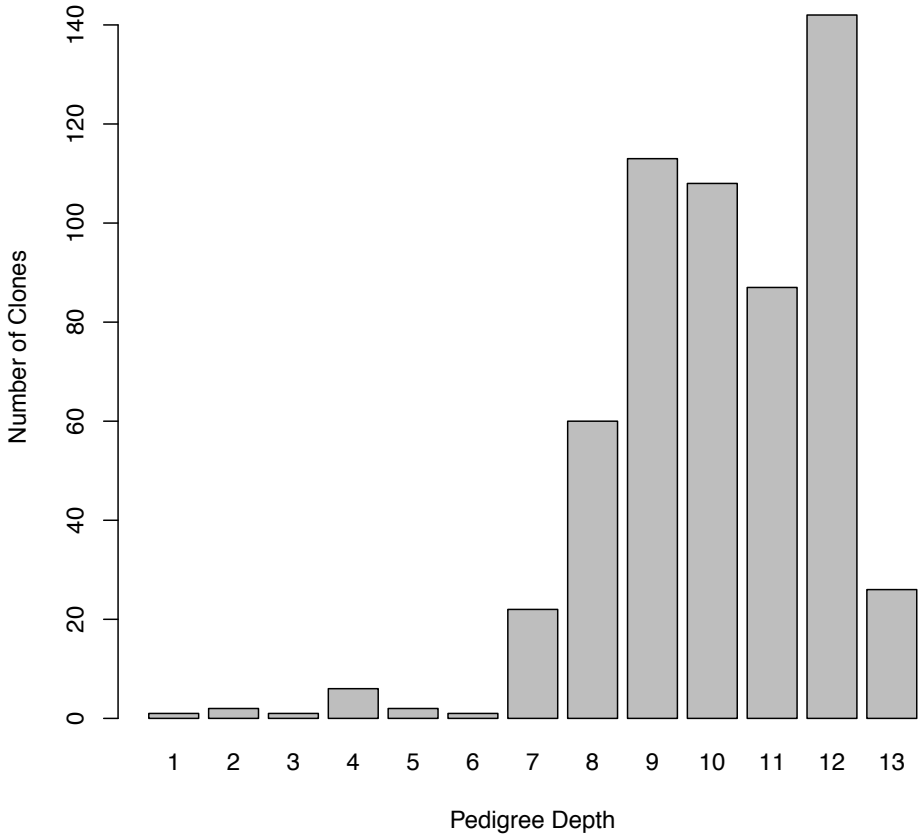

Figure S1 Distribution of pedigree depth for the training population of 571 clones. Pedigree depth is the maximum number of generations to a founder.

