Supporting Information

Xu et al. 10.1073/pnas.1413750111

SI Text

Converting Polygenic Effects into Marker Effects. ***Conditional mean approach.*** Let us reintroduce the polygenic model here,

$$y = X\beta + \xi + \varepsilon, \quad \text{[S1]}$$

where

$$\xi = \sum_{k=1}^{m} Z_k \gamma_k = Z\gamma \quad \text{[S2]}$$

is the polygenic effect. We replaced the summation by a compact form of matrix multiplication. The joint distribution of ξ and γ is multivariate normal with mean and variance given by

$$\mathrm{E}\begin{bmatrix}\xi\\ \gamma\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \quad \text{[S3]}$$

and

$$\mathrm{var}\begin{bmatrix}\xi\\ \gamma\end{bmatrix} = \begin{bmatrix}\mathrm{var}(\xi) & \mathrm{cov}(\xi,\gamma)\\ \mathrm{cov}(\gamma,\xi) & \mathrm{var}(\gamma)\end{bmatrix} = \begin{bmatrix}ZZ^T\phi^2/m & Z\phi^2/m\\ Z^T\phi^2/m & I_m\phi^2/m\end{bmatrix}, \quad \text{[S4]}$$

respectively. The converting formula is the conditional expectation of γ given ξ, which is

$$\begin{aligned}\mathrm{E}(\gamma|\xi) &= \mathrm{E}(\gamma) + \mathrm{cov}(\gamma,\xi)\mathrm{var}^{-1}(\xi)[\xi - \mathrm{E}(\xi)]\\ &= (Z^T\phi^2/m)(ZZ^T\phi^2/m)^{-1}\xi\end{aligned} \quad \text{[S5]}$$

The predicted genomic value is the conditional expectation of ξ given y,

$$\begin{aligned}\mathrm{E}(\xi|y) &= G(G+R)^{-1}(y - X\beta)\\ &= K\phi^2(K\phi^2 + I_n\sigma^2)^{-1}(y - X\beta)\\ &= (ZZ^T\phi^2/m)(ZZ^T\phi^2/m + I_n\sigma^2)^{-1}(y - X\beta)\\ &= (ZZ^T\phi^2)(ZZ^T\phi^2 + mI_n\sigma^2)^{-1}(y - X\beta)\\ &= (ZZ^T\phi^2/\sigma^2)(ZZ^T\phi^2/\sigma^2 + mI_n)^{-1}(y - X\beta)\end{aligned} \quad \text{[S6]}$$

Substituting ξ in Eq. **S5** by this conditional expectation in Eq. **S6** yields

$$\begin{aligned}\mathrm{E}(\gamma|\xi) &= (Z^T\phi^2/m)(ZZ^T\phi^2/m)^{-1}(ZZ^T\phi^2/\sigma^2)\\ &\quad\times(ZZ^T\phi^2/\sigma^2 + mI_n)^{-1}(y - X\beta)\\ &= (Z^T\phi^2/\sigma^2)(ZZ^T\phi^2/m)^{-1}(ZZ^T\phi^2/m)\\ &\quad\times(ZZ^T\phi^2/\sigma^2 + mI_n)^{-1}(y - X\beta)\\ &= (Z^T\phi^2/\sigma^2)(ZZ^T\phi^2/\sigma^2 + mI_n)^{-1}(y - X\beta)\\ &= \lambda Z^T(\lambda ZZ^T + mI_n)^{-1}(y - X\beta),\end{aligned} \quad \text{[S7]}$$

where $\lambda = \phi^2/\sigma^2$ and I_n is an identity matrix with order n. Therefore, the estimated marker effects are

$$\hat{\gamma} = \mathrm{E}(\gamma|\xi) = \hat{\lambda}Z^T\left(\hat{\lambda}ZZ^T + mI_n\right)^{-1}\left(y - X\hat{\beta}\right). \quad \text{[S8]}$$

Replacing m by

$$c_a = \mathrm{mean}\left[\mathrm{diag}\left(ZZ^T\right)\right] \quad \text{[S9]}$$

and noting that $K_a = ZZ^T/c_a$, we have an alternative expression for $\hat{\gamma}$,

$$\hat{\gamma} = \mathrm{E}(\gamma|\xi) = \frac{1}{c_a}\hat{\lambda}Z^T\left(\hat{\lambda}K_a + I_n\right)^{-1}\left(y - X\hat{\beta}\right) \quad \text{[S10]}$$

We can take advantage of the eigen-decomposition to solve the inverse involved in Eq. **S10**. This concludes the derivation of Eq. **5** in the main text.

