

## A Gene-Based Model to Simulate Soybean Development and Yield Responses to Environment

C. D. Messina,\* J. W. Jones, K. J. Boote, and C. E. Vallejos

### ABSTRACT

Realizing the potential of agricultural genomics into practical applications requires quantitative predictions for complex traits and different genotypes and environmental conditions. The objective of this study was to develop and test a procedure for quantitative prediction of phenotypes as a function of environment and specific genetic loci in soybean [*Glycine max* (L.) Merrill]. We combined the ecophysiological model CROPGRO-Soybean with linear models that predict cultivar-specific parameters as functions of *E* loci. The procedure involved three steps: (i) a field experiment was conducted in Florida in 2001 to obtain phenotypic data for a set of near-isogenic lines (NILs) with known genotypes at six *E* loci; (ii) we used these data to estimate cultivar-specific parameters for CROPGRO-Soybean, minimizing root mean square error (RMSE) between observed and simulated values; (iii) these parameters were then expressed as linear functions of the (known) *E* loci. CROPGRO-Soybean predicted various phenological stages for the same NILs grown in 2002 in Florida with a RMSE of about 5 d using the *E* loci-derived parameters. A second evaluation of the approach used phenotypic data from cultivar trials conducted in Illinois. Cultivars were genotyped at the *E* loci using microsatellites. The model predicted time to maturity in the Illinois variety trials with RMSE around 7.5 d; it also explained 75% of the time-to-maturity variance and 54% of the yield variance. Our results suggest that gene-based approaches can effectively use agricultural genomics data for cultivar performance prediction. This technology may have multiple uses in plant breeding.

RECENT ADVANCES in agricultural genomics promise a new era of yield maximization through genetic improvement and optimized crop management (Somerville and Somerville, 1999; Huang et al., 2002). The realization of the potential of agricultural genomics into practical applications will depend on our capacity to make accurate quantitative predictions of agronomic traits for different genotypes under varying environmental conditions. Most agronomic traits are genetically complex (Daniell and Dhingra, 2002; Stuber et al., 2003; Orf et al., 1999; Lark et al., 1995) and often show genotype  $\times$  environment interactions (Allard and Bradshaw, 1964) making their prediction a challenging task. Dynamic process-oriented crop models such as DSSAT (Jones et al., 2003) and APSIM (Keating et al., 2003) have the potential to link genetic architectures and whole-organism phenotypic expression (Boote et al.,

2001; White and Hoogenboom, 2003; Chapman et al., 2003; Cooper et al., 2002).

These crop models incorporate knowledge of environmental and management effects on crop growth and development by dynamically simulating the effects of climate and soils on physiological processes, soil water, and nutrient dynamics. Genotypic variation for morphological and physiological traits is taken into account by introducing cultivar-specific parameters (Hunt and Boote, 1998; Boote et al., 2003). Although these parameters are frequently referred to as “genetic coefficients” (GC), they are usually estimated directly through phenotypic data without considering genetic information. Few such coefficients are directly measured. Rather, their values are estimated using optimization algorithms and large data sets (Hunt et al., 1993; Mavromatis et al., 2001, 2002; Grimm et al., 1993). This limits the applicability of crop models. If instead we could estimate genetic coefficients using genetic information, we could then link the crop’s genetic architecture and plant phenotypes through the effects of genes on the regulation of physiological processes and how these respond to the environment. This parameterization scheme could enhance model applicability in crop management optimization, plant breeding and agricultural genomics.

The first attempt to implement this concept was published by White and Hoogenboom (1996) in Genegro, a process-oriented model that incorporated effects of seven genes affecting phenology, growth habit and seed size of common bean (*Phaseolus vulgaris* L.). Genetic coefficients in Genegro were based on the allelic configuration of a set of loci and a set of linear functions. Genegro accurately predicted bean development but poorly explained yield variations between sites (Hoogenboom et al., 1997). Recent improvements of Genegro included the simulation of the effects of temperature on photoperiod sensitivity regulated by the gene *Tip* and a new function to predict seed weight (Hoogenboom and White, 2003; Hoogenboom et al., 2004). Similar modeling approaches were used to incorporate quantitative trait loci effects on leaf elongation rate in maize (*Zea mays* L.) (Reymond et al., 2003); plant height, preflowering duration, carbon partitioning to spike, spike number, and radiation use efficiency in barley (*Hordeum vulgare* L.) (Yin et al., 2003); and transpiration efficiency in sorghum [*Sorghum bicolor* (L.) Moench] (Chapman et al., 2003). Collectively these methods are referred to as top-down methodologies (Hammer et al., 2004).

**Abbreviations:** cM, centimorgan; DOY: day of year; GC, genetic coefficients; ME, mean error; MEP, mean error of the prediction; NILs, near-isogenic lines; PCR, polymerase chain reaction; PTD, photothermal days; RMSE, root mean square error; RMSEP, root mean square error of prediction; SSR, simple sequence repeat; TD, thermal days.

C.D. Messina and J.W. Jones, Agric. & Biol. Eng. Dep., K.J. Boote, Dep. of Agronomy, and C.E. Vallejos, Horticultural Sci. Dep., Univ. of Florida, Gainesville, FL 32611. Florida Agricultural Experiment Station, Journal Series No. R-11017. Received 18 June 2004. \*Corresponding author (Charlie.Messina@pioneer.com).

Published in Crop Sci. 46:456–466 (2006).  
Crop Breeding, Genetics & Cytology  
doi:10.2135/cropsci2005.04-0372  
© Crop Science Society of America  
677 S. Segoe Rd., Madison, WI 53711 USA

**Table 1. List of soybean near-isogenic lines used for model development and for molecular marker evaluation.**

Near-isogenic line genotype†	Name‡	Molecular marker evaluation
<i>e1 e2 e3 E4 e5 E7</i>	L71-920	yes
<i>e1 e2 e3 E4 E5 E7</i>	L97-2076	
<i>e1 e2 E3 e4 e5 E7</i>	L92-21	yes
<i>e1 e2 E3 E4 e5 E7</i>	L63-3117	yes
<i>e1 e2 E3 E4 E5 E7</i>	L94-1110	
<i>e1 E2 e3 E4 e5 E7</i>	L63-2404	
<i>e1 E2 E3 E4 e5 E7</i>	Clark	yes
<i>e1 E2 E3 E4 E5 E7</i>	L92-1195	
<i>E1 e2 e3 E4 e5 E7</i>	L80-5914	
<i>E1 e2 E3 E4 e5 E7</i>	L66-432	
<i>E1 e2 E3 E4 E5 E7</i>	L97-4081	yes
<i>E1 E2 e3 E4 e5 E7</i>	L74-441	yes
<i>E1 E2 E3 E4 e5 E7</i>	L65-3366	yes
<i>E1 E2 E3 E4 E5 E7</i>	L98-2064	yes

† Refer to Bernard (1971), Buzzel (1971), Buzzel and Voldeng (1980), Saindon et al. (1989), McBlain and Bernard (1987), and Cober and Voldeng (2001) for genetic analysis of the *E* loci.

‡ Near isogenic lines were provided by Dr. Bernard, University of Illinois.

Photothermal models (e.g., Grimm et al., 1993) that predict time to flowering and flowering duration based on the genetic makeup of *E* loci were developed for soybean (Stewart et al., 2003; Cober et al., 2001; Upadhyay et al., 1994a; Summerfield et al., 1998). The *E* loci were shown to regulate time to flowering but also time to maturity: *E1* and *E2* (Bernard, 1971), *E3* (Buzzel, 1971), *E4* (Buzzel and Voldeng, 1980), *E5* (McBlain and Bernard, 1987), and *E7* (Cober and Voldeng, 2001). The dominant alleles lengthen time to flowering and maturity in response to photoperiod (Saindon et al., 1989; McBlain et al., 1987; Cober et al., 1996; Cober and Voldeng, 2001). These loci were mapped with simple sequence repeat (SSR) markers (Cregan et al., 1999; Molnar et al., 2003). Similar studies were conducted in other important agronomic traits (Lark et al., 1994, 1995; Orf et al., 1999; Mansur et al., 1993, 1996). These advances in trait mapping and in the understanding of the genetic basis of agronomic traits suggest the opportunity to develop a gene-based model that predicts growth and development of soybean. The objective of this study was to develop and test a procedure for quantitative prediction of phenotypes as a function of environment and specific genetic loci in soybean.

