RESEARCH

Genomewide Selection when Major Genes Are Known

Rex Bernardo*

ABSTRACT

Current methods for genomewide selection do not distinguish between known major genes and random genomewide markers. My objectives were to determine if explicitly modeling the effects of known major genes affects the response to genomewide selection, and to identify situations in which considering major genes as having fixed effects is helpful. Simulation experiments showed that having a fixed effect for a major gene became more advantageous as the percentage of genetic variance (V_G) explained by a major gene (R^2) increased and as the heritability on an entry-mean basis (h^2) increased. With R^2 = 50% and h^2 = 0.80, the relative efficiency (based on selection gains in Cycle 4) with a major gene having a fixed versus random effect was 112–121%. Specifying a fixed effect for a single major gene was never disadvantageous except with R^2 < 10%. With $h^2 \geq 0.50$, specifying a fixed versus random effect for a single major gene had little effect on prediction accuracy in Cycle 0. However, prediction accuracy in later cycles declined more rapidly when a major gene had a random effect instead of a fixed effect. The results with L = 2 or 3 major genes were similar to those with one major gene. In contrast, the usefulness of gene information was low with L = 10 major genes. Overall, major genes should be fitted as having fixed effects in genomewide selection when only a few major genes are present and each major gene accounts for ≥10% of V_G.

Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108. Received 16 May 2013. *Corresponding author (bernardo@umn.edu).

Abbreviations: DGAT, diacylglycerol acyltransferase; QTL, quantitative trait loci; RR-BLUP, ridge regression–best linear unbiased prediction.

The best ways of using molecular markers in selection largely depend on the genetic architecture of the trait (Bernardo, 2008). Some simpler traits, such as disease resistance or plant composition, may have major genes or quantitative trait loci (QTL) that account for a large percentage of the variation for the trait. Examples of major genes or QTL are the diacylglycerol acyltransferase (DGAT) gene for kernel oil concentration in maize (*Zea mays* L.) (Zheng et al., 2008), glutenin genes for dough quality in wheat (*Triticum aestivum* L.) (Weegels et al., 1996; Eagles et al., 2002), and several QTL for resistance to cyst nematode (*Heterodera glycines* Ichinohe) in soybean [*Glycine max* (L.) Merrill] (Concibido et al., 2004). For such traits, a useful breeding approach is to find markers for the major genes or QTL, confirm their effects in different genetic backgrounds, and widely introgress the validated major genes or QTL across elite germplasm (Bernardo, 2010).

But for more complex traits, such as grain yield in elite germplasm, the preferred approach that has emerged is to bypass QTL mapping altogether and to instead identify the best individuals in a population by genomewide selection (or genomic selection) (Meuwissen et al., 2001). Unlike QTL mapping, genomewide selection does not involve finding markers with significant effects on the trait but instead uses a large set of random, genomewide markers to predict performance. Genomewide prediction equations are developed from a training population that has been genotyped and

Published in Crop Sci. 54:68–75 (2014).
doi: 10.2135/cropsci2013.05.0315

phenotyped, and the prediction equations are used to select candidates in a test population that has been genotyped but not phenotyped.

Even when major genes or QTL are present, a substantial portion of the genetic variance (V_G) for the trait may be due to unknown background QTL with minor effects. For such traits, introgressing only the major genes or QTL will fail to capture the effects of minor QTL. On the other hand, current genomewide selection approaches do not explicitly model the effects of major genes or QTL versus background QTL. Marker-based selection methods that lead to genetic gains at both known major genes or QTL and unknown minor QTL would be useful.

Results in plants have shown that ridge regression–best linear unbiased prediction (RR-BLUP) is useful for obtaining genomewide predictions (Lorenzana and Bernardo, 2009; Heffner et al., 2009; Lorenz et al., 2011; Guo et al., 2012; Schulz-Streeck et al., 2012). For a given total number of markers (N_M), RR-BLUP assumes that each marker accounts for $(1/N_M)^{th}$ of V_G. Now suppose that one of the N_M markers corresponds to a known major gene. In this situation, the effect of the major gene is naturally included in the RR-BLUP model and genomewide selection will lead to gains at both the known major gene and the unknown background QTL. However, the assumption of a common variance for the known major gene and for each of the remaining $N_M - 1$ markers leads to an underestimation (i.e., overshrinkage towards zero) of the estimated effect of the major gene. Such underestimation may affect the response to several cycles of genomewide selection (Combs and Bernardo, 2013).

A straightforward alternative is to model any known major genes or QTL as having fixed effects and the unknown minor QTL as having random effects in RR-BLUP (Hayr et al., 2013). Suppose a number of major genes (L) are known to control a trait. Three approaches that differ in the number of major genes with fixed effects (K) can then be used. First, the L major genes may not be given any special treatment and their effects are then modeled as random effects, along with those of genomewide markers, in RR-BLUP ($K = 0$). Second, all L major genes may be specified as having fixed effects in the genomewide prediction model ($K = L$). Third, if some major genes are known to be more important than others, a subset of major genes with the largest effects can be specified as having fixed effects and the remaining major genes with smaller effects are not given any special treatment in RR-BLUP ($K < L$).

