Available online at www.sciencedirect.com
SCIENCE DIRECT®
Field Crops Research 90 (2004) 145–163
Field Crops Research
www.elsevier.com/locate/fcr

From genome to crop: integration through simulation modeling

Gerrit Hoogenboom[a,*], Jeffrey W. White[b], Carlos D. Messina[c]

[a]*Department of Biological and Agricultural Engineering, University of Georgia, Griffin, Georgia 30223-1797, USA*
[b]*U.S. Water Conservation Laboratory, USDA-ARS, 4331 E. Broadway Rd., Phoenix AZ 85040, USA*
[c]*Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida 32611, USA*

Abstract

Crop models use mathematical equations to simulate growth, development and yield as a function of weather, soil conditions and crop management. Such models integrate scientific knowledge from diverse agronomic disciplines, ranging from plant breeding to soil physics. Most crop models use one or more cultivar-specific parameters to identify differences in performance among cultivars. Until recently, however, there was little relation between cultivar-specific parameters and genotypes. The GeneGro model simulates the impact of seven genes on physiological processes in common bean (*Phaseolus vulgaris* L.), specifying cultivar differences through the presence or absence of the seven genes. The model was based on the bean model BEANGRO. GeneGro has now been incorporated into the cropping system model (CSM), which can simulate growth and development for more than 20 different crops, although the CSM-GeneGro version is currently implemented only for common bean and soybean [*Glycine max* (L.) Merr.]. Gene-based models can provide a well-structured linkage between functional genomics and crop physiology, especially as more genes are identified and their functions are clarified. Incorporating genetic information strengthens underlying physiological assumptions of the model, improving its utility for research in crop improvement, crop management, global change, and other fields. We first briefly review issues related to development of gene-based models, ranging from modeling approaches to data management. The CSM-GeneGro model is then used to show how specific genes can simulate both yield levels and yield variability for three locations in the USA. The model is also used to examine how single genes can affect crop response to global change. Gene-based modeling approaches could significantly enhance our ability to predict how global change will impact agricultural production, but modelers and physiologists will have to be proactive in accessing information and tools being developed in the plant genomics community.

Keywords: Common bean; Genomics; Global warming; Temperature; Crop simulation model

1. Introduction

Computer simulation models have been used in various agronomic disciplines for nearly 40 years, becoming more complex with advances in scientific knowledge and computer science. Such models are used to examine the underlying physical, chemical and

* Corresponding author. Tel.: +1 770 229 3438; fax: +1 770 228 7218.
E-mail address: gerrit@griffin.uga.edu (G. Hoogenboom).

0378-4290/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.fcr.2004.07.014

biological processes of phenomena that are studied, such as water movement in soils or evapotranspiration in the soil–plant–atmosphere continuum, and to express the known science as mathematical equations. Models can simulate known processes and predict responses of the system to changes in boundary conditions and inputs. Crop models have been well received in fields where describing physical processes is common, such as soil physics and agrometeorology. Applications include impacts of global warming (e.g. Mearns et al., 1999; Alexandrov and Hoogenboom, 2000; Carbone et al., 2003; Jones and Thornton, 2003), crop response to sowing dates (Singh et al., 1994a,b; Acosta-Gallegos and White, 1995; Hunt et al., 1996), crop response to spacing (Egli and Bruening, 1992), characterizations of production environments (White et al., 1995; Chapman et al., 2000), and regional targeting of technologies (Hartkamp et al., 2004). Simulation models have seen less enthusiastic reception in the general fields of agronomy and crop science, where the perception appears to prevail that the models are unable to simulate the soil–plant–atmosphere system due to inherent complexity (Boote et al., 1996; Monteith, 1996; Sinclair and Seligman, 1996).

Efforts have been made to indicate how models could be used to assist in crop improvement efforts (White, 1998). Examples include simulating differences in variety trials (Mavromatis et al., 2001, 2002), identifying unique physiological traits of potentially high yielding lines (Hammer et al., 1996; Boote et al., 2001; Banterng et al., 2004a,b) and suggesting breeding strategies for drought tolerance (Spitters and Schapendonk, 1990). Nonetheless, wider use of models in crop improvement is limited because cultivar differences are inadequately represented.

Rapid advances in characterizing plant genomes and resulting understanding of growth and developmental processes suggests that information from plant genomics will finally allow physiologists and modelers to implement a much-needed overhaul of how plant responses are represented in models. This promise is supported by successful use of genetic information in the common bean (*Phaseolus vulgaris* L.) model GeneGro both to specify cultivar differences (White and Hoogenboom, 1996; Hoogenboom et al., 1997) and to modify how physiological processes were represented (Hoogenboom and White, 2003). Surprisingly, although other crops are better characterized genetically, gene-based approaches have received limited attention in the modeling community. A symposium 'Crop Modeling and Genomics' was held at the 2000 ASA-CSSA-SSSA meetings in Minneapolis, MN (Weiss, 2003), but only three of the ten papers illustrated gene-based approaches (Stewart et al., 2003; Hoogenboom and White, 2003; Welch et al., 2003). Subsequently, Messina et al. (2002) developed a version of GeneGro for soybean and implemented this in CSM-GeneGro–Soybean. Based on 48 near-isogenic lines with cvs. Clark and Harosoy as genetic backgrounds, the effects of six *E* loci on growth and development were characterized and then used to estimate cultivar coefficients in CSM-GeneGro–Soybean. The prediction of phenology and yield with CSM-GeneGro–Soybean was improved by these changes (Messina, 2003).

Drawing from our experiences with GeneGro and initial efforts to apply gene-based approaches to other crops, this paper first reviews key issues in applying information from genomics in simulation models, building on the review presented in the 2000 symposium (White and Hoogenboom, 2003). We then illustrate how GeneGro can be used to explore impacts of global warming on production of common bean in different environments and specifically, whether warming would require adaptation in cultivar types, accounting both for mean response and risk due to variability.

2. Selected issues in applying genomics to simulation modeling

Genetics plays a key role in crop simulation models, since crop species and cultivars respond differently to temperature, day length, solar radiation, water, nitrogen, and other environmental factors (Fig. 1). However, simulation models vary greatly in their genetic complexity. Some models ignore genetics completely, some use a few parameters to deal with unique crop or cultivar responses, and others provide detailed response functions and parameters at both the crop and cultivar level. White and Hoogenboom (2003) identified six levels of genetic details in crop simulation models:

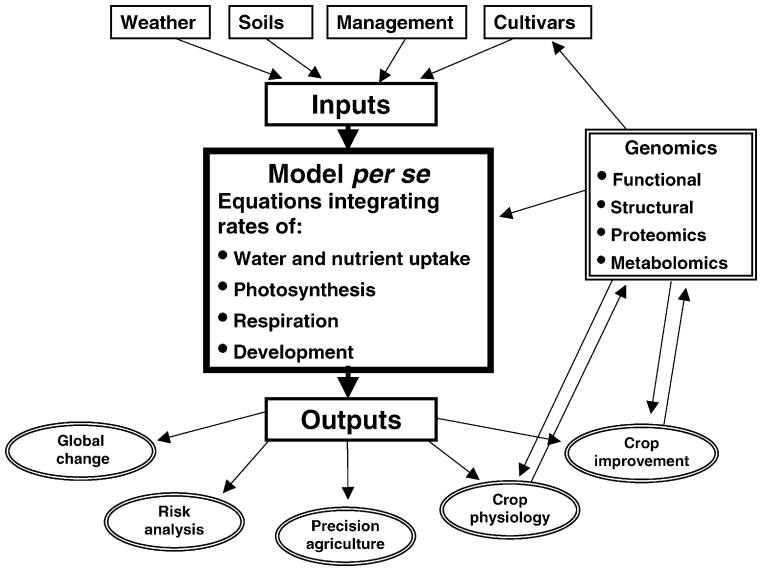

Fig. 1. Diagram indicating possible relations among modeling, genomics and common model applications.

1. Generic model with no reference to species.
2. Species-specific model with no reference to genotypes.
3. Genetic differences represented by cultivar-specific parameters.
4. Genetic differences represented by specific alleles, with gene action represented through linear effects on model parameters.
5. Genetic differences represented by genotypes, with gene action explicitly simulated based on knowledge of regulation of gene expression and effects of gene products.
6. Genetic differences represented by genotypes, with gene action simulated at the level of interactions of regulators, gene-products, and other metabolites.

Examples of levels four to five are found in papers from the 'Crop Modeling and Genomics' symposium (Stewart et al., 2003; Hoogenboom and White, 2003; Welch et al., 2003). Level six is currently limited to models for bacteria and yeast (McAdams and Arkin, 1998).

Operating at the fourth level, GeneGro version 1 incorporated effects of seven genes to simulate crop growth, development and ultimately, grain yield and components, using inputs of initial soil conditions, crop management, and daily weather (White and Hoogenboom, 1996; Hoogenboom et al., 1997). In comparisons, GeneGro simulated growth and development as well as BEANGRO (Hoogenboom et al., 1992, 1994), from which it was originally derived, for environments ranging from tropical conditions in Colombia to continental conditions in the northern US and Canada.

A major limitation to further development of GeneGro was lack of information on modes of actions of known genes or descriptions of new genes (Hoogenboom et al., 1997; Hoogenboom and White, 2003). One might expect a flood of useful insights in recent years, following the advances in plant genomics and related fields. For common bean, only marginal progress in understanding of gene action and physiological traits is found, this coming from two studies using QTL analyses. The *Phs* gene, which controls differences in the phaseolin seed storage protein, was shown to affect seed size (Johnson et al., 1996), suggesting that it is one of the *Ssz* genes used in GeneGro, and beans of Mesoamerican origin may carry a locus for indeterminate growth habit that is distinct to the *Fin* locus (Kolkman and Kelly, 2003). No genes useful in modeling have been identified since the *Tip* gene, which affects temperature sensitivity of the photoperiod response (White et al., 1996) and whose affect was examined in GeneGro version 2 (Hoogenboom and White, 2003).