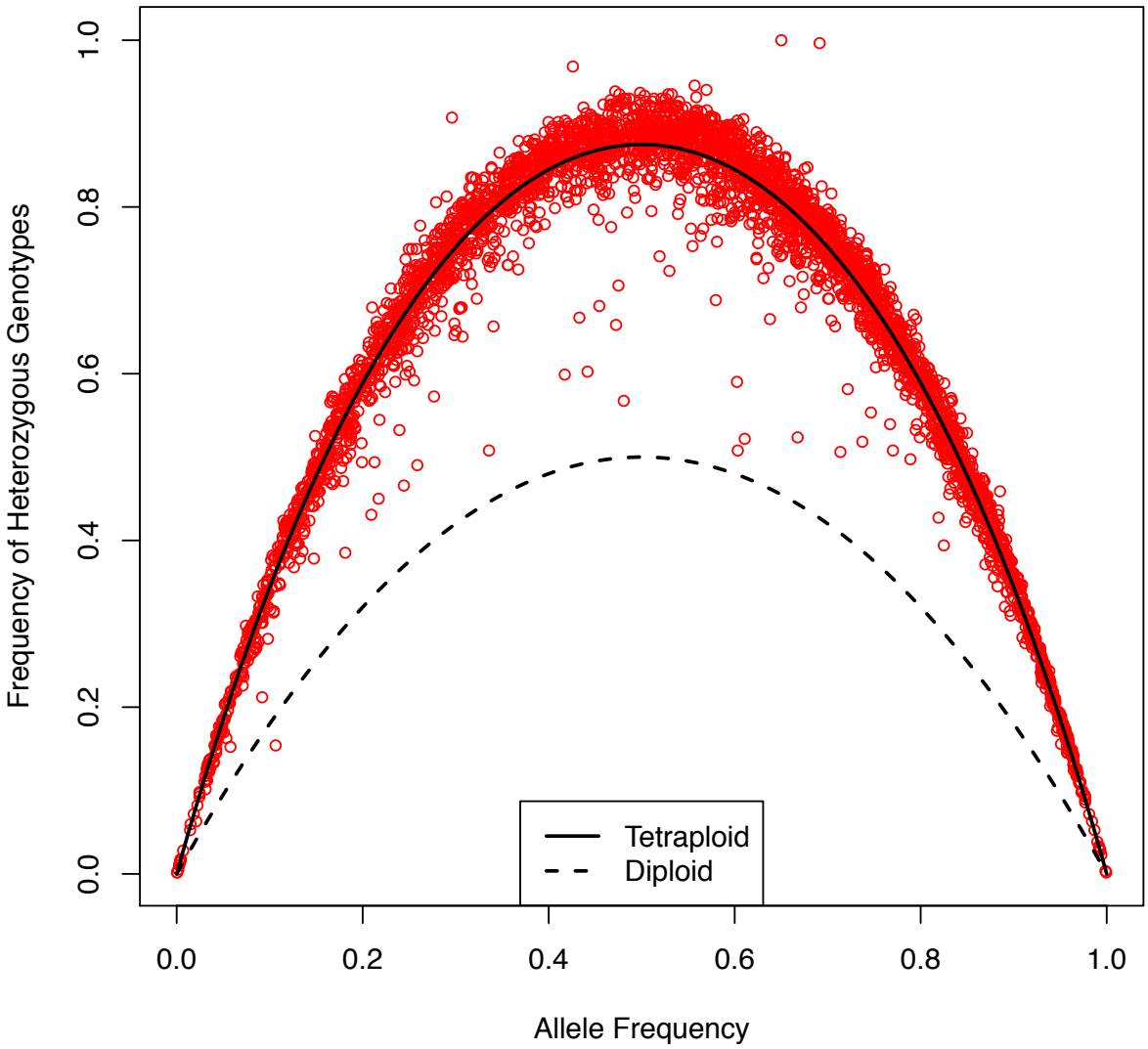

Figure S2 Comparison of the observed (red circle) vs. expected (black line) frequency of heterozygotes under random mating, for the 3895 SNPs used in this study. For allele frequency p (q = 1–p), the heterozygote frequency at panmictic equilibrium is $4p^3q + 6p^2q^2 + 4pq^3$ for tetraploids compared with $2pq$ for diploids (Gallais 2003).