## MATERIALS AND METHODS

### Overview

We developed the procedure for quantitative prediction of phenotypes as a function of environment and specific genetic loci in soybean by combining the ecophysiological framework of CROPGRO-Soybean (Boote et al., 1998) and linear models that predict cultivar-specific parameters in CROPGRO-Soybean as a function of *E* loci. The procedure involved three steps. First, a field experiment was conducted in Florida in 2001 to generate phenotypic data for a set of near isogenic lines (NILs) with known genotypes at six *E* loci (Table 1). Second, we used this data to estimate cultivar-specific parameters for CROPGRO-Soybean (Table 2) by minimizing the RMSE between field observations and simulated values. Third, these parameters were then estimated as linear functions of the *E* loci, and a variable NLOCI that accounts for the total number of dominant alleles at the *E* loci.

A first evaluation of the procedure was conducted using phenotypic data collected in 2002 in Florida. This data set contains the same phenotypic and genotypic information as the data set used for model development. It is independent, however, since the NILs were grown in a different year.

A second evaluation of the approach used phenotypic data collected in variety trials conducted in Illinois. SSR were analyzed to deduce the genotypes of *E* loci in a set of cultivars that were grown in Illinois.

### Simulation of Soybean Development

In the CROPGRO-Soybean model, the vegetative phase, comprising the time from emergence to flowering, is subdivided into three subphases: (i) juvenile, (ii) inductive, and (iii) flower initiation that ends when the first flower becomes visible (Boote et al., 1998). The duration of the vegetative phase varies between determinate and indeterminate soybeans, and this difference is coded by the time to flowering plus the genetic coefficient FL-VS, which determines the physiological time between flowering and the end of differentiation of nodes in the main stem (Boote et al., 2003). FL-VS generally coincides with the appearance of a seed greater than 3 mm in the upper four nodes in the main stem (R5). The reproductive R phase, covering from flowering to physiological maturity, includes (i) the onset of pod addition, (ii) the onset of seed addition, and (iii) the time from the onset of seed addition to physiological maturity. R stages are defined in Fehr and Caviness (1977).

CROPGRO uses a multiplicative function of photoperiod (*P*) and temperature (*T*) to model daily relative developmental progress [*R*(*t*)] during different growth phases with maximum rate standardized to 1.0 (Grimm et al., 1993, 1994).

$$R(t) = f(P) \times f(T) \quad [1]$$

At optimum temperature and photoperiod, the rate of progress in physiological days equals the rate of progress in calendar days. When conditions deviate from the optimum, the rate of development per day decreases, becoming a fraction of a physiological day. The multiplicative model holds for a period beginning after the juvenile phase. Before and during the juvenile phase, the plant is not sensitive to changes in day-length, and development depends only on temperature. Each phase has its own "developmental accumulator" starting at a unique point in time; when it reaches a threshold (genotype characteristic), the phase ends. This nonlinear model has been shown to have high predictive capabilities for soybean (Grimm et al., 1993, 1994; Mavromatis et al., 2001, 2002).

### Field Experiments

A set of soybean NILs (Table 1) was grown in the field at the University of Florida, Gainesville, FL (29°38' N; 82°22' W), in 2001 and 2002. Soybeans were planted on 22 May and 25 July 2001 and on 23 May and 7 Aug. 2002. Planting was done by hand at a density of 20 seed m<sup>-2</sup> and with a row spacing of 0.56 m. Plots were single rows 3 m long in 2001, and 1.5 m long in 2002. Weeds were controlled by a pre-emerge herbicide application of 1.0 g m<sup>-2</sup> of pendimethalin [*N*-(1-ethylpropyl)-3,4-dimethyl-2,6-dinitrobenzenamine] and by manual control during the growing season. Pests and diseases were controlled by applications of 1.17 mL m<sup>-2</sup> of daconil [tetrachloroisophthalonitrile] and 0.07 mL m<sup>-2</sup> of permethrin (3-phenoxyphenyl)methyl 3-(2,2-dichloroethenyl)-2,2-dimethylcyclopropanecarboxylate] as necessary, varying between seasons and planting dates. Fertilizer (14:6:12) was applied in bands at a rate of 16.8 g m<sup>-2</sup>, when the plants reached the

**Table 2. Selected genetic coefficients in CROPGRO-Soybean controlling plant development in soybean, potential associations with *E* loci, and variables used for parameter estimation.**

Genetic coefficient	Initial value†	Plant trait	Potential loci effects	Variables used for parameter estimation‡	Units
PPSEN	0.171	Photoperiod sensitivity	<i>E</i>	R1	h <sup>-1</sup>
CSDL	14.1	Critical photoperiod	<i>E</i>	R1	h
RIPRO	0.189	Reduction in CSDL after R1	<i>E</i>	R7	h
EM-FL	16.8	Emergence to flowering	<i>E</i>	R1	PTD§
V1-JU	0.0	Juvenile phase	<i>E1</i>	R1	TD§
FL-SD	13.0	Flowering to first seed	<i>E</i>	FL-VS, R5	PTD
FL-VS	26.0	First flower to last leaf on main stem	<i>E</i>	R5	PTD
SD-PM	30.0	First seed to physiological maturity	<i>E</i>	R7	PTD
FL-SH	5.0	Flowering to onset of pod addition	<i>E</i>	R1, OPA	PTD

† Initial values were based on values for a generic maturity group 0 soybean cultivar.

‡ R1, first flower; R5, presence of a seed greater than 3 mm in a pod in the upper four nodes; OPA, onset of pod addition (when 50% of the plants have a pod greater or equal to 5 mm anywhere on the plant); R7, beginning maturity, first pod achieving normal mature color anywhere on the plant.

§ PTD, photothermal days; TD, thermal days.

four-leaf stage. Plants were grown under irrigation. The soil of the experimental plots is loamy, siliceous, hyperthermic Grossarenic Paleudult. Detailed characteristics for this soil are available from DSSAT (Jones et al., 2003). Daily weather data were measured using an automated weather station 50 m from the experimental plots (temperature, rainfall, solar radiation). Within planting dates and years, NILs were planted in two complete randomized blocks. Time of first visible flower, onset of pod addition, first seed on the upper four nodes (R5), and physiological maturity (R7) were recorded at 2 to 4 d intervals.

### Gene-Based Parameterization of CROPGRO-Soybean

A systematic approach was used to estimate the genetic coefficients (Hunt and Boote, 1998; Boote et al., 2003) listed in Table 2 using a procedure similar to Mavromatis et al. (2001). The initial values for the genetic coefficients were those for maturity group 0 (Jones et al., 2003). These values (Table 2) were modified in sequence using selected NILs and planting dates to minimize the root mean square error,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad [2]$$

where  $n$  is the number of observations and  $x_i$  and  $y_i$  are predicted and observed values for the variables listed in Table 2.

Soybean NILs carrying recessive alleles for *E* are impaired in the perception or transduction of the photoperiodic signal. These loss-of-function lines grown under short photoperiod ensure the absence of photoperiodic effects on plant development, allowing us to estimate the thermal components of the photothermal time coefficient. The line L71–920 grown in late plantings (short photoperiod) were used to estimate the thermal component of the photothermal time between emergence to flowering and from flowering to the onset of pod addition, and between flowering and the end of canopy leaf area expansion (FL-LF).

Lines carrying the *E* alleles grown in early plantings responded to photoperiod. All lines but L71–920 were used to estimate the critical photoperiod and photoperiod sensitivity to predict time to flowering. The coefficient R1PPO, which accounts for increased photoperiod sensitivity after flowering (Piper et al., 1996) was estimated using time to physiological maturity date.

