Genomewide Selection with Minimal Crossing in Self–Pollinated Crops

Rex Bernardo★

ABSTRACT

Genomewide selection in plants has focused on cross-pollinated species, such as maize (*Zea mays* L.), in which selected plants can be easily crossed to obtain a large number of seeds for the next cycle of selection. Here I describe and present simulation results for a genomewide selection procedure that requires minimal crossing in self-pollinated crops. This procedure, referred to as GWS_{Self}, involves producing only a few F_1 seeds from crosses among the selected progenies in Cycle 0, allowing the F_1 seeds to naturally self-pollinate to produce F_2 seeds, conducting marker-based selection among the resulting F_2 plants, and repeating this select-recombine-self procedure in a year-round nursery or greenhouse. With barley (*Hordeum vulgare* L.) as a model species, I found from simulation experiments that selection responses with GWS_{Self} were 81 to 87% of the responses obtained if genomewide selection procedures appropriate for maize were used. These lower responses can be compensated for by more stringent selection.

Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Cir., St. Paul, MN 55108. Received 8 May 2009. ★Corresponding author (bernardo@umn.edu).

Unlike marker-based selection procedures that use only those markers with significant effects on the trait, genomewide selection does not involve significance tests but instead involves the prediction of breeding values for a large number of random markers (Meuwissen et al., 2001). Simulation studies in maize (*Zea mays* L.) have shown that the gains from genomewide selection were 18 to 43% larger than the gains from marker-assisted recurrent selection (Edwards and Johnson, 1994; Eathington et al., 2007) using a subset of markers with significant effects (Bernardo and Yu, 2007). Empirical results in maize, barley (*Hordeum vulgare* L.), and *Arabidopsis* showed consistently higher accuracies of genomewide predictions than of predictions based on significant markers only (Lorenzana and Bernardo, 2009).

The genomewide selection scheme proposed in maize involves selecting the best plants based on marker information, crossing these selected plants to form the next cycle of selection, and repeating this select-and-recombine procedure for several cycles in a year-round nursery or greenhouse (Bernardo and Yu, 2007; Bernardo, 2009). This scheme is feasible in species, such as maize or tomato (*Lycopersicon esculentum* L.), in which hand crossing to produce a large number of seeds is easy. However, this scheme is not feasible in many

self-pollinated species, such as barley and soybean [*Glycine max* (L.) Merrill], in which producing large numbers of seeds from hand crosses is laborious and time–consuming.

The purpose of this Short Communication is to describe a genomewide selection procedure that requires minimal crossing in self–pollinated species (denoted by GWS_{Self} in this paper) and to present simulation results on selection gains from this procedure.

MATERIALS AND METHODS
GWS_{Self} Scheme
In this procedure for a self-pollinated species, recombinant inbreds (Cycle 0) are developed by several generations of selfing from an F_2 between two parental inbreds. The recombinant inbreds are evaluated in multilocation yield trials in the summer of Year 1 (Table 1). The yield trials may be repeated during a second year if genotype × year interaction is substantial. This second year of yield trials, however, will lengthen the selection procedure.

The best recombinant inbreds are then selected based on their phenotypic performance only as explained by Bernardo and Yu (2007). To minimize hand crosses, cyclical improvement of the population by genomewide selection involves producing only a few F_1 seeds from crosses among the selected progenies in Cycle 0, allowing the F_1 seeds to naturally self-pollinate to produce F_2 seeds, obtaining genomewide predictions of genotypic value among the resulting F_2 plants and repeating this procedure in a year-round nursery or greenhouse (Table 1). Whereas genomewide selection in a cross-pollinated species involves a select-and-recombine procedure (denoted by GWS_{Cross} in this paper), genomewide selection in a self–pollinated species there–fore involves a select-recombine-self procedure.

Table 1. Genomewide selection with minimal hand crossing (GWS_{Self}) in a self-pollinated species for which three seasons per year can be grown in a year-round nursery or greenhouse.