Our initial efforts to implement gene-based models suggest that discovery of physiologically useful genes has also proven difficult for soybean and wheat (*Triticum aestivum* L.; Messina et al., 2002). Extrapolation from arabidopsis [*Arabidopsis thaliana* (L.) Heynh.] has yielded fewer successes than initially expected. For responses such as vernalization, arabidopsis and wheat appear to have evolved partially distinct mechanisms (Yan et al., 2003, 2004; Laurie et al., 2004). Besides evolutionary differences, comparisons of gene expression in arabidopsis plants from field and controlled environments suggest substantial differences in gene action occur in key processes such as photosynthesis (Külheim et al., 2002) and development (Weinig et al., 2002). Thus, progress through sequencing genomes of crop plants and characterizing gene expression patterns in crops grown under relevant field conditions may offer greater benefits than was expected during the early phase of optimism over direct application of information from the arabidopsis genome.

Assuming the rate of gene discovery does accelerate for traits and species of agricultural importance, the question remains whether models will soon have to deal with perhaps 100 genes, which is within the realm of conventional modeling approaches, or whether models will need to account for action of potentially thousands of interacting genes and gene products (White and Hoogenboom, 2003). The initial successes of GeneGro and the CSM-GeneGro–Soybean model argue that simple models of gene effects have much to offer. In contrast, an assessment of the impact of different genetic backgrounds of wheat lines on model parameterization found that model parameters varied greatly with genetic background of the cultivars, suggesting that there are cumulative effects of large numbers of genes (Hunt et al., 2003). Perhaps simplistic gene-based approaches will explain the easy 80% of variation related to photoperiod response, vernalization and gross differences in growth habit—variation that plant breeders and agronomists handle intuitively or through qualitative concepts such as growth habit and maturity groups. However, as we seek further refinements, the inherent complexity of plant physiology will require new modeling approaches—or simply prove intractable. The question of 'few or many?' represents a basic question for plant biology. The emergence of systems biology as a new discipline dealing with descriptive and mechanistic models of networks of genes and gene products (Aggarwal and Lee, 2003; Ge et al., 2003; Blanchard, 2004) implies that researchers from the genomics community expect complexity to prevail. Messina (2003) and Hammer et al. (2004), however, emphasize the formidable challenges implicit in scaling up a ''systems biology approach'' to whole plants.

Accepting the optimistic view that flow of usable data on new genes and their modes of action will increase rapidly in coming years, yet another challenge for modelers will be to devise strategies that allow for rapid modification and testing of model equations. A foundation of this approach is isolating physiological responses in discrete software units, such as ''modules'' in the sense of Porter et al. (1999) and implemented in Cropping System Model (CSM; Jones et al., 2003). A further step, however, is needed to allow users to modify systems of equations dynamically. Such a capacity is available in various simulation environments. For instance, in the genomics community there are systems for dynamic modeling of cell biochemistry, such as the E-Cell 3 Simulation Environment (Takahashi et al., 2002), that merit exploration by the crop modeling community.

Calls for improved data management usually elicit a lukewarm response, at best. Based on our experiences calibrating, evaluating and applying GeneGro, and noting how bioinformatics underpins all aspects of genomics, including applications to crop improvement, we reiterate the need for crop physiologists, agronomists, breeders and modelers to integrate data across disciplines. For most crops, current ability to link cultivars to specific genotypes is marginally more than using web-based search engines to locate research papers, the researcher then searching for needed information in each paper. Integration requires guidelines or standards for documenting field research and model inputs, such as promoted by the international consortium for agricultural systems applications (ICASA; Hunt et al., 2001). ICASA also recently established the ICASA Data Exchange (IDE) to promote and facilitate the exchange of experimental data sets among modelers, crop physiologists and agronomists (www.ICASA.net; Bostick et al., 2004). Initiatives to establish standard nomenclatures for plant biology and modeling metabolic and control networks (Blanchard, 2004) also merit consideration.

Apathy towards data management often extends to data analysis procedures. But again, looking toward successes in bioinformatics and the much broader fields of data mining and visualization, there is value in looking at novel approaches for examining potentially massive amounts of model outputs. GeneGro version 1 described effects of seven genes (White and Hoogenboom, 1996). From a potential of 128 (2^7) combinations of homozygous lines, 96 distinct phenotypes were possible. Extrapolating this to a model using 14 genes, one might encounter over 9000 phenotypically distinguishable lines. Considering different environments, management scenarios and response variables further increases the volume of data. Fortunately, advances in data visualization and data mining (Fayyad et al., 2002; Keim, 2002) suggest numerous options for analyzing data from large, complex modeling studies. We illustrate below one such approach as applied to global warming scenarios modeled using GeneGro.

3. Materials and methods

3.1. Model description

The previous versions of the GeneGro model (versions 1 and 2) were based on the dry bean model BEANGRO (White and Hoogenboom, 1996; Hoogenboom and White, 2003), which was created by adapting the SOYGRO model (Wilkerson et al., 1983). Recognizing physiological similarities among grain legumes and to speed model development, BEANGRO and SOYGRO were combined along with the peanut (*Arachis hypogea* L.) model into a generic model, CROPGRO, for the simulation of growth and development of grain legumes (Boote et al., 1998). Crops added subsequently include chickpea (*Cicer arietenum* L.), cowpea [*Vigna unguiculata* (L.) Walp.], faba bean (*Vicia faba* L.) and velvet bean [*Mucuna pruriens* (L.) DC; Singh and Virmani, 1996; Hartkamp et al., 2002a,b; Boote et al., 2002]. A similar generic model was developed for grain cereals and included the CERES-Maize, CERES-Wheat, CERES-Rice, CERES-Barley and CERES-Sorghum models (Ritchie et al., 1998). The grain cereal model CERES and the grain legume model CROPGRO shared common routines for the dynamic simulation of soil water and nitrogen, while CROPGRO also simulated nitrogen fixation (Godwin and Singh, 1998; Ritchie, 1998).

Capitalizing on advances in computer programming and to expand potential applications, the CROPGRO and CERES models, as well as the potato (*Solanum tuberosum* L.) model SUBSTOR, were further combined into one model, CSM (Jones et al., 2003). In CSM, soil water, nitrogen and carbon balances are simulated by common routines, while crop growth and development are simulated by individual plant modules specific for groups of crops. Thus, the CERES module handles maize, wheat, rice (*Oryza sativa* L.), barley (*Hordeum vulgare* L.), grain sorghum (*Sorghum bicolor* L.) and pearl millet (*Pennisetum americanum* L.), while the CROPGRO module handles soybean, peanut, dry bean and other grain legumes. In the CROPGRO module, computer code for growth and development processes is identical, and differences in genetics among species are specified through external data files, called species files (Table 1). These files define responses to temperature, solar radiation, CO_2, and photoperiod, drought and nitrogen stresses, as well plant composition and other functions and parameters. Within each species, ecotype files define coefficients for groups of cultivars that show similar behavior and responses to environment (Table 1). Cultivar differences are further specified in cultivar files including parameters for sensitivity to photoperiod, photothermal days to flowering and maturity, and seed size (Table 1). The distinction between ecotype and cultivar parameters is somewhat arbitrary; in many cases, coefficients that are constant among groups of cultivars are moved to the ecotype file.

To accommodate numerous changes from BEANGRO to CSM-CROPGRO, thus creating CSM-GeneGro–Common Bean, genetic effects had to be re-estimated. The genes considered were *Ppd* for basic photoperiod response, *Hr* for enhanced effect of *Ppd*, *Fin* for indeterminate versus determinate stem and growth habit, *Fd* for early flowering and *Ssz-1*, *Ssz-2*, and *Ssz-3* for seed size (Table 2). As in GeneGro version 1, most genetic effects were estimated with linear models and ordinary least squares regression, treating the individual ecotypic and cultivar coefficients (Table 1) as dependent variables and coding independent variables (genetic effects) with values of 1 or 0 for homozygous dominant or recessive,

Table 1
Examples of parameters used to represent species, ecotype and cultivar differences used for common bean in the CSM model

Parameter description	Abbreviation	Units
Species file		
Temperature effect on vegetative development	TB, TO1, TO2, TM	Thermal time
Temperature effect on early reproductive development	TB, TO1, TO2, TM	Thermal time
Temperature effect on late reproductive development	TB, TO1, TO2, TM	Thermal time
Vegetative partitioning to leaves, stems, and roots	XLEAF, YLEAF, YSTEM	g g^{-1}
Protein composition of leaves, stems and roots	PROLFI, PROSTI, PRORTI	g[protein] g[biomass]$^{-1}$
Temperature effect on canopy photosynthesis	FNPGT	[relative]
Nitrogen effect on canopy photosynthesis	FNPGN	[relative]
Ecotype file		
Minimum rate of reproductive development under long days and optimal temperature	THVAR	Photothermal days
Time between planting and emergence	PL-EM	Thermal days
Time required from emergence to first true leaf	EM-V1	Thermal days
Time required from first true leaf to end of juvenile phase	V1-JU	Thermal days
Time required for floral induction	JU-R0	Thermal days
Proportion of time between first seed and physiological maturity that the last seed can be formed	PM09	[relative]
Time required for growth of individual shells	LNGSH	Photothermal days
Time between physiological and harvest maturity	R7–R9	Days
Rate of leaf appearance of leaves on the main stem	TRIFL	Photothermal days
Maximum ratio of seed over pod weight at maturity	THRSH	Percent
Fraction protein in seeds	SDPRO	g[oil] g[seed]$^{-1}$
Cultivar file		
Critical short day length for reproductive development	CSDL	Hour
Relative response of reproductive development to photoperiod	PPSEN	h^{-1}
Time between plant emergence and flower appearance	EM-FL	Photothermal days
Time between first flower and first pod	FL-SH	Photothermal days
Time between first flower and first seed	FL-SD	Photothermal days
Time between first seed and physiological maturity	SD-PM	Photothermal days
Time between first flower and end of leaf expansion	FL-LF	Photothermal days
Maximum leaf photosynthesis rate at 30 °C, 350 vpm CO_2 and high light	LFMAX	mg [CO_2] m^{-2} s^{-1}
Specific leaf area under standard growth conditions	SLAVR	cm^2 g^{-1}
Maximum size of full leaf (three leaflets)	SIZLF	cm^{-2}
Maximum fraction of daily growth that is partitioned to seed and shell	XFRT	[relative]
Maximum weight per individual seed	WTPSD	g
Seed filling duration for pod cohort at standard growth conditions	SFDUR	Photothermal days
Average seed per pod under standard growing conditions	SDPDV	#/pod
Time required for cultivar to reach final pod load under optimal conditions	PODUR	Photothermal days