The usefulness of these three approaches for incorporating information on known major genes in genomewide selection has not been reported. My objectives in this study were to determine if explicitly modeling the effects of known major genes affects the response to genomewide selection, and to identify situations in which considering major genes as having fixed effects is helpful.

MATERIALS AND METHODS

Simulation Experiments

Each simulation experiment comprised a combination of L, K, percentage of V_G explained by a major gene (R^2), heritability on an entry-mean basis (h^2), and population size (N) used during genomewide selection. Along with the L major genes, 100 QTL with minor effects also controlled the trait. Genotypic values were defined relative to testcross performance, as appropriate for maize.

The scheme for genomewide selection was the same as that simulated by Bernardo and Yu (2007) and studied empirically in maize by Massman et al. (2013) and Combs and Bernardo (2013): a Cycle 0 population was phenotyped and genotyped, and the genomewide prediction equations developed in Cycle 0 were applied to several cycles of selection based on genomewide markers. In this study, each simulation experiment was repeated 1000 times. Each repeat differed in the location of the 100 minor QTL and in the genotypes, genotypic values, and phenotypic values at each cycle of genomewide selection.

Major Genes, Minor QTL, and Molecular Markers

The total number of major genes was $L = 1$, 2, 3, or 10. With $L = 1$, the value of R^2 was 1/2 (50%), 1/3 (33%), 1/4 (25%), 1/6 (17%), 1/10 (10%), and 1/20 (5%). With $L = 2$, the R^2 values were 33% for the first major gene and 17% for the second major gene. With $L = 3$, the R^2 values were 25% for the first major gene, 17% for the second major gene, and 8% for the third major gene. With $L = 10$, each major gene had $R^2 = 5\%$. With $L > 2$, the major genes therefore jointly accounted for 50% of V_G.

The L major genes and 100 minor QTL were located, along with $N_M = 438$ marker loci, on 10 chromosomes that corresponded to a 1749-cM maize linkage map (Senior et al., 1996). The choice of $N_M = 438$ markers was based on previous results showing that responses to multiple cycles of genomewide selection were similar with 256 to 768 markers (Bernardo and Yu, 2007). The genome was divided into N_M bins that were $1749/N_M = 4$ cM long. A marker was located at the midpoint of each bin. The L major genes were unlinked to each other, with the ith major gene being located on the ith chromosome. Furthermore, each of the L major genes corresponded to one of the N_M markers. Each major gene was therefore in perfect linkage with one known marker.

The 100 minor QTL were randomly located among the 10 chromosomes. Unlike the major genes, each minor QTL therefore was not constrained to correspond to one of the N_M markers. The sizes of minor QTL effects followed a geometric series (Lande and Thompson, 1990; Bernardo and Yu, 2007). Testcross means behave in an additive manner (Bernardo, 2010, p. 84), and dominance and epistasis were absent. For reference, the effect of the most important minor QTL ranged from 2% of V_G with $L > 2$, to 4% of V_G with $L = 1$ and $R^2 = 5\%$.

Cycle 0 of Genomewide Selection

Cycle 0 of genomewide selection was a population of $N = 100$ or 250 F_2 individuals derived from the cross between two inbreds; the impact of using doubled haploids instead of F_2 individuals or F_3 families in genomewide selection has been previously studied by Mayor and Bernardo (2009). At the marker loci that corresponded to the L major genes, the first parent had the favorable

allele at the odd-numbered loci whereas the second parent had the favorable allele at the even-numbered loci. Likewise, at the 100 minor QTL, the first parent had the favorable allele at the odd-numbered QTL whereas the second parent had the favorable allele at the even-numbered QTL.

Testcross genotypic values of the N Cycle 0 plants were obtained as the sum of genotypic values of an individual across the L major genes and 100 minor QTL. Phenotypic values were simulated for testcrosses of the N individuals in each of eight environments with one replication in each environment. Phenotypic values were obtained by adding a random nongenetic effect to the genotypic value of each individual in each environment. The nongenetic effects were normally and independently distributed with a mean of zero and a nongenetic variance scaled to achieve h^2 = 0.20, 0.50, or 0.80. To facilitate comparisons across different values of L, the nongenetic variance was calculated and h^2 was expressed relative to the genetic variance at minor QTL (V_{QTL}) only, which remained constant regardless of L. The V_{QTL} was estimated in each repeat as the variance among testcross genotypic values of 20,000 individuals segregating at the 100 minor QTL only (L = 0).

Genomewide Selection with All or a Subset of Known Major Genes

When all N_M markers were assumed as having random effects in RR-BLUP (K = 0), the covariance of marker effects was $V(\mathbf{m}) = \mathbf{I}V_M$ where $\mathbf{m}$ was an $N_M \times 1$ vector of marker effects, $\mathbf{I}$ was an $N_M \times N_M$ identity matrix, and V_M was the variance due to each marker (Meuwissen et al., 2001). When K out of the N_M markers were considered as having fixed effects because they corresponded to known major genes, the covariance of marker effects was $V(\mathbf{m}) = \mathbf{F}V_M$ where $\mathbf{F}$ was an $N_M \times N_M$ diagonal matrix with (i) diagonal elements of 0 for the K markers that corresponded to major genes and that were considered to have fixed effects, and (ii) diagonal elements of 1 for the $N_M - K$ markers with random effects. The $\mathbf{m}$ vector, which included both fixed (for $K > 0$) and random effects of markers, was then solved with the usual mixed-model equations for RR-BLUP (Bernardo, 2010, p. 294).