Ridge regression approach. Alternatively, we can use the ridge regression approach (1) to deriving the converting equation, provided that the ridge parameter is defined as the variance ratio

$$\eta = \sigma^2/\left(\phi^2/m\right) = m\sigma^2/\phi^2 = m\lambda^{-1}. \quad \text{[S11]}$$

The ridge regression coefficients for the following model

$$y - X\beta = Z\gamma + \varepsilon \quad \text{[S12]}$$

is

$$\begin{aligned}\gamma &= \left(Z^TZ + \eta I_m\right)^{-1}Z^T(y - X\beta)\\ &= \left(Z^TZ + mI_m\lambda^{-1}\right)^{-1}Z^T(y - X\beta)\\ &= \lambda\left(\lambda Z^TZ + mI_m\right)^{-1}Z^T(y - X\beta)\end{aligned} \quad \text{[S13]}$$

Therefore, the estimated marker effects are

$$\hat{\gamma} = \hat{\lambda}\left(\hat{\lambda}Z^TZ + mI_m\right)^{-1}Z^T\left(y - X\hat{\beta}\right). \quad \text{[S14]}$$

Placing the two converting equations together, we have

$$\hat{\gamma} = \begin{cases}\hat{\lambda}Z^T\left(\hat{\lambda}ZZ^T + mI_n\right)^{-1}\left(y - X\hat{\beta}\right) & \text{conditional mean}\\ \hat{\lambda}\left(\hat{\lambda}Z^TZ + mI_m\right)^{-1}Z^T\left(y - X\hat{\beta}\right) & \text{ridge regression}\end{cases} \quad \text{[S15]}$$

These two equations appear to have different forms but they give the same estimated marker effects. For the conditional mean approach, the matrix to be inverted has a dimension $n \times n$, whereas the matrix to be inverted for the ridge regression approach has a dimension $m \times m$. For computational convenience, the conditional mean approach should be used if $m > n$; otherwise, the ridge regression approach should be used.

We used both equations to convert the estimated genomic values into marker effects and they produced the same solutions. The converted marker effects are given in Dataset S2 for the four traits. For trait 1,000 grain weight (KGW), the large peaks of the estimated effect profile for GBLUP are consistent with the other two methods, although they are in quite different scales. The SSVS method has shrunken all small effects into zero and leaves only three large effects in the model. The LASSO method shows more nonzero effects. The GBLUP method shows small effects all of over the genome. Interestingly, the predictabilities of the three methods are very similar, regardless of the differences in the estimated marker effects. This result further supports the notion that accurate estimation of individual marker effects is not crucial to genomic selection. It is the pattern of the genetic effects distributed along the genome that plays an important role.

Restricted Maximum-Likelihood Method Incorporating Epistasis. For the multiple variance component model, the fast eigen-decomposition algorithm cannot be used here. Instead, the likelihood function must be evaluated in its original form. Once the Newton–Raphson iteration is converged, we get the estimates of variance components and the Hessian matrix as a by-product. The Hessian matrix then provides an approximate variance–covariance matrix using

$$\mathrm{var}\left(\hat{\theta}\right) \approx -\left[\frac{\partial^2 L\left(\hat{\theta}\right)}{\partial\theta\ \partial\theta^T}\right]^{-1} \qquad \textbf{[S16]}$$

This is a 7×7 covariance matrix whose elements are given below:

$$\mathrm{var}\left(\hat{\theta}\right) = \begin{bmatrix} \mathrm{var}\left(\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_a,\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2_a\right) & \mathrm{var}\left(\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_d,\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2_d\right) & \mathrm{var}\left(\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{aa},\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2_{aa}\right) & \mathrm{var}\left(\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{dd},\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2_{dd}\right) & \mathrm{var}\left(\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{ad},\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2_{ad}\right) & \mathrm{var}\left(\hat{\sigma}^2_{da}\right) & \mathrm{cov}\left(\hat{\sigma}^2_{da},\hat{\sigma}^2\right) \\ \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_a\right) & \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_d\right) & \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_{aa}\right) & \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_{dd}\right) & \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_{ad}\right) & \mathrm{cov}\left(\hat{\sigma}^2,\hat{\sigma}^2_{da}\right) & \mathrm{var}\left(\hat{\sigma}^2\right) \end{bmatrix} \qquad \textbf{[S17]}$$