Table 2
Effects of genes that are considered in the simulation model CSM-GeneGro

Gene	Status[a]	Effect	Reference
Ppd	N	Long days delay flowering (classic short-day response)	Wallace et al., 1993
Hr	I	Enhances effect of *Ppd* but requires *Ppd* to be present	Kornegay et al., 1993
Fin	N	Indeterminate stem type, which is associated with late flowering, vs. determinate stem type	Mendel, 1866
Fd	I	Early flowering and maturity	Coyne, 1970
Ssz-1	N	Seed size	Vallejos and Chase, 1991; Johnson et al., 1996
Ssz-2	I	Seed size	
Ssz-3	I	Seed size	

[a] N: named gene; I: gene inferred to exist based on indirect evidence.

Table 3
Genetic effects used to estimate ecotype or cultivar coefficients in the CSM-GeneGro–Drybean model

Cultivar parameter	Linear model of genetic effects	R^2
Relative response of reproductive development to photoperiod	PPSEN = 0.004 + 0.0154*Ppd + 0.036*Hr − 0.0104*Ppp*Hr	0.66
Time between plant emergence and flower appearance	EM-FL = 26.63 + 4.886*Fin − 5.188*Fd	0.58
Time between first flower and first pod	FL-SH = 4.63 + 0.972*Ssz-1 − 0.98*Ssz-2 − 1.8*Ssz-3	0.56
Time between first flower and first seed	FL-SD = 10.61 + 2.028*Ssz-2 − 2.1*Ssz-3	0.46
Time between first seed and physiological maturity	SD-PM = 21.027 − 0.11*Ssz-1 + 4.13*Hr	0.45
Time between first flower and end of main stem node formation	FL-VS = 7.00 + 4.76*Fin − 2.75*Ssz-2 − 1.02*Fin*Ssz-2	0.63
Time between first flower and end of leaf expansion	FL-LF = 18.0 + 3.8*Fd − 6.9*Ssz-2	0.61
Specific leaf area under standard growth conditions	SLAVR = 322 + 41*Ssz-1 − 38.0*Ssz-2 − 25.3*Ssz-2	0.49
Average seed per pod under standard growing conditions	SDPDVR = 5.14 − 0.2*Fin − 1.9*Ssz-1 + 0.24*Ssz-3	0.98
Maximum ratio of seed over pod weight at maturity	THRSH = 78 − 3.5*Ssz-2 + 1.5*Fin*Ssz-2	0.75
Maximum weight per individual seed	WTPSD = 0.22 + 0.21*Ssz-1 + 0.07*Ssz-2	0.90

Time required for cultivar to reach final pod load under optimal conditions[a]	IF (Hr. GE. 0.5) THEN PODUR = 5.33
	ELSE IF (Ppd. LE. 0.5) THEN PODUR = 8.45
	ELSE PODUR = 7.14
	ENDIF	

Genes are as described in Table 2, with values coded as 1 for homozygous dominant and 0 for homozygous recessive at each locus.
[a] PODUR was estimated using classification and regression trees (Venables and Ripley, 1997), a non-parametric method that does not provide an R^2-value.

respectively, at each locus. Based on expectations concerning the physiological effects of individual genes and tests of significance of regression coefficients, various combinations of genes were tested. The resulting equations and R^2-values are summarized in Table 3. Table 4 presents examples for six cultivars, showing the action of these seven genes as well as how this is translated into a digital input by coding the homozygous dominant condition as '1' and the recessive condition as '0'. The regression equations were based on experimental data presented previously by White and Hoogenboom (1996) and Hoogenboom et al. (1997). In general, physiological responses of these genes for developmental traits were lower, ranging from an R^2 of 0.45 for the number of photothermal days from first seed to physiological maturity to 0.66 for photoperiod sensitivity, than for growth traits, where R^2 ranged from 0.75 for threshing percentage to 0.98 for the number of seeds per pod.

3.2. Scenarios modeled

Four sites were selected to represent major bean production areas in the USA. The sites were Kellogg Biological Station, Michigan, Twin Falls, Idaho, Prosser, Washington, and Grand Forks, North Dakota (Table 5). Daily historical weather data were obtained from the National Climatic Data Center in Ashville, North Carolina, ranging from 39 years for Twin Falls to 77 years for Prosser. Prosser was the driest site, with an annual average total rainfall of 203 mm, while the Kellogg Biological Station (KBS) had an annual average rainfall of 930 mm (Fig. 2). Annual average maximum and minimum temperature were similar for the KBS, Twin Falls and Prosser. Grand Forks

Table 4
Example of genotypes or traits assumed for a few sample cultivars and their digital implementation in the CSM-GeneGro–Common Bean Model

Cultivars	Genes						
A 70	*Ppd*	*hr*	*Fin*	*fd*	*ssz-1*	*ssz-2*	*Ssz-3*
BAT 477	*ppd*	*hr*	*Fin*	*fd*	*ssz-1*	*ssz-2*	*Ssz-3*
Bayo Madero	*Ppd*	*Hr*	*Fin*	*Fd*	*Ssz-1*	*Ssz-2*	*Ssz-3*
Porrillo Sintetico	*Ppd*	*hr*	*Fin*	*fd*	*ssz-1*	*ssz-2*	*Ssz-3*
Rabia de Gato	*ppd*	*hr*	*Fin*	*Fd*	*ssz-1*	*ssz-2*	*Ssz-3*
Manitou	*Ppd*	*hr*	*fin*	*fd*	*Ssz-1*	*Ssz-2*	*Ssz-3*
C-20	*ppd*	*hr*	*Fin*	*fd*	*ssz-1*	*ssz-2*	*ssz-3*
	Digital Implementation						
	Ppd	*Hr*	*Fin*	*Fd*	*Ssz-1*	*Ssz-2*	*Ssz-3*
A 70	1	0	1	0	0	0	1
BAT 477	0	0	1	0	0	0	1
Bayo Madero	1	1	1	1	1	1	1
Porrillo Sintetico	1	0	1	0	0	0	1
Rabia de Gato	0	0	1	1	0	0	1
Manitou	1	0	0	0	1	1	1
C-20	0	0	1	0	0	0	0

See Table 2 for physiological effects of the seven genes.

Table 5
Locations used to examine effects of global warming on common bean using CSM-GeneGro

Site	Latitude (°)	Longitude (°)	Elevation (m)	First record	Number of years	T_{max} (°C)	T_{min} (°C)	Rain (mm)
Kellogg Biological Station, Michigan	42.40	−85.38	277	1930	73	17.6	3.9	930
Twin Falls Idaho	42.55	−114.35	1207	1963	39	17.2	1.9	271
Prosser, Washington	46.20	−119.75	253	1925	77	18.9	3.7	203
Grand Forks, North Dakota	48.17	−97.50	276	1949	53	15.6	−0.1	488

was the coldest site with an annual average minimum temperature of −0.1 °C and an annual average maximum temperature of 15.6 °C (Table 5). Monthly variations in temperatures are shown in Fig. 2 for each site; Prosser had somewhat higher minimum and maximum temperatures than Twin Falls and KBS.

