

A Genetic Neural Network Model of Flowering Time Control in *Arabidopsis thaliana*

Stephen M. Welch,* Judith L. Roe, and Zhanshan Dong

ABSTRACT

Crop simulation models incorporate many physiological processes within sophisticated mathematical frameworks. However, the control mechanisms for these processes tend to be ad hoc, empirical, and indirectly inferred from data and may lack realistic plasticity. Using model organisms like *Arabidopsis thaliana*, genomic scientists are rapidly disentangling the networks of genes that exert physiological control. As yet, however, these networks are qualitative in nature, depicting promotion and inhibition pathways but not supporting quantitative predictions of overall integrated effects. We believe (i) that neural networks can provide the quantification that current genetic networks lack and (ii) that taxonomic conservation of central genetic mechanisms will make networks developed for model plants also useful in crops. This paper presents evidence supporting the first point based on a neural network with eight nodes corresponding to *A. thaliana* genes controlling inflorescence timing. The nodes were linked into photoperiod and autonomous pathways abstracted from an existing qualitative genetic network model. Growth chamber data on transition timing were collected at 16 and 24°C for seven *A. thaliana* strains possessing loss-of-function mutations at the network loci. An eighth strain served as a common wild-type control. The neural network model reproduced the time course of the transition at both temperatures for all eight genotypes. Results included tracking a novel, temperature-dependent exchange in transition order exhibited by two mutants whose duplication is not possible by usual crop simulation methods. Furthermore, the ability to imitate the data appeared to have a desirable sensitivity to assumed network structure.

THE DEVELOPMENT of crop simulation models has been an intensely researched area for many years (Hanks and Ritchie, 1991). At the present time, simulation models with widely varying degrees of predictive skill exist for many major and minor crops (Tsuji et al., 1994). While these models differ according to the idiosyncrasies of the taxonomic groups involved, there are distinct commonalities among them as well. Model outputs are calculated by integrating daily or hourly changes in the rates of key physiological processes, including photosynthesis, C partitioning among plant organs, evapotranspiration, respiration, biomass accumulation and/or senescence, and phenological development. These rates are modeled as functions of both internal plant states and the external environment, with the latter including time-varying temperature, solar radiation, and soil water balance.

Varietal differences in these processes have been modeled by the introduction of *genetic coefficients*, so named because they quantify effects assumed to be of genetic origin. Recently, there has been a great deal of

research on developing efficient methods for estimating these parameters (Irmak et al., 2000), including attempts to relate them to specific plant genotypes (White and Hoogenboom, 1996). A valuable result of these studies has been a reduction in the number of quantities needing estimation and the work required to do so.

However, these efforts have had limited impact on the model algorithms used to control individual processes. Currently, modeled control mechanisms tend to be empirical and inferred indirectly, albeit with great ingenuity, from observations. Such empiricism may limit the accuracy of augmented models that attempt to combine physiological and physical models of the canopy and soil environments. It is not uncommon in model verification studies to find overestimation of low yields and underestimation of high yields (Heiniger et al., 1997; Roman-Paoli, 1997). While it is usually possible to find case-by-case adjustments to remove such effects, it is likely that current approaches to process-control modeling are simply less plastic than real plants. As a result, it has been suggested that improved representation of process controls should have modeling priority (Hammer, 1998).

Crop simulation models have much to gain from the exploding body of genomic science, which is unraveling the networks of interacting genes that actually control important plant processes. Unfortunately, experimental practicalities, such as long generation times, physical size, and other constraints, greatly complicate the elucidation of such networks in agronomic and horticultural crops. Also, the methods used to construct genetic networks, such as mutant screenings, epistasis detection experiments, and phenotype rescue treatments, yield qualitative relationships (e.g., promotion and inhibition) rather than the quantitative mathematical formulas needed in simulation models.

Our response to these issues is based on two premises. First, neural networks (Kasabov, 1996) can be used to supply the quantification that genetic networks currently lack (Mendoza and Alvarez-Buylla, 1998, 2000). Second, networks developed using model plants will prove directly adaptable to crops because genetic mechanisms of central importance are often widely conserved taxonomically.

This paper presents preliminary results supporting the first premise based on efforts to model the control of inflorescence transition in *A. thaliana*. The first section reviews current information on the genetics of this process. The next introduces neural networks, including one specifically structured to approximate inflorescence control. Succeeding sections detail a data collection experiment currently underway and successful efforts to

S.M. Welch and Z. Dong, Dep. of Agron., and J. Roe, Div. of Biol., Kansas State Univ., Manhattan, KS 66506. Contrib. no. 01-290-J from the Kansas Agric. Exp. Stn. (KAES), Kansas State Univ. This project was supported in part by KAES Project 0507 at Kansas State Univ. Received 1 May 2001. *Corresponding author (welchsm@ksu.edu).

Abbreviations: ES, Excel Solver (software); G × E, genotype × environment (interaction); LS, least squares; MQL, Marquardt-Levenberg (algorithm); RMSE, root mean square error; SCE, simulated complex evolution; SSE, sum of squared errors.

mimic data from two temperature treatments. Included with the latter are some additional results of possible relevance to bioinformatics as well as to crop simulation. The final section includes a general discussion and conclusions to date.

FLOWERING CONTROL IN *ARABIDOPSIS THALIANA*

Accurate mimicry of plant phenology (especially flowering time) is critical in crop modeling to establish the temporal limits within which growth and yield-generating processes operate. Understanding what controls transition to flowering is a fundamental concern in developmental biology as well. There is an enormous literature on the control of flowering, but a unified physiology model has not emerged (McDaniel, 1996). The genetic control of flowering has been extensively studied in *A. thaliana* due to its small genome, short generation time, self-compatibility, amenability to stable transformation, and the availability of numerous mutants. *Arabidopsis thaliana* is a facultative long-day plant, so the stimulus of a long day coupled with input from the circadian clock will promote flowering. Under short days, plants will flower but do so much later. Flowering in this species entails two transitions, one to form an inflorescence (or bolt) and the second to produce flowers. The two transition processes are distinct and can be distinguished genetically (Howell, 1998). The inflorescence transition is influenced by environmental signals, such as daylength, cold temperature (vernalization), and growth temperature. The transition from inflorescence to floral development does not appear to require additional environmental inputs. The former transition is the focus of the work reported here.

The interplay between environmental and genetic factors has been assembled into qualitative network models for the control of flowering in *A. thaliana* (Koorneef et al., 1998a, 1998b; Levy and Dean, 1998; Simpson et al., 1999; Theissen and Saedler, 1999; Blazquez, 2000; Blazquez and Weigel, 2000; Devlin and Kay, 2000). It is quite important to understand that the genes in these models control phenology directly. That is, observed temporal changes are not the byproducts of alterations in plant growth or vigor (Reeves and Coupland, 2000). The work reported here is based on the Simpson et al. (1999) model because the Blazquez (2000) version is too recent to have affected our research planning. Both of these models can be viewed on the web at the addresses given in the reference list.