Square roots of the diagonal elements are the SEs of the estimated parameters, denoted by a column vector $\mathrm{StdErr}(\hat{\theta}) = \sqrt{\mathrm{diag}[\mathrm{var}(\hat{\theta})]}$. The correlation matrix (standardized covariance matrix) was obtained using

$$r\left(\hat{\theta}\right) = \left\{\mathrm{diag}\left[\mathrm{var}\left(\hat{\theta}\right)\right]\right\}^{-\frac{1}{2}} \mathrm{var}\left(\hat{\theta}\right) \left\{\mathrm{diag}\left[\mathrm{var}\left(\hat{\theta}\right)\right]\right\}^{-\frac{1}{2}} \qquad \textbf{[S18]}$$

The model with six genetic variance components is the full model. Various reduced models were also evaluated. For example, if only the additive variance is included, the model is called the additive model or model 1 with a model size 1. The dominance model includes both the additive and dominance variances and is thus called model 2. The full model is called model 6. The model number represents the model size. Table S2 lists all six models evaluated in this study.

We wrote an R program to perform the REML variance component analysis. Results of the R analysis were confirmed by PROC MIXED in SAS (2). The estimated variance components, including all genetic variance components and the residual variance, are given in Table S2. We can see that as the model size increases, the estimated variance component of each type declines; this indicates that when more variance components are included in a single model, the total variation tends to be shared by all components. Reduction of the residual variance as the model grows is even more obvious. In the end, the residual variance becomes extremely small compared with the phenotypic variance.

Under the full model (model 6), yield (YIELD) had no main effects. Among the four types of epistatic effects, additive × dominance seemed to play a more important role. However, under models 1 and 2 where epistatic effects were excluded, neither the additive nor the dominance variance was zero. The additive and dominance variances actually captured part of the other types of genetic variances. Epistatic variances also contributed more than the additive and dominance variances for trait tiller number (TILLER). Each of the additive and epistatic variances contributed about half of the total genetic variance. The last trait, KGW, was almost exclusively controlled by the additive variance.

Fig. 1 *Upper* shows the goodness of fit for each trait under all six models. Clearly, adding epistatic variances had improved the goodness of fit for all traits. In the end, the full model fitted perfectly to the data for all traits. The perfect goodness of fit does not mean that all traits are 100% controlled by genetics. The perfect fit is an artifact caused by the model overfitting the data. However, the relative contribution of each component to the total genetic variance may be meaningful. For example, trait KGW is primarily controlled by the additive variance, which agrees with common knowledge of this trait. The predictability drawn from fivefold cross-validation analysis is much more reliable than the goodness of fit to evaluate the total genetic contribution. In the cross-validation analysis, lines predicted do not contribute to parameter estimation and thus will correct model overfitting. Fig. 1 *Lower* shows the predictability for each trait under each of the six models. Interestingly, none of the traits shows any improvement of predictability as the model grows. The simple additive model performed as well as the full model. For the purpose of predicting genomic values, there is no benefit to add epistatic variances. Of course, our sample size is small and thus this conclusion may only apply to small sample sizes.

There are at least two reasons to explain why adding more variances did not improve the predictability. First, the estimated parameters (variance components) may be highly correlated between different types of effects. We examined the correlations between the estimated variance components. Some of the variance components are indeed highly correlated. Table S6 lists the correlation matrices for all traits under the full model (model 6). We have observed some common patterns from the table. The additive variance often has very little correlation with the dominance variance, which means that these two parameters are separable. However, the additive variance is always correlated to the residual variance. The dominance variance has extremely high correlation with the dominance × dominance variance. The high correlation coefficients mean that the estimated parameters are not stable, especially when the model is large. Because of the high correlations, the variance components of small models can usually capture the variances excluded from

the models. Therefore, not much information had been actually lost for the small models. We now examine the additive variance and residual variance because they were estimated from all six models. Fig. S1 shows the plot of the estimated parameter (additive variance) against the model size. The estimated variance is labeled Estimate and the SE of the estimate is labeled StdErr. The estimated variance sharply decreases as the model size increases, but the SE has increased as the model grows. The same trends were found for the estimated residual variances, as shown in Fig. S2. Therefore, the high correlations and large SEs for estimated parameters (variance components) contribute to the lack of improvement by adding epistatic variances to the model. Increasing sample size is necessary to reduce the correlations and the SEs.