When only a subset of the major genes was explicitly specified in the model, those with the largest effects were successively considered as the ones with fixed effects. As previously mentioned, the R^2 values with L = 3 were 25% for the first major gene, 17% for the second major gene, and 8% for the third major gene. This meant that with L = 3 and K = 1, the marker for the first major gene was considered as having a fixed effect. With L = 3 and K = 2, the markers for the first and second major genes were considered as having fixed effects. With L = 10 and K < L, all major genes had an R^2 of 5% and the major genes on the lower-numbered chromosomes were first considered as those having fixed effects.

For the $N_M - K$ markers with random effects, V_M was equal to $V_G/(N_M - K)$, where V_G was the genetic variance that was not due to the K major genes with fixed effects. In other words, the V_G used in calculating V_M was due to the joint effects of the 100 minor QTL and the $L - K$ major genes that were not specified as having fixed effects in the model. For simplicity, the value of V_G was estimated in two steps. In the first step, the effects of the K major genes were estimated from the Cycle 0 phenotypic and marker data by simple linear regression (for K = 1) or multiple linear regression (for $K > 1$) as $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, where $\mathbf{b}$ was a $K \times 1$ vector of the estimated effect(s) of the K major gene(s), $\mathbf{X}$ was an $N \times K$ design matrix that related $\mathbf{b}$ to $\mathbf{y}$, and $\mathbf{y}$ was an $N \times 1$ vector of the mean phenotypic value, expressed as a deviation from the overall mean, of the N individuals across all eight environments. An element of $\mathbf{X}$ was 1 if the individual was homozygous for the marker allele for the major gene, 0 if the individual was heterozygous, and −1 otherwise. In the second step, adjusted phenotypic values were obtained by subtracting $\mathbf{Xb}$ from the vector of phenotypic values in each of the eight environments. The V_G and nongenetic variance were then estimated from ANOVA of the adjusted phenotypic values and in accordance with a one-factor design (Bernardo, 2010, p. 152). Simulations indicated that, as expected, the above two-step procedure led to good estimates of V_G (results not shown).

Cycles 1 to 4 and Data Analysis

The performance of the N = 100 or 250 individuals in Cycle 0 was predicted as $\mathbf{Xm}$, and the best 10 individuals were randommated to form Cycle 1. The same procedure was repeated in each cycle until Cycle 4 was obtained, with the $\mathbf{X}$ matrices being updated in each cycle but with the marker effects ($\mathbf{m}$) being the same as those in Cycle 0. The population size and number of individuals selected were constant in each cycle.

The accuracy of genomewide prediction was assessed as the correlation between marker-predicted and true genotypic values (r_{MG}). The selection response was obtained as the difference between the genotypic mean of the N individuals in a given cycle and the genotypic mean at Cycle 0, divided by the square root of V_{QTL}. Relative efficiency was calculated as the selection response when one or more major genes had a fixed effect ($K > 1$) divided by the selection response when none of the major genes had a fixed effect (K = 0) in RR-BLUP. The frequency of each major gene was calculated in each cycle. For each criterion (r_{MP}, selection response, relative efficiency, and gene frequency), the mean and standard error were calculated across the 1000 repeats of a simulation experiment. The standard errors were used to conduct pairwise z-tests or to calculate least significant differences (P = 0.05) for each criterion.

RESULTS AND DISCUSSION

Gains from Multiple Cycles of Genomewide Selection with a Single Major Gene

Having a fixed effect for a major gene became more advantageous in genomewide selection as the R^2 value of the major gene and the h^2 of the trait increased (Table 1). When a major gene accounted for R^2 = 50% of V_G and h^2 was 0.80, the relative efficiency (based on selection gains in Cycle 4) with a major gene having a fixed versus random effect was 112% with N = 250 and 121% with N = 100. The relative efficiency was always significantly greater (P = 0.05) than 100% with $R^2 \geq 25\%$ and $h^2 \geq 0.50$.

However, when h^2 was low (0.20), there was no advantage in considering a single major gene as having a fixed effect in genomewide selection. Even when the major gene accounted for R^2 = 50% of V_G, the relative efficiency with the major gene having a fixed versus random effect was not significantly different from 100%. These results suggest

Table 1. Relative efficiency (mean across 1000 repeats) of genomewide selection when a single major gene had a fixed effect versus a random effect.

N†	R^2	h^2		
		0.20	0.50	0.80
250	50	101‡	106*	112*
	33	99	103*	107*
	25	101	102*	105*
	17	100	101	103*
	10	99	100	102*
	5	99	100	101*
100	50	99	107*	121*
	33	99	103*	114*
	25	101	103*	111*
	17	100	100	106*
	10	99	100	103*
	5	97*	99	101

* Significantly different (P = 0.05) from 100% based on 1000 repeats.

† N, population size; R^2, percentage of V_G explained by a major gene; h^2, heritability on an entry-mean basis. The h^2 values were based on the genetic variance due to the minor quantitative trait loci only.