At KBS, a Kalamazoo loam soil was used, the crop was planted on July 9 at a plant population of 25 plants m^{-2}, and rainfed conditions were assumed. For Twin Falls, a Pourtneuf silt loam was used, the crop was planted on June 9 at a plant population of 25 plants m^{-2}, and the crop was irrigated when 50% of the plant extractable soil moisture in the top 0.3 m of the soil profile had been removed. For Prosser, a Quincy sand was used, the crop was planted on June 3 at a plant population of 25 plants m^{-2}, and the crop was

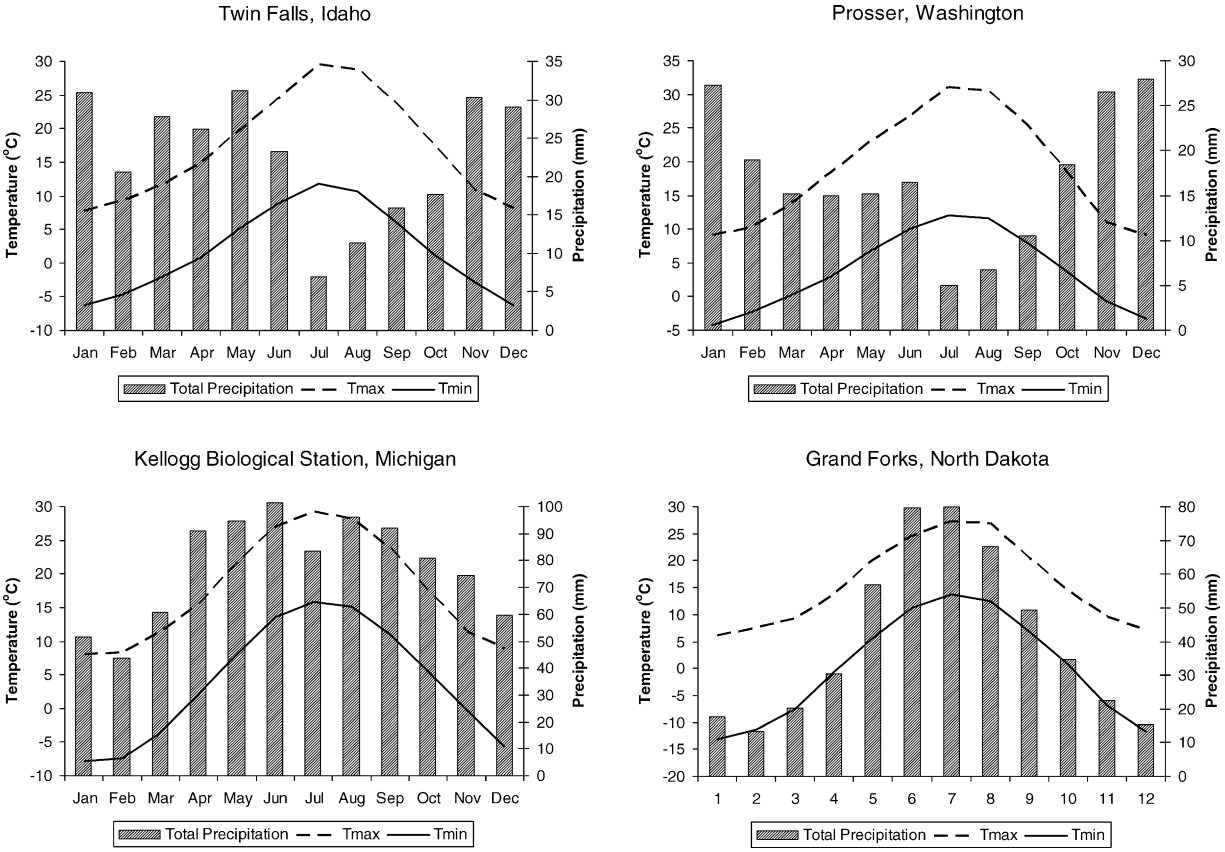

Fig. 2. Monthly average maximum and minimum temperature and total rainfall for Twin Falls, Prosser, Kellogg Biological Station, and Grand Forks for the time spans indicated in Table 5.

irrigated when 50% of the plant extractable soil moisture in the top 0.3 m of the soil profile had been removed. For Grand Forks, a frigid Typic Haplaquad was used, and the crop was planted on May 30 at a planting density of 20 plants m^{-2} with 100 kg [N]/ha applied at planting.

At each location, all genotypes for the seven genes were simulated that would result in distinct phenotypes over environments (Hoogenboom et al., 1997). These can be viewed as a set of congenic lines varying for the seven genes. Of potentially 128 different combinations of homozygous lines (2^7), only 96 combinations would result in unique phenotypes due to the epistatic effect of the *Ppd* gene on *Hr* (Tables 2 and 3; Kornegay et al., 1993).

To examine possible impacts of temperature increase, different temperature scenarios were simulated by incrementing the daily maximum and minimum temperature of the historical weather data up to a 4 °C increase. For KBS, the temperature was increased by increments of 1 °C and for Grand Forks, temperature was increased by +0.5 °C increments.

3.3. Data visualization

To facilitate examining multiple traits across the 96 lines and nine temperature regimes, simulation outputs for Grand Forks were plotted as pseudo-maps using ArcMap 8.2 (Environmental Systems Research Institute, Inc.; Redlands, California). For each line (combination of genotypes) and trait, the mean value over 56 years was calculated. Duration of grain filling was estimated as the difference between anthesis and maturity dates. Water use efficiency (WUE) and nitrogen use efficiency (NUE) were calculated as the ratio of grain yield to total evapotranspiration and crop nitrogen uptake, respectively. The means were standardized on a 0–1 scale, with a value of one corresponding to the maximum value across lines and temperature regimes.

Values for using *x*-coordinates of each line were generated based on gene combinations. Thus, the line that was dominant at all loci was located in position 1, and the line that was recessive at all loci, position 96. For *y*-coordinates, data for each trait were presented in separate blocks, with each row (*y*-value) within a block corresponding to a single temperature regime. To index the lines, lines were coded 0 (red) or 1 (blue) for recessive or dominant homozygous at each of the seven loci, this value being displayed as separate rows at the bottom of the main data area.

4. Results and discussion

The new model, CSM-GeneGro–Common Bean, performed better than previous versions. For example, in comparing simulated to observed values, R^2 for seed yield increased from 0.45 for GeneGro v. 2 to 0.79 for CSM-GeneGro–Common Bean, while the R^2 for seed size increased from 0.53 to 0.69 (all R^2-values significant at $P < 0.01$ level). Comparisons for CSM-GeneGro–Common Bean were based on six trials from four different locations in Florida, USA, Guatemala and Colombia, with 21 cultivars representing 65 treatment combinations, while GeneGro V. 2 simulations were based on eight trials from five locations, ranging from Canada to Colombia with 28 cultivars, representing approximately 200 treatment combinations.

4.1. Genotype and environmental interactions

For each location, performance of all lines was simulated for each available weather year (Table 5). With access to this large number of historical weather years, which ranged beyond the duration of many long-term trials, the probable impact of weather variability on common growth, development and yield could be assessed. Fig. 3a summarizes simulations for KBS. The median yield and 0 to 25th, 25th to 75th and 75th to 100th percentiles are shown for each line, with lines (genotype combinations) represented on the *x*-axis. Yield was as low as 23 kg ha^{-1} and as high as 3975 kg ha^{-1}. Median yield varied between 2200 and 2400 kg ha^{-1}; minimum yield did not vary much among lines, showing that for most environments a certain combination of weather and soil conditions causes crop failure. However, maximum yield varied significantly and was as low as 2500 kg $ha^{-1}$for line 0001100 (line no. 84; *Fd* and *Ssz-1* dominant) and as high as 4100 kg ha^{-1} for line 1110110 (line no. 10; *fd* and *ssz-3* recessive).

In crop improvement, one usually seeks to combine high yields with stable performance across environments, such as different weather years. An example of the relation between mean yield and variance is shown

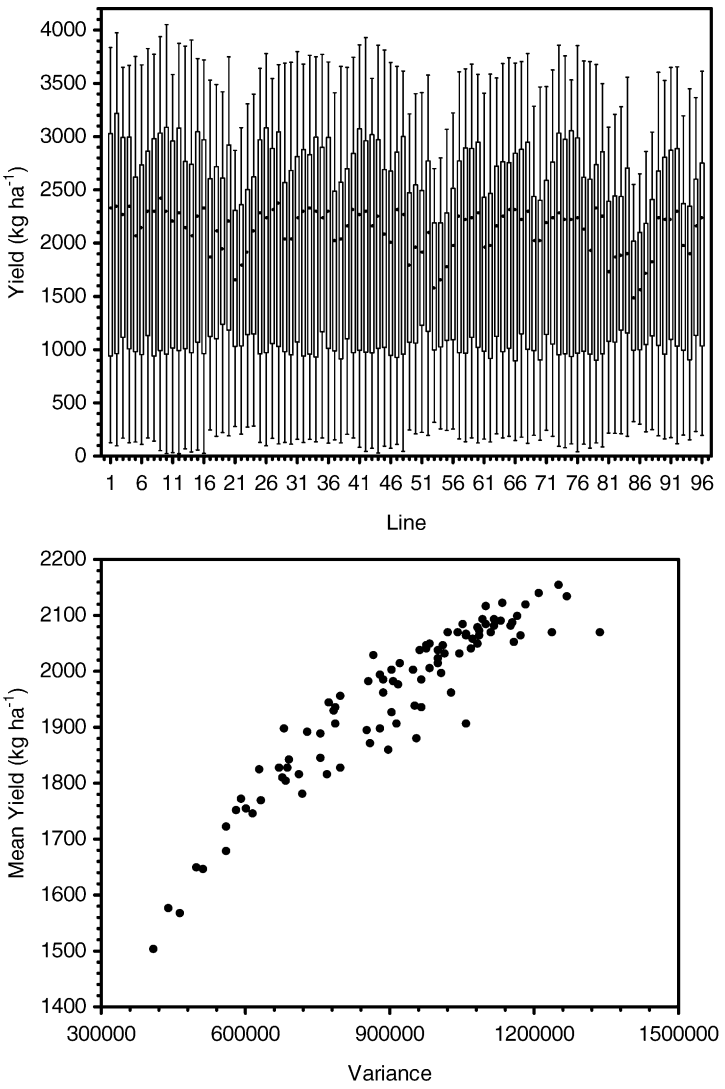

Fig. 3. Box plot for yield as a function of line (top); the box plot represents the median and the 0 to 25th, 25th to 75th, and 75th to 100th percentiles. Mean vs.variance for yield (bottom) for the Kellogg Biological Station.

in Fig. 3b. High yielding lines also were associated with high variability. Line 1111110 (line no. 2; *ssz-3* recessive) had the highest mean yield potential for all weather years. (Relations between line identifiers, genotypes and line numbers are summarized in Table 6).

