2019

Context-Specific Genomic Selection Strategies Outperform Phenotypic Selection for Soybean Quantitative Traits in the Progeny Row Stage

Christopher J. Smallwood
csmallw1@vols.utk.edu

Arnold M. Saxton
University of Tennessee, Knoxville

Jason D. Gillman
USDA, Agricultural Research Service

Hem S. Bhandari
University of Tennessee, Knoxville

Phillip A. Wadl
USDA, Agricultural Research Service

Published January 4, 2019

RESEARCH

Context-Specific Genomic Selection Strategies Outperform Phenotypic Selection for Soybean Quantitative Traits in the Progeny Row Stage

Christopher J. Smallwood,* Arnold M. Saxton, Jason D. Gillman, Hem S. Bhandari, Phillip A. Wadl, Benjamin D. Fallen, David L. Hyten, Qijian Song, Vincent R. Pantalone

ABSTRACT

Evaluating different breeding selection strategies for relative utility is necessary to choose those that maximize efficiency. Soybean [*Glycine max* (L.) Merr.] seed yield and fatty acid, protein, and oil contents are all commercially important traits that display complex quantitative inheritance. A soybean population consisting of 860 F_5–derived recombinant inbred lines (RILs), genotyped with 4867 polymorphic single nucleotide polymorphism (SNPs) was used to compare phenotypic and context specific genomic selection (GS) strategies. To simulate progeny rows, each RIL was grown in a single plot in 2010 in Knoxville, TN, and phenotype was recorded. A subset of 276 RILs with similar maturity was then grown in multilocation, replicated field trials in 2013 to compare the performance of each selection method in field conditions. Notably, the preferred method for each trait was GS. Of the GS approaches evaluated, Epistacy performed best for yield, and BayesB and/or genomic best linear unbiased prediction (G-BLUP) were preferred for each of the other traits. Yield was the only trait for which the predictions had a large change when the number of SNPs and the number of RILs were randomly reduced for the G-BLUP model, with the best predictions occurring when RILs with different maturity that were not grown in 2013 were removed from the training set. These findings provide important information on how soybean breeders can maximize selections from the progeny row stage for yield and fatty acid, protein, and oil contents by using appropriate selection strategies.

C.J. Smallwood, H.S. Bhandari, and V.R. Pantalone, Dep. of Plant Sciences, Univ. of Tennessee, 2431 Joe Johnson Dr., Knoxville, TN 37996; A.M. Saxton, Dep. of Animal Science, Univ. of Tennessee, 2506 River Dr., Knoxville, TN 37996; J.D. Gillman, Plant Genetic Research Unit, USDA-ARS, Univ. of Missouri, 110 Waters Hall, Columbia, MO 65211; P.A. Wadl, USDA-ARS, US Vegetable Lab., 2700 Savannah Hwy, Charleston, SC 29414; B.D. Fallen, Clemson Univ., Advanced Plant Technology, Clemson Pee Dee REC, 2200 Pocket Rd., Florence, SC 29506; D.L. Hyten, Dep. of Agronomy and Horticulture, Univ. of Nebraska-Lincoln, 322 Keim Hall, Lincoln, NE 68583; Q. Song, Soybean Genomics and Improvement Lab., USDA-ARS, Beltsville, MD 20705. Received 21 Mar. 2018. Accepted 16 Oct. 2018. *Corresponding author (csmallw1@vols.utk.edu). Assigned to Associate Editor Leah Mchale.

Abbreviations: CSM, context-specific marker-assisted selection; GBLUP, genomic best linear unbiased prediction; GS, genomic selection; MAS, marker-assisted selection; MG, maturity group; NIRS, near-infrared reflectance spectroscopy; PS, phenotypic selection; QTL, quantitative trait locus/loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism.

Soybean [*Glycine max* (L.) Merr.] is a major crop produced globally for a wide range of purposes. Protein and oil are major components of soybean seed that contribute to its high value. Historically, oil and protein in soybean seed are negatively correlated (Yaklich et al., 2002). Oil and yield share a positive relationship, and protein and yield have a negative relationship (Morrison et al., 2008). Because of this, increases in soybean oil and yield must be sought after while simultaneously seeking to maintain adequate protein levels (Cober et al., 2009).

Within soybean oil, there are five primary fatty acids: palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2), and linolenic (18:3). These typically occur in relative concentrations of 100, 40, 220, 540, and 100 g kg^{-1} of total lipids, respectively

Published in Crop Sci. 59:54–67 (2019).
doi: 10.2135/cropsci2018.03.0197

(Wilson, 2004). Improving the fatty acid profile in soybean has gained importance recently, particularly with the Food and Drug Administration recently ruling that partially hydrogenated oils are no longer generally recognized as safe (https://www.federalregister.gov/articles/2015/06/17/2015-14883/final-determination-regarding-partially-hydrogenated-oils). Due to this decision, a primary goal of fatty acid improvement is to reduce linolenic acid (<30 g kg^{-1}), thus reducing the need to partially hydrogenate soybean oil. Coinciding with this is the goal of increasing oxidatively stable, monounsaturated oleic acid (>800 g kg^{-1}). Oleic acid has been shown to lower cholesterol than saturated fatty acids in human consumption (Kris-Etherton and Yu, 1997). Additionally, soybean oil with increased oleic acid has higher oxidative stability, resulting in increased shelf life of soybean oil food products (Kinney, 1996) and biodiesel (Kinney and Clemente, 2005; Fallen et al., 2012). Although much recent work has occurred in the improvement of soybean fatty acids (Pantalone et al., 2002; Pham et al., 2010; Bilyeu et al., 2011; Boersma et al., 2012; Gillman et al., 2014), there is still a need for continued advancement.

In soybean cultivar development, after crossing segregating parents and developing inbred populations through naturally occurring self-pollination, it is common to evaluate progeny rows derived from inbred single plants based on appearance or phenotypic score for advancement into replicated testing and eventual cultivar release (Fehr, 1987). This approach has worked well for decades, with the average rate of yield increase in soybean from the 1920s to the 1980s estimated to be 15.1 kg ha^{-1} (0.6% yr^{-1}; Specht and William, 1984). However, soybean yield increase trends are still falling short of the levels needed to feed the predicted global population by the year 2050 (Ray et al., 2013). Because of this, it is necessary to explore other techniques for improving soybean yield and other complex traits to adequately provide for consumers of soybean products.

Targeted goals have been achieved for oleic acid and linolenic acid using mutant alleles from relatively few loci (Pham et al., 2010; Bilyeu et al., 2011). However, for oleic acid, there is still concern that environmental variation may result in levels that drop below the industry standard of 800 g kg^{-1} (Fallen et al., 2012; Lee et al., 2012). In such cases, it would be useful to evaluate breeding strategies that could account for a broader range of genetic effects, fine-tuning major effect genes to provide more consistent results. In addition to fatty acids, such approaches would also be worth exploring for oil, protein, and yield improvement.

Quantitative trait loci (QTL)-based selection strategies are inherently biased, as they only account for a limited amount of genetic information. A more robust method such as genomic selection (GS), which accounts for the entire genome (Nakaya and Isobe, 2012), would be worth investigating. First described by Meuwissen et al. (2001), GS is the simultaneous selection of many thousands of markers that densely cover the entire genome, with genes affecting the targeted trait expected to occur in linkage disequilibrium with a subset of genetic markers (Meuwissen, 2007). Numerous studies have explored the potential of GS in animal and plant breeding with evidence of success (Ødegård et al., 2009; Lillehammer et al., 2011; Poland et al., 2012; Resende et al., 2012; Sitzenstock et al., 2013; Crossa et al., 2014; Heslot et al., 2015). Given this potential, there is a need to evaluate the accuracy of GS over multiple generations, rather than only reporting cross-validation results from the same generation, as has been common in crop studies (Jonas and de Koning, 2013).

However, for complex traits with low heritability, GS may be prone to limited success (Nakaya and Isobe, 2012). For marker-assisted selection (MAS) in complex traits, a context specific MAS (CSM) approach can be beneficial for increasing the selection efficiency within target environments by reducing the potential for genotype × environment interaction (Sebastian et al., 2012). In a CSM breeding approach, biparental populations are ideal for training predictions due to the reduced genetic complexity and larger recombination blocks (Sebastian et al., 2012). The greater control exhibited with CSM using a biparental population grown in a limited number of environments can benefit the selection potential for complex traits that are otherwise difficult to improve (Sebastian et al., 2012).

Context-specific MAS has previously demonstrated potential for soybean yield in elite mother line populations (Sebastian et al., 2010). We were interested in trying a similar approach using context-specific GS for other soybean traits in addition to yield, with selections occurring in the progeny row stage, and evaluations occurring in replicated field trials. Thus, the purpose of this research was to evaluate the relative utility for soybean yield, fatty acids, protein, and oil with various GS methods using a CSM approach in comparison with phenotypic selection (PS).

MATERIALS AND METHODS

Plant Materials

In keeping with a CSM approach, a biparental population of 860 F_5–derived recombinant inbred lines (RILs) with both genotypic and phenotypic data was developed from the cross between 'Essex' and 'Williams 82' (hereafter known as E×W-50K). Essex is a Maturity Group (MG) V soybean cultivar with a determinate growth habit, purple flower, and gray pubescence (Smith and Camper, 1973), whereas Williams 82 is an MG III soybean cultivar with indeterminate growth habit, white flower, and tawny pubescence (Bernard and Cremeens, 1988). To provide highly homozygous parental lines for RIL development, a random single plant of each parental line was intentionally selfed for two additional generations. The population was advanced using single-seed descent (Brim 1966).