‡ Ratio between the response at Cycle 4 when the major gene had a fixed effect and the response at Cycle 4 when the major gene had a random effect, along with the remaining genomewide markers, in ridge regression best linear unbiased prediction.

that the R^2 value of a major gene is a meaningful criterion in genomewide selection only when h^2 is moderate to high. Furthermore, the h^2 values in Table 1 are based on the effects of the 100 minor QTL only and are lower than the h^2 for the trait as a whole. As previously mentioned in the Materials and Methods, h^2 was scaled according to the V_G at the minor QTL only so that the nongenetic variance for the trait was constant regardless of the contribution of the single major gene to the total V_G. When h^2 is calculated based on the V_G due to both the minor QTL and the single major gene, the h^2 of 0.20 in Table 1 corresponds to $h^2 = 0.33$ for the trait as a whole with $R^2 = 50\%$, to $h^2 = 0.25$ for the trait as a whole with $R^2 = 25\%$, and to $h^2 = 0.22$ for the trait as a whole with $R^2 = 10\%$. These higher values of h^2 should be considered when comparing the results in Table 1 to reports in the literature of R^2 values for a major gene and h^2 for the trait.

As N decreased, it became more advantageous to consider a major gene as having a fixed effect instead of a random effect. Consider a major gene with $R^2 = 50\%$ and $h^2 = 0.80$. The corresponding relative efficiencies were 112% with $N = 250$ and 121% with $N = 100$ (Table 1). The actual gains with $N = 250$ (in units of the square root of V_G due to the minor QTL) were 6.10 when the major gene had a fixed effect and 5.45 when the major gene had a random effect. The advantage of considering a fixed effect for the major gene was therefore 6.10 − 5.45 = 0.65 when expressed as a difference, and 6.10/5.45 = 112% when expressed as a percentage. The actual gains with $N = 100$ were 4.85 when the major gene had a fixed effect and 4.02 when the major gene had a random effect. Compared with the results for $N = 250$, the advantage of considering a fixed effect for the major gene was therefore larger when expressed both as a difference (4.85 − 4.02 = 0.83) and as a percentage (4.85/4.02 = 121%). It is speculated that when both R^2 and h^2 are high, a population size of $N = 100$ may be largely sufficient for estimating the effect of a single major gene. However, a larger population is known to lead to better predictions of genomewide marker effects (Daetwyler et al., 2008; Lorenzana and Bernardo, 2009). A minimal advantage of $N = 250$ over $N = 100$ for estimating the effect of a major gene and a substantial advantage of $N = 250$ over $N = 100$ for predicting genomewide marker effects would lead to a higher relative efficiency when N is low (Table 1).

While specifying a fixed effect for a single major gene was most advantageous when h^2 and R^2 were both high, doing so was never disadvantageous except when h^2, R^2, and N were all low. In particular, with $h^2 = 0.20$, $R^2 = 5\%$, and $N = 100$, the relative efficiency of specifying a fixed versus random effect for the major gene was 97% (Table 1). This 3% decline in the relative response was statistically significant and the underlying conditions of low h^2, low R^2, and small N are the same conditions that render QTL mapping ineffective (Lande and Thompson, 1990).

Overall, specifying a fixed effect for a single major gene was never disadvantageous except when $R^2 < 10\%$. The results for gains from multiple cycles of genomewide selection therefore suggest the following rule-of-thumb: assuming that the estimates of R^2 are accurate, a major gene can be safely specified as having a fixed effect when the R^2 value for the gene is at least 10%.

Accuracy of Genomewide Prediction with a Single Major Gene

If genomewide selection with a given prediction equation is to be performed for only one cycle, the correlation between marker-predicted and true genotypic values (r_{MG}) in Cycle 0 is more meaningful than the response to multiple cycles of genomewide selection. With $h^2 \geq 0.50$, specifying a fixed versus random effect for a single major gene had little effect on r_{MG}. With a moderate to high h^2, the r_{MG} values with fixed versus random effects for a single major gene differed by only 0.00 to 0.03 across different values of R^2 and N (Table 2). These small differences in prediction accuracy were consistent with previous results for the DGAT gene in dairy cattle (*Bos taurus*). The DGAT gene had an R^2 of 51% for milk fat in dairy cattle (Grisart et al., 2002), and the correlation between marker-predicted values and phenotypic values was 0.38 when the DGAT gene was assumed to have a random effect and 0.39 when the DGAT gene was assumed to have a fixed effect (Hayr et al., 2013).

Furthermore, the above results for the DGAT gene were obtained with a Bayes C model for genomewide marker effects (Hayr et al., 2013) whereas the results from the current study were obtained by RR-BLUP. The small differences in the correlations with fixed versus random effects

Table 2. Correlation between marker-predicted and true genotypic values (r_{MG}) in Cycle 0, frequency of the major gene in Cycle 1 (p_{C1}), and frequency of the major gene in Cycle 4 (p_{C4}) when a single major gene had a fixed effect versus a random effect. Results are the means of 1000 repeats.