Season	Procedure
Year 1	
Summer†	Multilocation yield trials of *N* recombinant inbreds developed from an F_2 population (Cycle 0)
	Genotype the *N* recombinant inbreds with markers
	Select the best N_{Sel} recombinant inbreds based on phenotypic data
September–December	Intercross the best N_{Sel} recombinant inbreds to obtain F_1 seed
Year 2	
January–April	Self F_1 plants to obtain F_2 seeds (Cycle 1)
May–August	Grow and genotype Cycle 1 F_2 plants
	Select the best N_{Sel} plants based on marker data
	Intercross the best N_{Sel} plants to form F_1 seeds (i.e., F_2 plant × F_2 plant → F_1 plant)
September–December	Self F_1 plants to obtain F_2 seeds (Cycle 2)
Year 3	
January–April	Grow and genotype Cycle 2 F_2 plants
	Select the best N_{Sel} plants based on marker data
	Intercross the best N_{Sel} plants to form F_1 seeds
Summer	Self F_1 plants to obtain F_2 seeds (Cycle 3)

†The yield trials may be repeated during a second year if genotype × year interaction is substantial. In this situation, GWS_{Self} will be completed in Year 4 instead of in Year 3.

The GWS_{Self} procedure outlined in Table 1 is appropriate for a self-pollinated species for which three generations can be grown per year in a year-round nursery or greenhouse. If the objective is to complete marker-based selection by the end of summer in Year 3, genomewide selection up to Cycle 3 can be conducted (Table 1). In contrast, if hand crossing were easy, five cycles of genomewide selection could be achieved by the end of summer in Year 3 (i.e., Cycle 1 to Cycle 2 in May to August of Year 2; Cycle 2 to Cycle 3 in September to December of Year 2; Cycle 3 to Cycle 4 in January to April of Year 3; and Cycle 4 to Cycle 5 in the summer of Year 3). An equal-time comparison in the response to selection is therefore between Cycle 3 of GWS_{Self} and Cycle 5 of GWS_{Cross}. Unless otherwise stated in the rest of this paper, responses to GWS_{Self} are based on the mean of Cycle 3, whereas responses to GWS_{Cross} are based on the mean of Cycle 5.

A total of $N = 144$ or 288 recombinant inbreds (in Cycle 0) or plants (in subsequent cycles) were evaluated in each cycle of selection, and the best $N_{Sel} = 4$ or 8 recombinant inbreds or plants were selected to form the next cycle. I assumed that the N_{Sel} plants were crossed in all $N_{Sel}(N_{Sel} - 1)/2$ possible combinations, and that only two F_1 seeds were produced from each cross. Equal or near-equal numbers of (selfed) F_2 seeds were then obtained from each F_1 plant to obtain a bulk of N plants for the next cycle of selection. To illustrate, suppose $N_{Sel} = 4$ plants were selected out of $N = 288$ Cycle 1 plants. A total of 12 F_1 seeds (i.e., two from each of the six crosses among the N_{Sel} plants) were obtained, and $N/[N_{Sel}(N_{Sel} - 1)] = 24$ F_2 seeds were then harvested from each of the 12 F_1 plants to form Cycle 2. In contrast, when $N_{Sel} = 8$ plants were selected out of $N = 288$, each of the 56 F_1 plants contributed five F_2 seeds, making a total of 280 plants. Eight additional F_2 seeds were then chosen at random to maintain a population size of 288 (Bernardo et al., 2006).

Linkage Map, Genetic Models, and Marker Effects
Barley was considered as the model self-pollinated species. The barley linkage map comprised 1069 cM (Varshney et al., 2007) across seven pairs of chromosomes. A total of $N_M = 128$ or 256 evenly spaced, codominant markers were used in genomewide selection. The trait of interest was controlled by 100 quantitative trait loci (QTL). The details of the genetic model used have been described in previous articles (Bernardo and Yu, 2007; Bernardo, 2009). Briefly, the 100 QTL were randomly located on the chromosomes according to a uniform distribution across the entire genome, with both coupling and repulsion linkages occurring at random. Dominance and epistasis were absent. The QTL effects followed a geometric distribution (Lande and Thompson, 1990; Bernardo and Yu, 2007). Heritability on an entry-mean basis was $h^2 = 0.20$, 0.50, or 0.80 in Cycle 0.