In 2010, each RIL was grown in Knoxville, TN (35°54′15″ N, 83°57′13″ W) in a single plot consisting of two adjacent rows 6.1 m in length, with the rows spaced 0.8 m apart. Along with the RIL and the parents, four checks with relevant maturities were included in the 2010 field test. The checks were 'LD00-3309' (MG IV-early) (Diers et al., 2006), 'IA4004' (MG IV-early), '5002T' (MG V-early) (Pantalone et al., 2004), and '5601T' (MG V-mid) (Pantalone et al., 2003). Flower color was determined at the R2 growth stage; pubescence color, plant height, and maturity were determined at the R8 growth stage (Fehr and Caviness, 1977).

The 2010 RIL maturity recorded in Julian calendar date ranged from 251 to 288 d (Table 1). To narrow the maturity range for replicated field testing, 276 RIL with maturities ranging from 266 to 273 d (approximately MG IV-mid to IV-late) were chosen for advancement into replicated field trials in 2013. This maturity-based selection fits in with the context-specific approach used in this study, by growing appropriate maturity soybeans for a targeted environment. The MG IV-mid to IV-late range is of primary importance to Tennessee soybean producers, as evidenced by the number of lines tested in this maturity range relative to others in the Soybean Variety Performance Tests in Tennessee (Allen et al., 2011, 2012, 2013).

In 2013, 276 RILs were tested in a randomized complete block design with three replications per environment at three environments (Knoxville, TN [35°54′15″ N, 83°57′13″ W]; Springfield, TN (36°28′12″ N, 86°50′31″ W); and Milan, TN (35°56′3″ N, 88°43′44″ W]), representative of the ecogeographic regions of east, middle, and west Tennessee, respectively. Soil type was primarily Shady loam (fine-loamy, mixed, subactive, thermic Typic Hapludults) and Shady–Whitwell (fine-loamy, siliceous, semiactive, thermic Aquic Hapludults) complex in Knoxville, Dickson (fine-silty, siliceous, semiactive, thermic Glossic Fragiudults) and Staser (fine-loamy, mixed, active, thermic Cumulic Hapludolls) silt loams in Springfield, and Loring (fine-silty, mixed, active, thermic Oxyaquic Fragiudalfs) and Routon (fine-silty, mixed, active, thermic Typic Epiaqualfs) silt loams in Milan. In addition to the RILs and parents, three maturity checks were included: LD00-3309 (MG early-IV), 'LD00-2817P' (MG mid-IV) (Diers et al., 2010), and 'Ellis' (MG late-IV) (Pantalone et al., 2017). As in the 2010 field test, flower color was determined at the R2 growth stage, and pubescence color, plant height, and maturity were determined at the R8 growth stage (Fehr and Caviness, 1977). For both field seasons, plots were harvested at maturity. Yield was measured in kilograms per hectare after adjusting the plot weight to 13% moisture.

Seed Quality Trait Detection

Fatty acid estimates for each plot from the 2010 and 2013 field tests for 16:0, 18:0, 18:1, 18:2, and 18:3 were done using gas chromatography with a procedure described by Spencer et al. (2004). This analysis was performed using an HP 6890 series gas chromatograph (Agilent Technologies) system equipped with a 7683 auto sampler, a 7673 flame ionization detector, and an immobilized 30-m × 0.53-mm-i.d. Agilent DB-23 capillary column with a 0.5-μm fused stationery phase. Fatty acid estimates were obtained as percentage of total seed oil and converted to grams per kilogram of seed oil.

After harvest from the 2010 growing season, ∼25 g of seed from each plot was ground for 20 s in a Knifetec 1095 sample mill (FOSS Tecator) to produce ground whole soybean with a uniform consistency and particle size. Samples were analyzed for protein and oil content using the near-infrared reflectance spectroscopy (NIRS) instrument (NIR 6500, FOSS North America) as described by Panthee et al. (2006), except that for this study, the ground samples were scanned using updated ISIscan software version 2.85 (Infrasoft International, 2007). Plots from the 2013 season were scanned as whole bean samples using a Perten DA 7200 diode array NIRS instrument in collaboration with the University of Minnesota. The calibration equations used for analysis were developed through a cooperative effort between Perten and the University of Minnesota (Bolon et al., 2011). For each NIRS analysis, values for protein and oil concentration were adjusted to grams per kilogram of seed on a dry weight basis.

SNP Genotyping

Single nucleotide polymorphism (SNP) genotyping was performed as described by Smallwood et al. (2017). Briefly, in 2009, samples of DNA were collected from crushed leaves of each F_5 greenhouse single plant from this population at the Soybean Genomics Laboratory at the USDA Beltsville Agricultural Research Center (USDA-ARS) in Beltsville, MD. The DNA samples were analyzed using the Illumina Infinium

Table 1. Simple statistics for soybean population E×W-50K (with parental line Essex and Williams 82) consisting of 860 F_5–derived recombinant inbred lines (RILs) planted in single-replication plots in 2010 in Knoxville, TN. This dataset was used to make performance predictions for traits of interest in a subset of the population (276 RILs) grown in replicated field trials in 2013 at three locations (Knoxville, TN; Springfield, TN; and Milan, TN).

Trait	Essex	Williams 82	Min.	Mean	Max.	SD†
Maturity (Julian d)	278.0	262.0	251.0	270.2	288.0	7.3
Height (cm)	53.3	61.0	25.4	78.5	132.1	20.0
Yield (kg ha^{-1})	2548.8	1566.9	686.0	2137.5	3591.2	528.3
Palmitic (g kg^{-1} seed oil)	107.2	100.3	90.5	106.6	165.0	9.3
Stearic (g kg^{-1} seed oil)	48.5	44.0	32.5	42.4	79.9	4.9
Oleic (g kg^{-1} seed oil)	233.5	237.3	158.4	242.5	353.0	27.7
Linoleic (g kg^{-1} seed oil)	534.4	551.3	436.4	535.6	601.1	22.0
Linolenic (g kg^{-1} seed oil)	76.4	67.1	53.9	72.9	116.6	7.3
Protein (g kg^{-1} seed dry wt.)	430.5	417.0	366.3	412.9	459.5	16.2
Oil (g kg^{-1} seed dry wt.)	217.8	232.8	200.5	226.0	247.4	7.4

† SD of least square means.

beadchip SoySNP50K (Song et al., 2013), with marker positions obtained from the genetic map estimated in Song et al. (2016). Imputations with the 'codeGeno' function in the 'synbreed' package (Wimmer et al., 2012) in R (R Core Team, 2015), using imputation type "beagle" (Beagle Genetic Analysis Software Package version 3.3.1; Browning and Browning 2007, 2009) were used to address missing marker data. Potential genotyping errors were screened using the 'calc.errorlod' function within the 'qtl' package (Broman et al., 2003) in R. To limit the influence of duplicate SNPs, markers were screened for variation among RILs using the 'findDupMarkers' function in the R 'qtl' package (Broman et al., 2003), with one marker randomly chosen from each duplicate set to remain for analysis. After removing duplicate markers, 4867 SNPs remained.

Selection Methods and Statistical Analysis

Genomic selections were performed using the 'BGLR' package (Pérez and de los Campos, 2014) in R (R Core Team, 2015). The 860 RILs planted in single-replication plots simulating progeny rows in Knoxville served as the training population and were used to generate predictions for yield, fatty acids (oleic and linolenic), protein, and oil. Because there were no replicates, plots with missing phenotypic data were dropped from the analysis; thus yield, fatty acids, and protein and oil were tested with 860, 855, and 826 RILs, respectively.

Since this population segregates for maturity at the *E1* locus (Glyma.06g207800; Xia et al., 2012; Wolfgang and An, 2017) and growth habit at the *Dt1* locus (Glyma.19g194300; Tian et al., 2010), SNPs located adjacent to (<25 kbp) *E1* (ss715593840) and *Dt1* (ss715635422 and ss715635423, confirmed by field calls) loci based on the Wm82.a2.v1 genome sequence were used to predict the parental allele. The *E1* and *Dt1* loci were included as covariates in the GS models to minimize any associated variability.

The GS models chosen for analysis were genomic best linear unbiased prediction (G-BLUP) and BayesB (Meuwissen et al., 2001), because they are commonly used and have performed well in previous studies (de los Campos et al., 2013). Both GS and BayesB included the *E1* and Dt1 covariates as fixed effects, as well as 40,000 iterations and a burn in of 10,000. Cross-validations were replicated 50 times for each trait. In each replication, a randomly chosen 20% of the population had phenotypic data removed (test set), whereas phenotypic and genotypic information were retained for the remaining 80% of the population (training set). Since both prediction methods shared the training and test set partitioning for each of the 50 cross-validations, the prediction accuracies (Pearson correlation coefficients) were compared using a paired-*t* test (Pérez and de los Campos, 2014).

An additional selection model was performed using the Epistacy macro version 2.0 (Holland, 1998) in SAS 9.4 (SAS Institute, 2013), with modifications provided by Arnold Saxton. This model was included in the analysis as an effort to account for significant ($P < 0.001$) epistatic interactions that influence yield, fatty acids, protein, and oil. Deviations due to these interactions for each RIL were then summed, divided by the number of SNPs (4867), and added to the mean to predict expected performance.