Average days to harvest maturity ranged from 72 (day of year 262) to 101 days (day of year 291; Fig. 4a), with the lowest value being 68 days and the highest 118 days among all combinations of lines and years. Line 0001001 (line no. 87; *Fd* and *Ssz-3* dominant) was the earliest line with the lowest variation, but its average mean yield and standard deviation (S.D.) were only 1.648 and 707 kg ha^{-1}, respectively (Fig. 3b). The average number of days to harvest maturity for the highest yielding line was 93 days (S.D. of 4.5 days).

Seed size differentiated into three groups, small-seeded types with 150–190 mg per seed, medium-seeded types of 320–350 mg with a relative small variance, and large-seeded types of 340–390 mg with a large variance (Fig. 4b). These groups corresponded

Table 6
Examples of genotypes of the 96 hypothetical homozygous bean lines evaluated with GeneGro

Line number	Binary identifier of line	Genotype
1	1111111	*Ppd Ppd Hr Hr Fin Fin Fd Fd Ssz1 Ssz1 Ssz2 Ssz2 Ssz3 Ssz3*
2	1111110	*Ppd Ppd Hr Hr Fin Fin Fd Fd Ssz1 Ssz1 Ssz2 Ssz2 ssz3 ssz3*
3	1111101	*Ppd Ppd Hr Hr Fin Fin Fd Fd Ssz1 Ssz1 ssz2 ssz2 Ssz3 Ssz3*
9	1110111	*Ppd Ppd Hr Hr Fin Fin fd fd Ssz1 Ssz1 Ssz2 Ssz2 Ssz3 Ssz3*
10	1110110	*Ppd Ppd Hr Hr Fin Fin fd fd Ssz1 Ssz1 Ssz2 Ssz2 ssz3 ssz3*
12	1110100	*Ppd Ppd Hr Hr Fin Fin fd fd Ssz1 Ssz1 ssz2 ssz2 ssz3 ssz3*
32	1100000	*Ppd Ppd Hr Hr fin fin fd fd ssz1 ssz1 ssz2 ssz2 ssz3 ssz3*
64	1000000	*Ppd Ppd hr hr fin fin fd fd ssz1 ssz1 ssz2 ssz2 ssz3 ssz3*
81	0001111	*ppd ppd hr hr fin fin Fd Fd Ssz1 Ssz1 Ssz2 Ssz2 Ssz3 Ssz3*
85	0001011	*ppd ppd hr hr fin fin Fd Fd ssz1 ssz1 Ssz2 Ssz2 Ssz3 Ssz3*
86	0001010	*ppd ppd hr hr fin fin Fd Fd ssz1 ssz1 Ssz2 Ssz2 ssz3 ssz3*
95	0000001	*ppd ppd hr hr fin fin fd fd ssz1 ssz1 ssz2 ssz2 Ssz3 Ssz3*
96	0000000	*ppd ppd hr hr fin fin fd fd ssz1 ssz1 ssz2 ssz2 ssz3 ssz3*

The binary identifier indicates homozygous dominant (1) or recessive for each of the seven loci represented in CSM-GeneGro and as described in Table 2. The line number is as used in Figs. 3–6 and equals the column position in Fig. 8.

to those expected from the hypothesized gene action. For example, the smallest seeded group consisted of lines that were recessive at all three seed size loci (i.e. *ssz-1*, *ssz-2* and *ssz-3*) or were dominant at only one seed-size locus. Line 0001111 (line no. 81; Fd, *Ssz-1*, *Ssz-2* and *Ssz-3* dominant) had the largest seed, while the highest yielding line had a seed size of 339 mg.

The relation between mean yield and variance for Twin Falls, Idaho (Fig. 5a) differed from KBS (Fig. 3b) due to use of irrigation at Twin Falls. Mean yields for Twin Falls varied from 2914 kg ha^{-1} to 5080 kg ha^{-1}, with the highest yield obtained by line 1110111 (line no. 9; *fd* recessive; Fig. 5a). There was a range of approximately 30 days between the earliest and latest maturing lines, similar to the range at KBS. The highest yielding line took an average of 139 days to mature.

Examining variation in resource use, total season irrigation ranged from 400 to 570 mm (Fig. 5b). Where water is expected to be a limited resource under climate change conditions, breeders might select not only for stable, high-yielding lines, but also for reduced water requirements. Unfortunately, the highest yielding line had the second highest irrigation requirements, 556 mm (S.D. of 58 mm).

Prosser is another dry environment, similar to Twin Falls, and yields for rainfed conditions at Prosser were extremely low, with a maximum yield of 569 kg ha^{-1}. The strength of the relation between mean yield and variance for irrigated conditions at Prosser was considerably weaker than at Twin Falls. At the latter site, there was a somewhat asymptotic relation and in most cases, similar yielding lines had similar variances (Fig. 5a). In Prosser, lines with similar variances differed in yield by as much as 800 kg ha^{-1} (Fig. 6a). Conversely, line 1010110 (line no. 42, *hr*, *fd* and *ssz-3* recessive) and line 1110100 (line no. 12, *fd*, *ssz-2* and *ssz-3* recessive) had similar yields, but the variance in yield for line 100110 was about twice that as line 1110100. Thus, changing two loci (*Hr* and *ssz-2*) was predicted to have a dramatic effect on yield stability at Prosser, but not at Twin Falls.

The highest yielding line at Prosser was 1110110 (line no. 10, *fd, ssz-3* recessive) with an average yield of 4617 kg ha^{-1}. This line also was the latest maturing (Fig. 6b), requiring 111 days (S.D. of 3.3). Similar to yield, there was a wide range among lines for variance of maturity.

4.2. High and low yielding lines

Although most breeders select for a limited target population of environments, a gene-based modeling approach also allows us to identify lines that might perform well across different environments. All three locations shared two of the top three high-yielding lines, i.e. 1111110 (line no. 2, *ssz-3* recessive) and 1110111 (line no. 9, *fd* recessive) (Table 7). Line 1111110 was the highest yielding line for KBS, ranked second in Twin Falls and third in Prosser. Line 1110111 was the highest yielding line for Twin Falls,

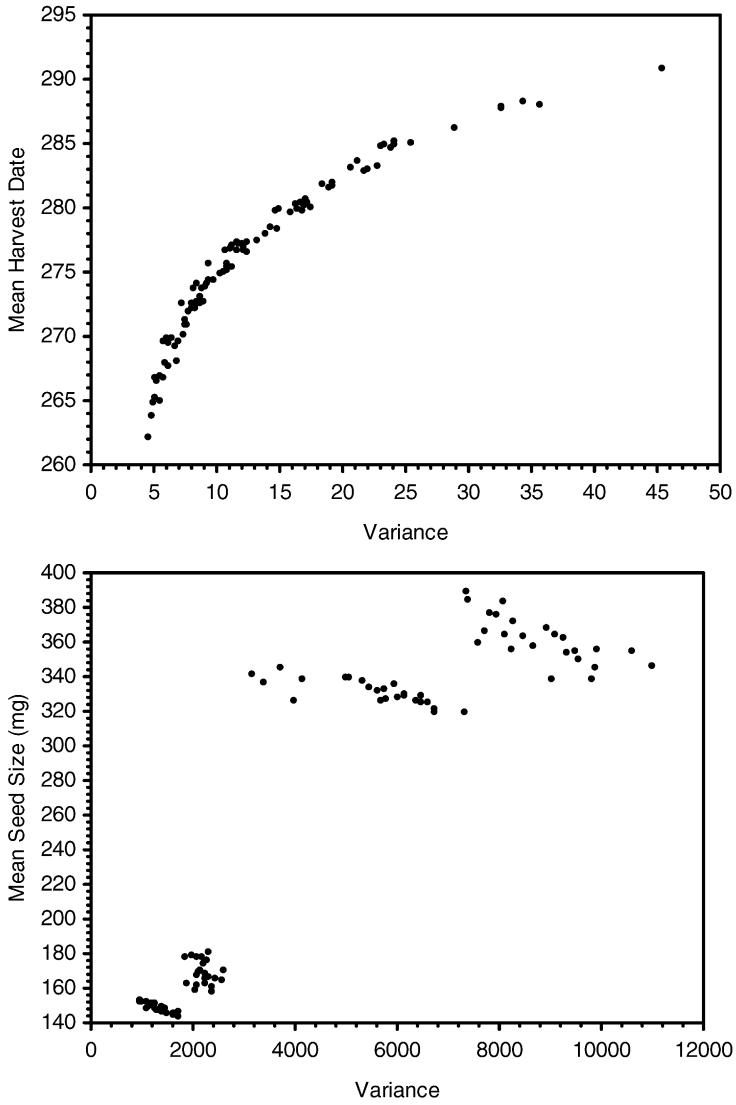

Fig. 4. Mean vs. variance for harvest date (day of year; top) and seed size/unit grain weight (bottom) for simulations at the Kellogg Biological Station.

ranked second in Prosser and third at KBS. The unique characteristics of these lines are that all genes are dominants, except for *ssz-3*, which is recessive for genotype 2 and *fd*, which is recessive for genotype 9. *Ssz-3* relates to seed size and *Fd* relates to early flowering, late flowering thus increasing yield potential for these conditions.

The three locations also shared two of three poorest performing lines, i.e. 0001011 (line no. 85, *Fd, Ssz-2* and *Ssz-3* dominant) and 0001010 (line no. 86, *Fd* and *Ssz-2* dominant; Table 6). Line 0001011 was the lowest-yielding line for all three sites, while 0001010 ranked as the second lowest-yielding line in Twin Falls and the third lowest-yielding line at KBS and Prosser. These lines share *Fd* and *Ssz-3* as dominant, while all other genes are recessive, except for *Ssz-2* for genotype 85. Thus, loci associated with the best performance were also implicated in poor performance.