The next question is why some parameters tend to have high correlations than other parameters. We examined the similarities between all pairwise kinship matrices, a total of $6(6-1)/2=15$ comparisons. We used the Pearson correlation coefficient to measure the similarity between a pair of kinship matrices. The correlation matrix is given in Table S4. The correlation coefficients do vary across different pairs of comparisons. Matrices K_a and K_d have a low correlation. The highest correlation occurs between K_d and K_{dd}. Our conclusion is that the high correlations of the estimated parameters (variance components) are also caused by the high similarities between the kinship matrices. Increasing sample sizes does not help to reduce the similarity between pair of kinship matrices.

1. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67.

2. SAS Institute Inc (2009) *SAS/STAT: Users' Guide, Version 9.3* (SAS, Cary, NC).

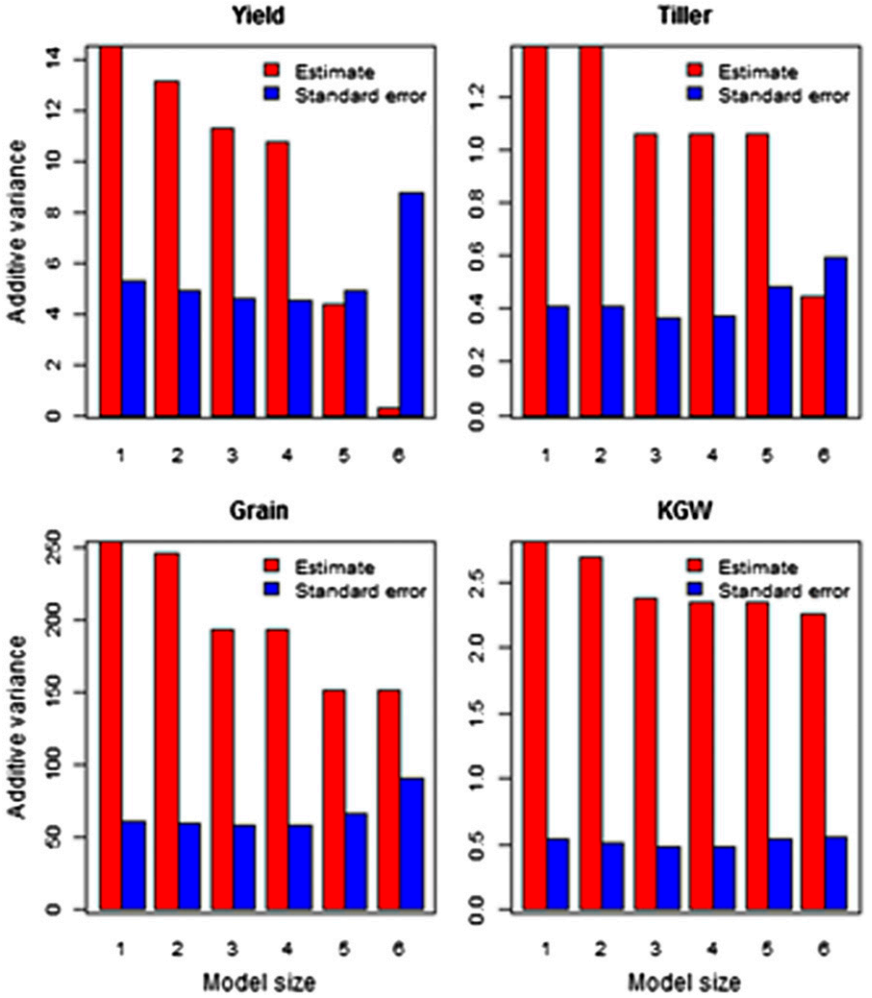

Fig. S1. Estimated additive variance and its SE plotted against model size.

