

## Flowering time control: gene network modelling and the link to quantitative genetics

Stephen M. Welch<sup>A,E</sup>, Zhanshan Dong<sup>A,B</sup>, Judith L. Roe<sup>C</sup>, and Sanjoy Das<sup>D</sup>

<sup>A</sup>Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA.

<sup>B</sup>Current address: Pioneer Hi-Bred International, Inc., 7300 NW 62nd Ave, Johnston, IA 50131, USA.

<sup>C</sup>Division of Biology, Kansas State University, Manhattan, KS 66506, USA.

<sup>D</sup>Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506, USA.

<sup>E</sup>Corresponding author. Email: welchsm@ksu.edu

**Abstract.** Flowering is a key stage in plant development that initiates grain production and is vulnerable to stress. The genes controlling flowering time in the model plant *Arabidopsis thaliana* are reviewed. Interactions between these genes have been described previously by qualitative network diagrams. We mathematically relate environmentally dependent transcription, RNA processing, translation, and protein–protein interaction rates to resultant phenotypes. We have developed models (reported elsewhere) based on these concepts that simulate flowering times for novel *A. thaliana* genotype–environment combinations. Here we draw 12 contrasts between genetic network (GN) models of this type and quantitative genetics (QG), showing that both have equal contributions to make to an ideal theory. Physiological dominance and additivity are examined as emergent properties in the context of feed-forwards networks, an instance of which is the signal-integration portion of the *A. thaliana* flowering time network. Additivity is seen to be a complex, multi-gene property with contributions from mass balance in transcript production, the feed-forwards structure itself, and downstream promoter reaction thermodynamics. Higher level emergent properties are exemplified by critical short daylength (CSDL), which we relate to gene expression dynamics in rice (*Oryza sativa*). Next to be discussed are synergies between QG and GN relating to the quantitative trait locus (QTL) mapping of model coefficients. This suggests a new verification test useful in GN model development and in identifying needed updates to existing crop models. Finally, the utility of simple models is evinced by 80 years of QG theory and mathematical ecology.

**Additional keywords:** regulation, differential equations, photothermal, pathways.

### Introduction

An organism's genome is a functional control system that, *inter alia*, responds to inputs from the external environment, yielding a sequence of states whose observable features are phenotypes. Prediction of crop plant phenotypes in differing environments is of critical importance to all aspects of agriculture, including new variety development (breeding and marketing), crop production management (variety selection and cultural practices), and utilisation (grain quality and quantity forecasting). The need for high-quality feeds creates a demand for crop phenotype data even within animal agriculture. Three technologies for predicting phenotypes are (in order of age) quantitative genetics (QG),

the basis of scientific breeding-program design (Walsh 2001); physiological crop simulation (CS) modelling, which has been used in research (Hanks and Ritchie 1991), policy analysis (Rosenzweig *et al.* 1996; Tubiello *et al.* 1999), and decision support (McCown *et al.* 2002); and genetic network (GN) theory. As the papers in this Special Issue show, all 3 methods are reacting to, and/or exploiting, the explosive advances within genomic studies. Modelling the links between genotypes and phenotypes as influenced by the environment is a major issue in computational biology that Cooper *et al.* (2002) refer to as the 'GP problem'.

Our research group has been particularly interested in (1) gene network modelling and, more recently, (2) the

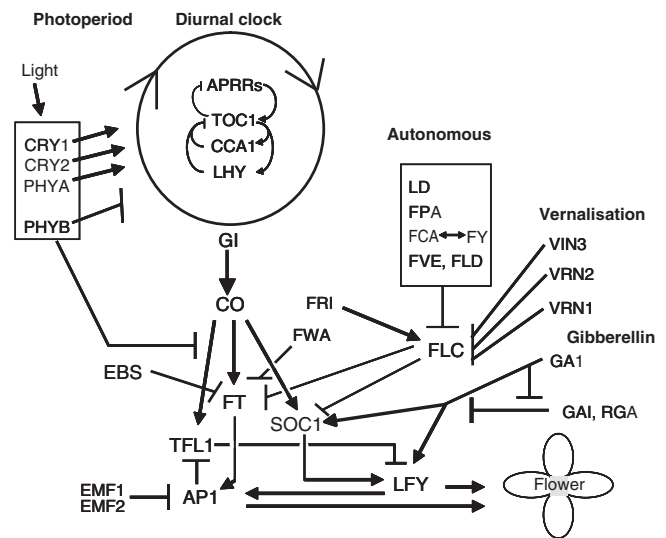
Abbreviations used: CS, crop simulation; CSDL, critical short day length; DNA, deoxyribonucleic acid; Gbp, 10<sup>9</sup> base pairs; GN, genetic network; LER, leaf expansion rate; MAS, marker assisted selection; Mbp, 10<sup>6</sup> base pairs; QG, quantitative genetics; QTL, quantitative trait locus; RNA, ribonucleic acid; RNAP, RNA polymerase; SLA, specific leaf area; TF, transcription factor. Abbreviations for genes with accepted names are given in context.