A final step used both planting dates in 2001 and photoperiod-sensitive lines to estimate the photoperiod component of the photothermal times between flowering and last leaf formed in the main stem (FL-VS), between flowering and first seed (FL-SD), and time from first seed to physiological maturity (SD-PM). Photothermal time from flowering to first seed in

photoperiod-sensitive lines was estimated as a fraction of the photothermal time FL-VS. We estimated FL-VS through an analysis of observations made for stage R5, which typically coincides with the expansion of the last leaf on the main stem. The average ratio of FL-SD to FL-VS duration is 0.56 for several standard cultivars of different maturity groups in CROPGRO-Soybean (Jones et al., 2003). FL-SD was then estimated as 0.56(FL-VS). Results from reciprocal transplant experiments suggest that the locus *E1* is involved in extending the juvenile phase (Upadhyay et al., 1994b). Near-isogenic lines carrying the *E1* allele were used to estimate the duration of the juvenile phase.

After obtaining cultivar-specific parameters for CROPGRO-Soybean, using only the data from the 2001 field experiment, we estimated them from *E* loci information using multiple linear regression. The linear functions used allelic information at the loci *E1*, *E2*, *E3*, *E4*, *E5*, and *E7*. For each loci, a value of 0 or 1 was assigned, depending on whether the allele present in the near isogenic line was recessive or dominant, respectively. A variable NLOCI was introduced to account for the total number of dominant alleles at the *E* loci. Stewart et al. (2003) showed a linear relationship between photoperiod sensitivity in time to flowering and the number of dominant alleles at selected *E* loci.

### Model Evaluation

Quantitative measurements for model evaluation include RMSE, slope and intercept from the regression between simulated and independently observed values (Hunt and Boote, 1998), and mean error (ME),

$$ME = \frac{1}{n} \sum_{i=1}^n (x_i - y_i) \quad [3]$$

which evaluates model bias.

The procedure was evaluated using two datasets. First, phenotypic data from the field experiment conducted in 2002 season at the University of Florida were used to evaluate the gene-based model developed using data collected during 2001 season at the same location and for the same NILs.

Second, we used the gene-based model to simulate growth and development of a set of public soybean varieties grown in a variety trial network in Illinois. The trial network consisted of eight locations: Belleville, Urbana, Dekalb, Dixon Springs, Dwight, Monmouth, Carbondale, and Perry, where soybeans were grown between 1995 and 1999. Yield, time to maturity, and crop management data were from the University of Illinois (University of Illinois, 2005). Weather data were from the Midwestern Regional Climate Center (Midwestern Regional Climate Center, 2005). Soil parameters were provided by Dr.



**Table 3. Cultivar specific parameters not estimated from genotypic information but required to simulate yield for soybean varieties grown in eight locations in Illinois (1995–1999). Data from Dr. T. Mavromatis (T. Mavromatis, personal communication, 2003).**

Cultivar specific parameters	Cultivar						
	Linford	Nile	Omaha	Savoy	Vinton 81	Williams82	Yale
Maximum threshing percentage, %	74.5	76.0	74.7	76.7	75.5	76.5	77.0
Fraction of protein in seeds, g g <sup>-1</sup>	0.440	0.380	0.420	0.405	0.405	0.405	0.405
Fraction of oil in seeds, g g <sup>-1</sup>	0.189	0.180	0.200	0.205	0.205	0.205	0.205
Maximum leaf photosynthesis rate at 30°C, 350 vpm CO <sub>2</sub> , and high light, mg CO <sub>2</sub> m <sup>-2</sup> s <sup>-1</sup>	1.070	0.930	1.030	1.030	1.030	0.970	0.970
Seed filling duration for pod cohort at standard growth conditions, PTD†	22.70	26.00	23.0	23.0	23.0	21.50	25.10
Time required for cultivar to reach final pod load under optimal conditions, PTD	11.40	7.00	10.0	10.0	10.0	13.00	8.20

† PTD, photothermal days.

T. Mavromatis (T. Mavromatis, personal communication, 2003). Genetic coefficients controlling plant development were estimated as functions of *E* loci after genotyping (see next section), while the remaining genetic coefficients were from Mavromatis (Table 3) using a calibration procedure described in Mavromatis et al. (2001).

### Molecular Marker Analysis

To determine the general applicability of the gene-based model, we performed an independent test to determine how well the model would simulate the Illinois varieties. To run the test, the allelic constitution at the selected *E* loci had to be determined for these varieties. This procedure was performed in two steps. First, we characterized the *E<sub>n</sub>*-linked SSR marker loci in the NILs, and then analyzed the same SSR loci in the Illinois varieties to deduce the genotypes at the selected *E* loci. The deduced *E* genotypes were used to assign the proper genetic coefficients to each variety.

*E1*, *E2*, and *E3* along with several number of SSR markers have been incorporated into the integrated (classical and molecular maps) soybean linkage map (Cregan et al., 1999). The following flanking markers around *E1*, *E2*, and *E3* were considered for evaluation and genotyping: *Satt557–E1–Satt319*; *Satt518–E2–Sat\_038*; and *Satt513–Satt229–E3–Satt006* (Cregan et al., 1999; Grant and Shoemaker, 2005). More recently, Abe et al. (2003) mapped *E4* 5 centimorgans (cM) apart from *Satt496* and Molnar et al. (2003) mapped *E4* at loci *Satt354*, located 11.9 cM from *Satt496*. Molnar et al. (2003) mapped *E7* at loci *Satt319*.

Seven NILs were selected to investigate the presence of length polymorphism of SSR at the loci *E1*, *E2*, and *E3* (Table 1). Plants were grown in a greenhouse for 2 wk. Soybean DNA was isolated from upper node leaves by a modified procedure of Murray and Thompson (1980) as reported in Vallejos et al. (1992). Seven soybean cultivars, Yale, Williams 82, Vinton 81, Savoy, Omaha, Nile, and Linford, were genotyped at *E1* to *E4* loci. This soybean germplasm is from GRIN Germplasm-Soybean Collection and provided by Dr. Randal Nelson. DNA was extracted from 50 mg seed tissue flour in 150 µL TES (Tris 0.1 M, pH 8; EDTA 5 µM; NaCl 50 mM) and 800 µL of 1.25× extraction buffer [125 mM Tris-HCl (pH 7.8); 12.5 mM EDTA-Na (pH 8.0); 1.4 M NaCl; 1.25% CTAB; 0.5% NaSO<sub>2</sub>]. Samples were incubated for 50 min at 65°C, and then extracted with chloroform-octanol (400 µL). Phases were separated by centrifugation at 13c000 RPM for 15 min. Due to high concentration of polysaccharides and oils, this step was repeated two to three times. DNA was precipitated in isopropanol (600 µL) for 30 min, incubated for an hour in a solution containing 76% ethanol and 0.2 M sodium acetate, followed by a 30-s rinse in a solution containing 76% ethanol and 10 mM ammonium acetate, and finally air dried for 1 h.

The pellet was resuspended in 800 µL of TE buffer (10 mM Tris-HCl; 1.0 mM EDTA-Na).

Polymerase chain reaction mixes contained 1× PCR buffer [20 mM Tris-HCl (pH 8.4), 50 mM KCl] (Cat. No. 10342-020, Invitrogen, Carlsbad, CA), 1.5 mM MgCl<sub>2</sub>, 200 µmol of each nucleotide, 0.1 µL of 300 Ci mmol<sup>-1</sup> α-<sup>32</sup>PdATP, 0.1 µmol of 3' and 5' end primers, 0.5 unit *Taq* DNA polymerase, and 30 ng of soybean genomic DNA in a total volume of 20 µL. Thermocycling consisted of a 30 s denaturation at 94°C, a 30 s annealing at 50°C, and a 30 s extension at 72°C for 35 cycles on a Geneamp PCR System 9600 (PE Applied Biosystems, Foster City, CA). The thermocycle had an initial denaturation phase of 2 min at 94°C, and a final extension phase of 5 min at 72°C. Samples were denatured at 72°C in formamide for 5 min and quenched on ice before loading. Polymerase chain reaction products (3 µL) were separated on a DNA vertical gel containing 6% Long Ranger (Cat. No. 50611, Cambrex Bio Science Rockland, Inc., Rockland, ME), 0.5× TBE (0.5 M Tris, 0.445 M boric acid, 10 mM EDTA) and 6 M urea, at 50 W constant power for 90 min. Polymerase chain reaction amplification products were visualized by autoradiography on a Kodak X-OMAT film (Cat. No. 1651512, Kodak, Rochester, NY).