The performance of each GS method (BayesB, G-BLUP, and Epistacy), along with PS, was then evaluated in the 276 RIL population subset. Many GS studies have sought to evaluate predictions in one growing season, without testing the performance of predictions over time. This is commonly done by subsetting a portion of the population to serve as a training set, predicting the performance of another portion of the population, and then evaluating the predictions using cross-validations (Duhnen et al., 2017). In contrast, we sought to evaluate the accuracy of predictions over multiple generations by making predictions using the whole 860-RIL population from 2010 and validating in a subset of the population using 276 RILs grown in replicated field trials in 2013. To visualize the degree of relationship with the 2013 observed phenotypes, a regression was plotted for each selection against the observed 2013 values in R (R Core Team, 2015). Additionally, the Spearman correlations between each selection method with the observed phenotypes in 2013 were obtained using the 'Hmisc' package in R (Harrell, 2018). In addition, 15% (41 RILs) tail selections chosen using each selection method were evaluated for performance in the 2013 field season by calculating the realized gain compared with the population mean. Finally, the selection efficiency was estimated by comparing the 15% tail selection from each method with the 15% tail selection based on the observed 2013 phenotype using the formula displayed below, where S is the selection efficiency, B is the number selected in the alternate system (2010 selection method), C is the number expected by chance, and A is the chosen selection system (2013 observed rankings) (Hamblin and Zimmerman, 1986):

$$S = \left(\frac{B - C}{A - C}\right)100$$

To determine the impact of marker density and population size on selection accuracy, additional G-BLUP analyses were performed with randomly chosen SNPs and/or RILs removed from the prediction model. The SNP marker densities used were 4867, 3867, 2867, 1867, and 867. The population sizes chosen were 860, 714, 568, 422, and 276 for yield; 855, 709, 566, 420, and 275 for fatty acids; and 826, 686, 551, 405, and 271 for protein and oil. The different population sizes for each of the traits were due to missing data from the 2010 field season. Each combination of marker and RIL density was used for a separate G-BLUP analysis, for a total of 25 analyses per trait. The results from these G-BLUP predictions were then compared with the 2013 phenotypic results using Spearman correlations.

Least square means were obtained from statistical analysis performed in SAS PROC GLIMMIX (SAS Institute, 2013). The model used for analysis was a randomized complete block design, with RIL as a fixed term; location, replicate (location), and RIL × location as random terms; and denominator df method set to residual. Contrast statements were used to compare different genotypic classes for the *Dt1* stem termination locus and the *E1* maturity locus. In addition, a model with no fixed terms and RIL, location, replicate (location), and RIL × location as random terms was run to obtain the variance for each term. These variances were then used to estimate heritability on an entry-means basis (Nyquist, 1991).

RESULTS

Yield, fatty acids, protein, oil, maturity, and height were all measured in the E×W-50K 860-RIL soybean population

grown in Knoxville in 2010. Variability among RILs within the population was observed for each trait, although this was not supported by statistical analysis, since only a single field replication was performed (Table 1). However, the 2010 nonreplicated field test provided an ideal opportunity to evaluate GS and PS methods for advancement from a progeny row stage into multilocation replicated field trials, which is a routine practice in soybean breeding. A subset of 276 RILs selected based on maturity were advanced into replicated field trials in 2013 to minimize the effect of maturity on the traits analyzed in this research (Table 1). The simple statistics for the 2013 field test are displayed in Table 2, with each trait displaying a significant difference ($P < 0.05$) among RILs. Estimates of the effect of the *Dt1* stem termination locus and *E1* maturity locus were also performed. There were 149 determinate, 114 indeterminate, and 13 segregating RILs in this population for stem termination. The differences between indeterminate and determinate RILs were significant ($P < 0.001$) for each trait. After the population was subset based on maturity, there were 245 RILs with the *E1* genotype, four with the *e1-as* genotype, and 27 segregating at the *E1* maturity locus. This segregation distortion was due to selection for appropriate photoperiod response imposed on the original 860 RILs. With the exception of palmitic, stearic, and linolenic acids ($P > 0.05$), the differences between the *E1* and *e1-as* genotypes were significant ($P < 0.05$) for each trait.

Although the magnitude of the phenotypic variances were smaller than seen with induced mutant studies (Pham et al., 2010; Bilyeu et al., 2011, 2018; Boersma et al., 2012; Gillman et al., 2014), they were still of interest for evaluating the whole-genome selection methods in this study. Of the traits chosen for selection strategy evaluation, three different phenotyping methods were used: recorded seed mass per unit area (yield), gas chromatography (fatty acids), and NIRS predictions of seed components (protein and oil). These traits displayed a range of heritability values (Table 2), with yield the lowest at 0.63, followed by NIRS traits (0.87), and finally gas chromatography traits (0.90–0.94), indicating differing potential of gains from selection. Thus, it was of interest to evaluate different selection strategies for yield, fatty acids, protein, and oil.

For the four different selection methods chosen (PS, BayesB, G-BLUP, and Epistacy), an initial comparison was performed using BayesB and G-BLUP with a cross-validation approach. Following this approach, BayesB was able to more accurately match ($P < 0.001$) the 2010 phenotypic

Table 3. Comparison of cross-validations for BayesB and genomic best linear unbiased prediction (G-BLUP) methods of genomic selection for soybean population E×W-50K (with parental line Essex and Williams 82) consisting of 860 F_5–derived recombinant inbred lines grown in 2010 at Knoxville, TN. Cross-validations were replicated 50 times for each trait. In each replication, a randomly chosen 1/5 of the population had phenotypic data removed (test set), whereas phenotypic and genotypic information were retained for the remaining 4/5 of the population (training set). The values displayed for BayesB and G-BLUP are the mean prediction accuracies (Pearson correlation coefficients) for the predicted and observed values in the test set.

Trait	BayesB	G-BLUP	Difference	*P* value
Yield	0.5145	0.4862	0.0283	***
Oleic	0.6442	0.6417	0.0025	***
Linolenic	0.5029	0.5020	0.0009	***
Protein	0.6718	0.6715	0.0003	NS†
Oil	0.5694	0.5583	0.0111	***

*** Significant at the 0.001 probability level.

† NS, not significant at the 0.05 probability level.

Table 2. Simple statistics for soybean population E×W-50K subset (with parental line Essex and Williams 82) consisting of 276 F_5–derived recombinant inbred lines (RILs) planted in replicated field trials at three locations in 2013 (Knoxville, TN; Springfield, TN; and Milan, TN). Information from this dataset was compared with performance predictions for traits of interest in the full population (860 RILs) grown in 2010 in single-replication plots planted at Knoxville.

Trait	Genotype *P* value	G×E† *Z* value	Essex	Williams 82	Min.	Mean	Max.	SD‡	LSD value	h^2§
Maturity (Julian)	***	***	272.2	262.6	259.4	270.4	276.9	2.8	3.5	0.79
Height (cm)	***	***	75.6	94.0	37.8	89.1	133.5	18.8	11.7	0.95
Yield (kg ha^{-1})	***	***	3588.9	3002.9	1371.6	3222.9	4087.6	395.3	663.4	0.63
Palmitic (g kg^{-1} seed oil)	***	NS¶	108.9	98.4	92.7	105.3	117.3	4.6	4.2	0.90
Stearic (g kg^{-1} seed oil)	***	**	38.9	36.6	31.5	37.6	47.2	3.2	2.2	0.94
Oleic (g kg^{-1} seed oil)	***	***	233.3	279.1	178.1	240.7	358.5	35.7	24.8	0.94
Linoleic (g kg^{-1} seed oil)	***	***	545.0	521.6	447.6	545.6	592.4	28.8	21.8	0.93
Linolenic (g kg^{-1} seed oil)	***	*	73.9	64.4	54.9	70.8	87.8	6.4	5.4	0.91
Protein (g kg^{-1} seed dry wt.)	***	***	423.6	421.6	376.3	410.5	444.0	12.0	12.2	0.87
Oil (g kg^{-1} seed dry wt.)	***	***	212.0	227.4	200.5	218.3	238.3	5.9	6.0	0.87

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.† G×E, genotype × environment.

‡ SD of least square means.

§ Heritability calculated using entry-means basis (Nyquist, 1991).

¶ NS, not significant at the 0.05 probability level.

data for each trait except protein ($P > 0.05$) (Table 3). The cross-validation accuracies ranged from 0.5029 to 0.6718 for BayesB and 0.4862 to 0.6715 for G-BLUP, with none of the differences between the estimates from the two methods for individual traits exceeding 0.0283.

Regression plots are displayed for each trait and selection method (Fig. 1–5) to visualize the relationship between the 2010 predictions and the 2013 observed phenotypes. For yield, each selection method displayed a weak relationship with the 2013 observed phenotypes, with R^2 values ranging from 0.041 (BayesB) to 0.058 (Epistacy) (Fig. 1). This trend did not continue for the other traits, with Epistacy displaying the lowest R^2 values in relation to the 2013 observed phenotypes (Fig. 2–5). The fatty acid predictions with BayesB and G-BLUP for oleic ($R^2 = 0.73$, Fig. 2) and linolenic ($R^2 = 0.68$, (Fig. 3) were much more closely aligned with the observed 2013 phenotypes in comparison with yield. Phenotypic selection was not able to predict the 2013 phenotypes as well as BayesB and G-BLUP for oleic ($R^2 = 0.58$) or linolenic ($R^2 = 0.42$) fatty acids (Fig. 2–3). The R^2 values for oil and protein were somewhat lower than those for the fatty acids, with values for protein (Fig. 4) and oil (Fig. 5) ranging from 0.095 to 0.29 and 0.16 to 0.33, respectively. There was not much difference between PS, BayesB, and G-BLUP for these traits, with Epistacy as the least capable predictor (Fig. 4–5).