According to Simpson et al. (1999), there are four pathways that operate in parallel to control the inflorescence transition. The two major ones are the autonomous pathway and the photoperiod pathway. The prevailing view is that the latter exerts the promotive effect of long days, whereas the former is an endogenous photoperiod-independent, flower-inhibiting pathway. Autonomous path repression of flowering during vegetative growth must be overcome to convert the meristem to a reproductive state. Repression can be accomplished by vernalization, which injects repressive signals into

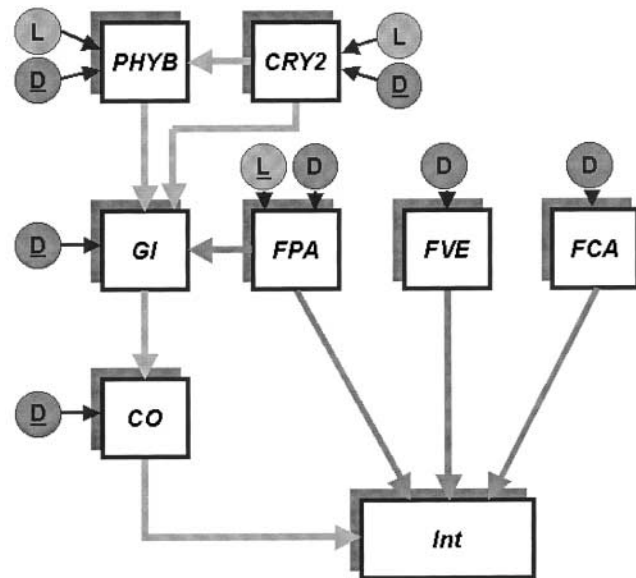


Fig. 1. Initial genetic neural network. The nodes correspond to genes controlling flowering time. Darker arrows denote the autonomous pathway, and the lighter tone indicates the photoperiod path. The inputs are daylength (L) and days after planting (D). Inputs with underlines were later dropped (see text).

the autonomous path at the FLC (Flowering Locus C) gene.¹ Simpson et al. (1999) accords vernalization the status of a separate pathway although Blazquez (2000) considers it a part of the autonomous path. Additionally, gibberellin is required for flowering in short days and also participates in flowering promotion under long days. The genes mediating gibberellin's influence constitute the fourth independent path. In all models, signal integration of the two main paths feeds into a transduction cascade ultimately upregulating flowering genes and stimulating conversion to the reproductive phase.

A GENETIC NEURAL NETWORK

Neural networks (Kasabov, 1996) were initially developed to model brain function with the intent of enabling computers to mimic various cognitive feats common in the animal kingdom, such as learning or pattern recognition. They consist of interlinked nodes that mathematically abstract neuron function. Each node typically has a number of inputs and one output that serves as an input to multiple succeeding nodes. Nodal outputs are usually calculated as some sigmoidal function applied to an arithmetic combination of the inputs. Thus, like the genes in existing qualitative network models, these abstract neurons are capable of both ON and OFF and graded activity in response to the net balance of stimulatory and inhibitory inputs (Burstein, 1995; Mjolsness et al., 1991). All nodes in a neural network use the same sigmoidal formula, which is called the nodal *transfer function*. Finally, the links between nodes typically modify the values they transmit. Almost always, this modifi-

¹ In this paper, GENE denotes a particular locus, *GENE* is the corresponding neural network node, and *gene* is a plant with loss-of-function mutant alleles at that locus.

Table 1. Genotype details of experimental materials.

Stock no.	Allele	Symbol, position	Phenotype, gene product, and citations
CS172	<i>fca-6</i>	FCA, 4-32.0	<ul style="list-style-type: none"> • Late flowering; vernalization affects flowering time under long and short days; strong short-day influence on flowering time; flowers ca. 19 d after wild type • RNA binding protein • Macknight et al. (1997)
CS166	<i>fpa-2</i>	FPA, 2-67.0	<ul style="list-style-type: none"> • Late flowering; vernalization affects flowering time under long and short days; flowers ca. 13 d after wild type • Unknown • Koornneef et al. (1991)
CS174	<i>fve-2</i>	FVE, 2-32.0	<ul style="list-style-type: none"> • Late flowering; vernalization affects flowering time under long and short days; strong short-day influence on flowering time; flowers ca. 26 d after wild type • Unknown • Koornneef et al. (1991)
CS179	<i>co-6</i>	CO, 5-13.0	<ul style="list-style-type: none"> • Late flowering (long days only); no effect of vernalization; incompletely dominant; flowers ca. 21 d after wild type. • Zn-finger transcription factor • Putterill et al. (1995)
CS108	<i>fha-1</i>	CRY2, 1-12.0	<ul style="list-style-type: none"> • Late flowering; unaffected by short days or vernalization; increased number of rosette leaves. • Cryptochrome 2, a blue-light photoreceptor • Ahmad and Cashmore (1993); Guo et al. (1998)
CS183	<i>gi-6</i>	GI, 1-33.0	<ul style="list-style-type: none"> • Late flowering (long days only); no effect of vernalization; flowers ca. 22 d after wild-type. • Transmembrane protein • Park et al. (1999); Fowler et al. (1999)
CS6211	<i>phyB-1</i>	PHYB, 2-35.0	<ul style="list-style-type: none"> • Early flowering under short and long days; long hypocotyl in red and white but not in far-red light; visibly reduced chlorophyll level; abnormally long petioles, stems, and root hairs; smaller leaves and fewer in rosette; greater apical dominance (longer main inflorescence, fewer lateral branches, stem termini, and main inflorescence siliques). • Phytochrome B, a red-light photoreceptor • Reed et al. (1993)
CS20 (Ler-0)	<i>er-1</i>	ER, 2-48.0	<ul style="list-style-type: none"> • Common background for all mutants; blunt fruits, short petioles, usually upright stems, compact inflorescence; Ler-0, ecotype derived from Landsberg (La-0), but genetically distinct; carries erecta gene; blunt siliques, upright stems, bunched flowers; suitable wild type, provides control for many mutants.

cation is to multiply the value by some constant weight. Thus, the function of a neural network is completely determined by four items: the structure of its nodes and links, the method of combining nodal inputs for substitution into the transfer function, the transfer function itself, and the weights applied to each link. The network designer determines the first three, and the weights are found from experimental data via a *training* process.

If functioning genes can, in fact, be identified with nodes, then removing a node should equate to a loss-of-function mutation. To test this concept, a set of genes and mutant alleles was chosen according to several criteria. First, the set contained representatives from both the photoperiod and autonomous pathways. Second, all genes possessed readily available mutant alleles with strong phenotypic effects that achieved maximal loss of function. Third, the alleles were obtainable in a common genetic background, providing proper experimental control. Finally, the selection balanced the number of genes, the number of plants per genotype, and the available growth chamber space.

Figure 1 shows the selected genes and their connection into a neural network abstracted from the Simpson et al. (1999) model. Both the photoperiod and autonomous pathways are present. These feed into a single integration node (*Int*) whose 0 to 1 output is interpreted as the fraction of the transition completed. Inputs to the system are photoperiod and days after planting. The latter input was used as a surrogate for growth, which is thought to be of relevance to the autonomous pathway

(Blazquez, 2000). Initial concerns existed that daylength might somehow influence the autonomous pathway or, conversely, that growth-related information might leak onto the photoperiod path by some yet undiscovered genetic mechanism. For this reason, extra inputs were added to each pathway (underlined symbols in Fig. 1) with little justification beyond the authors' desire not to be victimized by limited thinking. Later, cause arose to remove them (see below).

Table 1 shows the specific mutant alleles and background used. All seed was obtained from the *Arabidopsis* Biological Resource Center (Columbus, OH). Seed stock numbers are listed in the first column of the table. The next two columns list the alleles, gene symbols, and map positions (chromosome number and location² in centimorgans). The gene symbol can be used to locate the position of the gene in our simplified network. The final column qualitatively describes the phenotype and gives a molecular characterization of the gene product, if known, and relevant literature citations. It should be understood that phenotype descriptions pertain to nominal rearing conditions and may not reflect the results herein.

All of the chosen mutant alleles were available in the Landsberg erecta (Ler) background. In addition to the traits listed in the table, Ler has reduced functionality at *FLC*. Therefore, although Fig. 1 does not include an *FLC* node, it is assumed that the effects of upstream genes propagate as shown to the pathway integration node.

² Look for more detail genetic information at <http://www.arabidopsis.org/chromosomes/> (verified 16 Sept. 2002).