relationship of GN to quantitative genetics. In pursuit of the first of these interests, we have modelled the GN control of floral initiation in *Arabidopsis thaliana*. This choice of test system was motivated by 2 considerations: (1) intensive research during the 1990s elucidated the structure of the control network and (2) many skillful empirical flowering-time models already exist. Both of these factors suggested a likelihood of success for the undertaking, a rationale that has proved felicitous. The resulting model not only reproduces its calibration data, but also simulates (from first principles) the inflorescence bud dates for many mutant and engineered genotypes that differ widely from those used for model calibration (Dong 2003).

As is often the case in science, however, one answer leads to another question. In the preface to their book, Lynch and Walsh (1998) relate that to some the advent of genomics marks the end of quantitative genetics. In contrast, we marvel at the long-term success of QG and wonder how to account for it. The theory does have shortcomings. Among these is that it can overestimate the progress resulting from artificial selection (Cooper et al. 2005). In this Special Issue, Cooper et al. (2005) and van Eeuwijk et al. (2005) seek to amend and/or extend QG, whereas Walsh (2005) shows ways to use the theory to greater effect. These efforts might be characterised as improvement from within. Although their results are impressive, our approach is different and might be classified as improvement from outside. As Lynch and Walsh (1998) document, QG is a theory grounded in Mendelian principles that emerged in response to certain observed regularities in biological behaviour. We seek (1) gene and gene network features that may account for those regularities and (2) ways to apply GN and GC in concert that may suggest further points of contact. Meeting such goals will undoubtedly resemble climbing Mt Everest. Below we attempt to leverage our flowering-time work to at least discern a few foothills.

### Gene network control of flowering time in *Arabidopsis*

The transition to flowering is influenced by both endogenous and exogenous signals. The underlying genetic regulatory network that processes these signals has been elucidated in *Arabidopsis* over the last 15 years (Levy and Dean 1998). Several graphical network models have been developed that position genes in several interacting pathways which, depending on environmental or autonomous conditions, promote or repress flowering (Martinez-Zapater et al. 1994; Haughn et al. 1995; Blazquez 2000; Boss et al. 2004). A summary view is presented in Fig. 1. Four major pathways (photoperiod, autonomous, vernalisation, and gibberellin) converge on the meristem identity genes *LEAFY* (*LFY*) and *APETALA1* (*AP1*). Under appropriate conditions, their activities elevate *LFY* and *AP1* expression, which then mediates the switch to reproductive development in the shoot meristem.



**Fig. 1.** The genetic regulatory network controlling flowering time in *Arabidopsis*. The positive (arrows) and negative (bars) regulatory relationships are from reviews cited in the text.

#### Photoperiod pathway

*Arabidopsis* is a facultative or quantitative long day (LD) plant that can flower, albeit much later, in short days (SD). Key regulatory genes seem to be conserved between *Arabidopsis* and rice, a SD plant (Blazquez et al. 2001; Samach and Gover 2001; Yano et al. 2001; Goff et al. 2002; Mouradov et al. 2002; Shimamoto and Koyuzuka 2002), suggesting that common pathways are used. Photoperiod is perceived by the plant and transduced to a downstream signalling system by the interaction of photoreception mechanisms with the endogenous diurnal clock (reviewed in Hayama and Coupland 2003). The light- and clock-regulated expression of the flowering time gene *CONSTANS* (*CO*) (*Hd1* in rice) is critical to the timing of flowering.