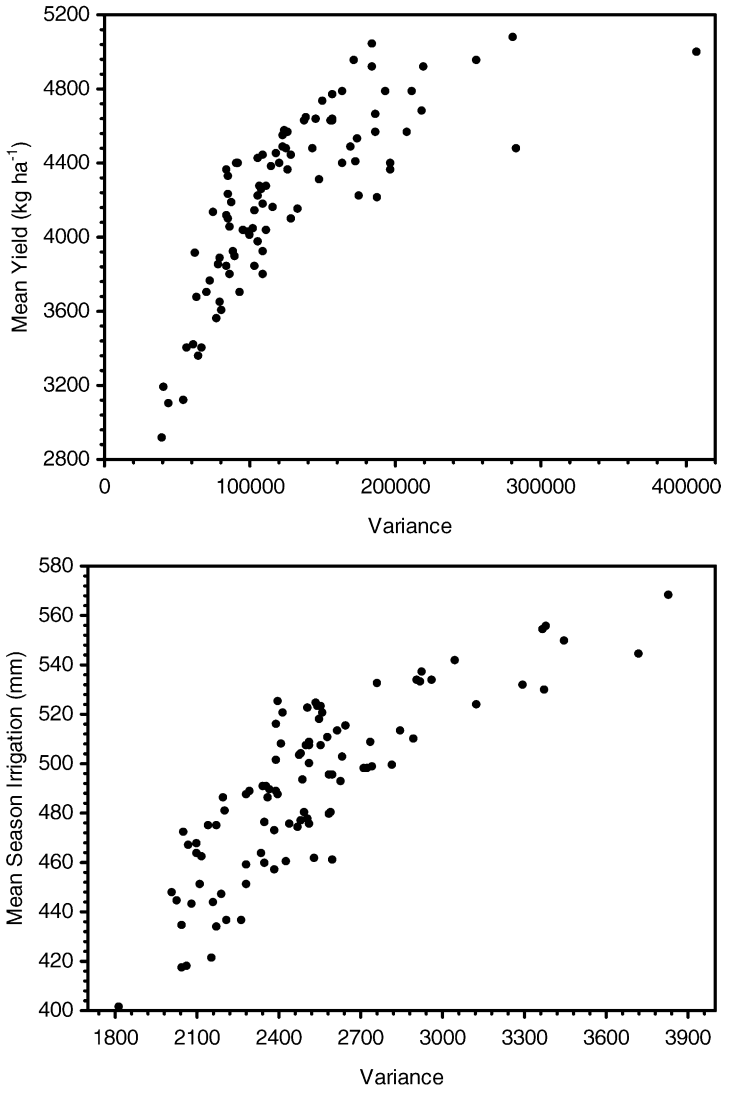

Fig. 5. Mean vs. variance for yield (top) and total water use for irrigation (bottom) for simulations at Twin Falls.

4.3. Response to increase in temperatures

For the climate change scenarios modeled for KBS, irrigated conditions were assumed to eliminate the confounding effect of drought stress on temperature stress under climate change conditions. Increasing the temperature by 4 °C decreased the number of days to anthesis by approximately 2–3 days, while the number of days to harvest maturity was reduced by 6–7 days. Due to a shorter growing season and other detrimental impacts of high temperatures, yield declined for all lines, in some cases as much as 40%. For the highest yielding line, 1111110 (line no. 2, *Ssz-3* recessive), under standard conditions, yield dropped from 3525 to 2471 kg ha^{-1}. In Fig. 7, a comparison is shown between yield based on historic weather data and yield based on increasing temperatures up to 4 °C. While yields declined at higher temperatures, certain lines were more stable than others. Line 1110110 performed relatively well under both temperature regimes: it yielded 3436 kg ha^{-1} under historic conditions and 2779 kg ha^{-1} under increased-tem-

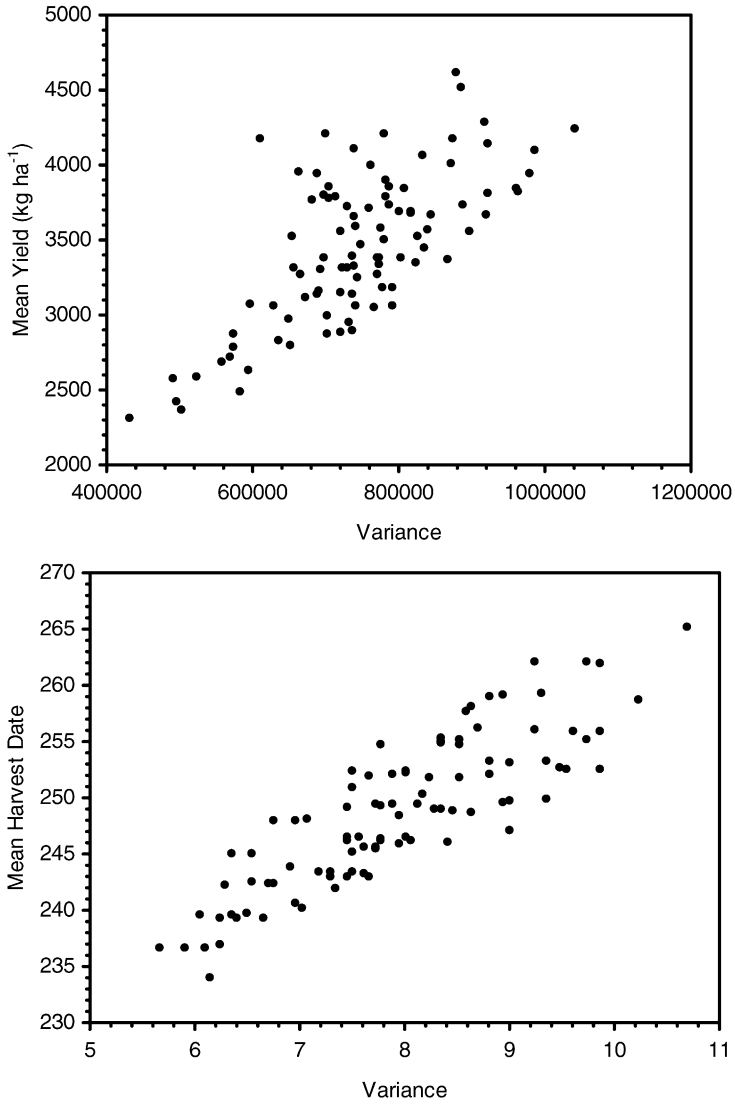

Fig. 6. Mean vs. variance for yield (top) and harvest date (day of year; bottom) for simulations at Prosser.

perature conditions, where it was also the highest ranking line.

Key loci were *Fd* and *Ssz-3*, while the other genes were dominant. The same genes that controlled stability across different environments also controlled yield stability under increased temperature regimes, keeping in mind the limited number of genes implemented in CSM-GeneGro–Common bean. The gene *Fd* plays a major role in defining the number of photothermal days from emergence to flowering, and with *fd* as recessive increases the number of days to flowering (Table 3). The gene*Ssz-3* has a major impact on the number of photothermal days from flowering to first shell and flowering to first seed (Table 3). With *ssz-3* as recessive increases the number of days to first pod and first seed. Other traits that could be examined include the various yield components, as well as resource use of water for supplemental irrigation, and environmental impact, such as through nitrogen leaching.

Table 7
Extremes of performance as simulated by GeneGro for the 96 lines and assuming historic weather conditions

Site	High yielding lines								
	Highest			Second			Third		
	Genotype		Yield kg ha^{-1}	Genotype		Yield kg ha^{-1}	Genotype		Yield kg ha^{-1}
Kellogg Biological Station, Michigan	Line 2	1111110	2153	Line 1	1111111	2138	Line 9	1110111	2134
Twin Falls, Idaho	Line 9	1110111	5080	Line 2	1111110	5039	Line 10	1110110	4994
Prosser, Washington	Line 10	1110110	4617	Line 9	1110111	4512	Line 2	1111110	4281
	Low yielding lines								
	Lowest			Second			Third		
	Genotype		Yield kg ha^{-1}	Genotype		Yield kg ha^{-1}	Genotype		Yield kg ha^{-1}
Kellogg Biological Station, Michigan	Line 85	0001011	1503	Line 53	1001011	1567	Line 86	0001010	1574
Twin Falls, Idaho	Line 85	0001011	2913	Line 86	0001010	3104	Line 87	0001001	3118
Prosser, Washington	Line 85	0001011	2306	Line 87	0001001	2362	Line 86	0001010	2414

4.4. Integrated view of response

As discussed previously, a further challenge facing researchers is how to interpret the large data sets that undoubtedly will result from simulating scenarios for global change when numerous genotypes or management options are involved. Examining mean values of six traits grown over nine temperature regimes using all 96 lines involves over 5000 data points.

Using simulations at Grand Forks as a test case, the overall trend again was that grain yield declined with warmer temperatures, but the individual lines varied in their response (Fig. 8). Days to flowering or anthesis also declined, although the lines with *fd* recessive showed less of a response than with *Fd* dominant;

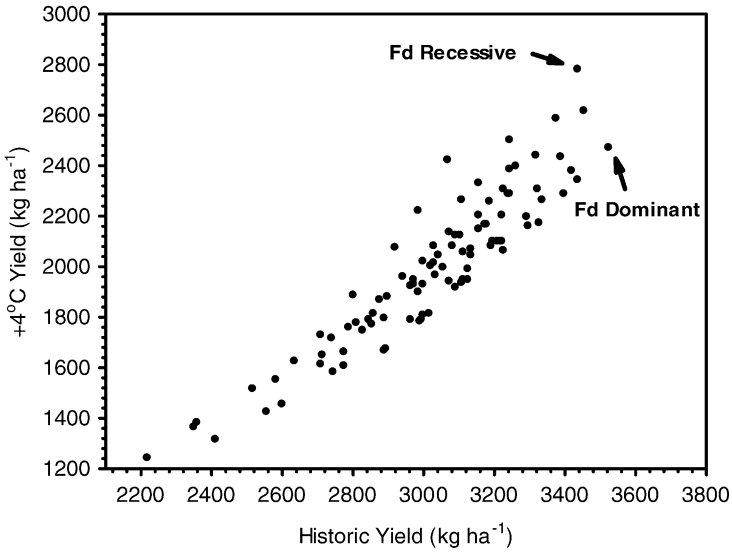

Fig. 7. Grain yield as a function of temperature change vs. yield as simulated for historical weather conditions for the Kellogg Biological Station.