			h^2 = 0.20		h^2 = 0.50		h^2 = 0.80	
N[a]	R^2	Criterion	Fixed	Random	Fixed	Random	Fixed	Random
250	50	r_{MG}	0.86[b]	0.82	0.98	0.96	1.00	0.99
		p_{C1}	1.00	0.90	1.00	1.00	1.00	1.00
		p_{C4}	1.00	1.00	1.00	1.00	1.00	1.00
	33	r_{MG}	0.80	0.78	0.96	0.94	0.99	0.98
		p_{C1}	0.98	0.83	1.00	0.98	1.00	1.00
		p_{C4}	1.00	0.98	1.00	1.00	1.00	1.00
	25	r_{MG}	0.77	0.77	0.95	0.92	0.99	0.98
		p_{C1}	0.95	0.80	0.99	0.96	1.00	1.00
		p_{C4}	1.00	0.97	1.00	1.00	1.00	1.00
	17	r_{MG}	0.75	0.75	0.93	0.91	0.98	0.97
		p_{C1}	0.91	0.73	0.98	0.92	0.99	0.98
		p_{C4}	0.98	0.93	1.00	1.00	1.00	1.00
	10	r_{MG}	0.72	0.74	0.91	0.90	0.97	0.97
		p_{C1}	0.85	0.69	0.94	0.86	0.97	0.94
		p_{C4}	0.96	0.87	1.00	0.99	1.00	1.00
	5	r_{MG}	0.70	0.73	0.89	0.88	0.96	0.96
		p_{C1}	0.76	0.62	0.86	0.77	0.91	0.88
		p_{C4}	0.87	0.78	1.00	0.96	1.00	1.00
100	50	r_{MG}	0.81	0.73	0.96	0.94	0.99	0.99
		p_{C1}	0.99	0.81	1.00	0.98	1.00	1.00
		p_{C4}	1.00	0.98	1.00	1.00	1.00	1.00
	33	r_{MG}	0.73	0.69	0.94	0.91	0.99	0.98
		p_{C1}	0.96	0.74	1.00	0.94	1.00	1.00
		p_{C4}	1.00	0.94	1.00	1.00	1.00	1.00
	25	r_{MG}	0.68	0.67	0.92	0.89	0.98	0.97
		p_{C1}	0.93	0.71	0.99	0.90	1.00	0.98
		p_{C4}	0.98	0.90	1.00	1.00	1.00	1.00
	17	r_{MG}	0.63	0.65	0.89	0.87	0.97	0.97
		p_{C1}	0.88	0.66	0.96	0.85	0.98	0.96
		p_{C4}	0.95	0.84	1.00	0.99	1.00	1.00
	10	r_{MG}	0.59	0.64	0.86	0.85	0.96	0.95
		p_{C1}	0.81	0.63	0.91	0.79	0.95	0.90
		p_{C4}	0.89	0.77	1.00	0.97	1.00	1.00
	5	r_{MG}	0.55	0.63	0.84	0.83	0.95	0.94
		p_{C1}	0.72	0.58	0.83	0.71	0.87	0.82
		p_{C4}	0.79	0.69	0.98	0.91	1.00	0.99

[a] N, population size; R^2, percentage of V_G explained by a major gene; h^2, heritability on an entry-mean basis. The h^2 values were based on the genetic variance due to the minor quantitative trait loci only.

[b] The approximate least significant difference (P = 0.05) was 0.01 for r_{MG}, p_{C1}, and p_{C4}.

for a major gene, in both the Hayr et al. (2013) study and in the current study, suggested that the method for calculating random effects of genomewide markers has little influence on the impact of having a fixed effect for the major gene.

With $h^2 \geq 0.50$, the r_{MG} when the major gene had a fixed effect was never less than the r_{MG} when the major gene had a random effect (Table 2). With $h^2 = 0.20$, specifying a fixed versus random effect for a single major gene sometimes had a large effect on r_{MG}. With $h^2 = 0.20$, $R^2 = 50\%$, and $N = 100$, r_{MG} was 0.81 when the major gene had a fixed effect and 0.73 when the major gene had a random effect. When both h^2 and R^2 were low, r_{MG} was lower when the major gene had a fixed effect instead of a random effect. For example, the r_{MG} for a major gene with $h^2 = 0.20$, $R^2 = 10\%$, and $N = 100$ was 0.59 when the major gene had a fixed effect and 0.64 when the major gene had a random effect (Table 2). Such differences in r_{MG} were reflected in the response to the first cycle of genomewide selection: with $h^2 = 0.20$, $R^2 = 5\%$, and $N = 100$, the gain from the first cycle genomewide selection was slightly but significantly lower when the major gene had a fixed effect (gain of 1.06) than when the major gene had a random effect (gain of 1.13).

The results also indicated that the r_{MG} in Cycle 0 could not be used as the basis for deciding if a major gene should have a fixed effect or random effect during multiple cycles of genomewide selection. Consider a major gene with $R^2 = 50\%$ and $h^2 = 0.80$. Fitting a fixed effect for the major gene led to high relative efficiencies of 112 to 121% under these conditions (Table 1). However, the r_{MG} in Cycle 0 under these conditions was 0.99 to 1.00 regardless of whether or not the

major gene had a fixed effect. The r_{MG} values then declined with each cycle of genomewide selection. With $N = 250$, the r_{MG} values when the major gene ($R^2 = 50\%$ and $h^2 = 0.80$) had a fixed effect were 1.00 in Cycle 0, 0.83 in Cycle 1, 0.74 in Cycle 2, and 0.65 in Cycle 3. The corresponding r_{MG} values when the major gene had a random effect were 1.00 in Cycle 0, 0.66 in Cycle 1, 0.56 in Cycle 2, and 0.47 in Cycle 3. The r_{MG} values therefore declined more rapidly when the major gene had a random instead of a fixed effect.