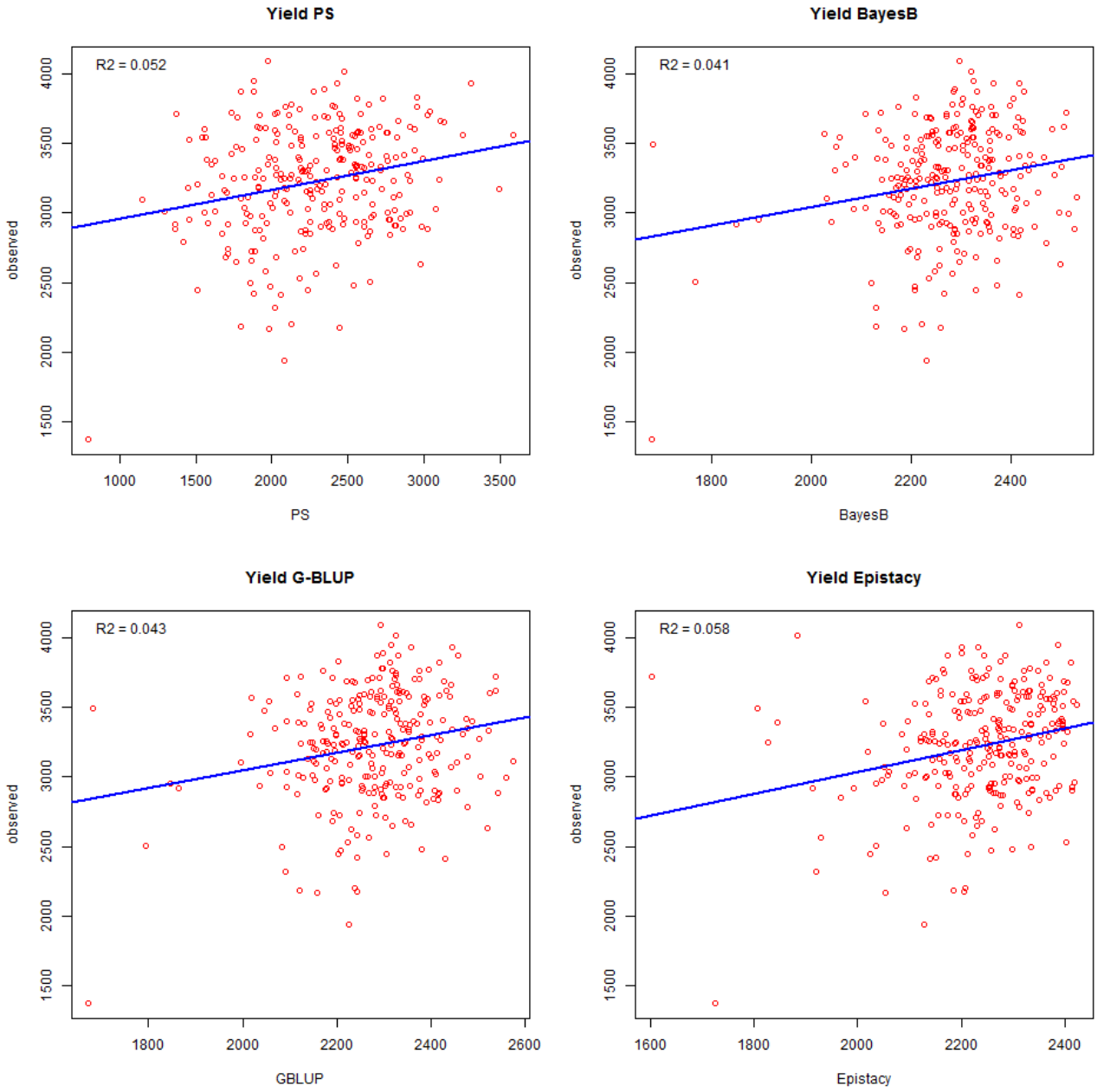

Fig. 1. Yield (kg ha^{-1}) performance comparisons between 2010 predictions (*x* axis) and 2013 phenotypes (*y* axis) in a soybean population E×W-50K subset consisting of 276 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) strategies. Predictions with higher R^2 were more closely related to 2013 observed phenotypes.

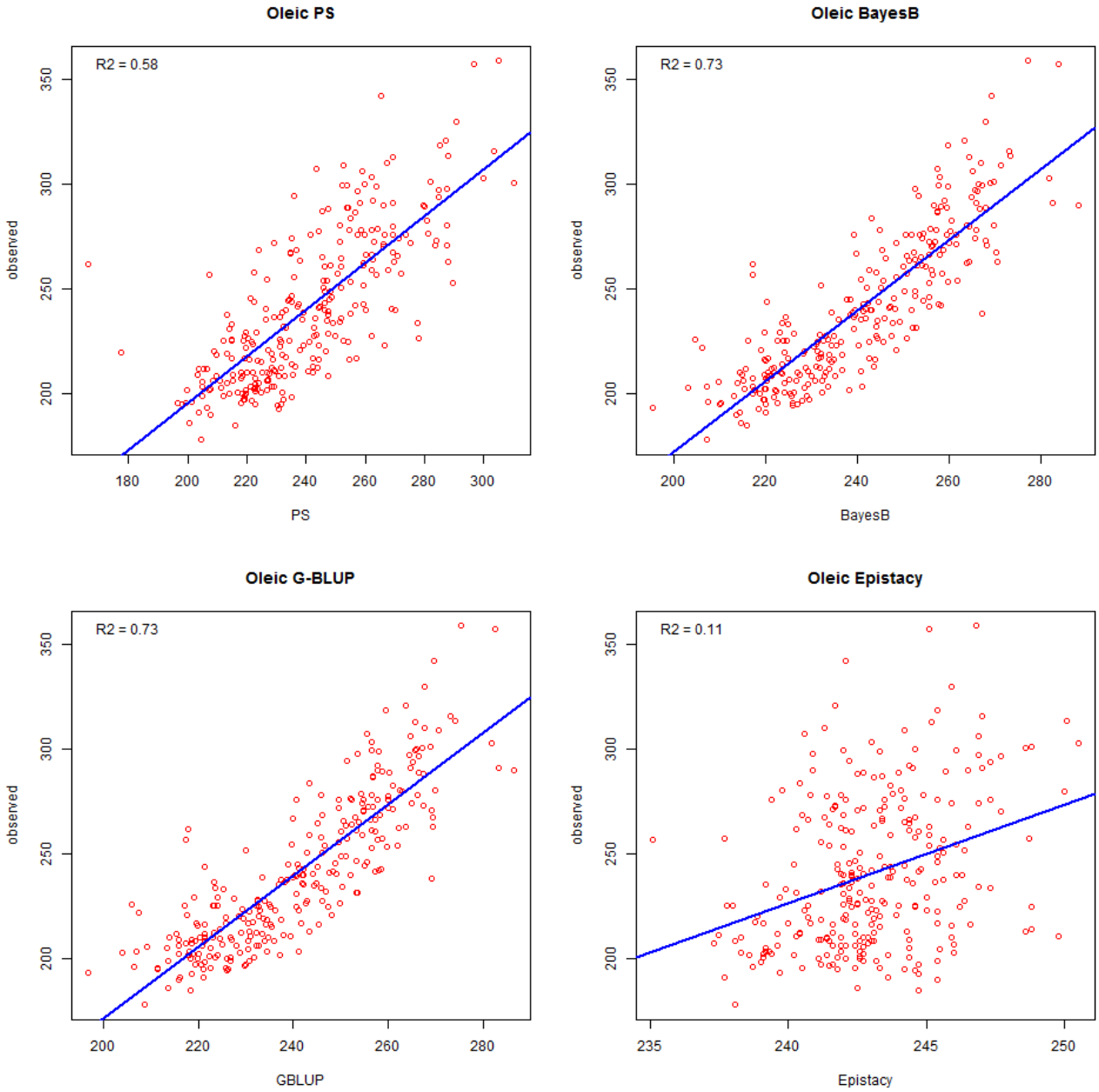

Fig. 2. Oleic acid (g kg^{-1}) performance comparisons between 2010 predictions (*x* axis) and 2013 phenotypes (*y* axis) in a soybean population E×W-50K subset consisting of 275 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) strategies. Predictions with higher R^2 were more closely related to 2013 observed phenotypes.

Spearman correlations were performed between the 2013 observed phenotypes and each selection method based on 2010 data for yield, oleic acid, linolenic acid, protein, and oil (Table 4). For yield, the Spearman correlations ranged from 0.13 (BayesB) to 0.22 (Epistacy). For each other trait, Epistacy had the lowest correlation with the 2013 phenotype. BayesB (0.87) had the highest correlation with 2013 phenotype for oleic acid, closely followed by G-BLUP (0.86). BayesB and G-BLUP tied for the highest correlation with the 2013 phenotype for both linolenic acid (0.83) and protein (0.49), whereas PS (0.59) had the highest correlation with the 2013 phenotype for oil.

An additional comparison of selection methods was performed by calculating the realized gain based on 15% tail selections for yield, oleic acid, linolenic acid, protein, and oil (Table 5). The tail selections were performed in the direction appropriate for improvement of each trait, with high tail selections for yield, oleic acid, protein, and oil, and low tail selections for linolenic acid. For yield, the realized gains ranged from 0.5 (G-BLUP) to 4.5% (Epistacy). As with the other comparisons, Epistacy was only successful as a predictor for yield, ranking last for each other trait. For oleic acid (21.7%), linolenic acid (−11.6%), and oil (2.4%), G-BLUP was able to achieve the most realized gain, whereas for protein, BayesB (2.6%) had the most realized gain (Table 5).