DATA COLLECTION METHODS

Once the structure of a neural network is set, further quantification requires data. To obtain such, plants were reared in a growth chamber at two constant temperatures, 16 and 24°C. The temperatures were selected to give quite different development rates while avoiding stressful extremes. To ensure that genes on both pathways were active, data were collected under long-day conditions (16 h light and 8 h dark). For each temperature run, the eight genotypes were individually sown on the soil surface in square pots (8.9 by 7.2 cm, catalog no. 12-1350, Hummert Int., Earth City, MO) with nine replications for a total of 72 pots. The potting soil was Fafard Mix no. 2 from Hummert International. Within each pot, single seeds were placed approximately 1 cm inside every corner. After germination, plants were thinned to two per pot, almost always diagonal to each other to minimize interplant effects. Lacking a fully functional FLC gene, Ler does not require vernalization. However, to help synchronize germination, the pots were placed in darkness at 4°C for 72 h.

The growth chamber used in the study was a Conviron Model E8VH (Controlled Environ., Winnipeg, MB, Canada). Lighting was provided by 20 fluorescent tubes (F48T12/CW/VHO, 110 W, Philips Lighting Co., Somerset, NJ) and two incandescent bulbs (Philips Standard Frost Incandescent Lamps, 60 W). Light intensity at plant level was 216.1 $\mu\text{mol m}^{-2} \text{s}^{-1}$ as measured with a Quantum sensor (LI-COR, Lincoln, NE). Chamber relative humidity was maintained between 70 and 80%.

The 72 pots were randomly assigned to one of four trays and placed in the chamber. To limit systematic biases resulting from any residual within-chamber environmental gradients, trays were moved daily in an orderly fashion so that pots spent nearly equal times in every chamber location. In addition, thermistors monitored the canopy-level temperature within each tray at hourly intervals, and data were recorded by a CR-21 data logger from Campbell Scientific (Logan, UT). To provide the seedlings with an initial high humidity, transparent domes (Catalog no. 14-2568-1, Hummert Int., Earth City, MO) were placed over the trays from planting until 6 to 7 d after emergence. Although standard practice in work with *A. thaliana*, the hoods created a transient average temperature increase of 0.38 to 0.74°C. This increase is included in all temperature accumulations, which yielded treatment-long average temperatures of 24.16 and 16.37°C. Plants were watered at about 3-d intervals, which is adequate to avoid stress.

The pots were observed daily and records made of emergence date, bolting date, flowering date, leaf number (including rosette and inflorescence leaves), and plant height. Emergence date was taken as the first visible green, but this is not particularly reliable because of the small seed size. Bolting date, when flower buds are first visible on the main meristem, marks the transition to inflorescence, which is the focus of the current report. Flowering date is the opening of the first bud and denotes the second transition discussed above.

GENETIC NEURAL NETWORK WEIGHTS

Neural network weights encode the system's storehouse of knowledge. They are found through a training process during which weight estimates are refined until the network can produce correct, known outputs when supplied the corresponding inputs. Larger networks with more weights are presumed necessary to retain knowledge of a more complicated nature. If initial training efforts are not successful, the typical assumption is

that the original network had insufficient storage, and more nodes are added. Although this strategy is often successful, in reality, another principle may also be at work.

Training a neural network is an example of a nonlinear optimization problem where the objective function is a goodness-of-fit criterion. A common difficulty in such work is that initial estimates of the solution may be so distant from the best answer that the optimizer may converge prematurely to some inferior result. Because of network complexity and perhaps because of the opacity of brain function, designers neither demand nor expect any clear relationship between the values of the weights and the knowledge represented. Good initial estimates are therefore considered unobtainable, and zeros are commonly used. As a result, subsequent success with a larger network may have nothing to do with storage capacity. It may simply be that the goodness-of-fit landscape has been altered so as to provide the optimizer with a more congenial path from arbitrary initial estimates to an acceptable endpoint.

In the current study, the strategy of adding nodes disappeared once the choice was made to work with a particular gene subnet. Therefore, when the first training attempts failed, several tactics were employed to search more efficiently and to modify the goodness-of-fit landscape directly. The modifications included (i) alternative goodness-of-fit functions and search algorithms, (ii) reducing problem dimensionality (i.e., the number of weights) whenever possible, (iii) increasing dimensionality in a controlled fashion whenever unavoidable, (iv) reparameterizing, and (v) expanding model scope incrementally. Candidly, exploitation of these strategies was informed by some science, some intuition, and a great deal of error.

To date, two different goodness-of-fit measures have been used in training: ordinary least squares (LS) and a maximum likelihood (ML) method that parallels the Kaplan–Meier product-limit estimator well known in survival analysis (Cox and Oakes, 1984). Three search procedures were used: simulated complex evolution (SCE) (Duan et al., 1992, 1994; Thyer et al., 1999), the Marquardt–Levenberg (MQL) algorithm (Press et al., 1992), and the variant of Newton's method (Press et al., 1992) embodied in the Excel Solver software (ES). The SCE procedure is readily adaptable to a parallel computing environment and therefore potentially useful for large-scale problems. The MQL algorithm is the current method of choice for nonlinear LS (Press et al., 1992). In terms of execution speed, robustness, flexibility, and numerical charisma, ES has limited long-term potential compared with MQL and SCE. However, the ease of Excel programming has permitted rapid testing of new ideas, the determinative success factor to date. This report therefore contains results obtained by combining LS and ES using an adaptation of the methods of Wraith and Or (1998).

MODEL SPECIFICS AND RESULTS

In what follows, it will be useful to have a concise notation for a neural network model (Fig. 2). The formula

<a: *GENE* := expression

denotes that the expression on the right is computed yielding, e.g., u , which is substituted into the transfer function for the *GENE* node. The common sigmoidal transfer function is used for all genes. The function result is multiplied by the quantity \mathbf{a} , which may be either a scalar or a vector as needs warrant. The operator (a multiplication by 0 for loss-of-function mutants and by 1 otherwise) calculates the penultimate nodal output. If L represents photoperiod and D is days after planting, the starting network in Fig. 1 can be written as

$$<1: CRY2 := a_1L + a_2D + b_{CRY2}$$

$$<1: PHYB := a_7CRY2 + a_8L + a_9D + b_{PHYB}$$

$$<1: FPA := a_3L + a_4D + b_{FPA}$$

$$<1: GI := a_{10}D + a_{11}CRY2 + a_{12}PHYB + a_{19}FPA + b_{GI}$$

$$<1: FVE := a_5D + b_{FVE}$$

$$<1: CO := a_{13}GI + a_{14}D + b_{CO}$$

$$<1: FCA := a_6D + b_{FCA}$$

$$\ni 1: Int := a_{15}CO + a_{16}FPA + a_{17}FVE + a_{18}FCA + b_{Int}$$

[1]

where *Int* represents the integration pathway, \ni indicates the final network output, and the a 's and b 's are weights to be found. The b values control the time until the onset of nodal switching while the a 's affect the time it takes to complete a transition, once started. Our initial training efforts focused on this model and failed utterly.

Once it became clear that the training process was going to be nontrivial, an incremental modeling approach was adopted. First, training only the autonomous pathway was attempted, just using the 24°C data alone. Once this was achieved, the lessons learned were applied to an enlarged, two-path model, still for one temperature. The final step was a two-path model including both temperatures. The next sections recapitulate the latter two steps.