The rate of decline in linkage disequilibrium from Cycle 0 to 4 is expected to be largely due to N_M and the number of individuals selected in each cycle. Because N_M and the number of selected individuals were the same regardless of whether a major gene had a fixed versus random effect, the difference in the rate of decline in r_{MG} could not be attributed to a difference in the rate of decline in linkage disequilibrium. The r_{MG} values therefore reflected differences in the ability to capture the effects due to minor QTL at the later cycles of selection when a major gene had a fixed versus random effect, particularly after the major gene had become homozygous or near homozygous at an earlier cycle of selection (Combs and Bernardo, 2013). However, the breeding scheme simulated in this study and implemented in maize by Massman et al. (2013) and Combs and Bernardo (2013) involved constructing a genomewide prediction equation in Cycle 0 followed by successive generations of marker-based selection in a greenhouse or year-round nursery. Because of the lack of phenotyping in each cycle, the decline in r_{MG} cannot be monitored in practice under the breeding scheme used.

Changes in Frequency of a Single Major Gene

A major gene has a larger estimated effect if the major gene has a fixed effect instead of a random effect. The larger effect when the major gene is assumed fixed then leads to a stronger selection pressure on the major gene, and this stronger selection pressure led to larger changes in gene frequency when the major gene had a fixed versus random effect. The frequency of the major gene in Cycle 1 (p_{C1}) indicated that one cycle of genomewide selection led to homozygosity or near-homozygosity of the major gene when R^2 and h^2 were both high, regardless of whether the major gene had a fixed or random effect (Table 2). Genomewide selection is a form of index selection on multiple markers, and index selection is expected to be superior to tandem selection (Hazel and Lush, 1942). Tandem selection by first selecting for a major gene followed by genomewide selection for minor QTL (among individuals fixed for the major gene) was therefore not considered in this study. The high values of p_{C1} nevertheless indicated that in terms of fixing a known major gene, genomewide selection was largely equivalent to tandem selection when R^2 and h^2 were both high.

With $R^2 \geq 10\%$ and $h^2 \geq 0.50$, the frequency of the major gene in Cycle 4 (p_{C4}) approached or was equal to 1.00. These high p_{C4} values indicated that the differences in selection responses in Cycle 4 when a major gene had a fixed versus random effect (Table 1) were not largely due to the major gene itself, but were due to the genomewide markers being able to better capture the effects at minor QTL when the effect of the major gene was modeled separately from the effects of the background QTL (Combs and Bernardo, 2013).

Both p_{C1} and p_{C4} were lower with $h^2 = 0.20$ than with $h^2 \geq 0.50$ (Table 2). Across different values of R^2 and N, p_{C1} with $h^2 = 0.20$ was 0.72 to 1.00 when the major gene had a fixed effect and 0.58 to 0.90 when the major gene had a random effect. The corresponding p_{C4} values were 0.79 to 1.00 when the major gene had a fixed effect and 0.69 to 1.00 when the major gene had a random effect. As previously mentioned, assuming a major gene had a fixed instead of random effect led to a 3% decrease in relative efficiency with $N = 100$, $R^2 = 5\%$, and $h^2 = 0.20$. Under these conditions, p_{C4} was 0.79 when the major gene had a fixed effect and 0.69 when the major gene had a random effect. When N, h^2, and R^2 were all low, the higher selection gain when the major gene had a random effect was therefore largely due to the minor QTL.

Genomewide Selection with Multiple Major Genes

The general results obtained for genomewide selection with a single major gene were also obtained for genomewide selection with multiple major genes. However, the results also indicated difficulty in simultaneously estimating the fixed effects of more than a few major genes.

When $L > 1$ major genes were present, the L major genes jointly accounted for 50% of V_G in the simulation experiments. It is therefore meaningful to compare the relative efficiency with $L > 1$ major genes versus the relative efficiency with $L = 1$ major gene that explained $R^2 = 50\%$ of V_G. With $N = 250$, the highest relative efficiency with $L = 1$ major gene (112%, $R^2 = 50\%$, $h^2 = 0.80$; Table 1) was equal to the highest relative efficiency with $L = 2$ or 3 major genes (112%, $h^2 = 0.80$, $K = L$; Table 3). This result indicated that when N was large and h^2 was high, it did not matter in terms of relative efficiency whether the major gene effects were concentrated in a single major gene or were partitioned across $L = 2$ or 3 unlinked major genes.

The above result was not obtained when the population size decreased to $N = 100$. In this situation, the highest relative efficiencies with $L = 2$ or 3 major genes (116–119%, $K = L$, $h^2 = 0.80$; Table 3) were slightly lower than the highest relative efficiency with a single major gene (121%, $R^2 = 50\%$, $h^2 = 0.80$; Table 1). The above result was also not obtained when $L = 10$ unlinked major genes controlled the trait. The $L = 10$ major genes likewise jointly accounted for 50% of V_G, yet the relative efficiency did not exceed 105% when all or a subset of the 10 major genes had a fixed effect (Table 3). The usefulness of having fixed versus random effects for multiple major genes decreased further as h^2 decreased.

Table 3. Relative efficiency (mean across 1000 repeats) of genomewide selection when multiple major genes had fixed effects versus random effects.