## RESULTS AND DISCUSSION

### Parameterization of CROPGRO-Soybean

CROPGRO-Soybean accurately simulated soybean phenology during the 2001 season after calibration. Precision of estimates for the timing of developmental processes varied between 1.4 and 3.2 d according to RMSE calculations (Table 4). These RMSE values are close to the observational error, defined by the interval between field observations (2–4 d). When calibrated GC were used, there were a high correlations between observed and simulated values for time to flowering, time to maturity, time to onset of pod addition, and time to last leaf on main stem node for 2001 (Fig. 1, Table 4). However, there was a tendency for the correlation to deteriorate as development progressed toward physiological maturity. This decrease in predictive ability is apparently related to both the propagation of errors as the simulation of later stages progresses, and the uncertainty in measuring late stages of development. Comparable observations have been reported by Grimm et al. (1993, 1994). The calibration procedure provided genetic coefficient estimates with little or no bias as reflected in ME, and a slope not significantly different from 1 (Table 4). The model accounted for 88 to 96% of the variance in ob-

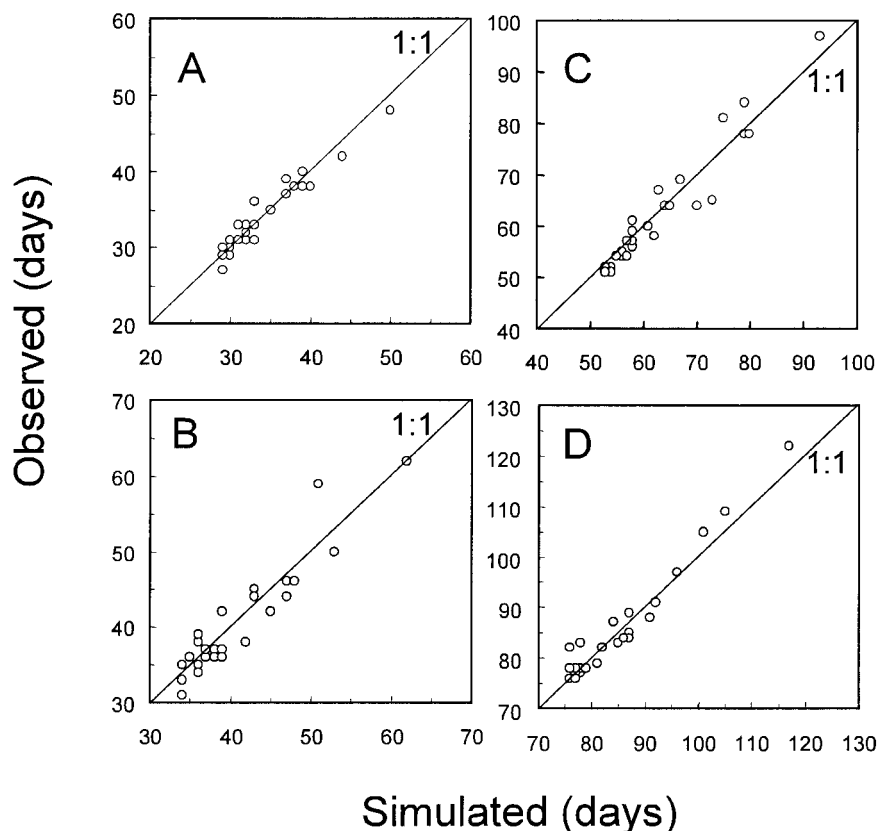
**Table 4.** Comparisons between CROPGRO-Soybean parameterized directly from data and CROPGRO-Soybean parameterized using *E* loci information using data collected in Gainesville, FL, during 2001.

Physiological processes	CROPGRO-Soybean				CROPGRO- <i>E</i> loci			
	RMSE†	ME‡	R²	b§	RMSE	ME	R²	b
Time to flowering, d	1.4	-0.17	0.92	0.89 (0.05)	2.7	-0.10	0.77	0.7 (0.08)
Time to onset of pod addition, d	2.5	0.51	0.88	1.00 (0.07)	4.0	0.32	0.72	0.83 (0.1)
Time to last mainstem node, d	3.2	0.69	0.93	1.05 (0.06)	4.2	1.14	0.87	0.99 (0.08)
Time to maturity, d	2.5	-0.45	0.96	1.08 (0.05)	3.9	-0.25	0.87	0.99 (0.08)

† RMSE, root mean square error.

‡ ME, mean error.

§ Slope of the linear regression between simulated and observed values. Standard errors are shown in parentheses (df = 27).

**Fig. 1.** Relation between observed and predicted (A) time to flowering, (B) time to onset of pod addition, (C) time to last mainstem node (R5), and (D) time to physiological maturity. Soybean development simulated using CROPGRO-Soybean and cultivar-specific parameters calibrated for each near-isogenic line. Data collected in Gainesville, FL, during 2001.

served values (Table 4), which were within the range of results reported in other soybean studies (Mavromatis et al., 2001, 2002; Grimm et al., 1994; Elizondo et al., 1994) and in common bean studies (Hoogenboom et al., 1997; White and Hoogenboom, 1996).

### Estimation of Genetic Coefficients for CROPGRO-Soybean from *E* loci

Genetic coefficients were estimated from *E* loci with different levels of accuracy (Table 5). The proportion of the total variance explained by the linear models varied between 32 and 88%, similar to values reported for Genegro (White and Hoogenboom, 1996). The number of dominant alleles (NLOCI) in a NIL affected the coefficients CSDL, PPSEN, and SD-PM. Because CSDL and PPSEN mediate the influence of photoperiod on rate of development, their relationship with *E* loci was expected and was consistent with previous

models of time to flowering based on *E* loci (Stewart et al., 2003; Upadhyay et al., 1994a). However, the prediction of SD-PM from *E* loci had higher RMSE than other phases, indicating that photoperiodic effects indirectly amplify differences in phase duration. Therefore, this study indicates two modes of action of *E* loci on soybean development. One mode acts by the modulation of the critical photoperiod and photoperiod sensitivity. A second mode regulates the number of physiological days required for phase duration.

*E1* alone showed major control on PPSEN, EM-FL, V1-JU, and SD-PM and, in interaction with other loci, on PPSEN. This result confirmed the hypothesis that *E1* affects the juvenile phase as was inferred from Upadhyay et al. (1994b). It also confirmed previous evidence of epistatic effects between *E1* and *E3* on the regulation of time to flowering (Upadhyay et al., 1994a). Notably, *E1* had a negative effect on physiological days from first

**Table 5. Associations between *E* loci and genetic coefficients in CROPGRO-Soybean. Dominant and recessive alleles take values of 1 and 0, respectively. NLOCI denotes the sum of dominant alleles.**

Genetic coefficient†	Linear model	R <sup>2</sup>
CSDL	CSDL = 14.33 - 0.44 NLOCI + 0.27 <i>E3</i> - 0.48 <i>E5</i> + 0.18 NLOCI <i>E5</i>	0.88
PPSEN	PPSEN = 0.11 + 0.063 NLOCI + 0.58 <i>E1</i> - 0.13 <i>E1</i> NLOCI	0.70
EM-FL	EM-FL = 20.77 + 2.1 <i>E1</i> + 1.8 <i>E3</i>	0.78
FL-SD	FL-SD = 0.56 FL-VS	—
FL-VS	FL-VS = 20.9 + 0.67 NLOCI	0.47
SD-PM	SD-PM = 35.2 - 1.0 NLOCI - 9.2 <i>E1</i> + 2.0 NLOCI <i>E1</i>	0.57
V1-JU	V1-JU = 4.16 <i>E1</i>	0.71
RIPRO	RIPRO = 0.1 + 0.066 NLOCI	0.32

† CSDL, critical photoperiod; PPSSEN, photoperiod sensitivity; RIPRO, reduction in CSDL after flowering; EM-FL, photothermal time between emergence and flowering; V1-JU, juvenile phase; FL-SH, photothermal time between flowering and onset of pod addition; FL-SD, photothermal time between flowering and first seed; FL-VS, photothermal time between flowering and appearance of the last leaf on the main stem; FL-LF, photothermal time between flowering and end of canopy expansion; PODUR, photothermal time for a cultivar to reach final pod load under optimal conditions; SD-PM, photothermal time between first seed and physiological maturity. FL-SH, FL-LF, and PODUR did not significantly vary with *E* loci, they were set equal to 5.0, 26, and 13 photothermal days respectively for all NILs.

seed to maturity, consistent with previous observations indicating that *E1* hastened soybean development during the reproductive period (McBlain et al., 1987).