The 24°C Model

To reduce dimensionality, all extraneous D and L inputs were removed, eliminating $a_2, a_3, a_9, a_{10},$ and a_{14} . Removal is easily justified because it actually makes the neural network more similar to the underlying Simpson et al. (1999) model. Also, given data on only one photoperiod, there is no way to unambiguously determine the parameters a_1 and b_{CRY2} (for *CRY2*) or a_8 and b_{PHYB} (for *PHYB*). Therefore, the *CRY2* output was fixed at 1 for long days and 0 for short days (although no use has yet been made of the latter). The a_8 parameter was absorbed into b_{PHYB} and renamed b'_{PHYB} . According to the data, once the transition starts, all mutant populations can progress from 10 to 90% completion in roughly similar intervals (3 to 6 d). Thus, as an approximation, $a_4, a_5, a_6, a_7, a_{11}, a_{12}, a_{13},$ and a_{19} were replaced by a single parameter, a . However, this simplification did not mimic *co* data very well, so a_{13} was released to vary indepen-

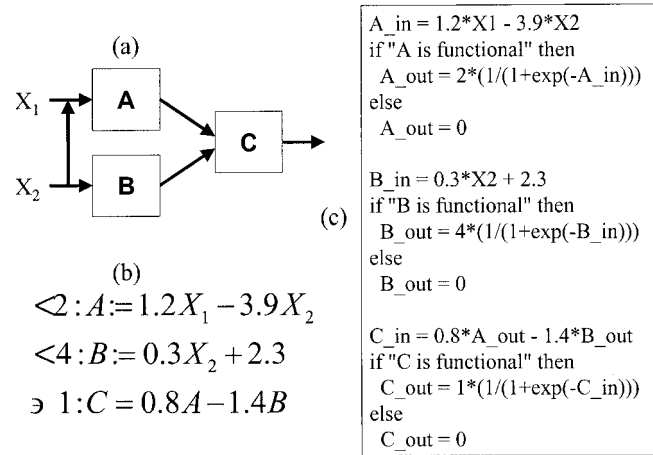


Fig. 2. Neural network notation example: (a) a simple neural network structure, (b) its corresponding notation with example weights, and (c) equivalent computer code. The network inputs are the two variables X_1 and X_2 . The final output is from the *C* node as indicated by the \ni symbol (*C_out* in the computer code).

dently. A second parameter, a' , was used for $a_{15}, a_{16}, a_{17},$ and a_{18} in the *Int* node.

Another change was quite ad hoc but helpful. It seemed reasonable that if no information was propagating along a path, then the input to *Int* should be zero. However, with the given transfer function, a zero input to an upstream node propagates a signal of 0.5 to the next downstream node. This effect was removed by making b_{Int} proportional to a' and to -0.5 times the number of upstream nodes. In addition to addressing the issue just mentioned, this change further reduced dimensionality by one parameter. It has another advantage as well. Mathematically, b_{Int} would be expected to influence transition onset time as previously mentioned. However, the *Int* node is a synthetic model construct that replaces several actual genes with roles in pathway integration. By fixing b_{Int} , the model is forced to more closely relate timing to the b values associated with real genes that actually affect temporal shifts.

Together, these modifications yielded the nine-parameter model

$$<2: CRY2 := 0$$

$$<1: PHYB := aCRY2 + b'_{PHYB}$$

$$<1: FPA := aD + b_{FPA}$$

$$<1: GI := a(CRY2 + PHYB + FPA) + b_{GI}$$

$$<1: FVE := aD + b_{FVE}$$

$$<1: CO := a_{13}GI + b_{CO}$$

$$<1: FCA := aD + b_{FCA}$$

$$\ni 1: Int := a'(CO + FPA + FVE + FCA - 2) \quad [2]$$

This model was fit using ES. Figure 3 plots predicted vs. observed values for those observations falling in the actual transition interval. Except for *fve*, predictions are close to the actual data. The overall R^2 of all predicted and observed values used in training was 0.9693.

An important question is whether genomic structure actually played a part in these results or whether they are

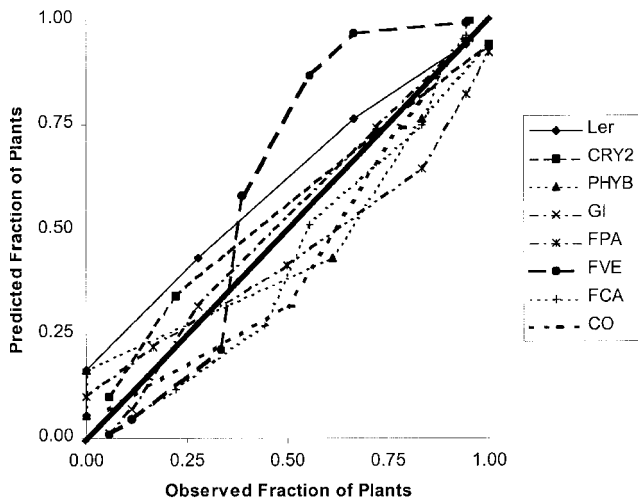


Fig. 3. Predicted vs. observed completion of the inflorescence transition. To focus on the accuracy of the transition, only those data are plotted for which either the predicted or actual values are between 0.05 and 0.95.

only due to the well-known ability of neural networks to fit complex data. To assess this question, 300 random networks were synthesized and fit with ES. To toughen the evaluation, each of the networks had the same number of nodes, number of inputs per node, and redundancy of coefficients as Eq. [2]. The data, search method, initial conditions, and convergence criterion were the same for all 300 runs. Figure 4 shows the cumulative distribution of final sums of squared errors (SSE).³ As the graph indicates, SSE values ranged from 6.50 to 31.13. By comparison, the SSE of the genomically constructed network is, at 0.59, slightly more than 11 times better. The position of the genomic network on the extreme tail of the SSE distribution strongly suggests the value of genetic information.

³Because the model output is the fraction of plants completing the transition, all reported SSE and root mean square error (RMSE) values are unitless.

A common situation in genomics is to know that a group of genes may form a path but not be able to tell the order in which they fall. For example, one might be uncertain as to whether a portion of the photoperiod pathway runs PHYB→GI→CO, as in the Simpson et al. (1999) model, or the reverse. Typically, a set of single and double mutant experiments will be performed to resolve such issues. Although no such experiments were performed here, there was a clear distinction between these two alternatives—the reversed order yielded an SSE of 6.68.

The Two-Temperature Model

It has often been suggested that temperature effects on development may be related to underlying, enzyme-catalyzed, biochemical reaction rates (Pradhan, 1946; Sharpe and DeMichelle, 1977; Wagner et al., 1984; Bridges et al., 1989). Furthermore, the activity curves of many regulatory enzymes are allosteric (i.e., S-shaped) with increasing substrate concentration (Segel, 1975, p. 353 onward) just as neural node outputs are sigmoidal with increasing weighted input sums. In combination, these ideas suggested a simple analogy in which the amplitude of nodal outputs could be multiplied by temperature-dependent factors (henceforth denoted q_i).

Applied completely, this idea would add 16 parameters to the model (eight nodes times two temperatures). To limit this increase in dimensionality while maximizing consistency with Eq. [2], the temperature factors for 24°C were assigned to be unity. Under this scheme, the 16°C factors are interpreted as relative changes. Furthermore, because the *Int* output must range from 0 to 1, no temperature multiplier was used there. It was also found desirable to relax the assumption of common a and a' values. The weights, a_i , were estimated by parameters of the form av_i or $a'v_i$ where a and a' are as above and v_i ($i = 11, 12, 15, 16, 17, 18, \text{ and } 19$) values are new with initial estimates of 1.0. This parameterization allowed the search engine to find the approximate solu-