N	No. of major genes (L)	No. of major genes with fixed effects (K)	h^2 0.20	h^2 0.50	h^2 0.80
250	2	2	100	103	112
	2	1	100	103	107
		LSD	1.2	1.0	1.0
	3	3	100	104	112
	3	2	99	103	109
	3	1	100	102	105
		LSD	1.2	1.0	1.0
	10	10	102	100	103
	10	9	102	101	104
	10	8	102	101	103
	10	7	102	101	103
	10	6	102	102	103
	10	5	101	101	102
	10	4	101	101	102
	10	3	101	101	101
	10	2	101	100	101
	10	1	101	100	101
		LSD	1.5	1.2	1.2
100	2	2	102	106	119
	2	1	100	104	110
		LSD	1.7	1.4	1.4
	3	3	103	103	116
	3	2	102	103	112
	3	1	100	101	106
		LSD	1.7	1.3	1.4
	10	10	103	101	103
	10	9	100	102	105
	10	8	100	103	104
	10	7	99	103	104
	10	6	99	103	104
	10	5	99	102	103
	10	4	100	102	102
	10	3	99	101	102
	10	2	100	101	101
	10	1	100	101	100
		LSD	2.3	1.6	1.5

N, population size; h^2, heritability on an entry-mean basis. The h^2 values were based on the genetic variance due to the minor quantitative trait loci only.

Ratio between the response at Cycle 4 when K major genes had fixed effects and the response at Cycle 4 when all L major genes had random effects, along with the remaining genomewide markers, in ridge regression best linear unbiased prediction.

Approximate least significant difference at P = 0.05.

These results regarding the low relative efficiency of genomewide selection with $L = 10$ major genes with fixed effects were consistent with the known difficulties in QTL mapping and in incorporating gene information to predict performance when both N and h^2 are low. In a simulation experiment with 10 unlinked QTL controlling the trait (with no background QTL) and $h^2 = 0.30$, the power to detect QTL was only 12% with $N = 100$ and the effects of the QTL were overestimated (Beavis, 1994). With $N = 500$ and $L = 50$ QTL controlling the trait, selection based on gene information for all 50 QTL was 5% less efficient than phenotypic selection, to the extent that it was more advantageous to ignore the genes with smaller effects even when the identity of such genes was known (Bernardo, 2001).

Application in a Breeding Program

The main conclusion from this study is that when a few (1–3) major genes are present for a quantitative trait and each major gene accounts for ≥10% of V_G, these major genes should be fitted as having fixed effects instead of random effects in the genomewide prediction model. The effects of the remaining genomewide markers can then be fitted by RR-BLUP, with the variance of marker effects being adjusted for the contributions of the major genes to V_G. Having fixed effects for major genes may not always increase the response to genomewide selection, particularly when h^2 is low. However, having fixed effects for major genes would lead to no harm because the gains from genomewide selection would at least be as large as those obtained when the major genes are not treated any differently from the genomewide markers.

Having known major genes implies that prior estimates of their effects are also available, and these prior estimates could be used in the first step of the procedure used in this study for estimating V_G. On the other hand, obtaining ad hoc estimates of the effects of known major genes, as was done in this study, would help account (i) for any differences in the effects of major genes across different populations or genetic backgrounds (Pumphrey et al., 2007) and (ii) for correlations between the effects of major genes and of unknown minor QTL that are in linkage disequilibrium with the major genes. While a two-step procedure was used to estimate V_G in this study, any procedure that leads to good estimates of V_G should suffice. For example, a one-step, mixed-model approach can be used to estimate V_G when known major genes are present (Kennedy et al., 1992).

In this study, the R^2 values for major genes were known without error and were expressed in terms of the V_G explained by each major gene. In practice, the effect of a major gene is usually expressed as the percentage of phenotypic variance (V_P) explained by the gene and needs to be estimated from experimental data. Suppose the h^2 is 0.60 for the trait as a whole (i.e., based on the contributions of both the minor QTL and a single major gene to V_G). In this situation, a major gene with $R^2 = 33\%$ in this study explained 0.60(0.33) = 20% of V_P. This result means that an estimated R^2 value, reported in the literature and expressed in terms of V_P, theoretically would be lower than the corresponding R^2 values in this study. On the other hand, the effects of genes are often overestimated in QTL mapping experiments (Beavis, 1994; Schön et al., 2004) and estimates in the literature of the percentage of V_P explained by each major gene may be inflated.

Major genes for quantitative traits in crops are not uncommon. In wheat, the *Fhb1* QTL for Fusarium head blight resistance explained of 25 to 42% of V_P (Anderson et al., 2001) and reduced disease severity by 25 to 32% among

near-isogenic lines (Pumphrey et al., 2007; Agostinelli et al., 2012). In maize, the DGAT gene for kernel oil explained 26% of V_p (Garcia, 2008). The *rd1* maize dwarfing gene explained 63% of V_p and, in homozygous form, decreased plant height by about 31 cm (Combs and Bernardo, 2013). In soybean, the *rhg1* gene for resistance to soybean cyst nematode explained 35 to 54% of V_p (Concibido et al., 1996). The results from this study suggest that when genomewide selection involves any of these major genes or other major genes that similarly have large effects, the markers for such genes should have fixed instead of random effects.