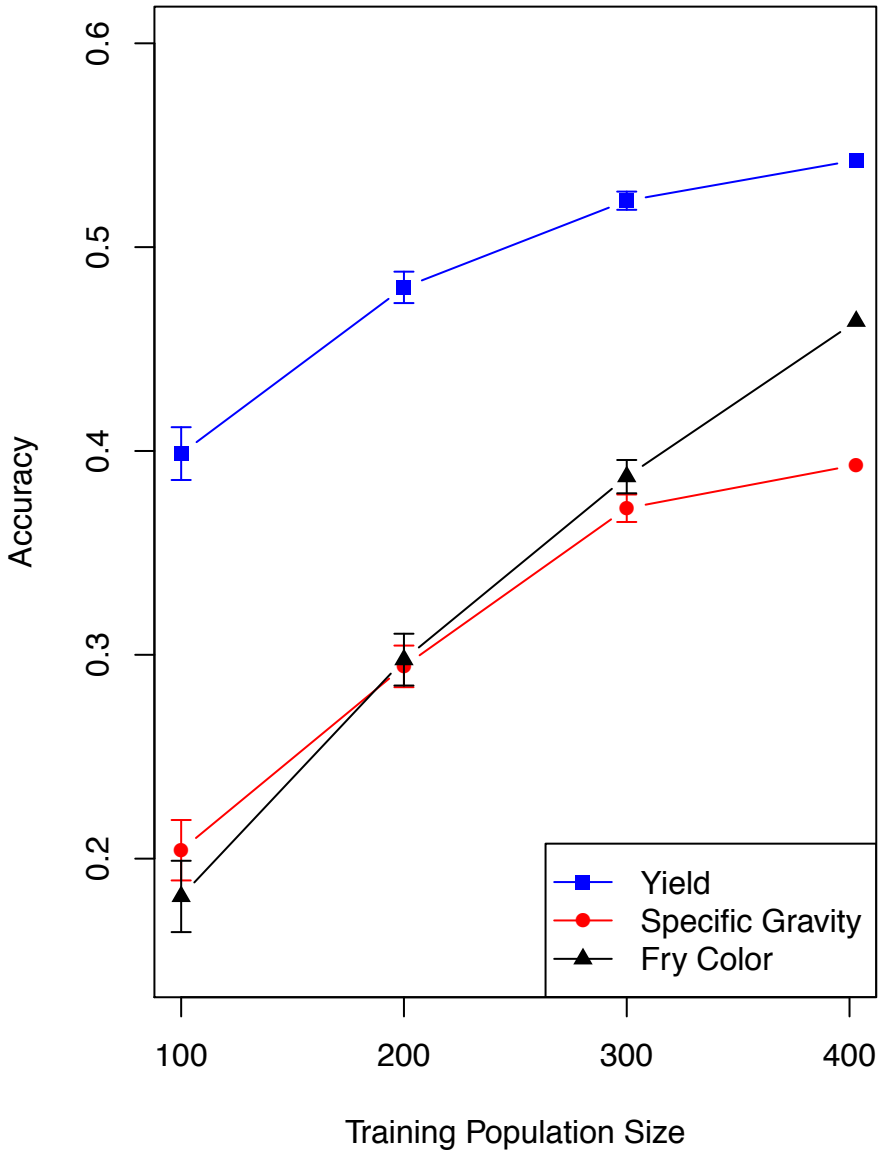

Figure S3 Effect of population size (N) on prediction accuracy within the training set. The points at N = 403 correspond to using all clones with pedigree depth < 12 to predict clones with pedigree depth ≥ 12, using the G+GG+D model. The points at N = 100, 200, 300 are the mean accuracy for 200 random subsets, and the error bars show ± 1 standard error.

Supporting Methods

Normal Equations for the Additive Effects

Taking the derivative of Eq. 2 with respect to an arbitrary allele *t*, and setting the result equal to zero:

$$\begin{aligned} 0 = -2\sum_{jkl} p_t p_j p_k p_l \left[z_{tjkl} - \left(\alpha_t + \alpha_j + \alpha_k + \alpha_l\right)\right] \\ - 2\sum_{ikl} p_i p_t p_k p_l [z_{itkl} - (\alpha_i + \alpha_t + \alpha_k + \alpha_l)] \\ - 2\sum_{ijl} p_i p_j p_t p_l \left[z_{ijtl} - \left(\alpha_i + \alpha_j + \alpha_t + \alpha_l\right)\right] \\ - 2\sum_{ijk} p_i p_j p_k p_t \left[z_{ijkt} - \left(\alpha_i + \alpha_j + \alpha_k + \alpha_t\right)\right] \end{aligned} \quad \text{(S1)}$$

Because genotypic value is unchanged upon permutation of the indices, each of the terms in Eq. S1 is identical, which leads to the result:

$$0 = \sum_{jkl} p_j p_k p_l \left[z_{tjkl} - \left(\alpha_t + \alpha_j + \alpha_k + \alpha_l\right)\right] \quad \text{(S2)}$$

Rearranging Eq. S2, and using the identity $\sum_j p_j = 1$, leads to

$$\sum_{jkl} p_j p_k p_l z_{tjkl} = \alpha_t + 3\sum_j p_j \alpha_j \quad \text{(S3)}$$

Multiplying Eq. S3 by p_t, summing over *t*, and using the interchangeability of the indices, generates

$$\sum_{tjkl} p_t p_j p_k p_l z_{tjkl} = 4\sum_j p_j \alpha_j \quad \text{(S4)}$$

From the definition of z (Eq. 1), the left side of Eq. S4 is zero, so the right side is also zero, which completes the proof of Eq. 3.