Locus *E5* showed an association with CSDL (Table 5). In previous experiments it was shown that *E5* has major control over the duration of pod addition (Messina, 2003). This study shows that *E5* affects pod addition duration by delaying the transition of the apex from vegetative to reproductive. Longer duration of pod addition would be the result of a longer time to sink development allowing the plant to set more pods (Wardlaw, 1990). A strong interaction was also shown between *E3* and *E5* in regulating pod number (Messina, 2003). Our results suggest that the interaction between *E3* and *E5* is via their effects on CSDL. A reduction in CSDL would delay the onset and rate of seed growth under long photoperiods, increasing pod addition duration and seed number.

CROPGRO-Soybean simulated phenology well (RMSE = 2.7–4.0 d) for the 2001 data where cultivar coefficients were estimated from the genetic makeup of the NILs (Table 4). The RMSE increased and the proportion of the explained variance of the observations decreased relative to the calibrated results, which did not use *E* loci information. This result was expected due to the propagation of errors intrinsic to linear models used to estimating genetic coefficients, although this characteristic was not observed in the development of Genegro (White and Hoogenboom, 1996).

With the exception of time to flowering, all predictions showed good agreement with observed results with little deviations from the 1:1 relationship (Table 4). CROPGRO-Soybean insensitivity in predicting time to flowering when parameters were estimated from genotypes ( $b < 1$ ;  $P < 0.05$ ) was due to one extreme value (50 d), which on removal, the slope of regression be-

tween observed and simulated values was not significantly different from one.

### Model Evaluation with Independent Data from 2002

We tested the model capabilities to predict crop development using an independent data set collected during 2002. The model accurately predicted reproductive development with low bias. Root mean square error of prediction (RMSEP) ranged between 2.6 and 7.6 d, and mean error of prediction (MEP) varied between -5.9 and 1.1 d (Table 6). These values are slightly higher than RMSE of the calibration (Table 4) data set. Despite the small decrease in the model precision relative to the calibration values, these RSMEP are comparable in magnitude with the precision of the measurements (2–4 d) and previous modeling results (Grimm et al., 1993).

CROPGRO-Soybean, run either with genetic coefficients estimated using data collected during 2001 or estimated from *E* genotypes (also 2001), showed poor accuracy when predicting physiological maturity. The slope of the regression between observed and simulated values was different from one ( $P < 0.01$ ) (Table 6). The model consistently underpredicted the time of physiological maturity for long cycle NILs grown in early plantings. After removing six simulations for early plantings and long life cycle (>105 d) the model predicted physiological maturity without systematic bias ( $b = 1$ ;  $\alpha = .01$ ). The removed data points had values greater than the upper quartile plus 1.5 times the interquartile range. The simulation error can arise from measurement errors in R7.

### SSR Length Polymorphisms Linked to *E* Loci and Cultivar Genotyping

Polymorphisms were detected at *E*-linked SSR marker loci in the NILs (Fig. 2). Results obtained from this study support the proposed location of the *E* loci on the molecular map and were used to genotype soybean varieties (Fig. 2). Because of the linkage between the SSR and *E* loci position, we can infer with varying degrees of certainty the presence of the dominant allele in each public soybean cultivar. Due to the close linkage between *Satt557* and *E1*, the expected uncertainty in determining the presence of the dominant allele is in the order of 1% since there is about 1.2 cM from *Satt557* to *E1* (Cregan et al., 1999; Abe et al., 2003). The uncertainty increases for *E4*, which was mapped 5.0 cM apart from *Satt496* (Abe et al., 2003), and it is highest for *E2* as the distances between *E2* and *Satt518* and *Sat\_038* were estimated as 17.2 and 18.3 cM, respectively. Locus *E3* was located within a bracket of 14 cM between *Satt006* and *Satt513*. However, within this bracket only *Satt229* showed length polymorphism between the near isogenic lines.

Given the uncertainty in the determination of each allele at a given locus, we inferred from the PCR fragment sizes (Fig. 2) the *E* loci makeup for each cultivar as follows: Linford, *e1E2E3E4*; Nile, *e1E2E3E4*; Omaha, *e1E2E3E4*; Savoy, *e1e2e3e4*; Vinton 81, *e1e2E3e4*; Wil-



**Table 6. Evaluation of CROPGRO-Soybean and CROPGRO-Soybean parameterized using *E* loci information using independent data collected in Gainesville, FL, during 2002.**

Physiological processes	CROPGRO-Soybean				CROPGRO- <i>E</i> loci			
	RMSEP†	MEP‡	R <sup>2</sup>	b§	RMSEP	MEP	R <sup>2</sup>	b
Time to flowering, d	2.6	0.1	0.65	0.79 (0.11)	2.8	0.38	0.62	0.73 (0.12)
Time to onset of pod addition, d	2.7	0.69	0.85	0.73 (0.06)	2.9	0.55	0.81	0.71 (0.07)
Time to last mainstem node, d	3.31	0.34	0.91	0.97 (0.06)	3.8	1.11	0.89	0.97 (0.07)
Time to maturity, d	7.6	-3.7	0.88	1.42 (0.09)	7.3	-5.9	0.89	1.39 (0.09)

† RMSEP denotes root mean square error of prediction since it was calculated using data collected during 2002, which were not used to estimate genetic coefficients, or to fit the relationships between *E* loci and genetic coefficients.

‡ MEP, mean error of prediction.

§ Slope of the linear regression between simulated and observed values. Standard errors are shown in parentheses (*df* = 27).

liams 82, *e1E2E3E4*; Yale, *e1E2E3E4*. At locus *Satt496* the PCR fragment sizes differ from previous reports (Abe et al., 2003). *E4* alleles were determined by comparing the fragment size at the locus *Satt496* relative to the NIL for Clark carrying the dominant allele *E4*. We assume that deviations from this size indicate the absence of *E4*, hence the cultivar might have the *e4* genotype.

All genotypes had the recessive allele at the marker locus *Satt557*, therefore, its genotype is *e1* and probably *e7* since these loci are tightly linked (Cober and Voldeng, 2001). The cultivar Vinton 81, however, has gray pod pubescence. The alternative allele, tawny color, was found associated with earlier maturity (Cober and Voldeng, 2001). The locus *T* controlled this trait and is tightly linked to *E1* (1.4 cM) and *E7* (4.0 cM). Fragment size for *Satt557* suggests that the genotype is *e1*, hence *e7*. However, from the color of the pubescence we can infer the genotype as being *E7*. Finally, due to the lack of markers for *E5* and based on the maturity groups we assumed that all cultivars carried the *e5* allele at this locus.

### Predicting Soybean Yield and Maturity in Variety Trials

We used *E* loci-linked SSR markers to deduce the genotypes of a set of cultivars for which actual data were already available for variety trials performed at eight locations in the state of Illinois over a period of 5 yr. The use of variety trials for model evaluation added realism to the test but restricted the genotypic diversity. Nonetheless, the dataset included a wide range of maturity dates and yield variation. The model was able to predict maturity, which varied from about 240 and 275 d of the year, and yields, which varied from 1.5 to 5.0 Mg ha<sup>-1</sup> (Fig. 3). The model predicted 75 and 54% of the observed maturity date and yield variances. These values are within the lower range of values obtained in previous modeling studies predicting yield and development in variety trials (Mavromatis et al., 2001, 2002; Hoogenboom et al., 1997). Prediction errors are also similar to those found in predicting leaf expansion in maize (Reymond et al., 2003) and phenology and yield in beans (Hoogenboom et al., 2004).