Fits of Random Networks

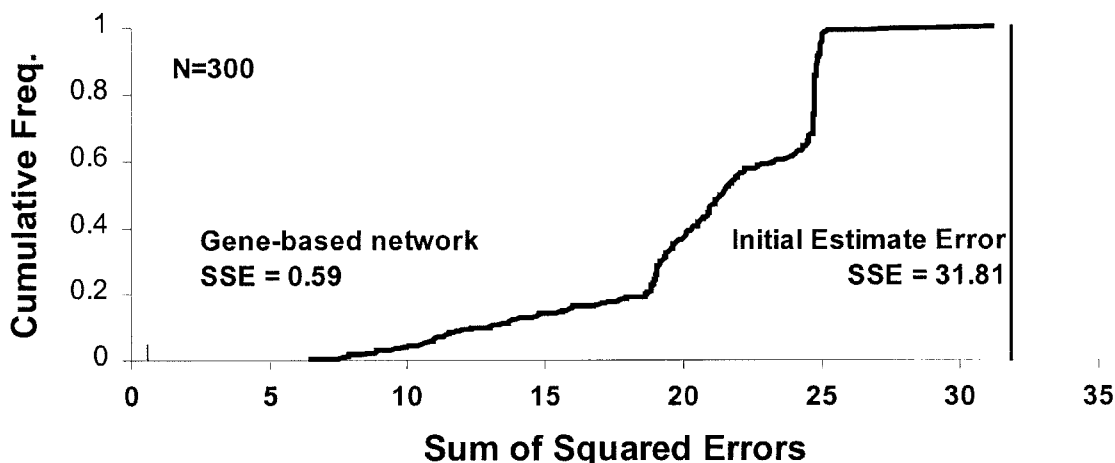


Fig. 4. Cumulative distribution of sums of squared errors (SSEs) for 300 random networks. Training for all networks began with parameter values generating an SSE of 31.81. Final SSEs ranged from 6.50 to 31.13. The genetically based network had a final SSE of 0.59.

tion via a , which was then particularized by the offset v_i parameters. The weights a_4 , a_5 , a_6 , and a_7 were varied freely. Again, it should be emphasized that these reparameterizations were tactical and intended to facilitate training. No scientific significance is attached to their specific form.

In what follows, $[r,s]$ is a vector of temperature multipliers for 24 and 16°C, respectively. Taken together, these changes yield the combined temperature model

$$\begin{aligned} <2[1,q_{CRY2}] : CRY2 : = 0 \\ <1[1,q_{PHYB}] : PHYB : = a_7CRY2 + b'_{PHYB} \\ <1[1,q_{FPA}] : FPA : = a_4D + b_{FPA} \\ <1[1,q_{GI}] : GI : = a(v_{11}CRY2 + v_{12}PHYB + v_{19}FPA) + b_{GI} \\ <1[1,q_{FVE}] : FVE : = a_5D + b_{FVE} \\ <1[1,q_{CO}] : CO : = a_{13}GI + b_{CO} \\ <1[1,q_{FCA}] : FCA : = a_6D + b_{FCA} \\ \exists[1,1] : Int : = a'(v_{15}CO + v_{16}FPA + v_{17}FVE + v_{18}FCA - 2) \end{aligned} \quad [3]$$

Column I of Table 2 gives the resulting parameter values and the root mean square error (RMSE). There is an immediately noteworthy pattern among the temperature factors (q_i 's). Four of the seven nodes (all of the autonomous pathway plus the closely linked GI) have temperature factors between 0 and 1, plausibly suggesting a general plant slowdown with cooling. Perhaps not surprisingly, the remaining photoperiod sensory pathway nodes ($CRY2$, $PHYB$, and CO) depart from these values. A final analysis of the sensory path should await the addition of short-day data to the training set. Even so, there are some comments about the light receptors that can be made now.

Interestingly, the $CRY2$ factor is quite close to 1. This gene is known to interact quite closely with the diurnal-clock genes (Simpson et al., 1999; Blazquez, 2000), so it would not be unreasonable to hypothesize some form of temperature stabilization in this part of the network. With this possibility in mind, the network was retrained under the assumption that q_{CRY2} was identically 1. The very similar results are shown in Column II of Table 2.

The blue-light receptor encoded by $CRY2$ is known to mediate the red light-dependent flowering-control actions of phytochrome B (Guo et al., 1998). The network was able to reproduce this to the extent that, neglecting the temperature constant, the $PHYB$ node exactly tracked the $CRY2$ output. However, in the absence of specific training data under red and blue lighting, the network was unable to resolve the precise antagonistic relationship that exists between these receptors in reality. Under the white light in this experiment, the $phyB$ inflorescence transition preceded that of the Ler wild type by 1 to 2 d regardless of temperature. The temperature-stabilized $CRY2$ was apparently a much more reliable indicator of training set behavior than was $PHYB$. As a result, v_{12} , the weight applied to $PHYB$ outputs, diminished by almost four orders of magnitude (Table 2, Column II). For this reason, training was re-

Table 2. Parameter values and goodness of fit for three variations of the two-temperature model.

Node	Symbol	Values by version		
		I	II	III
<i>CRY2</i>	q_{CRY2}	1.116	1.000	1.000
<i>PHYB</i>	q_{PHYB}	-34.993	-50.367	-
	a_7	18.556	28.312	-
	b'_{PHYB}	-27.682	-18.222	-
<i>FPA</i>	q_{FPA}	0.036	0.033	0.121
	a_4	0.233	0.241	0.132
	b'_{FPA}	-8.833	-8.724	-5.390
<i>GI</i>	q_{GI}	0.128	0.125	0.130
	a	1.625	4.244	1.408
	v_{11}	1.492	0.640	1.099
	v_{12}	10.986	.003	-
	v_{19}	142.686	46.855	44.580
	b_{GI}	-2.311	-2.285	-2.235
<i>FVE</i>	q_{FVE}	0.415	0.410	0.382
	a_5	0.070	0.077	0.093
	b_{FVE}	-2.093	-2.192	-2.627
<i>CO</i>	q_{CO}	54.288	88.176	86.032
	a_{13}	5.209	5.265	5.584
	b_{CO}	-4.778	-5.708	-5.495
<i>FCA</i>	q_{FCA}	0.331	0.321	0.288
	a_6	0.060	0.061	0.066
	b_{FCA}	-3.726	-3.811	-4.046
<i>Int</i>	a'	23.777	23.426	21.280
	v_{15}	1.470	2.296	1.903
	v_{16}	-0.863	-0.758	-2.040
	v_{17}	2.272	2.621	2.860
	v_{18}	8.860	9.850	11.720
RMSE†		7.34×10^{-2}	7.33×10^{-2}	6.57×10^{-2}

†RMSE, root mean square error.

peated (Table 2, Column III) after deleting $PHYB$ from the network and excluding the corresponding mutant data. Not only is the aggregate RMSE value better, but the bulk of the improvement is associated with the wild type, whose prediction errors were reduced by 33% (from RMSE = 0.152 to 0.102).

Predicted transition curves and observed data for both temperatures using the values from Table 2, Column III, are plotted in Fig. 5. The ability of the model to track multiple genotypes accurately as temperature changes is evident. However, the figure also reveals an interesting feature, namely that the order of inflorescence transition for fve and co differs between the two temperatures (arrows). Moreover, the scale of the exchange is large. Comparing median transition times (estimated by linear interpolation), co switches from 7.7 d ahead of fve at 24°C to 9.3 d behind at 16°C for a *switching magnitude* (i.e., net change) of 17 d. To the authors' knowledge, this observation regarding co and fve is novel.

Crop simulation models traditionally use either thermal time (Hesketh et al., 1973) or photothermal day (Grimm et al., 1993) systems to calculate flowering times. These approaches cannot reproduce the phenological order switch seen here unless base temperatures differ between varieties, an uncommon assumption among crop models.⁴ Thus, the observed order switch was unexpected, its magnitude startling, and its associa-

⁴A reviewer noted the Sirius model (Jamieson et al., 1998) can switch in some situations because it alters thermal time to flowering by making main-stem final leaf number responsive to daylength and vernalization. However, in the experiment reported here, both photoperiod and the short, initial chilling, were common across treatments.