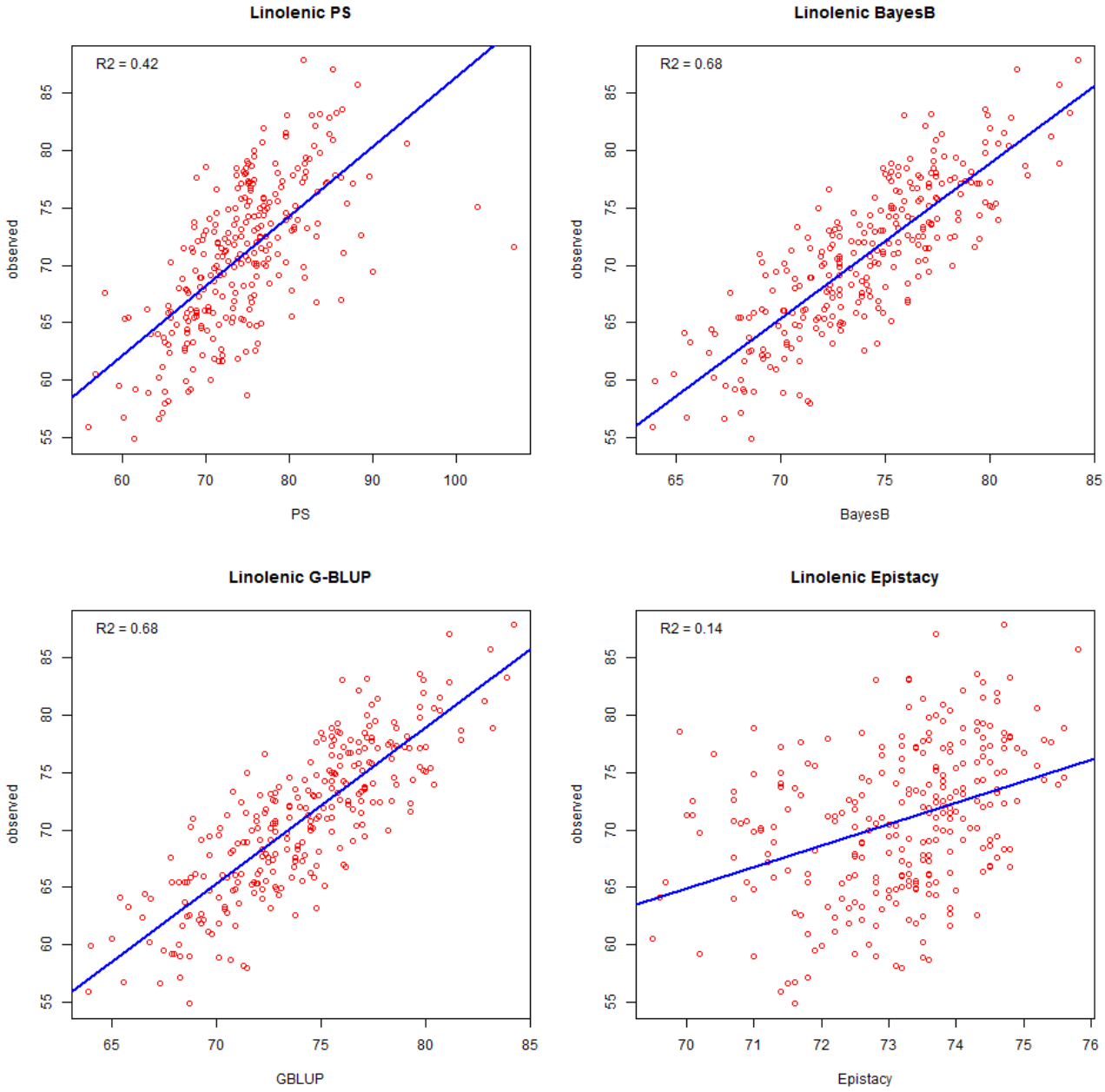

Fig. 3. Linolenic acid (g kg^{-1}) performance comparisons between 2010 predictions (*x* axis) and 2013 phenotypes (*y* axis) in a soybean population E×W-50K subset consisting of 275 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) selection strategies. Predictions with higher R^2 were more closely related to 2013 observed phenotypes.

Finally, a comparison of selection efficiency was done to compare the 15% tail selection from each selection method with the 15% tail based on the observed 2013 phenotype (Table 5). For each trait, Epistacy had the lowest or tied for the lowest selection efficiency. For yield, PS and BayesB were tied with the highest selection efficiency at 8.6%. For oleic acid (62.9%) and protein (40%), GBLUP had the highest selection efficiency, whereas for linolenic acid (57.1%) and oil (28.6%), BayesB and GBLUP were tied with the highest selection efficiency (Table 5).

Additional G-BLUP analyses were performed to determine the impact of marker density and population sized on selections. Based on Spearman correlations between these predictions and the 2013 phenotypic data, there were minimal differences between predictions with varying marker densities and population sizes for oleic acid, linolenic acid, protein, and oil (Table 6). However, for yield, there was a noticeable trend toward higher correlations when lower population sizes were evaluated, with the best predictions occurring when only the RILs grown in 2013 were used in the 2010 prediction (Table 6). For each trait, the correlation differences based on marker density were minimal, with each SNP density producing similar predictions (Table 6).

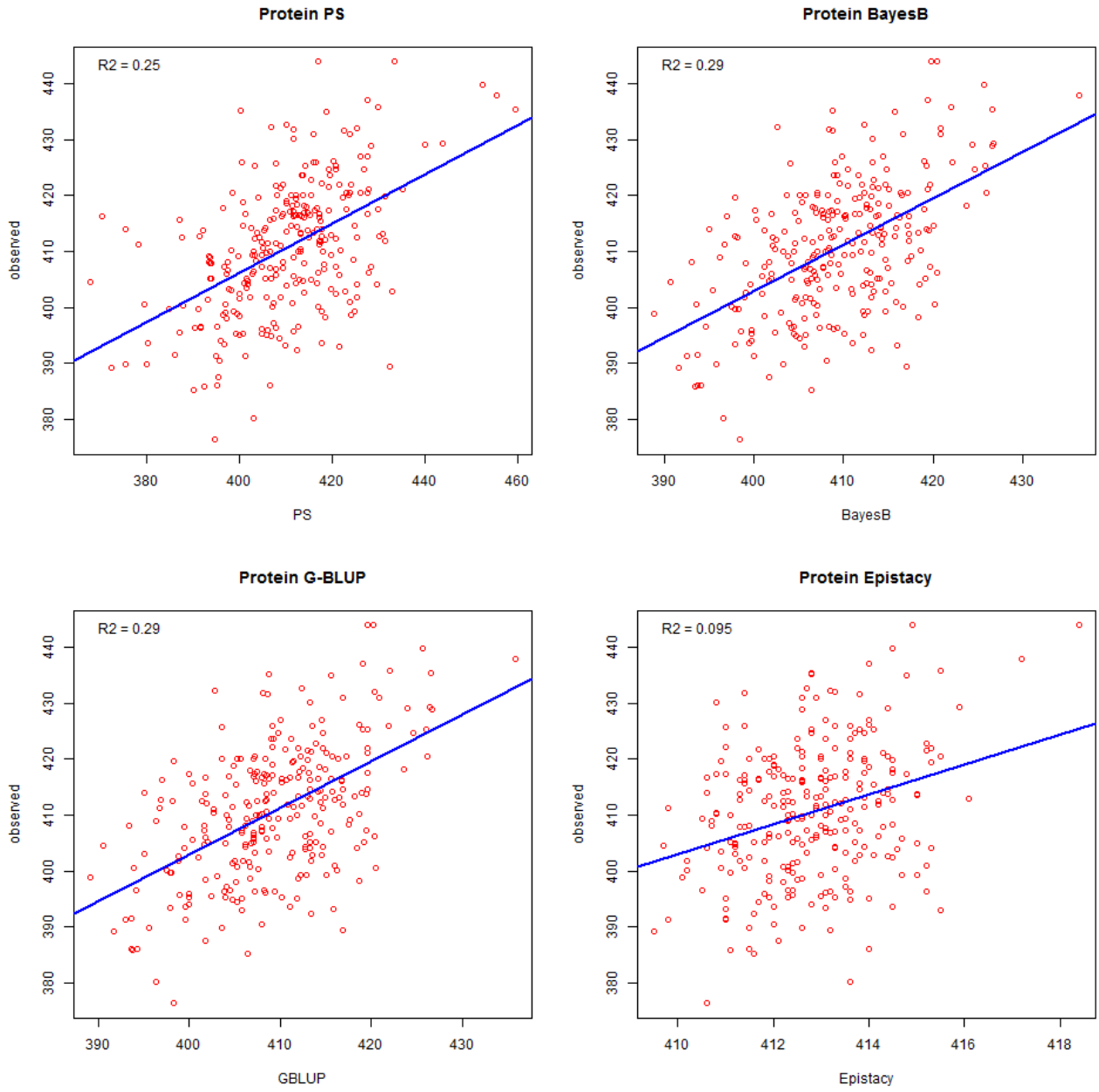

Fig. 4. Protein (g kg^{-1}) performance comparisons between 2010 predictions (*x* axis) and 2013 phenotypes (*y* axis) in a soybean population E×W-50K subset consisting of 275 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) selection strategies. Predictions with higher R^2 were more closely related to 2013 observed phenotypes.

DISCUSSION

Evaluating selection methods for soybean traits of interest is an important strategy for determining how best to make improvements. In this study, four selection methods (PS, BayesB, G-BLUP, and Epistacy) were evaluated for soybean yield, oleic acid, linolenic acid, protein, and oil. Duhnen et al. (2017) found similar results using G-BLUP cross-validations to the values reported in this study for yield (0.49) and protein (0.67) (Table 3), with prediction accuracies for yield ranging from 0.45 to 0.63 and for protein from 0.45 to 0.59. Using a similar cross-validation approach, Jarquín et al. (2014) estimated a 0.64 prediction accuracy for soybean yield. Although many crop studies have evaluated GS in the same generation with cross-validations (Jarquín et al., 2014; Duhnen et al., 2017), this study tested the effect of selections across generations as recommended by Jonas and de Koning (2013). By doing so, valuable insight was gained into which selection methods were preferable for moderate- (yield), high- (protein and oil), and very high-heritability (oleic and linolenic acids) traits (Tables 2 and 4–5).