We can identify some causes contributing to the slightly higher prediction errors in our study. This study used genetic coefficients calculated from *E* loci genotypes. Even though the model based on six loci ac-

counted for as much as 75% of the variance in maturity date, which demonstrates the importance of these loci on the regulation of soybean development, other loci involved in the regulation of soybean development and yield were not included in our model (Mansur et al., 1993, 1996; Orf et al., 1999; Tasma and Shoemaker, 2003; Tasma et al., 2001).

The selection of terms in multiple regression linear models was an iterative process. Errors could arise from the choice of variables that defined the statistical model (Pinheiro and Bates, 2000) and from the estimation of parameters.

Bias and systematic errors of prediction of time to maturity suggests that these patterns can be due to errors in predicting certain genotypes, locations, or years. Table 7 shows that CROPGRO-Soybean significantly underestimated time to maturity for the variety with the shortest predicted lifecycle (Savoy) but classified as maturity group II. This suggests that there is an error in the estimated genotype. Alternatively, other loci regulate development in Savoy.

We determined soybean genotypes based on the linkage between SSR markers and the *E* loci. Recombination between loci can occur and there is a risk of inferring the absence of the dominant allele when it is present. If this would have been the case for any of the marker loci, the genotype for Savoy would have been more sensitive to photoperiod, hence the model would have not underestimated the time to maturity and yield (Table 7). In addition, because soybean cultivars were not derived strictly from Clark genetic background, alleles could have mistakenly diagnosed as the sequence surrounding the actual gene could vary among varieties. This is noticeable for Savoy, which alleles at *Satt496* differed from the other varieties and the NILs. In contrast, at loci *Satt557* all varieties had Clark allele, and at *Satt229* all cultivars except Savoy carried the same allele as Clark (Fig. 2). Error of prediction for Savoy illustrates well important limitations of the method. However, recent advances in the identification and development of single nucleotide polymorphisms in soybean (Zhu et al., 2003) will help reduce the uncertainties associated with cultivar genotyping, and should increase the prediction skill of the model.

The statistics ME and RMSE calculated for time to maturity for all soybean cultivars excepting Savoy compares well with previous results (Table 4, Table 6, Mavromatis et al., 2001; Mavromatis et al., 2002). Errors in the

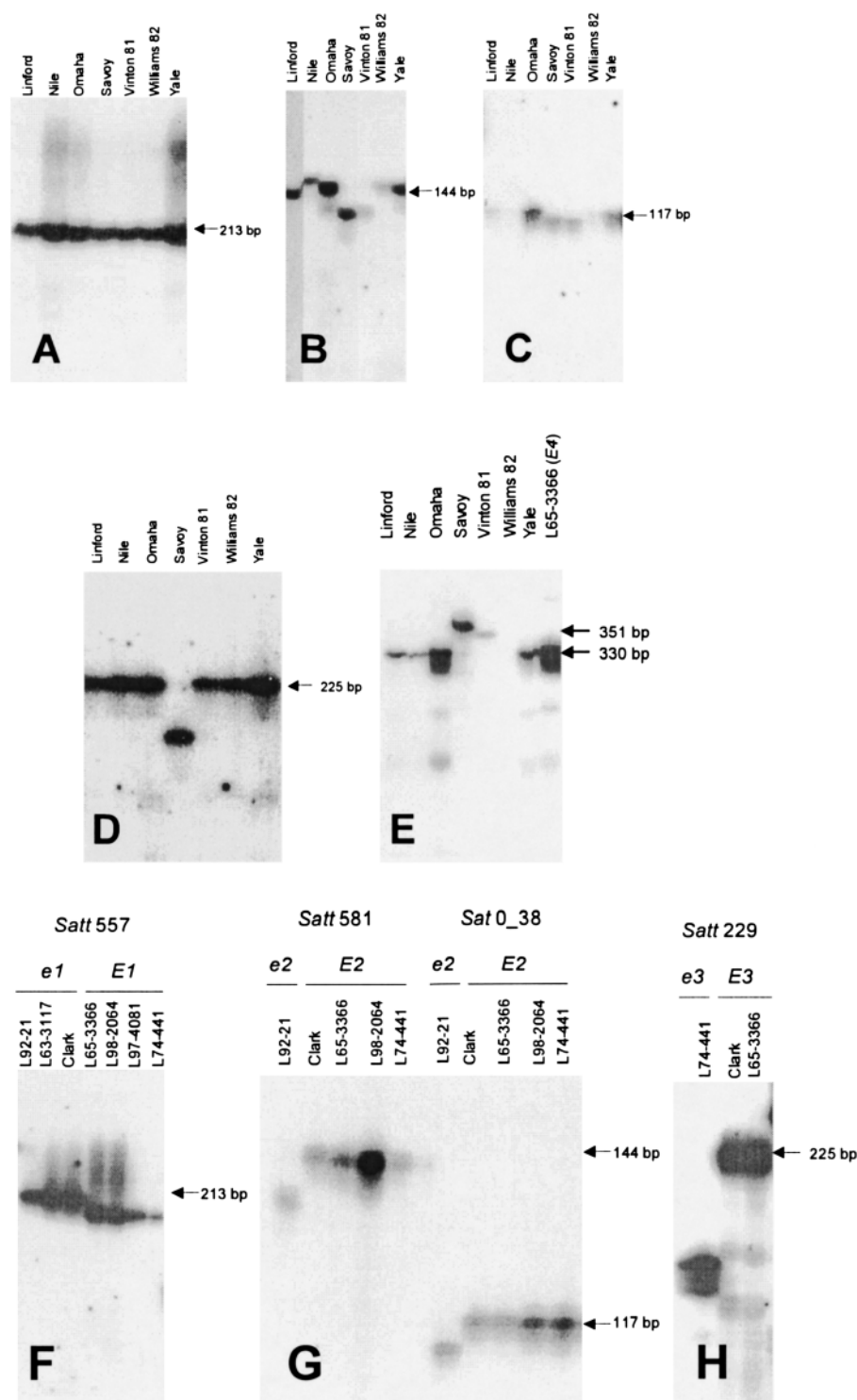


Fig. 2. Electrophoretic separation of SSR markers linked to four *E* loci. Selected soybean cultivars from Illinois are depicted in panels A–E, and near-isogenic lines with 'Clark' background in panels F–H. (A and E) *E1*-linked *Satt557*; (B, C, and G) *E2*-linked *Satt581* and *Satt038*; (D and H) *E3*-linked *Satt229*; and (E) *E4*-linked *Satt496*.

simulation of time to maturity for cultivar Savoy led to errors in the simulation of yield, which was underestimated ( $ME = -390 \text{ kg ha}^{-1}$ ). Simulated yield RMSE for all but Savoy and Vinton 81 cultivars varied from 9 to 15%, and ME varied between  $-3$  to  $5\%$  of average yields. It should be noted that yield predictions used a

subset of genetic coefficients characterizing growth parameters not estimated from *E* loci but fitted by Mavromatis (T. Mavromatis, personal communication, 2003; Table 3). Thus, yield predictions (Fig. 3) are not totally independent from the observed yields. Model skill to predict time to maturity and yield is considered accept-