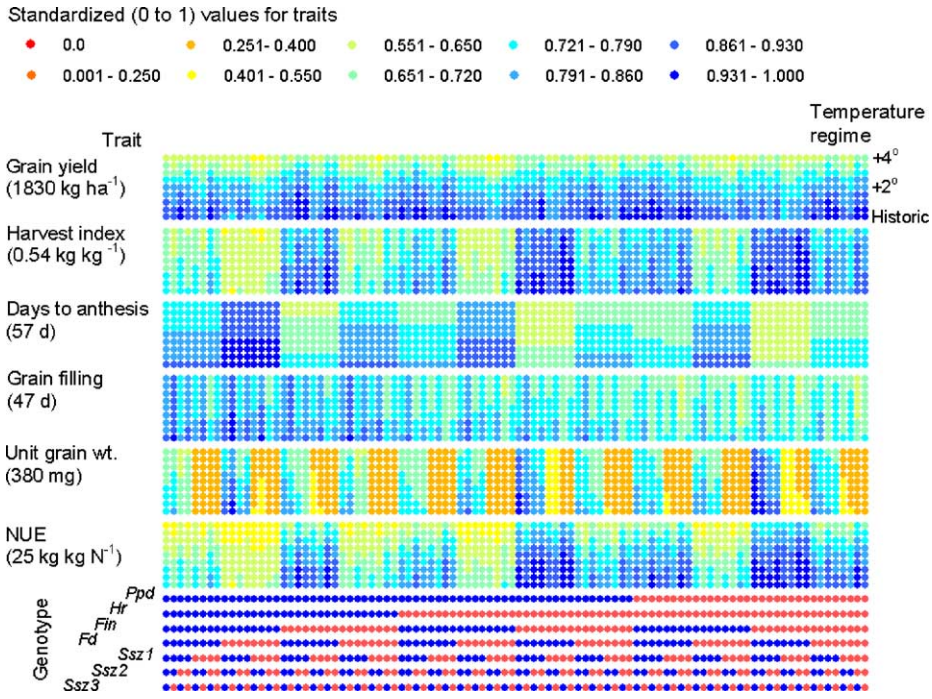

Fig. 8. Pseudo-map of showing relative variation in simulated grain yield, harvest index, days to anthesis, duration of grain filling, unit grain weight, and nitrogen use efficiency simulated with CSM-GeneGro for 96 possible homozygous lines of common bean for Grand Forks from 1949 to 2002. Each row within a trait grouping represents one of nine temperature regimes, ranging from historic value (1949–2002) to +4 °C increase using 0.5 °C increments. The lowermost rectangle on each array indicates the 96 lines, with each row corresponding to a different locus and values of 0 or 1, to recessive or dominant alleles.

there was little impact on the number of days to anthesis for gene combinations *ppd*, *hr* and *fin* all recessive and *Fd* dominant. The genes *Ssz-1*, *Ssz-2* and *Ssz-3* affected the duration of grain filling as implemented in equations shown in Table 3. Harvest index (HI) declined under higher temperatures, and similar to grain filling duration, HI was mainly affected by the *Fd* gene. For *ssz-1* recessive, there was no impact of temperature on unit grain weight or seed size, while for *Ssz-1* dominant seed size decreased under higher temperatures. The general response of nitrogen use efficiency to temperature changes was a decrease, but there did not seem to be a clear trend among the different genes.

The traits considered in Fig. 8 are only a small set of those that can be simulated through gene-based simulation. However, what is clearly needed is to identify and characterize additional genes that affect physiological traits, especially those influencing temperature effects on growth processes. Based on the modeling approach presented in this paper, this will exponentially expand the number of simulations that can be conducted for one or more environments and help determine which gene combinations confer both high and stable yields.

With the addition of more genes, however, the modeling community will face challenges in integrating these new genes and especially modes of gene-action in crop models. New modeling strategies may be needed to facilitate a seamless integration of genetic information into the crop models. There will continue to be a need for experimental data to evaluate these new modeling approaches, reiterating the value of databases of experiments that are shared among modelers and experimentalists (Bostick et al., 2004). In addition, there is a need for a routine genetic characterization of cultivars or lines used in physiological studies for evaluation of the current models, as demonstrated using CSM-GeneGro–Soybean (Messina, 2003).

5. Summary

The knowledge and insights emerging from plant genomics can greatly strengthen the assumptions of crop simulation models. Better integration of genomics into models is among promising avenues for reducing uncertainty relating to differences in physiological responses of cultivars. This benefit alone justifies a major rethinking of strategies in crop modeling, but even more important is the prospect of strengthening the fundamental physiological assumptions underlying model equations. Global change research emerges as a perhaps unexpectedly important beneficiary of this integration due to the need for models that are robust over wide geographic regions, with accompanying diversity in actual and potential cultivar types.

How well this potential will be realized remains alarmingly uncertain. Integration of modeling and genomics requires collaboration among disciplines that have seen, at best, minimal interaction, a problem also noted by Hammer et al. (2004). Substantial differences in terminology and research cultures must be overcome. At least until recently, the genomics community has seemed sufficiently self-confident of success that such high-risk partnerships may have held little attraction. Fortunately, the increasing realization that practical application of genomics in crop research requires greater emphasis on interactions of genes with environmental factors provides an important window of opportunity for developing collaborations among crop modeling, genomics, physiology and application fields such as global change research (Fig. 1).

Acknowledgements

The authors thank Greg McMaster for his insightful comments during the preparation and editing of this paper.

References

Acosta-Gallegos, J., White, J.W., 1995. Phenological plasticity as an adaptation by common bean to rainfed environments. Crop. Sci. 35, 199–204.

Aggarwal, K., Lee, K.H., 2003. Functional genomics and proteomics as a foundation for systems biology. Brief. Funct. Genomics Proteomics 2, 175–184.

Alexandrov, V.A., Hoogenboom, G., 2000. Vulnerability and adaptation assessments of agricultural crops under climate change in the Southeastern USA. Theor. Appl. Clim. 67, 45–63.

Banterng, P., Pannangpetch, K., Jogloy, S., Hoogenboom, G., Patanothai, A., 2004a. Potential applicability of the CROPGRO-Peanut model in assisting multi-location evaluation of peanut breeding. Thai J. Agric. Sc. 36 (in press).

Banterng, P., Patanothai, A., Pannangpetch, K., Jogloy, S., Hoogenboom, G., 2004b. Determination of genetic coefficients for peanut lines for breeding applications. Eur. J. Agron. (in press).

Blanchard, J., 2004. Bioinformatics and systems biology, rapidly evolving tools for interpreting plant response to global change. Field Crops Res. 90, 117–131.

Boote, K.J., Jones, J.W., Hoogenboom, G., 1998. Simulation of crop growth: CROPGRO Model. In: Peart, R.M., Curry, R.B. (Eds.), Agricultural Systems Modeling and Simulation, Marcel Dekker, Inc., New York, USA, pp. 651–692.

Boote, K.J., Jones, J.W., Pickering, N.B., 1996. Potential uses and limitation of crop models. Agron J. 88, 704–716.

Boote, K.J., Kropff, M.J., Brindraban, P.S., 2001. Physiology and modeling of traits in crop plants: implications for genetic improvement. Agric. Syst. 70, 395–420.

Boote, K.J., Mínguez, M.I., Sau, F., 2002. Adapting the CROPGRO-legume model to simulate growth of faba bean. Agron J. 94, 743–756.

Bostick, W.M., Koo, J., Walen, V.K., Jones, J.W., Hoogenboom, G., 2004. A web-based data exchange system for crop model applications. Agron J. 96 (3), 853–856.

Carbone, G.J., Mearns, L.O., Mavromatis, T., Sadler, E.J., Stooksbury, D., 2003. Evaluating CROPGRO-Soybean performance for use in climate impact studies. Agron J. 95, 537–544.

Chapman, C.S., Hammer, G.L., Butler, D.G., Cooper, M., 2000. Genotype by environment interactions affecting grain sorghum. III: Temporal sequences and spatial patterns in the target population of environments. Aust. J. Agric. Res. 51, 223–233.

Coyne, D.P., 1970. Genetic control of a photoperiod-temperature response for time of flowering in beans (*Phaseolus vulgaris* L.). Crop. Sci. 10, 246–248.

Egli, D.B., Bruening, W., 1992. Planting date and soybean yield: evaluation of environmental effect with a crop simulation model: SOYGRO. Agric. For. Meteorol. 62, 19–29.

Fayyad, U., Grinstein, G.G., Wierse, A. (Eds.), 2002. Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, San Francisco, 407 pp.

Ge, H., Walhout, J.M., Vidal, M., 2003. Integrating 'omic' information: a bridge between genomics and systems biology. Trends Gen. 19, 551–560.

Godwin, D.C., Singh, U., 1998. Nitrogen balance and crop response to nitrogen in upland and lowland cropping systems. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, Netherlands, p. 55–77.