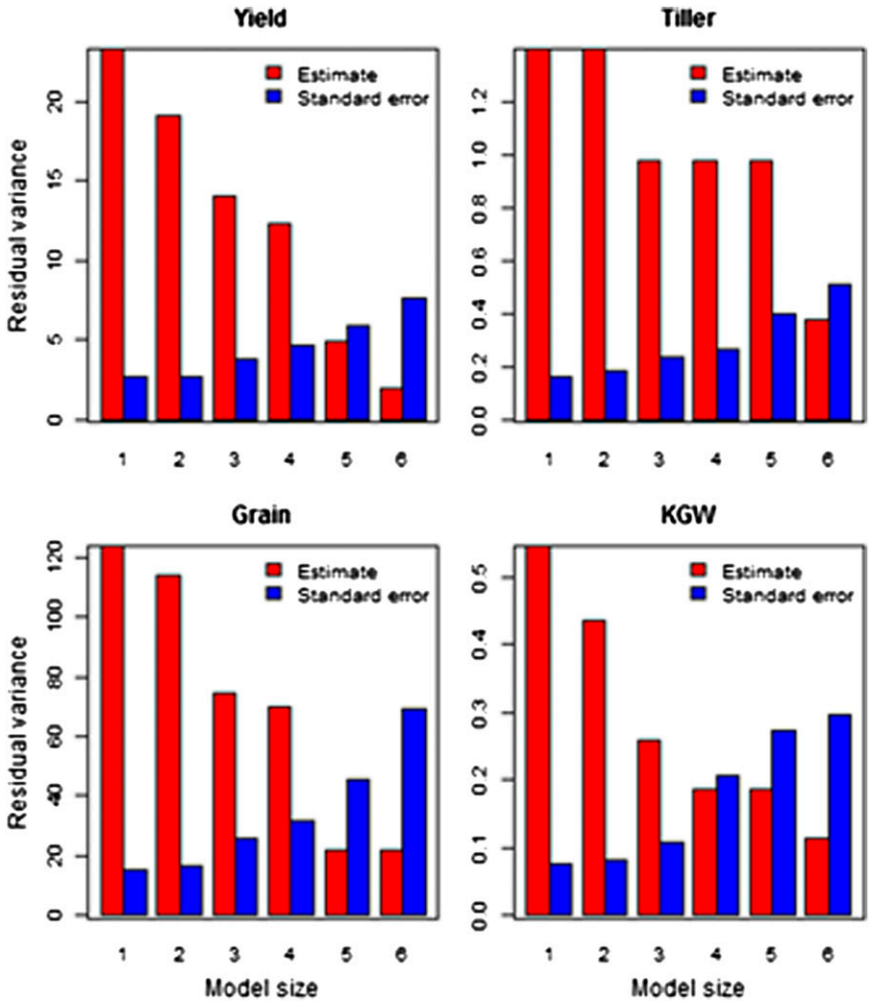

Fig. S2. Estimated residual variance and its SE plotted against model size.

Table S1. Six different models evaluated in the study, where an "X" in a cell means that the corresponding variance component (column header) is included in that model

Model	σ^2_a	σ^2_d	σ^2_{aa}	σ^2_{dd}	σ^2_{ad}	σ^2_{da}	σ^2
1	X						X
2	X	X					X
3	X	X	X				X
4	X	X	X	X			X
5	X	X	X	X	X		X
6	X	X	X	X	X	X	X

Table S2. Variance components estimated from six polygenic models for the four quantitative traits of rice

Trait	Model	a	d	aa	dd	ad	da	e
YIELD	1	14.4911						23.3308
	2	13.1129	9.0443					19.2088
	3	11.2609	7.73516	7.38906				14.0566
	4	10.7365	0.00001	7.80710	7.23373			12.2930
	5	4.32190	0.00001	7.27218	5.442035	16.7565		4.90684
	6	0.00001	0.00001	7.00964	5.271305	18.2771	5.54845	1.96303
TILLER	1	1.38792						1.39981
	2	1.38792	0.00001					1.39981
	3	1.05890	0.00001	0.61650				0.97550
	4	1.05895	0.00001	0.61645	0.00001			0.97549
	5	1.05896	0.00001	0.61650	0.00001	0.00001		0.97545
	6	0.44516	0.00001	0.58511	0.00001	0.12044	1.13048	0.37376
GRAIN	1	254.636						124.165
	2	245.599	24.3225					113.847
	3	193.179	14.9761	69.4831				74.4618
	4	192.990	0.00001	68.5104	17.41830			69.7084
	5	150.877	0.00001	66.8473	6.55935	110.184		21.5337
	6	150.913	0.00001	66.8373	6.579376	110.180	0.00001	21.5147
KGW	1	2.82000						0.54720
	2	2.69618	0.26283					0.43694
	3	2.38064	0.21624	0.32088				0.25818
	4	2.34714	0.00001	0.32697	0.245812			0.18642
	5	2.34717	0.00001	0.32699	0.245813	0.00001		0.18641
	6	2.26561	0.00001	0.31201	0.227324	0.00001	0.18803	0.11309

Table S3. Estimated variance components (estimate), the SEs (StdErr), and the correlation coefficient matrix $r(\hat{\theta})$ of the estimated parameters under the full model (model 6)