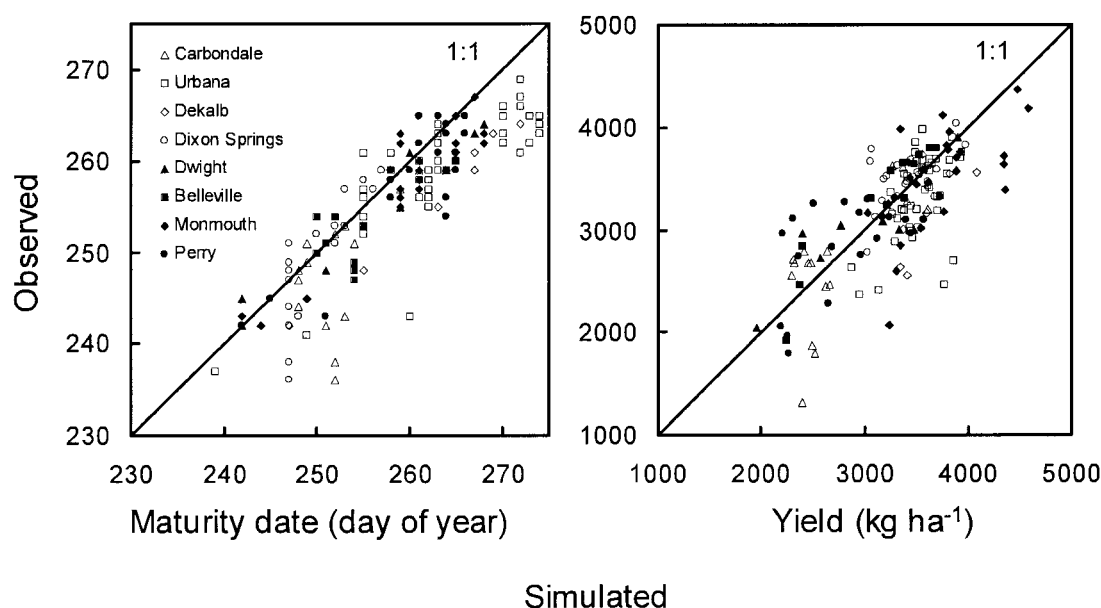


Fig. 3. Simulated and observed time to maturity (day of the year) and yield ( $\text{kg ha}^{-1}$ ) of seven soybean public varieties grown at eight locations and 5 yr in Illinois (1995–1999). Regression equations for time to maturity,  $y = -50.7 (\pm 9.7) + 0.8x (\pm 0.04)$ ,  $R^2 = 0.75$ ; and yield,  $y = 689 (\pm 192) + 0.77x (\pm 0.07)$ ,  $R^2 = 0.54$ . Time to maturity and yield RMSE were 5.2 (days) and 393 ( $\text{kg ha}^{-1}$ ), respectively. Yield RMSE is 12.3% of the average observed yields.

Table 7. Prediction errors in time to maturity and yield for soybean public varieties grown in eight locations in Illinois (1995–1999).

Cultivar	Linford	Nile	Omaha	Savoy	Vinton 81	Williams82	Yale
<b>Time to maturity</b>							
ME†, d	2.7	2.6	3.2	-8.5	3.7	2.5	3.7
RMSE‡, d	5.0	4.6	5.2	10.0	6.0	5.1	5.5
Observed mean, DOY§	256	257	258	248	242	254	256
<b>Yield</b>							
ME, $\text{kg ha}^{-1}$	129	108	-113	-390	373	173	-3.45
RMSE, $\text{kg ha}^{-1}$	355	359	329	537	567	485	338
Observed mean, $\text{kg ha}^{-1}$	3303	3110	3465	3611	2603	3160	3357

† ME, mean error.

‡ RMSE, root mean square error.

§ DOY, day of year.

able for applications of crop models in agricultural production.

## CONCLUSIONS

We developed a gene-based model for soybean, based on the CROPGRO-Soybean, which contributes toward linking crop's genetic architectures and whole organism phenotypic expression. The model accurately predicted time to flowering and post-flowering development phases and yield. This is significant advance with respect to previous models in predicting only time to flowering for soybean from *E* loci information. More importantly, we tested a gene-based model for its ability to reproduce yields and development at the field scale and for different environmental conditions and genetic materials as the ones used for model development.

The prediction skill showed by CROPGRO-Soybean linked to *E* loci was comparable with that of Genegro for dry bean demonstrating that the potential for the implementation of gene-based approaches are not conditioned to a particular species. However, prediction skill can vary with the level of ploidy, genetic background,

and zygosity of the organism. Our results also demonstrate the ability of gene-based models implemented on the basis of a top-down methodology to predict overall system behavior. The prediction errors using a gene-based approach are comparable with those using conventional parameter estimation. From a model application perspective, gene-based approaches can help reduce the requirements for expensive and intensive experimentation to provide up-to-date genetic coefficients.

Failure to simulate yield for Savoy and Vinton 81 cultivars shows there is potential for improvement and thus to reduce uncertainties and errors involved in the development and implementation of gene-based models. Genotyping for Savoy indicated the absence of dominant alleles at the *E* loci. This is large difference when compared to the other cultivars, and it is inconsistent with its maturity group. Major gain in predictability is conditioned on the development of allele specific markers and high-density molecular maps. Advances in agricultural genomics will help the identification and incorporation of novel genes regulating important physiological processes expanding the capabilities of current gene-based models. Increasing the density of markers around *E* and

other loci will reduce uncertainties during genotyping thus improving the simulation of yield. Improved gene-based models could enhance the simulation of breeding program.

### ACKNOWLEDGMENTS

The authors are grateful for the valuable comments received from three anonymous reviewers. Special thanks to Jeffrey W. White, Rafael A. Ferreyra, and Associate Editor Elroy Cober for their valuable suggestions during the preparation of the manuscript. Near isogenic lines were provided by Dr. Bernard. Thanks to Dr. T. Mavromatis for providing parameters to run CROPGRO.