For yield, Epistacy was the preferred selection method based on each comparison except for selection efficiency (Tables 4–5, Fig. 1). This differs sharply from each of the other traits, for which Epistacy was the least effective selection method (Tables 4–5, Fig. 2–5). Duhnen et al. (2017)

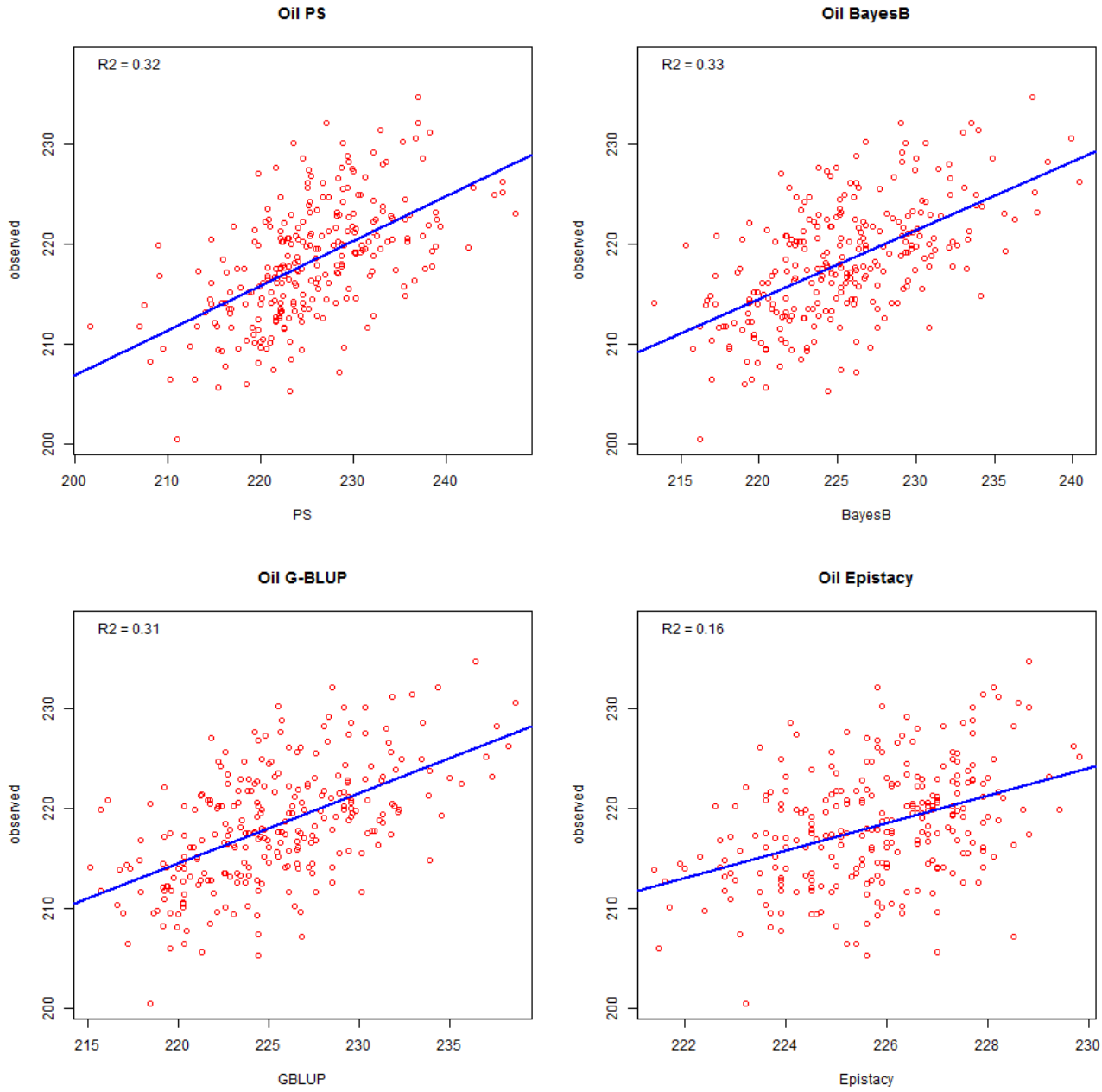

Fig. 5. Oil (g kg^{-1}) performance comparisons between 2010 predictions (*x* axis) and 2013 phenotypes (*y* axis) in a soybean population E×W-50K subset consisting of 275 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) selection strategies. Predictions with higher R^2 were more closely related to 2013 observed phenotypes.

noted that prediction accuracy for yield was improved by including epistasis into a G-BLUP model. With yield having a lower heritability than protein and oil (Table 2), it is surprising to note that Epistacy produced a greater realized gain for yield than any selection method for protein or oil (Table 5). Given these findings, along with the many known QTL that influence soybean yield, further testing using an epistatic approach for yield is warranted.

Soybean heritability for yield has been demonstrated to be lower than for protein and oil (Wiggins et al., 2018), as well as for fatty acids (Smallwood et al., 2017). These differences in heritability, along with the possibility of greatly influencing fatty acid traits based on few loci (Pantalone et al., 2002; Pham et al., 2010; Bilyeu et al., 2011; Boersma et al., 2012; Gillman et al., 2014) with no comparable studies for yield, demonstrate the highly quantitative nature of soybean yield and subsequently highlight the extreme challenge in making selections for yield improvement. As noted by Nakaya and Isobe (2012), GS methods for low-heritability traits may be prone to limited success, which matched our findings in this study, with BayesB and G-BLUP as the worst selection methods for yield (Tables 4–5; Fig. 1).

In contrast with yield, little previous work has been done testing GS for soybean fatty acids. As oleic acid and linolenic acid displayed the highest heritability of the traits

Table 4. Spearman correlations between 2013 phenotypic data and 2010 predictions for soybean yield, linolenic acid, oleic acid, protein, and oil in a population E×W-50K subset (with parental line Essex and Williams 82) consisting of 276 F_5–derived recombinant inbred lines. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) strategies.

Trait	Observed	PS	BayesB	GBLUP	Epistacy
Yield					
Observed	–	0.18	0.13	0.14	0.22
PS	**	–	0.38	0.40	0.30
BayesB	*	***	–	0.99	0.32
GBLUP	*	***	***	–	0.32
Epistacy	***	***	***	***	–
Oleic acid					
Observed	–	0.78	0.87	0.86	0.31
PS	***	–	0.82	0.81	0.39
BayesB	***	***	–	1.00	0.38
GBLUP	***	***	***	–	0.38
Epistacy	***	***	***	***	–
Linolenic acid					
Observed	–	0.68	0.83	0.83	0.41
PS	***	–	0.74	0.74	0.52
BayesB	***	***	–	1.00	0.55
GBLUP	***	***	***	–	0.55
Epistacy	***	***	***	***	–
Protein					
Observed	–	0.48	0.49	0.49	0.25
PS	***	–	0.66	0.66	0.55
BayesB	***	***	–	1.00	0.49
GBLUP	***	***	***	–	0.49
Epistacy	***	***	***	***	–
Oil					
Observed	–	0.59	0.56	0.54	0.38
PS	***	–	0.75	0.72	0.54
BayesB	***	***	–	0.99	0.66
GBLUP	***	***	***	–	0.67
Epistacy	***	***	***	***	–

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.

tested in this study (Table 2), it is not surprising that GS predictions were most accurate for these traits (Nakaya and Isobe, 2012). Notably, for every indicator analyzed in this study, BayesB and G-BLUP outperformed PS for these fatty acid traits (Tables 4–5, Fig. 2–3). This indicates strong potential for GS methods for advancing fatty acid traits at the progeny row stage.

For protein and oil, there was little difference in selection accuracy for PS, BayesB, and G-BLUP (Tables 4–5, Fig. 4–5). These findings are concordant with the findings of Duhnen et al. (2017), where little difference was observed between Bayesian and G-BLUP models. However, this differs from Clark et al. (2011), in which BayesB was noted to predict more accurately than G-BLUP. Although PS was comparable with the GS methods for protein and oil, it should be noted that GS methods offer the opportunity to increase gain more rapidly by making selections in

Table 5. Realized gain and selection efficiency for yield, oleic acid, linolenic acid, protein, and oil from 15% tail selections (41 recombinant inbred lines [RILs]) from 2010 predictions as measured in 2013 from soybean population E×W-50K subset (with parental line Essex and Williams 82) consisting of 276 F_5–derived RILs. The 2010 predictions were estimated with phenotypic (PS) and genomic selection (GS) (Epistacy, BayesB, and genomic best linear unbiased prediction [G-BLUP]) strategies. The method with the largest gain in the desired direction for each trait is displayed with bolded text.

		PS			BayesB			G-BLUP			Epistacy		
Trait	Mean	15% tail selection	Realized gain†	Selection efficiency‡	15% tail selection	Realized gain†	Selection efficiency‡	15% tail selection	Realized gain†	Selection efficiency‡	15% tail selection	Realized gain†	Selection efficiency‡
			%			%			%			%	
Yield (kg ha^{-1})	3222.9	3284.6	1.9	8.6	3272.3	1.5	8.6	3238.7	0.5	5.7	3368.4	**4.5**	0.0
Oleic (g kg^{-1} seed oil)	240.7	284.7	18.3	40.0	292.7	21.6	60.0	293.1	**21.7**	62.9	264.0	9.7	25.7
Linolenic (g kg^{-1} seed oil)	70.8	62.9	–11.2	48.6	62.7	–11.4	57.1	62.6	**–11.6**	57.1	68.2	–3.6	0.0
Protein (g kg^{-1} seed dry wt.)	410.5	418.7	2.0	25.7	420.9	**2.6**	37.1	420.7	2.5	40.0	415.6	1.3	11.4
Oil (g kg^{-1} seed dry wt.)	218.3	223.3	2.3	22.9	223.3	2.3	28.6	223.6	**2.4**	28.6	222.1	1.7	22.9

† Percentage increase of tail selection in comparison with the population mean.