Trait		a	d	aa	dd	ad	da	e	Estimate	StdErr
YIELD	a	1	0.029	−0.032	0.014	−0.548	−0.799	0.711	0.000	8.765
	d	0.029	1	−0.168	−0.916	−0.041	0.005	0.265	0.000	10.707
	aa	−0.032	−0.168	1	0.109	−0.077	−0.090	−0.293	7.010	4.798
	dd	0.014	−0.916	0.109	1	−0.058	−0.031	−0.242	5.271	9.253
	ad	−0.548	−0.041	−0.077	−0.058	1	0.246	−0.677	18.277	10.277
	da	−0.799	0.005	−0.090	−0.031	0.246	1	−0.634	5.548	9.406
	e	0.711	0.265	−0.293	−0.242	−0.677	−0.634	1	1.963	7.655
TILLER	a	1	−0.164	−0.081	0.140	−0.708	−0.520	0.714	0.445	0.592
	d	−0.164	1	−0.008	−0.864	0.142	−0.041	0.086	0.000	0.476
	aa	−0.081	−0.008	1	−0.142	−0.189	−0.076	−0.121	0.585	0.337
	dd	0.140	−0.864	−0.142	1	−0.079	0.028	−0.177	0.000	0.494
	ad	−0.708	0.142	−0.189	−0.079	1	0.172	−0.670	0.120	0.675
	da	−0.520	−0.041	−0.076	0.028	0.172	1	−0.665	1.130	0.648
	e	0.714	0.086	−0.121	−0.177	−0.670	−0.665	1	0.374	0.507
GRAIN	a	1	0.117	−0.241	−0.014	−0.481	−0.686	0.724	150.913	90.846
	d	0.117	1	−0.079	−0.936	−0.077	−0.188	0.369	0.000	64.553
	aa	−0.241	−0.079	1	0.011	−0.020	−0.046	−0.253	66.837	41.571
	dd	−0.014	−0.936	0.011	1	−0.063	0.090	−0.252	6.579	59.507
	ad	−0.481	−0.077	−0.020	−0.063	1	0.231	−0.655	110.180	84.790
	da	−0.686	−0.188	−0.046	0.090	0.231	1	−0.759	0.000	104.073
	e	0.724	0.369	−0.253	−0.252	−0.655	−0.759	1	21.515	69.356
KGW	a	1	0.200	−0.117	−0.200	−0.434	−0.228	0.481	2.266	0.553
	d	0.200	1	−0.165	−0.969	−0.118	−0.069	0.653	0.000	0.646
	aa	−0.117	−0.165	1	0.128	−0.202	−0.123	−0.156	0.312	0.182
	dd	−0.200	−0.969	0.128	1	0.134	0.010	−0.662	0.227	0.639
	ad	−0.434	−0.118	−0.202	0.134	1	−0.153	−0.506	0.000	0.422
	da	−0.228	−0.069	−0.123	0.010	−0.153	1	−0.446	0.188	0.383
	e	0.481	0.653	−0.156	−0.662	−0.506	−0.446	1	0.113	0.296

Table S4. Correlation matrix between six different kinship matrices (K)

	Ka	Kd	Kaa	Kdd	Kad	Kda
Ka	1	0.1501	0.6056	0.2078	0.8441	0.8439
Kd	0.1501	1	0.2565	0.9751	0.2681	0.2649
Kaa	0.6056	0.2565	1	0.3552	0.7278	0.7225
Kdd	0.2078	0.9751	0.3552	1	0.3652	0.3628
Kad	0.8441	0.2681	0.7278	0.3652	1	0.7382
Kda	0.8439	0.2649	0.7225	0.3628	0.7382	1

Table S5. Spearman rank correlation coefficients of predicted genomic values between three methods of genomic prediction

Trait		LASSO	SSVS	GBLUP
YIELD	LASSO	1	0.8105	0.9124
	SSVS	0.8105	1	0.65439
	GBLUP	0.9124	0.65439	1
TILLER	LASSO	1	0.89106	0.92744
	SSVS	0.89106	1	0.99327
	GBLUP	0.92744	0.99327	1
GRAIN	LASSO	1	0.9171	0.96796
	SSVS	0.9171	1	0.88387
	GBLUP	0.96796	0.88387	1
KGW	LASSO	1	0.98161	0.98387
	SSVS	0.98161	1	0.97939
	GBLUP	0.98387	0.97939	1

Other Supporting Information Files

Dataset S1 (XLXS)
Dataset S2 (XLXS)
Dataset S3 (DOC)