Hammer, G.L., Butler, D.G., Muchow, R.C., Meinke, H., 1996. Integrating physiological understanding and plant breeding via crop modeling and optimization. In: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Genetic Improvement. CAB International, New York, p. 419–422.

Hammer, G.L., Sinclair, T.R., Chapman, S.C., Van Oosterom, E., 2004. On systems thinking, systems biology, and the in silico plant. Plant Phys. 134, 909–911.

Hartkamp, A.D., Hoogenboom, G., Gilbert, R., Benson, T., Tarawali, S.A., Gijsman, A., Bowen, W., White, J.W., 2002a. Adaptation of the CROPGRO growth model to velvet bean as a green manure cover crop: II: Model evaluation and testing. Field Crops Res. 78, 27–40.

Hartkamp, A.D., Hoogenboom, G., White, J.W., 2002b. Adaptation of the CROPGRO growth model to velvet bean as a green manure cover crop: I: Model development. Field Crops Res. 78, 9–25.

Hartkamp, A.D., White, J.W., Rossing, W.A.H., van Ittersum, M.K., Bakker, E.J., Rabbinge, R., 2004. Regional application of a cropping systems simulation model: crop residue retention in maize production systems of Jalisco. Mexico. Agric. Syst. 82 (2), 117–138.

Hoogenboom, G., White, J.W., 2003. Improving physiological assumptions of simulation models by using gene-based approaches. Agron. J. 95, 82–89.

Hoogenboom, G., Jones, J.W., Boote, K.J., 1992. Modeling growth, development, and yield of grain legumes using SOYGRO, PNUTGRO and BEANGRO. A review. Trans. ASAE 35, 2043–2056.

Hoogenboom, G., White, J.W., Acosta-Gallegos, J., Gaudiel, R., Myers, J.R., Silbernagel, M.J., 1997. Evaluation of a crop simulation model that incorporates gene action. Agron. J. 89, 613–620.

Hoogenboom, G., White, J.W., Jones, J.W., Boote, K.J., 1994. BEANGRO: A process-oriented dry bean model with a versatile user interface. Agron. J. 86, 182–190.

Hunt, L.A., Pararajasingham, S., Wiersma, J.V., 1996. Effects of planting date on the development and yield of spring wheat: Simulation of field data. Can. J. Plant Sci. 76, 51–58.

Hunt, L.A., Reynolds, M.P., Sayre, K.D., Rajaram, S., White, J.W., Yan, W., 2003. Crop modeling and the identification of stable coefficients that may reflect significant groups of genes. Agron. J. 95, 20–31.

Hunt, L.A., White, J.W., Hoogenboom, G., 2001. Agronomic data: advances in documentation and protocols for exchange and use. Agric. Syst. 70, 477–492.

Johnson, W.C., Menéndez, C., Nodari, R., Koinange, E.M.K., Magnusson, S., Singh, S.P., Gepts, P., 1996. Association of a seed weight factor with the phaseolin seed storage protein locus across genotypes, environments, and genomes in *Phaseolus-Vigna* spp.: Sax (1923) revisited. J. Quant. Trait. Loci. 2, 5.

Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT Cropping System Model. Eur. J. Agron. 18, 235–265.

Jones, P.G., Thornton, P.K., 2003. The potential impacts of climate change on maize production in Africa and Latin America in. 13, 51–59.

Keim, D.A., 2002. Information visualization and visual data mining. IEEE Trans. Vis. Comp. Graph. 8 (1), 1–8.

Kolkman, J.M., Kelly, J.D., 2003. QTL conferring resistance and avoidance to white mold in common bean. Crop Sci. 43, 539–548.

Kornegay, J., White, J.W., Dominguez, J.R., Tejada, G., Cajiao, C., 1993. Inheritance of photoperiod response in Andean and Mesoamerican common bean. Crop Sci. 33, 977–984.

Külheim, C., Agren, J., Jansson, S., 2002. Rapid regulation of light harvesting and plant fitness in the field. Science 297, 91–93.

Laurie, D.A., Griffiths, S., Christodoulou, V., 2004. Comparative genetic approaches to the study of control of flowering time in temperate cereals. Field Crops Res. 90, 87–99.

Mavromatis, T., Boote, K.J., Jones, J.W., Irmak, A., Shinde, D., Hoogenboom, G., 2001. Developing genetic coefficients for crop simulation models with data from crop performance trials. Crop Sci. 41, 40–51.

Mavromatis, T., Boote, K.J., Jones, J.W., Wilkerson, G.G., Hoogenboom, G., 2002. Repeatability of model genetic coefficients derived from soybean performance trails across different states. Crop Sci. 42, 76–89.

McAdams, H.H., Arkin, A., 1998. Simulation of prokaryotic genetic circuits. Annu. Rev. Biophys. Biomol. Struct. 27, 199–224.

Mearns, L.O., Mavromatis, T., Tsvetsinskaya, E., 1999. Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios.. J. Geophys. Res. 104 (D6), 6623–6646.

Mendel, G., 1866. Experiments in plant hybridisation. (English Trans.). Edited by J.H. Bennett, with commentary by R.A. Fisher. Oliver & Boyd, Edinburgh, 1965.

Messina, C.D., 2003. Gene-based systems approach to simulate soybean growth and development and application to ideotype design in target environments. Thesis (Ph.D.). University of Florida, Gainesville, Florida.

Messina, C.D., Boote, K.J., Jones, J.W., 2002. Basis for modeling genetic controls of pod addition duration in soybean. Annual Meetings Abstracts (CD-ROM). American Society of Agronomy, Madison, WI.

Monteith, J.L., 1996. The quest for balance in crop modeling. Agronomy J. 88, 695–697.

Porter, C.H., Braga, R., Jones, J.W., 1999. An Approach for Modular Crop Model Development. Agricultural and Biological Engineering Department Research Report No 99-0701, University of Florida, Gainesville, Florida. http://www.icasa.net/modular/pdf/modular.pdf.

Ritchie, J.T., 1998. Soil water balance and plant stress. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 41–54.

Ritchie, J.T., Singh, U., Godwin, D.C., Bowen, W.T., 1998. Cereal growth, development and yield. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production., Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 79–98.

Sinclair, T., Seligman, N.G., 1996. Crop modeling: from infancy to maturity. Agron. J. 88, 698–704.

Singh, P., Virmani, S.M., 1996. Modeling growth and yield of chickpea (*Cicer arietinum* L.). Field Crops Res. 46, 1–29.

Singh, P., Boote, K.J., Rao, A.Y., Iruthayaraj, M.R., Sheikh, A.M., Hundal, S.S., Narang, R.S., Singh, P., 1994a. Evaluation of the groundnut model PNUTGRO for crop response to water availability, sowing dates, and seasons. Field Crops Res. 39, 147–162.

Singh, P., Boote, K.J., Virmani, S.M., 1994b. Evaluation of the groundnut model PNUTGRO for crop response to plant population and row spacing. Field Crops Res. 39, 163–170.

Spitters, C.J.T., Schapendonk, A.H.C.M., 1990. Evaluation of breeding strategies for drought tolerance in potato by means of crop growth simulation. Plant Soil. 123, 193–203.

Stewart, D.W., Cober, E.R., Bernard, R.L., 2003. Modeling genetic effects on the photothermal response of soybean phonological development. Agron. J. 95, 65–70.

Takahashi, K., Yugi, K., Hashimoto, K., Yamada, Y., Pickett, C.J.F., Tomita, M., 2002. Computational challenges in cell simulation: a software engineering approach. IEEE Intell. Syst. 17, 64–71.

Vallejos, C.E., Chase, C.D., 1991. Linkage between isozyme markers and a locus affecting seed size in *Phaseolus vulgaris* L. Theor. Appl. Genet. 81, 413–419.

Venables, W.N., Ripley, B.D., 1997. Modern applied statistics with S-Plus,second ed. Springer-Verlag, New York.

Wallace, D.H., Yourstone, K.S., Masaya, P.N., Zobel, R.W., 1993. Photoperiod gene control over partitioning between reproductive and vegetative growth. Theor. Appl. Genet. 86, 6–16.

Weinig, C., Ungerer, M.C., Dorn, L.A., Kane, N.C., Toyonaga, Y., Halldorsdottir, S.S., Mackay, T.F.C., Purugganan, M.D., Schmitt, J., 2002. Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. Genetics 162, 1875–1884.

Weiss, A., 2003. Symposium Papers. Introduction Agron. J. 95, 1–3.

Welch, S.M., Roe, J.L., Dong, Z., 2003. A genetic neural network model of flowering time control in *Arabidopsis thaliana*. Agron J. 95, 71–81.

White, J.W., 1998. Modeling and crop improvement. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 179–188.

White, J.W., Hoogenboom, G., 1996. Simulating effects of genes for physiological traits in a process-oriented crop model. Agron. J. 88, 416–422.

White, J.W., Hoogenboom, G., 2003. Gene-based approaches to crop simulation: past experiences and future opportunities. Agron. J. 95, 52–64.

White, J.W., Hoogenboom, G., Jones, J.W., Boote, K.J., 1995. Evaluation of the dry bean model Beangro V1.01 for crop production research in a tropical environment. Exper. Agric. 31, 241–254.

White, J.W., Kornegay, J., Cajiao, C., 1996. Inheritance of temperature sensitivity of the photoperiod response in common bean (*Phaseolus vulgaris*). Euphytica 91, 5–8.

Wilkerson, G.G., Jones, J.W., Boote, K.J., Ingram, K.T., Mishoe, J.W., 1983. Modeling soybean growth for management. Trans. ASAE 26, 63–73.

Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T., Dubcovsky, J., 2003. Positional cloning of the wheat vernalization gene VRN1. PNAS 100, 6263–6268.

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J.L., Echenique, V., Dubcovsky, J., 2004. The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303, 1640–1644.