Based on the genotypic and phenotypic data in Cycle 0, breeding values associated with each of the N_M markers were obtained by best linear unbiased prediction as described by Meuwissen et al. (2001) and by Bernardo and Yu (2007). Each simulation experiment comprised a combination of N_M, N, N_{Sel}, and h^2. Each simulation experiment was repeated 1000 times, and the results were analyzed across repeats as described by Bernardo and Yu (2007). Responses were expressed as a deviation from the Cycle 0 mean, divided by the square root of the genetic variance among recombinant inbreds in Cycle 0.

For reference, simulation experiments for GWS_{Cross} in barley were conducted assuming that hand crossing in barley is not a limitation. These simulation experiments for GWS_{Cross} in barley were conducted with the same values of N_M, N, N_{Sel}, and h^2 used in GWS_{Self} and with the same procedures described by Bernardo and Yu (2007) for maize. The Cycle 0 recombinant inbreds were evaluated for their per se performance (rather than testcross performance, as is appropriate for maize) in both GWS_{Cross} and GWS_{Self}.

RESULTS AND DISCUSSION

The responses to GWS_{Self} were 81 to 87% of the corresponding responses to GWS_{Cross} (Table 2). This decrease in equal–time selection response reflected the fewer cycles of selection in GWS_{Self} than in GWS_{Cross} owing to the intervening generations required for selfing in GWS_{Self}. However, the responses to GWS_{Self} (i.e., Cycle 3) were 104 to 112% of the corresponding responses to GWS_{Cross} in Cycle 3 (Table 2). While this advantage in GWS_{Self} over GWS_{Cross} in Cycle 3 is not meaningful in practice, this equal–cycle comparison suggested that the process of selfing helped increase the overall response to GWS_{Self}. Specifically, this result indicated that when nonhomozygous plants in Cycle 1 are crossed to create F_1 plants, self-pollination releases genetic variation that is expressed among F_2 plants, and this variation can then be exploited in selection.

As expected for any selection procedure (Falconer, 1981), the responses to both GWS_{Self} and GWS_{Cross} increased as population size (N) and heritability (h^2) increased and as the number of selected progenies (N_{Sel}) decreased (Table 2). The results provided information on the extent to which N should be increased to compensate for the lower responses to GWS_{Self} than to GWS_{Cross}. Consider the experiments with $N_M = 256$ and $N_{Sel} = 4$. When h^2 was 0.20, the response to GWS_{Cross} was 3.23 with $N = 144$, whereas the response to GWS_{Self} was 3.38 when the population size was doubled to $N = 288$ (Table 2). When h^2 was 0.80, the response to GWS_{Cross} with $N = 144$ was higher (5.29) than the response to GWS_{Self} with $N = 288$ (4.95). These results indicated that N in GWS_{Self} should be roughly twice as large as N in GWS_{Cross} if the objective is to have roughly the same response to both selection procedures.

With a 1069 cM barley linkage map (Varshney et al., 2007), the mean distance between adjacent markers was about 4.2 cM with $N_M = 256$ markers and 8.4 cM with $N_M = 128$ markers. Previous studies have indicated little advantage in having markers spaced less than about 10 cM apart in QTL mapping (Doerge, 2002) or in predicting the accuracy of genotypic value in biparental plant populations (Lorenzana and Bernardo, 2009). The nonsignificant differences in the responses to GWS_{Self} with 128 and 256 markers (Table 2; $LSD_{0.05} = 0.08$ to 0.11) were consistent with these previous results. A larger number of markers, however, would likely be needed for species such as wheat (*Triticum aestivum* L.) that have larger genomes.

An alternative scheme for genomewide selection in a self–pollinated species involves obtaining genomewide predictions during inbred development. In this scheme, the plants are allowed to naturally self–pollinate for several generations, genomewide predictions are obtained in each selfing generation, and recombinant inbreds are developed from the selfed plants deemed superior based on genomewide predictions. However, the cyclical-improvement approach in GWS_{Self} is expected to be superior to selection among recombinant inbreds in Cycle 0. Consider that the proportions selected in the simulation experiments were about 1 to 6%. With the stringent selection differential of 2.67 for 1% selected (Falconer, 1981), the predicted standardized mean of the best recombinant inbreds in Cycle 0 is 2.67h, where h is the square root of heritability among recombinant inbreds. The predicted standardized responses to selection among recombinant inbreds are therefore 1.19 with $h^2 = 0.20$, 1.89 with $h^2 = 0.50$, and 2.49 with $h^2 = 0.80$. These responses are much lower than the corresponding responses to GWS_{Self} (Table 2). Even though the responses to GWS_{Self} were lower than the responses to GWS_{Cross}, the GWS_{Self} procedure would therefore be superior to noncyclical selection among recombinant inbreds in Cycle 0.