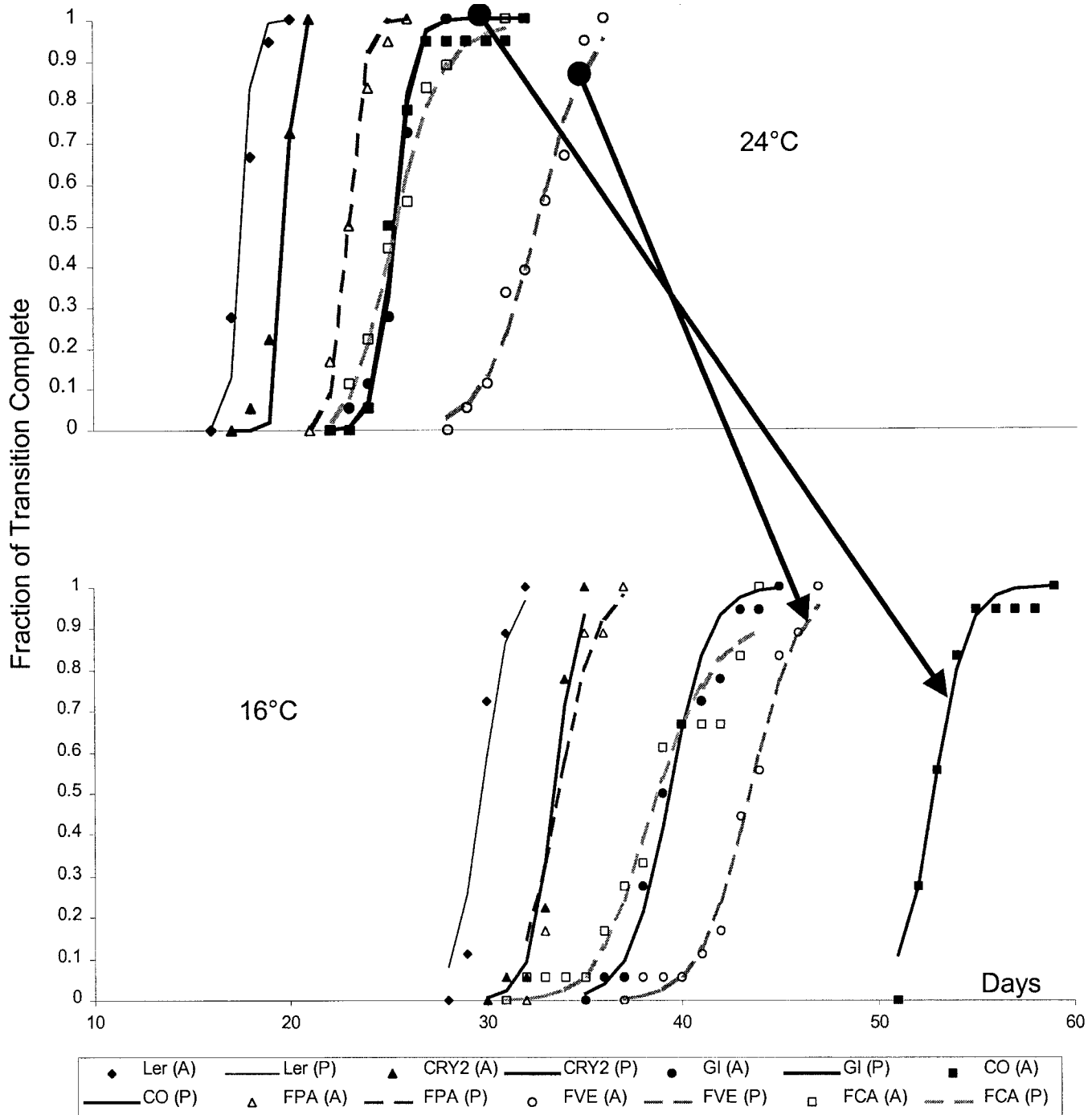


Fig. 5. Predictions from the final two-temperature model. To simplify the plot, each series is truncated 1 d before (after) the first (last) observed plant transition. Symbols are actual data, and lines are model predictions (denoted A and P in the legend, respectively). The black arrows indicate the switch in the transition order of FVE and CO with cooling (see text).

tion with specific point mutations intriguing. It raised the question as to the occurrence of order switching in actual crops. A side study, detailed in the Appendix, provided statistical evidence that order switching does occur in soybean [*Glycine max* (L.) Merr.]. Beyond statistics, Hoogenboom and White (2003) have reported a specific gene in common bean (*Phaseolus vulgaris* L.) that confers photoperiod sensitivity in warmer environments but not in cooler ones. Under appropriate circumstances, this gene would also produce order switching.

The researchers were able to incorporate its effect into a crop simulation model by appropriate, gene-specific modifications. Genetic neural networks, however, can potentially subsume a wide range of such unique behaviors within a single, general mechanism.

DISCUSSION AND CONCLUSIONS

Although preliminary, the above results indicate that neural networking has potential for converting qualita-

tive genetic networks into quantitative predictive tools, the first of the two long-term research objectives presented at the beginning of this paper. Not only can neural networks reproduce complex genetic data, but they can also do so parsimoniously. The authors are aware of only one other attempt to predict flowering time with a neural network, that of Elizondo et al. (1992), who studied soybean. They were successful, but lacking genomic information, they used a traditional three-layer network that required a minimum total of 25 nodes and 104 weights for acceptable prediction. In contrast, our final model has seven nodes and 22 weights.

A natural desideratum for applied models is that they be no more complex than necessary to achieve specific goals. In this regard, estimates such as approximately 26 000 genes for *A. thaliana* (Riechmann et al., 2000) can be quite intimidating. However, fears of the impracticality of genomic-level modeling are probably groundless because small numbers of genes often control broad developmental and physiological processes. The genes used in this study top a floral development hierarchy (Theissen and Saedler, 1999) that ultimately generates all reproductive organs. More generally, mathematical analysis suggests that the accrual of disproportionate influence to tiny gene minorities may be an inevitable consequence of network self-organization over evolutionary time (Barabasi and Albert, 1999). If so, significant crop simulation improvements may only require the modeling of a relatively few genes.

Furthermore, applied models, including genetic neural networks, will continue to benefit from sensitivity analysis and other traditional simplification techniques. The ability to delete *PHYB* under white-light conditions is one illustration. In addition, it may not be necessary to model the complete dynamics of all genes. White and Hoogenboom (1996) have had success melding genomics with existing simulation models using simple regression equations to predict traditional genetic coefficients based on the presence or absence of selected dominant alleles.

Beyond applied issues, one exciting aspect of this work is that it immediately suggests further researchable questions. Like all good models, neural networks seem capable of providing guidance in data interpretation. An apparent sensitivity to assumed structure was demonstrated in the reversed photoperiod pathway and randomized network examples above. It may also have played a part in the inability to fit the original model with its extraneous inputs of daylength and plant age. If such sensitivity is widespread, automated tools may, with appropriate human guidance, be able to screen large numbers of network structures for compatibility with existing data.

To be realized, such tools will require powerful numerical estimation methods. One of the most useful features of neural networks is the extensive body of research results on training techniques (Fine, 1999; Grieron and Hajela, 1996). Still, it remains to be seen what will prove to be the most efficacious ways to apply these methods to genomic data. It is obvious, however, that training will need to utilize data from multiple experi-

ments conducted in different laboratories. This need is illustrated by the *PHYB* results, which from the perspective of absolute realism, would have been improved by data from red- and blue-light environments. Clearly, no single research group can perform all the treatments necessary to obtain complete data.

Complicating factors are (i) the propensity of various *A. thaliana* investigators to use different sets of alleles in separate genetic backgrounds and (ii) the possible impacts of subtle interlab variations in rearing environments. While these issues will no doubt cause initial problems, quantitative models using neural networks or any other successful approach can provide common reference points against which effect magnitudes can be judged. This may be an area in which the experience of crop modelers can be of direct benefit to genomic scientists.