‡ Percentage of tail selection for this selection strategy overlapping noncoincidentally with tail selection based on observed 2013 phenotypes.

nontarget winter nurseries. Additionally, in a progeny row scenario, selections using GS can be made prior to harvest, improving the efficiency in comparison with PS.

When considering GS approaches, it is important to determine functional levels of marker and population densities to make the best predictions. Although this study used the Infinium beadchip SoySNP50K (Song et al., 2013) for genotyping, many soybean studies have begun genotyping with the less dense BARCSoySNP6k array (Song et al., 2014). When using very dense marker arrays, many of the markers could map to the same genetic location based on limited recombination. This redundancy occurred in this population, with an initial 11,633 SNPs being reduced to 4867 after removing those in duplicate locations. Given the findings in this study, G-BLUP was not largely affected by dropping from 4867 SNPs down to 867 SNPs (Table 6). Likewise, with the exception of yield, decreasing the number of RILs in the training population had little effect on prediction accuracy (Table 6). For yield, the predictions were most accurate when the training population and the test population were identical. In contrast with this study, Zhang et al. (2017) observed increased GS prediction accuracy for maize (*Zea mays* L.) with increased training population and marker densities. Continued refinement of training population and marker densities will be essential for maximizing the efficiency of GS in soybean breeding operations.

Table 6. Spearman correlations between 2013 phenotypic data and 2010 genomic best linear unbiased prediction (G-BLUP) predictions for soybean yield, linolenic acid, oleic acid, protein, and oil in a population E×W-50K subset consisting of 276 F_5–derived recombinant inbred lines (RILs). The G-BLUP analyses were performed with randomly chosen single nucleotide polymorphisms (SNPs) and/or RILs removed from the prediction model. The marker densities used were 4867, 3867, 2867, 1867, and 867 SNPs. The RIL densities chosen were 860, 714, 568, 422, and 276 for yield; 855, 709, 566, 420, and 275 for fatty acids; and 826, 686, 551, 405, and 271 for protein and oil.

Trait	No. of SNPs	No. of RILS				
		860	714	568	422	276
Yield	4867	0.14	0.19	0.28	0.25	0.37
	3867	0.14	0.19	0.28	0.24	0.37
	2867	0.14	0.19	0.27	0.24	0.36
	1867	0.14	0.19	0.28	0.24	0.37
	867	0.13	0.19	0.28	0.24	0.38
		855	**709**	**566**	**420**	**275**
Oleic acid	4867	0.86	0.87	0.86	0.86	0.86
	3867	0.86	0.87	0.86	0.86	0.86
	2867	0.86	0.86	0.86	0.86	0.86
	1867	0.87	0.87	0.86	0.86	0.86
	867	0.86	0.86	0.85	0.85	0.85
		855	**709**	**566**	**420**	**275**
Linolenic acid	4867	0.83	0.84	0.83	0.82	0.83
	3867	0.83	0.84	0.83	0.82	0.83
	2867	0.83	0.84	0.83	0.82	0.83
	1867	0.83	0.84	0.83	0.82	0.83
	867	0.82	0.84	0.83	0.81	0.83
		826	**686**	**551**	**405**	**271**
Protein	4867	0.49	0.47	0.49	0.51	0.52
	3867	0.48	0.47	0.49	0.50	0.51
	2867	0.48	0.47	0.49	0.50	0.51
	1867	0.49	0.47	0.49	0.51	0.52
	867	0.48	0.46	0.49	0.50	0.52
		826	**686**	**551**	**405**	**271**
Oil	4867	0.54	0.53	0.54	0.53	0.54
	3867	0.54	0.52	0.53	0.52	0.53
	2867	0.52	0.51	0.53	0.52	0.53
	1867	0.55	0.53	0.54	0.53	0.54
	867	0.51	0.51	0.52	0.53	0.53

CONCLUSIONS

Breeding method evaluation is an important strategy in maximizing genetic gains from selection. Given the importance of yield, fatty acids, protein, and oil in soybean production, it is necessary to determine the most useful approaches for trait improvement. Additionally, it is of interest to evaluate different selection strategies from the progeny row stage, as this is a critical step in the soybean breeding pipeline. In this study, we compared the relative utility of both PS and context-specific GS methods (BayesB, G-BLUP, and Epistacy). Although there was not a consensus best strategy for all traits tested, it is notable that for each trait, the preferred approach was a GS strategy. Epistacy was the best method for yield, which may indicate the importance of epistatic interactions for this trait. BayesB and/or G-BLUP were the preferred methods for all other, higher heritability traits. Yield was the only trait for which the predictions had a large change when the number of SNPs or RILs was reduced for the G-BLUP model, with the best predictions occurring when the training population and the test population were identical (Table 6). These findings provide important information on how soybean breeders can maximize selections from the progeny row stage for yield, fatty acids, protein, and oil by using appropriate breeding strategies.

Conflict of Interest

The authors declare that there is no conflict of interest.

Disclaimer

Mention of any trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that may also be suitable. The USDA is an equal opportunity provider and employer.

References

Allen, F.L., R. Johnson, R.C. Williams, Jr., A.T. McClure, M. Newman, and P. Donald. 2011. Soybean variety performance tests in Tennessee. Univ. Tennessee, Knoxville.

Allen, F.L., R. Johnson, R.C. Williams, Jr., A.T. McClure, M. Newman. H. Young-Kelly, and P. Donald. 2012. Soybean variety performance tests in Tennessee. Univ. Tennessee, Knoxville.

Allen, F.L., V.R. Sykes, R.C. Williams, Jr., A.T. McClure, H. Young-Kelly, and P. Donald. 2013. Soybean variety performance tests in Tennessee. Univ. Tennessee, Knoxville.

Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. Crop Sci. 28:1027–1028. doi:10.2135/cropsci1988.0011183X002800060049x

Bilyeu, K., J.D. Gillman, and A.R. LeRoy. 2011. Novel *FAD3* mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Sci. 51:259–264. doi:10.2135/cropsci2010.01.0044

Bilyeu, K., M. Skrabisova, D. Allen, I. Rajcan, D. Palmquist, A. Gillen, et al. 2018. The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. J. Am. Oil Chem. Soc. 95:39–49. doi:10.1002/aocs.12025

Boersma, J.G., J.D. Gillman, K.D. Bilyeu, G.R. Ablett, C. Grainger, and I. Rajcan. 2012. New mutations in a *Delta-9-Stearoyl-Acyl Carrier Protein Desaturase* gene associated with enhanced stearic acid levels in soybean seed. Crop Sci. 52:1736–1742. doi:10.2135/cropsci2011.08.0411

Bolon, Y., W.J. Haun, W.W. Xu, D. Grant, M.G. Stacey, R.T. Nelson, et al. 2011. Phenotypic and genomic analyses of fast neutron mutant population resource in soybean. Plant Physiol. 156:240–253. doi:10.1104/pp.110.170811

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220. doi:10.2135/cropsci1966.0011183X000600020041x

Broman, K.W., H. Wu, Ś. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. doi:10.1093/bioinformatics/btg112

Browning, B.L., and S.R. Browning. 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210–223. doi:10.1016/j.ajhg.2009.01.005

Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987

Clark, S.A., J.M. Hickey, and J.H.J. van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol. 43:18. doi:10.1186/1297-9686-43-18

Cober, E.R., S.R. Cianzio, V.R. Pantalone, and I. Rajcan. 2009. Soybean. In: J. Vollman and I. Rajcan, editors, Oil crops: Handbook of plant breeding. Vol. 4. Springer, New York. p. 57–90. doi:10.1007/978-0-387-77594-4_3

Crossa, J., P. Pérez, J. Hickey, J. Burgeño, L. Ornella, J. Cerón-Rojas, et al. 2014. Genomic prediction in CYMMIT maize and wheat breeding programs. Heredity 112:48–60. doi:10.1038/hdy.2013.16

de los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler, and M.P.L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. doi:10.1534/genetics.112.143313

Diers, B.W., T.R. Cary, D.J. Thomas, A. Colgrove, and T. Niblack. 2010. Registration of 'LD00-2817P' germplasm line with resistance to soybean cyst nematode from PI 437654. J. Plant Reg. 4:141–144. doi:10.3198/jpr2009.09.0546crg

Diers, B.W., T.R. Cary, D.J. Thomas, and C.D. Nickell. 2006. Registration of 'LD00-3309' soybean. Crop Sci. 46:1384. doi:10.2135/cropsci2005.06.0164

Duhnen, A., A. Gras, S. Teyssèdre, M. Romestant, B. Claustres, J. Daydé, and B. Mangin. 2017. Genomic selection for yield and seed protein content in soybean: A study of breeding program data and assessment of prediction accuracy. Crop Sci. 57:1325–1337. doi:10.2135/cropsci2016.06.0496

Fallen, B.D., K. Rainey, C.E. Sams, D.A. Kopsell, and V.R. Pantalone. 2012. Evaluation of agronomic and seed characteristics in elevated oleic acid soybean lines in the south-eastern US. J. Am. Oil Chem. Soc. 89:1333–1343. doi:10.1007/s11746-012-2026-x

Fehr, W.R. 1987. Soybean. In: W.R. Fehr, editor, Principles of cultivar development, Vol. 2: Crop species. Macmillan Publ. Co., New York. p. 533–576

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 3-1977. Agric. Home Econ. Exp. Stn., Iowa State Univ., Ames.

Gillman, J.D., M.G. Stacy, Y. Cui, H.R. Berg, and G. Stacey. 2014. Deletions of the *SACPD-C* locus elevate seed stearic acid but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol. 14:143. doi:10.1186/1471-2229-14-143

Harrell, F.E., Jr. 2018. Hmisc: Harell Miscellaneous. R package version 4.1-1. R Found. Stat. Comput. https://CRAN.R-project.org/package=Hmisc (accessed 16 Feb. 2018).