Normal Equations for the Digenic Dominance Effects

Taking the derivative of Eq. 5 with respect to an arbitrary allele pair (*s*,*t*), setting the result equal to zero, and using the interchangeability of the indices, generates

$$0 = \sum_{kl} p_k p_l [y_{stkl} - (\beta_{st} + \beta_{sk} + \beta_{sl} + \beta_{tk} + \beta_{tl} + \beta_{kl})] \quad \text{(S5)}$$

Upon rearranging and using the interchangeability of indices, Eq. S5 becomes

$$\sum_{kl} p_k p_l y_{stkl} = \beta_{st} + 2\sum_{k} \beta_{sk} p_k + 2\sum_{k} \beta_{tk} p_k + \sum_{kl} \beta_{kl} p_k p_l \quad \text{(S6)}$$

The left side of this equation equals $\bar{g}_{st\cdot\cdot} - \mu - \alpha_s - \alpha_t$. Multiplying by p_t and summing over t (at fixed s), the left side becomes zero, and the right side becomes

$$0 = 3\sum_{k} \beta_{sk} p_k + 3\sum_{kl} \beta_{kl} p_k p_l \quad \text{(S7)}$$

Multiplying Eq. S7 by p_s and summing over s leads to

$$0 = 6\sum_{k} \beta_{sk} p_s p_k \quad \text{(S8)}$$

Eq. S8 shows that the second term in Eq. S7 is zero, which implies the first term is also zero, which completes the proof of Eq. 6.

Derivation of Eq. 12

For a bi-allelic locus, the constraint Eq. 6 is a system of two equations with three digenic dominance parameters:

$$\begin{aligned} p\beta_{BB} + q\beta_{Bb} &= 0 \\ p\beta_{Bb} + q\beta_{bb} &= 0 \end{aligned} \tag{S9}$$

To uniquely determine the solution, we introduce the parameter $\beta \equiv \beta_{BB} - 2\beta_{Bb} + \beta_{bb}$, which in combination with Eq. S9 produces the linear system:

$$\begin{bmatrix} 1 & -2 & 1 \\ p & q & 0 \\ 0 & p & q \end{bmatrix} \begin{bmatrix} \beta_{BB} \\ \beta_{Bb} \\ \beta_{bb} \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \\ 0 \end{bmatrix} \tag{S10}$$

Using Gaussian elimination, Eq. S10 can be reduced to

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & q+2p & -p \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_{BB} \\ \beta_{Bb} \\ \beta_{bb} \end{bmatrix} = \begin{bmatrix} \beta \\ -p\beta \\ -pq\beta \end{bmatrix} \tag{S11}$$

and back-substitution produces Eq. 12.

Derivation of Eq. 13

Using Eq. 12, the total digenic dominance v_X for a clone with dosage X of the B allele is

$$\begin{aligned} v_4 &= 6\beta_{BB} = 6q^2\beta \\ v_3 &= 3\beta_{BB} + 3\beta_{Bb} = (3q^2 - 3pq)\beta \\ v_2 &= 1\beta_{BB} + 4\beta_{Bb} + 1\beta_{bb} = (q^2 - 4pq + p^2)\beta \\ v_1 &= 3\beta_{Bb} + 3\beta_{bb} = (-3pq + 3p^2)\beta \\ v_0 &= 6\beta_{bb} = 6p^2\beta \end{aligned} \tag{S12}$$

Replacing all instances of q with $1-p$, Eq. S12 becomes

$$\begin{aligned} v_4 &= (6p^2 - 12p + 6)\beta \\ v_3 &= (6p^2 - 9p + 3)\beta \\ v_2 &= (6p^2 - 6p + 1)\beta \\ v_1 &= (6p^2 - 3p)\beta \\ v_0 &= 6p^2\beta \end{aligned} \tag{S13}$$

which is equivalent to Eq. 13.