Because genetic network construction involves a consideration of phenotypes, there are multiple possible mechanisms underlying the resulting links. The relationships may be quite immediate, as when one gene codes for a transcription factor directly controlling the expression of another. *CO*, whose sequence includes the DNA-binding, Zn-finger motif (Lewin, 1997, p. 850), is a probable example. Alternatively, the effects may be indirect and mediated by significant physiological processes. Even so, the construction method still guarantees that the genes responsible will appear at their proper place in the network diagram. The cryptochrome and phytochrome receptors illustrate this situation. In spite of this diversity, the neural network approach appears able to span these differences in mechanism, at least to some degree.

Current crop simulation models, in contrast, emphasize physiological mechanisms to the complete exclusion of genomics. As noted earlier, this has led to a very empirical representation of process-control mechanisms that may, in the final analysis, lack plasticity. Even so, the extensive utilization of crop simulation models in areas as diverse as policy analysis (Rosenzweig et al., 1996; Tubiello et al., 1999); research (Hanks and Ritchie, 1991); and, increasingly, production management (Welch et al., 2002) shows that the physiological approach is not without its strengths.

However, if taken to extremes, genetic neural networks could run the risk of committing the reverse error, that of subsuming too much under the heading of genomics. Future plant models will need to craft a careful balance between traditional methods and new genomic approaches like, or derived from, the one presented here. Thus, the most productive way forward will undoubtedly entail a synthesis of crop modeling and genomic methods rather than the mere adoption of a new approach by one existing research community. Among all the exciting aspects of this work, this is the most exciting of all.

APPENDIX

Ten years (1987–1996) of anthesis-date (R1) data from soybean performance trials at Tifton, GA (obtained from G.

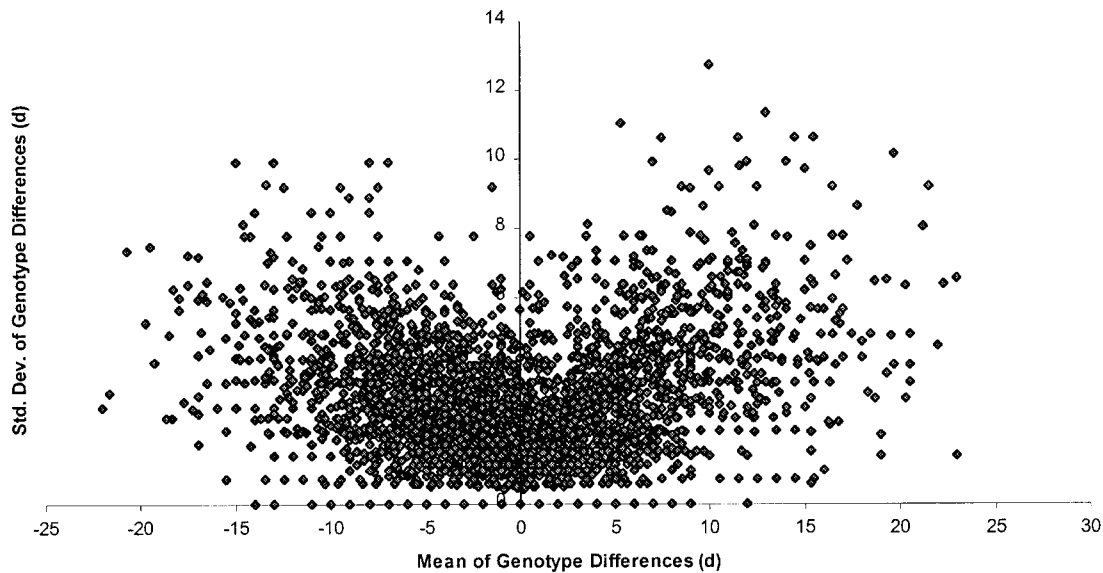


Fig. 6. Relation between genotype \times environment interaction magnitude and differences in genotype means. The shape of the scatter is consistent with both the rarity and existence of exchanges in anthesis date order among pairs of soybean varieties grown in different environments (see text).

Hoogenboom), were examined for evidence of order switching. There were early and late plantings in each year for a total of 20 different environments. Only irrigated treatments were used to limit the possible effects of water stress. The data contained a total of 901 observations (i.e., treatment means) representing 141 varieties. These were screened for all pairs of varieties that co-occurred in two or more environments. A subset of 137 varieties comprising 873 observations met this criterion.

A case was defined as two specific varieties compared in two particular environments. Among the 37 395 cases in the data were 3470 (9.3%) with apparent switches. However, the effect cannot be declared real without a statistical evaluation. The following reasoning was developed by Dr. T. Loughin (Dep. of Stat., Kansas State Univ., personal communication, 2000). Consider the distribution of differences in mean genotype effects between pairs of varieties. If no attention is paid to the order in which the subtractions are performed, then this distribution will span 0. In this situation, there are only two ways in which switches can fail to occur. The first is if there are no genotype \times environment ($G \times E$) interactions, a premise that is undeniably false. The second is if the magnitudes of $G \times E$ interactions become small when genotype differences are small. Only in this way can pairs of varieties have genotype differences that do not fall on opposite sides of the origin in diverse environments.

There were 5201 pairs of varieties found in two or more environments. The means and variances of the intervarietal differences in treatment means were calculated across environments. The variances estimate $G \times E$ interaction effects and were converted to standard deviations so as to have the same units as the means. Figure 6 is a plot of all 5201 (mean, standard deviation) pairs. As is evident, there is some reduction in the standard deviation as the means approach 0, consistent with the relative rarity of switching. However, there are points quite near 0 with standard deviations as large as 4 d or greater, making it very difficult to assert that all treatment differences actually have the same sign. These deviations also exceed the 1-d sampling interval (whose impact is clear along both axes), thus eliminating quantizing effects as a source of the observed switching.

The observed soybean switches have smaller magnitudes

(as defined in the main text) than seen in *A. thaliana*; 90% are 8 d or less. However, the consistent, 8°C separation between chamber runs is a wider deviation than would be expected between two temporally distinct field studies at a single site. It is not surprising, therefore, that the cumulative effect would be less in soybean plots. Dr. William Schapaugh, an experienced breeder (Dep. of Agron., Kansas State Univ.), has also indicated (personal communication, 2000) that switches of this size are not unusual in irrigated soybean plots. Nor is it remarkable, given rarity and small magnitude, that switching has gone unnoticed in comparisons of actual data with the predictions of simulation models incapable of the effect. Switching may have just been overlooked as noise.

REFERENCES

- Ahmad, M., and A.R. Cashmore. 1993. HY4 gene of *A. thaliana* encodes a protein with characteristics of a blue-light photoreceptor. *Nature (London)* 366:162–166.
- Barabasi, A., and R. Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.
- Blazquez, M.A. 2000. Flower development pathways. *J. Cell Sci.* 113:3547–3548. Available at <http://www.biologists.org/JCS/113/20/jcs8511.html>.
- Blazquez, M.A., and D. Weigel. 2000. Integration of floral inductive signals in *Arabidopsis*. *Nature (London)* 404:889–892.
- Bridges, D.C., H.I. Wu, P.J.H. Sharpe, and J.M. Chandler. 1989. Modeling distributions of crop and weed seed germination time. *Weed Sci.* 37:724–729.
- Burstein, Z. 1995. A network model of the developmental gene hierarchy. *J. Theor. Biol.* 174:1–11.
- Cox, D.R., and D. Oakes. 1984. *Analysis of survival data*. Chapman and Hall, London.
- Devlin, F.P., and S.A. Kay. 2000. Flower arranging in *Arabidopsis*. *Science* 288:1600–1602.
- Duan, Q., S. Sorooshian, and V. Gupta. 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.* 28:1015–1031.
- Duan, Q., S. Sorooshian, and V. Gupta. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol. (Amsterdam)* 158:265–284.
- Elizondo, D.A., R.W. McClendon, and G. Hoogenboom. 1992. Neural network models for predicting crop phenology. *Proc. ASAE Int. Winter Meet., Nashville, TN. 15–18 Dec. 1992. Paper 92–3596.* ASAE, St. Joseph, MI.