### REFERENCES

- Abe, J., D. Xu, A. Miyano, K. Komatsu, A. Kanazawa, and Y. Shimamoto. 2003. Photoperiod-insensitive Japanese soybean landraces differ at two maturity loci. *Crop Sci.* 43:1300–1304.
- Allard, R.W., and A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. *Crop Sci.* 4:503–508.
- Bernard, R.L. 1971. Two major genes for time of flowering and maturity in soybeans. *Crop Sci.* 11:242–244.
- Boote, K.J., J.W. Jones, W.D. Batchelor, E.D. Nafziger, and O. Myers. 2003. Genetic coefficients in the CROPGRO-Soybean model: Links to field performance and genomics. *Agron. J.* 95:32–51.
- Boote, K.J., J.W. Jones, and G.H. Hoogenboom. 1998. Simulation of crop growth: CROPGRO Model. p. 651–693. *In* R.M. Peart and R.B. Curry (ed.) *Agricultural systems modeling and simulation*. Marcel Dekker, New York.
- Boote, K.J., M.J. Krotz, and P.S. Brindaban. 2001. Physiology and modeling of traits in crop plants: Implications for genetic improvement. *Agric. Syst.* 70:395–420.
- Buzzel, R.I. 1971. Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can. J. Genet. Cytol.* 13:703–707.
- Buzzel, R.I., and H.D. Voldeng. 1980. Inheritance of insensitivity to daylength. *Soybean Genet. Newsl.* 7:26–29.
- Chapman, S., M. Cooper, D. Podlich, and G. Hammer. 2003. Evaluating plant breeding strategies by simulating gene action in dryland environment effects. *Agron. J.* 95:99–113.
- Cober, E.R., D.W. Stewart, and H.D. Voldeng. 2001. Photoperiod and temperature responses in early-maturing near-isogenic soybean lines. *Crop Sci.* 41:721–727.
- Cober, E.R., J.W. Tanner, and H.D. Voldeng. 1996. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. *Crop Sci.* 36:601–605.
- Cober, E.R., and H.D. Voldeng. 2001. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. *Crop Sci.* 41:698–701.
- Cooper, M., S.C. Chapman, D.W. Podlich, and G.L. Hammer. 2002. The GP problem: Quantifying gene-to-phenotype relationships [Online]. Available at [www.bioinfo.de/isb/2002/02/0013/](http://www.bioinfo.de/isb/2002/02/0013/) (verified 4 Sept. 2005). *In Silico Biol.* 2:0013.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. vanToai, D.G. Lohnes, L. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci.* 39:1464–1490.
- Daniell, H., and A. Dhirga. 2002. Multigene engineering: Dawn of an exciting new era in biotechnology. *Curr. Opin. Biotechnol.* 13:136–141.
- Elizondo, D.A., R.W. McClendon, and G. Hoogenboom. 1994. Neural network models for predicting flowering and physiological maturity of soybean. *Trans. ASAE* 37:981–988.
- Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Spec. Rep. 80. Iowa Agric. Home Econ. Exp. Stn., Iowa State Univ., Ames.
- Grant, D., and R.C. Shoemaker. 2005. SoyBase, the USDA-ARS soybean genome database [Online]. Available at [soybase.org](http://soybase.org) (verified 4 Sept. 2005).
- Grimm, S.S., J.W. Jones, K.J. Boote, and D.E. Herzog. 1994. Modeling the occurrence of reproductive stages after flowering for 4 soybean cultivars. *Agron. J.* 86:31–38.
- Grimm, S.S., J.W. Jones, K.J. Boote, and J.D. Hesketh. 1993. Parameter estimation for predicting flowering date of soybean cultivars. *Crop Sci.* 33:137–144.
- Hammer, G.L., T.R. Sinclair, S.C. Chapman, and E. van Oosterom. On systems thinking, systems biology, and the in silico plant. 2004. *Plant Physiol.* 134:909–911.
- Hoogenboom, G., and J.W. White. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agron. J.* 95:82–89.
- Hoogenboom, G., J.W. White, J. Acosta-Gallegos, R.G. Gaudiel, J.R. Myers, and M.J. Silbernagel. 1997. Evaluation of a crop simulation model that incorporates gene action. *Agron. J.* 89:613–620.
- Hoogenboom, G., J.W. White, and C.D. Messina. 2004. From genome to crop: Integration through simulation modeling. *Field Crops Res.* 90:145–163.
- Huang, J., C. Pray, and S. Rozelle. 2002. Enhancing the crops to feed the poor. *Nature (London)* 418:678–684.
- Hunt, L.A., and K.J. Boote. 1998. Data for model operation, calibration and evaluation. p. 41–54. *In* Tsuji, G.Y. et al. (ed.) *Understanding options for agricultural production*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Hunt, L.A., S. Pararajasingham, J.W. Jones, G. Hoogenboom, D.T. Imamura, and R.M. Ogoshi. 1993. GENCALC: Software to facilitate the use of crop models for analyzing field experiments. *Agron. J.* 85:1090–1094.
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijssman, and J.T. Ritchie. 2003. The DSSAT cropping system model. *Eur. J. Agron.* 18:235–265.
- Keating, B.A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, and C.J. Smith. 2003. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* 18:267–288.
- Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon specific allele at another. *Proc. Natl. Acad. Sci. USA* 92:4656–4660.
- Lark, K.G., J. Orf, and L.M. Mansur. 1994. Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max*, (L) Merr] determined by QTL association with RFLP alleles. *Theor. Appl. Genet.* 88:486–489.
- Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max*, L.). *Theor. Appl. Genet.* 86:907–913.
- Mansur, L.M., J. Orf, K. Chase, T. Jarvik, P. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci.* 36:1327–1336.
- Mavromatis, T., K.J. Boote, J.W. Jones, A. Irmak, D. Shinde, and G. Hoogenboom. 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. *Crop Sci.* 41:40–51.
- Mavromatis, T., K.J. Boote, J.W. Jones, G.G. Wilkerson, and G. Hoogenboom. 2002. Repeatability of model genetic coefficients derived from soybean performance trials across different states. *Crop Sci.* 42:76–89.
- McBlain, B.A., and R.L. Bernard. 1987. A new gene affecting the time of flowering and maturity in soybeans. *J. Hered.* 78:160–162.
- McBlain, B.A., J.D. Hesketh, and R.L. Bernard. 1987. Genetic effects on reproductive phenology in soybean isolines differing in maturity genes. *Can. J. Plant Sci.* 67:105–116.
- Messina, C.D. 2003. Gene-based systems approach to simulate soybean growth and development and application to ideotype design in target environments. Ph.D. diss. Univ. of Florida, Gainesville, FL.
- Midwestern Regional Climate Center. 2005. Weather data [Online]. Available at [mcc.sws.uiuc.edu/](http://mcc.sws.uiuc.edu/) (verified 25 Aug. 2005).
- Molnar, S.J., S. Rai, M. Charette, and E.R. Cober. 2003. Simple sequence repeat (SSR) markers linked to *E1*, *E2*, *E4* and *E7* maturity genes in soybean. *Genome* 46:1024–1036.

- Murray, M.G., and W.F. Thompson. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 19:4321-4325.
- Orf, J.H., K. Chase, F.R. Adler, L.M. Mansur, and K.G. Lark. 1999. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. *Crop Sci.* 39:1652-1657.
- Pinheiro, J.C., and D.M. Bates. 2000. *Mixed-Effects Models in S and S-Plus*. Springer-Verlag, New York.
- Piper, E.L., K.J. Boote, J.W. Jones, and S.S. Grimm. 1996. Comparison of two phenology models for predicting flowering and maturity date of soybean. *Crop Sci.* 36:1606-1614.
- Reymond, M., B. Muller, A. Leonardi, A. Charcosset, and F. Tardieu. 2003. Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 131:664-675.
- Saindon, G., W.D. Beversdorf, and H.D. Voldeng. 1989. Adjustment of the soybean phenology using the *E4* locus. *Crop Sci.* 29:1361-1365.
- Somerville, C., and S. Somerville. 1999. Plant functional genomics. *Science* (Washington, DC) 285:380-383.
- Stewart, D.W., E.R. Cober, and R.L. Bernard. 2003. Modeling genetic effects on the photothermal response of soybean phenological development. *Agron. J.* 95:65-70.
- Stuber, C.W., M. Polacco, and M.L. Senior. 2003. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci.* 39:1571-1583.
- Summerfield, R.J., H. Asumadu, R.H. Ellis, and A. Qi. 1998. Characterization of the photoperiodic response of post-flowering development in maturity isolines of soybean [*Glycine max* (L.) Merrill] 'Clark'. *Ann. Bot. (London)* 82:765-771.
- Tasma, I.M., L.L. Lorenzen, D.E. Green, and R.C. Shoemaker. 2001. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Mol. Breed.* 8:25-35.
- Tasma, I.M., and R.C. Shoemaker. 2003. Mapping flowering time gene homologs in soybean and their association with maturity (*E*) loci. *Crop Sci.* 43:319-328.
- University of Illinois. 2005. Soybeans in Illinois. Variety testing data [Online]. Available at [vt.cropsci.uiuc.edu/soybean.html](http://vt.cropsci.uiuc.edu/soybean.html) (Verified 25 Aug. 2005).
- Upadhyay, A.P., R.H. Ellis, R.J. Summerfield, E.H. Roberts, and A. Qi. 1994a. Characterization of photothermal flowering responses in maturity isolines of soybean [*Glycine max* (L.) Merrill] cv. Clark. *Ann. Bot. (London)* 74:87-96.
- Upadhyay, A.P., R.J. Summerfield, R.H. Ellis, E.H. Roberts, and A. Qi. 1994b. [*Glycine max* (L.) Merrill] Variations in the durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering among eight maturity isolines of soybean. *Ann. Bot. (London)* 74:97-101.
- Vallejos, C.E., N.S. Sakiyama, and C.D. Chase. 1992. A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* 131:733-740.
- Wardlaw, I.F. 1990. The control of carbon partitioning in plants. *New Phytol.* 116:341-381.
- White, J.W., and G. Hoogenboom. 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. *Agron. J.* 88:416-422.
- White, J.W., and G. Hoogenboom. 2003. Gene-based approaches to crop simulation: Past experiences and future opportunities. *Agron. J.* 95:52-64.
- Yin, X., P. Stam, M.J. Kropff, and A.H.C.M. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agron. J.* 95:90-98.
- Zhu, Y.L., Q.J. Song, D.L. Hyten, C.P. Van Tassell, L.K. Matukumalli, D.R. Grimm, S.M. Hyatt, E.W. Fickus, N.D. Young, and P.B. Cregan. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123-1134.