Overall, the results suggested that the GWS_{Self} procedure is useful for genomewide selection in a species in which

Table 2. Simulation responses to genomewide selection in barley with minimal hand crossing (GWS_{Self}) and with a scheme appropriate for a cross-pollinated species (GWS_{Cross}).

N_M[†]	N	N_{Sel}	h^2	Total response to selection[‡]		
				GWS_{Self}	GWS_{Cross}	Cycle 3 of GWS_{Cross}
256	144	4	0.20	2.72	3.23	2.54
			0.50	3.80	4.45	3.54
			0.80	4.39	5.29	4.15
256	144	8	0.20	2.67	3.13	2.42
			0.50	3.68	4.29	3.32
			0.80	4.23	5.04	3.86
256	288	4	0.20	3.38	4.15	3.26
			0.50	4.33	5.21	4.11
			0.80	4.95	5.98	4.70
256	288	8	0.20	3.36	4.06	3.13
			0.50	4.35	5.06	3.93
			0.80	4.87	5.77	4.43
128	144	4	0.20	2.71	3.16	2.50
			0.50	3.75	4.41	3.52
			0.80	4.42	5.22	4.13
128	144	8	0.20	2.62	3.15	2.45
			0.50	3.67	4.23	3.28
			0.80	4.20	4.96	3.81
128	288	4	0.20	3.39	3.99	3.11
			0.50	4.32	5.10	4.07
			0.80	4.92	5.88	4.67
128	288	8	0.20	3.39	3.96	3.07
			0.50	4.22	5.02	3.91
			0.80	4.82	5.67	4.36

[†]Genetic models differed in the number of markers (N_M), population size (N), number of plants selected (N_{Sel}), and trait heritability (h^2).

[‡]Responses were based on the mean of Cycle 3 in GWS_{Self} and of Cycle 5 in GWS_{Cross}. Responses are in units of the genetic standard deviation among Cycle 0 recombinant inbreds. The LSD0.05 ranged from 0.08 to 0.11.

obtaining large numbers of seed from hand crosses is difficult. The tradeoff, however, with the reduced number of hand crosses is a reduction in selection response of up to about 20% when compared with genomewide-selection procedures without selfing. This reduction in response can be overcome by the use of a more stringent selection differential. To reduce genetic drift, this would imply maintaining the number of plants selected (N_{Sel}) but increasing the number of plants evaluated (N) in each cycle of selection in GWS$_{Self}$. While increasing N would lead to higher costs for genotyping, genomewide selection assumes that markers are abundant and cheap and that an increase in N would be feasible.

ACKNOWLEDGMENTS

REFERENCES

Bernardo, R. 2009. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. 49:419–425.

Bernardo, R., L. Moreau, and A. Charcosset. 2006. Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection. Crop Sci. 46:1972–1980.

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

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3:43–52.

Eathington, S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, and J.K. Bull. 2007. Molecular markers in a commercial breeding program. Crop Sci. 47(S3):S154–S163.

Edwards, M., and L. Johnson. 1994. RFLPs for rapid recurrent selection. p. 33–40. *In* Analysis of molecular marker data. Joint Plant Breeding Symposium Series. Am. Soc. Hort. Sci., CSSA, Madison, WI.

Falconer, D.S. 1981. Introduction to quantitative genetics. 2nd ed. Longman, London.

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

Lorenzana, R.E., and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151–161.

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.

Varshney, R.K., T.C. Marcel, L. Ramsay, J. Russell, M.S. Röder, N. Stein, R. Waugh, P. Langridge, R.E. Niks, and A. Graner. 2007. A high density barley microsatellite consensus map with 775 SSR loci. Theor. Appl. Genet. 114:1091–1103.