- Fine, T.L. 1999. Feedforward neural network methodology. Springer-Verlag, New York.
- Fowler, S., K. Lee, H. Onouchi, A. Samach, K. Richardson, B. Morris, G. Coupland, and J. Putterill. 1999. GIGANTEA: A circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* 18:4679–4688.
- Grierson, D.E., and P. Hajela (ed.) 1996. Emergent computing methods in engineering design: Applications of genetic algorithms and neural networks. NATO Asi Ser. F. Springer-Verlag, New York.
- 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.
- Guo, H., H. Yang, T.C. Mockler, and C. Lin. 1998. Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* 279:1360–1363.
- Hammer, G.L. 1998. Crop modeling: Current status and opportunities to advance. *Acta Hortic.* 456:27–36.
- Hanks, J., and J.T. Ritchie. 1991. Modeling plant and soil systems. *Agron. Monogr.* 31. ASA, CSSA, and SSSA, Madison, WI.
- Heiniger, R.W., R.L. Vanderlip, and S.M. Welch. 1997. Developing guidelines for replanting grain sorghum: I. Validation and sensitivity analysis of the SORKAM sorghum growth model. *Agron. J.* 89:75–83.
- Hesketh, J.D., D.L. Myhre, and C.R. Willey. 1973. Temperature control of time intervals between vegetative and reproductive events in soybeans. *Crop Sci.* 13:250–254.
- Hoogenboom, G., and J.W. White. 2003. Improving physiological assumptions of simulation models by using gene-based approaches. *Agron. J.* 95:82–89 (this issue).
- Howell, S.H. 1998. Molecular genetics of plant development. Cambridge Univ. Press, Cambridge, UK.
- Irmak, A., J.W. Jones, T. Mavromatis, S.M. Welch, K.J. Boote, and G.G. Wilkerson. 2000. Evaluating methods for simulating soybean cultivar responses using cross validation. *Agron. J.* 92:1140–1149.
- Jamieson, P.D., M.A. Semenov, I.R. Brooking, and G.S. Francis. 1998. Sirius: A mechanistic model of wheat response to environmental variation. *Eur. J. Agron.* 8:161–179.
- Kasabov, N.K. 1996. Foundations of neural networks, fuzzy systems, and knowledge engineering. MIT Press, Cambridge, MA.
- Koornneef, M., C. Alonso-Blanco, H. Blankstijn-de Vries, C.J. Hanhart, and A.J.M. Peeters. 1998a. Genetic interactions among late flowering mutants of *Arabidopsis*. *Genetics* 148:885–892.
- Koornneef, M., C. Alonso-Blanco, A.J.M. Peeters, and W. Soppe. 1998b. Genetic control of flowering time in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49:345–370.
- Koornneef, M., C.J. Hanhart, and J.H. Van der Veen. 1991. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 229:57–66.
- Levy, Y.Y., and C. Dean. 1998. The transition to flowering. *Plant Cell* 10:1973–1989.
- Lewin, B. 1997. *Genes VI*. Oxford Univ. Press, Oxford, UK.
- Macknight, R., I. Bancroft, T. Page, C. Lister, R. Schmidt, K. Love, L. Westphal, G. Murphy, S. Sherson, C. Cobbett, and C. Dean. 1997. FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* 89:737–745.
- McDaniel, C.N. 1996. Developmental physiology of floral initiation in *Nicotiana tabacum* L. *J. Exp. Bot.* 47:465–475.
- Mendoza, L., and E.R. Alvarez-Buylla. 1998. Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. *J. Theor. Biol.* 193:307–319.
- Mendoza, L., and E.R. Alvarez-Buylla. 2000. Genetic regulation of root hair development in *Arabidopsis thaliana*: A network model. *J. Theor. Biol.* 204:311–326.
- Mjolsness, E., D.A. Sharp, and J.A. Reinitz. 1991. A connectionist model of development. *J. Theor. Biol.* 192:429–453.
- Park, D.H., D.E. Somers, Y.S. Kim, Y.H. Choy, H.K. Lim, M.S. Soh, H.J. Kim, S.A. Kay, and H.G. Nam. 1999. Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis* GIGANTEA gene. *Science* 285:1579–1582.
- Pradhan, S. 1946. Insect population studies: IV. Dynamics of temperature effect on insect development. *Proc. Natl. Inst. Sci. India* 12: 385–404.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. Numerical recipes in C: The art of scientific computing. Cambridge Univ. Press, Cambridge, UK.
- Putterill, J., F. Robson, K. Lee, R. Simon, and G. Coupland. 1995. The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* 80:847–857.
- Reed, J.W., P. Nagpal, D.S. Poole, M. Furuya, and J. Chory. 1993. Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *Plant Cell* 5:147–157.
- Reeves, P.H., and G. Coupland. 2000. Response of plant development to environment: Control of flowering by daylength and temperature. *Curr. Opin. Plant Biol.* 3:37–42.
- Riechmann, J.L., J. Heard, G. Martin, L. Reuber, C.-Z. Jiang, J. Keddie, L. Adam, O. Pineda, O.J. Ratcliffe, R.R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J.Z. Zhang, D. Ghandehari, B.K. Sherman, and G.-L. Yu. 2000. *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* 290:2105–2110.
- Roman-Paoli, E. 1997. Maize performance in Kansas: A CERES-Maize simulation. Ph.D. diss. Kansas State Univ., Manhattan (Diss Abstr. Int. 58–08B:AAI9804375).
- Rosenzweig, C., J. Philips, R. Goldberg, J. Carroll, and T. Hodges. 1996. Potential impacts of climate change on citrus and potato production in the US. *Agric. Syst.* 52:455–479.
- Segel, I.H. 1975. Enzyme kinetics. John Wiley & Sons, New York.
- Sharpe, P.J.H., and D.W. DeMichelle. 1977. Reaction kinetics of poikilotherm development. *J. Theor. Biol.* 64:649–670.
- Simpson, G.G., A.R. Gendall, and C. Dean. 1999. When to switch to flowering. *Ann. Rev. Cell Dev. Biol.* 15:519–550. Available at <http://cellbio.annualreviews.org/cgi/content/full/15/1/519>.
- Theissen, G., and H. Saedler. 1999. The golden decade of molecular floral development (1990–1999): A cheerful obituary. *Develop. Genet. (NY)* 25:181–193.
- Thyer, M., F. Kuezera, and B.C. Batas. 1999. Probabilistic optimization for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms. *Water Resour. Res.* 35:767–773.
- Tsuji, G.Y., G. Uehara, and S. Balas (ed.) 1994. Decision Support System for Agrotechnology Transfer (DSSAT), Version 3. Univ. of Hawaii, Honolulu.
- Tubiello, F.N., C. Rosenzweig, B.A. Kimball, P.J. Pinter, Jr., G.W. Wall, D.J. Hunsaker, R.L. LaMorte, and R.L. Garcia. 1999. Testing CERES-wheat with free-air carbon dioxide enrichment (FACE) experiment data: CO₂ and water interactions. *Agron. J.* 91:247–255.
- Wagner, T.L., H.I. Wu, P.J.H. Sharpe, R.M. Schoolfield, and R.N. Coulson. 1984. Modeling insect development rates: A literature review and application of a biophysical model. *Ann. Entomol. Soc. Am.* 77:208–225.
- Welch, S.M., J.W. Jones, M.W. Brennan, G. Reeder, and B.M. Jacobson. 2002. PCYield: Model-based decision support for soybean production. *Agric. Syst.* 74:79–98.
- 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.
- Wraith, J., and D. Or. 1998. Nonlinear parameter estimation using spreadsheet software. *J. Nat. Resour. Life Sci. Educ.* 27:13–19.