Acknowledgment

This research was funded in part by a USDA grant on Sorghum Biomass Genomics and Phenomics.

References

Agostinelli, A.M., A.J. Clark, G. Brown-Guedira, and D.A. Van Sanford. 2012. Optimizing phenotypic and genotypic selection for Fusarium head blight resistance in wheat. Euphytica 186:115–126. doi:10.1007/s10681-011-0499-6

Anderson, J.A., R.W. Stack, S. Liu, B.L. Waldron, A.D. Fjeld, C. Coyne, B. Moreno-Sevilla, J. Mitchell Fetch, Q.J. Song, P.B. Cregan, and R.C. Frohberg. 2001. DNA markers for Fusarium head blight resistance QTLs in two wheat populations. Theor. Appl. Genet. 102:1164–1168. doi:10.1007/s001220000509

Beavis, W.D. 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. In: D.B. Wilkinson, editor, Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference, Chicago, IL. 7–8 Dec. 1994. Am. Seed Trade Assoc., Washington, DC. p. 250–266.

Bernardo, R. 2001. What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci. 41:1–4. doi:10.2135/cropsci2001.4111

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1664. doi:10.2135/cropsci2008.03.0131

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury, MN.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082–1090. doi:10.2135/cropsci2006.11.0690

Combs, E., and R. Bernardo. 2013. Genomewide selection to introgress semidwarf corn germplasm into U.S. Corn Belt inbreds. Crop Sci. 53:1427–1435. doi:10.2135/cropsci2012.11.0666

Concibido, V.C., R.L. Denny, D.A. Lange, J.H. Orf, and N.D. Young. 1996. RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. Crop Sci. 36:1643–1650. doi:10.2135/cropsci1996.0011183X003600060038x

Concibido, V.C., B.W. Diers, and P.R. Arelli. 2004. A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci. 44:1121–1131. doi:10.2135/cropsci2004.1121

Daetwyler, H.D., B. Villanueva, and J.A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395. doi:10.1371/journal.pone.0003395

Eagles, H.A., G.J. Hollamby, N.N. Gororo, and R.F. Eastwood. 2002. Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. Aust. J. Agric. Res. 53:367–377. doi:10.1071/AR01074

Garcia, N.S. 2008. Mapping QTLs for seed oil, and embryo size in corn using Korean high oil germplasm. M.S. thesis, Univ. of Minnesota, Saint Paul.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222–231. doi:10.1101/gr.224202

Guo, Z., D. Tucker, J. Lu, V. Kishore, and G. Gay. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124:261–275. doi:10.1007/s00122-011-1702-9

Hayr, M., M. Saatchi, D. Johnson, and D. Garrick. 2013. Increasing the accuracy of genomic estimated breeding values of milk traits in New Zealand dairy cattle using DGAT genotypes. Poster presented at: Plant and Animal Genome Conference, San Diego, CA. 12–16 Jan. 2013. Scherago Int., Jersey City, NJ. Poster 577.

Hazel, L.N., and J.L. Lush. 1942. The efficiency of three methods of selection. J. Hered. 33:393–399.

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Kennedy, B.W., M. Quinton, and J.A. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70:2000–2012.

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.-L. Jannink. 2011. Genomic selection in plant breeding: Knowledge and prospects. Adv. Agron. 113:77–123. doi:10.1016/B978-0-12-385531-2.00002-5

Lorenzana, R., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151–161. doi:10.1007/s00122-009-1166-3

Massman, J.M., H.-J.G. Jung, and R. Bernardo. 2013. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53:58–66.

Mayor, P.J., and R. Bernardo. 2009. Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F_2 populations. Crop Sci. 49:1719–1725. doi:10.2135/cropsci2008.10.0587

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Pumphrey, M.O., R. Bernardo, and J.A. Anderson. 2007. Validating the *Fhb1* QTL for Fusarium head blight resistance in near-isogenic wheat lines developed from breeding populations. Crop Sci. 47:200–206. doi:10.2135/cropsci2006.03.0206

Schön, C.C., H.F. Utz, S. Groh, B. Truberg, S. Openshaw, and A.E. Melchinger. 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498. doi:10.1534/genetics.167.1.485

Schulz-Streeck, T., J.O. Ogutu, Z. Karaman, C. Knaak, and H.P. Piepho. 2012. Genomic selection using multiple populations. Crop Sci. 52:2453–2461. doi:10.2135/cropsci2012.03.0160

Senior, M.L., E.C.L. Chin, M. Lee, J.S.C. Smith, and C.W. Stuber. 1996. Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. Crop Sci. 36:1676–1683. doi:10.2135/cropsci1996.0011183X003600060043x

Weegels, P.L., R.J. Hamera, and J.D. Schofield. 1996. Functional properties of wheat glutenin. J. Cereal Sci. 23:1–17. doi:10.1006/jcrs.1996.0001

Zheng, P., W.B. Allen, K. Roesler, M.E. Williams, S. Zhang, J. Li, K. Glassman, J. Ranch, D. Nubel, W. Solawetz, D. Bhattramakki, V. Llaca, S. Deschamps, G.Y. Zhong, M.C. Tarczynski, and B. Shen. 2008. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat. Genet. 40:367–372. doi:10.1038/ng.85