Hamblin, J., and M.J.O. Zimmerman. 1986. Breeding common bean for yield in mixtures. In: J. Janick, editor, Plant breeding reviews. Vol. 4. AVI Publ. Co., Westport, CT. p. 245–272. doi:10.1002/9781118061015.ch8

Heslot, N., J.-L. Jannink, and M.E. Sorrells. 2015. Perspectives for genomic selection applications and research in plants. Crop Sci. 55:1–12. doi:10.2135/cropsci2014.03.0249

Holland, J.B. 1998. EPISTACY: A SAS program for detecting two-locus epistatic interactions using genetic marker information. J. Hered. 89:374–375. doi:10.1093/jhered/89.4.374

Infrasoft International. 2007. ISIscan software. Release 2.85. Infrasoft Int., State College, PA.

Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15:740. doi:10.1186/1471-2164-15-740

Jonas, E., and D.J. de Koning. 2013. Does genomic selection have a future in plant breeding? Trends Biotechnol. 31:497–504. doi:10.1016/j.tibtech.2013.06.003

Kinney, A.J. 1996. Development of genetically engineered soybean oils for food application. J. Food Lipids 3:273–292. doi:10.1111/j.1745-4522.1996.tb00074.x

Kinney, A.J., and T.E. Clemente. 2005. Modifying soybean oil for enhanced performance in biodiesel blends. Fuel Process. Technol. 86:1137–1147. doi:10.1016/j.fuproc.2004.11.008

Kris-Etherton, P.M., and S. Yu. 1997. Individual fatty acid effects on plasma lipids and lipoproteins: Human studies. Am. J. Clin. Nutr. 65:1628S–1644S. doi:10.1093/ajcn/65.5.1628S

Lee, J.D., K.D. Bilyeu, V.R. Pantalone, A.M. Gillen, Y.S. So, and J.G. Shannon. 2012. Environmental stability of oleic acid concentration in seed oil for soybean lines with *FAD2-1A* and *FAD2-1B* mutant genes. Crop Sci. 52:1290–1297. doi:10.2135/cropsci2011.07.0345

Lillehammer, M., T.H.E. Meuwissen, and A.K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. J. Dairy Sci. 94:493–500. doi:10.3168/jds.2010-3518

Meuwissen, T. 2007. Genomic selection: Marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. 124:321–322. doi:10.1111/j.1439-0388.2007.00708.x

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.

Morrison, M.J., E.K. Cober, M.F. Saleem, N.B. McLaughlin, J. Fregeau-Reid, B.L. Ma, et al. 2008. Changes in isoflavone concentration with 58 years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 48:2201–2208. doi:10.2135/cropsci2008.01.0023

Nakaya, A., and S.N. Isobe. 2012. Will genomic selection be a practical method for plant breeding? Ann. Bot. 110:1303–1316. doi:10.1093/aob/mcs109

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235–322. doi:10.1080/07352689109382313

Ødegård, J., A.K. Sonesson, M.H. Yazdi, and T.H.E. Meuwissen. 2009. Introgression of a major QTL from an inferior into a superior population using genomic selection. Genet. Sel. Evol. 41:38. doi:10.1186/1297-9686-41-38

Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of '5601T' soybean. Crop Sci. 43:1123–1124. doi:10.2135/cropsci2003.1123

Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2004. Registration of '5002T' soybean. Crop Sci. 44:1483–1484. doi:10.2135/cropsci2004.1483a

Pantalone, V.R., C.J. Smallwood, and B.D. Fallen. 2017. Development of 'Ellis' soybean with high soymeal protein, resistance to stem canker, southern root knot nematode, and frogeye leaf spot. J. Plant Reg. 11:250–255. doi:10.3198/jpr2016.12.0071crc

Pantalone, V.R., R.F. Wilson, W.P. Novitzky, and J.W. Burton. 2002. Genetic regulation of elevated stearic acid concentration in soybean oil. J. Am. Oil Chem. Soc. 79:549–553. doi:10.1007/s11746-002-0520-8

Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, J.H. Orf, and A.S. Killam. 2006. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor. Appl. Genet. 112:546–553. doi:10.1007/s00122-005-0161-6

Pérez, P., and G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198:483–495. doi:10.1534/genetics.114.164442

Pham, A.T., J.D. Lee, J.G. Shannon, and K.D. Bilyeu. 2010. Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biol. 10:195. doi:10.1186/1471-2229-10-195

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, et al. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113. doi:10.3835/plantgenome2012.06.0006

R Core Team. 2015. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Ray, D.K., N.D. Mueller, P.C. West, and J.A. Foley. 2013. Yield trends are insufficient to double global crop production by 2050. PLoS One 8:e66428. doi:10.1371/journal.pone.0066428

Resende, M.F.R., Jr., P. Muñoz, J.J. Acousta, G.F. Peter, J.M. Davis, D. Grattapaglia, et al. 2012. Accelerating the domestication of trees using genomic selection: Accuracy of prediction models across ages and environments. New Phytol. 193:617–624. doi:10.1111/j.1469-8137.2011.03895.x [erratum: 193(4):1099].

SAS Institute. 2013. The SAS system for Windows. Release 9.4. SAS Inst., Cary, NC.

Sebastian, S.A., L. Feng, and L.C. Kuhlman. 2012. Accelerated Yield TechnologyTM: A platform for marker assisted selection of simple and complex traits. In: R.F. Wilson, editor, Designing soybean for 21st century markets. Am. Oil Chem. Soc. Press, Urbana, IL. p. 297–305. doi:10.1016/B978-0-9830791-0-1.50020-0

Sebastian, S.A., L.G. Streit, P.A. Stephens, J.A. Thompson, B.R. Hedges, M.A. Fabrizius, et al. 2010. Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci. 50:1196–1206. doi:10.2135/cropsci2009.02.0078

Sitzenstock, F., F. Ytournel, A.R. Sharifi, D. Cavero, H. Täubert, R. Preisinger, and H. Simianer. 2013. Effeciency of genomic selection in an established commercial layer breeding program. Genet. Sel. Evol. 45:29. doi:10.1186/1297-9686-45-29

Smallwood, C.J., J.D. Gillman, A.M. Saxton, H.S. Bhandari, P.A. Wadl, B.D. Fallen, et al. 2017. Identifying and exploring significant genomic regions for soybean yield, fatty acids, protein, and oil. J. Crop Sci. Biotechnol. 20:243–253. doi:10.1007/s12892-017-0020-0

Smith, T.J., and H.M. Camper. 1973. Registration of Essex soybean (Reg. no. 97). Crop Sci. 13:495. doi:10.2135/cropsci1973.0011183X001300040033x

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8:e54985. doi:10.1371/journal.pone.0054985

Song, Q., J. Jenkins, G. Jia, D.L. Hyten, V. Pantalone, S.A. Jackson, et al. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genomics 17:33. doi:10.1186/s12864-015-2344-0

Song, Q., G. Jia, C.V. Quigley, E.W. Fickus, D.L. Hyten, R.L. Nelson, and P.B. Cregan. 2014. Soybean BARCSoySNP6K Beadchip: A tool for soybean genetics research. In: Proceedings of the 22th International Plant and Animal Genome Conference, San Diego, CA. 10–15 Jan. 2014. Scherago Int., Jersey City, NJ. Poster P306.

Specht, J.E., and J.H. Williams. 1984. Contribution of genetic technology to soybean productivity: Retrospect and prospect. In: W.R. Fehr, editor, Genetic contributions to yield gains of five major crop plants. ASA, Madison, WI. p. 49–74. doi:10.2135/cssaspecpub7.c3

Spencer, M.M., D. Landau-Ellis, E.J. Meyer, and V.R. Pantalone. 2004. Molecular markers associated with linolenic acid content in soybean. J. Am. Oil Chem. Soc. 81:559–562. doi:10.1007/s11746-006-0941-4

Tian, Z., X. Wang, R. Lee, Y. Li, J. Specht, R. Nelson, et al. 2010. Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. USA 107:8563–8568. doi:10.1073/pnas.1000088107

Wiggins, B.T., S. Wiggins, M. Cunicelli, C.J. Smallwood, F. Allen, D. West, and V. Pantalone. 2018. Genetic gain from soybean seed protein, oil, and yield in a recombinant inbred line population. J. Am. Oil. Chem. Soc. doi:10.1002/aocs.12166 (in press).

Wilson, R.F. 2004. Seed composition. In: H.R. Boerma and J.E. Specht, editors, Soybeans: Improvement, production, and uses. 3rd ed. ASA, CSSA, and SSSA, Madison, WI. p. 621–678.

Wimmer, V., T. Albrecht, H.J. Auinger, and C.C. Schön. 2012. Synbreed: A framework for the analysis of genomic prediction using R. Bioinformatics 28:2086–2087. doi:10.1093/bioinformatics/bts335

Wolfgang, G., and Y. An. 2017. Genetic separation of southern and northern soybean breeding programs in North America and their associated allelic variation at four maturity loci. Mol. Breed. 37:8. doi:10.1007/s11032-016-0611-7

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, et al. 2012. Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. USA 109:E2155–E2164. doi:10.1073/pnas.1117982109

Yaklich, R.W., B. Vinyard, M. Camp, and S. Douglass. 2002. Analysis of seed protein and oil from soybean northern and southern region uniform tests. Crop Sci. 42:1504–1515. doi:10.2135/cropsci2002.1504

Zhang, A., H. Wang, Y. Beyene, K. Semagn, Y. Liu, S. Cao, et al. 2017. Effect of trait heritability, training population size and marker density on genomic prediction accuracy in 22 bi-parental tropical maize populations. Front. Plant Sci. 8:1916. doi:10.3389/fpls.2017.01916