#### (a) Photoreceptors

The 2 main classes of photoreceptors in *Arabidopsis* are the phytochromes and the cryptochromes that are involved in sensing red/far red, and blue and ultraviolet wavelengths, respectively (reviewed in Hudson 2000; Lin 2000; Smith 2000; Devlin 2002). The products of the 5 phytochrome genes, *PHYTOCHROME A* (*PHYA*) through *PHYTOCHROME E* (*PHYE*), and the 2 cryptochrome genes, *CRYPTOCHROME 1* (*CRY1*) and *CRY2*, play critical roles in sensing light and entraining the diurnal clock. Both *PHYA* and *CRY2* can promote flowering (Guo et al. 1998; Mockler et al. 1999), and mutations in these genes cause a late-flowering phenotype (Johnson et al. 1994). On the other hand, *phyB* mutants flower earlier than wild type (Goto et al. 1991; Reed et al. 1993; Bagnall et al. 1995), implying that *PHYB* represses flowering (Mockler et al. 1999). But the early flowering phenotype of *phyB* and other phytochrome

double mutants was found to be temperature dependent, and can be suppressed by low temperature (Halliday *et al.* 2003). This suggests that certain phytochromes, including PHYB, PHYA, and PHYD, act as temperature-influenced floral repressors and that this activity increases as temperature rises.

#### (b) Endogenous circadian clock

The details of the *Arabidopsis* circadian clock are reviewed elsewhere (Millar 1999; Somers 1999; Samach and Coupland 2000; Alabadi *et al.* 2001; Johnson 2001; McClung 2001; Devlin 2002; Eriksson and Millar 2003). The central oscillator houses a negative feedback loop whereby *TIMING OF CAB 1 (TOC1)* stimulates expression of *LATE ELONGATED HYPOCOTYL (LHY)* and *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, which then feed back and repress *TOC1* expression (Alabadi *et al.* 2001) (Fig. 1). *TOC1* is a member of an *ARABIDOPSIS PSEUDO RESPONSE REGULATOR (APRR)* gene family (*TOC1/APPR1, APRR3, APRR5, APRR7, and APRR9*) whose staggered wave pattern of circadian expression may indicate a negative feedback loop and/or an oscillator (Makino *et al.* 2000, 2001, 2002; Matsushika *et al.* 2000, 2002a, 2002b; Murakami-Kojima *et al.* 2002). Eriksson and Millar (2003) depict the circadian system in *Arabidopsis* as the 2 just-mentioned loops linked by *TOC1*.

#### (c) Photoperiod measurement, transduction, and integration of pathways

Photoperiod is measured by an interaction of light with the circadian clock, and the result is conveyed to the meristem identity genes. Several important flowering time genes operate directly downstream of the circadian clock in the light-signalling pathway. *GIGANTEA (GI)* is involved in phytochrome signalling, with expression levels that are regulated by the circadian clock (Fowler *et al.* 1999; Park *et al.* 1999; Huq *et al.* 2000). In turn, it regulates some clock components in a complex relationship (Fowler *et al.* 1999; Park *et al.* 1999).

*CO* (Putterill *et al.* 1995) is tightly regulated by the circadian clock and is critical for photoperiod perception (Yanovsky and Kay 2002). *CO* accelerates flowering in long days by promoting the expression of the downstream genes *SOC1* and *FT* (Suárez-López *et al.* 2001; Simpson and Dean 2002). Roden *et al.* (2002) altered the timing of circadian rhythms of gene expression relative to dawn and dusk and found that a cycle was perceived as a long-day condition when elevated *CO* expression coincided with daylight, consistent with the external coincidence model of photoperiodism (Samach and Coupland 2000; Carre 2001; Davis 2002). This model says that only light coincident with a certain phase of the diurnal clock will promote flowering (Bunning 1936; Pittendrigh 1972).

*SOC1* and *FT* integrate signals from the photoperiod pathway just described (Kardailsky *et al.* 1999; Kobayashi *et al.* 1999; Samach *et al.* 2000; Suárez-López *et al.* 2001; Blazquez *et al.* 2002; Hepworth *et al.* 2002), the autonomous pathway discussed next (Simpson *et al.* 1999; Sheldon *et al.* 2000; Rouse *et al.* 2002) and, in the case of *SOC1*, gibberellin (Blazquez and Weigel 1999). *SOC1* response has also been linked to physiological age (Samach *et al.* 2000). *SOC1* and *FT* promote expression of the meristem identity genes *LFY* and *API*, respectively, which then cause the transition to inflorescence development in the shoot meristem (reviewed in Reeves and Coupland 2000; Mouradov *et al.* 2002; Sung *et al.* 2003).

#### Autonomous pathway

*FLOWERING LOCUS C (FLC)* is a floral repressor gene (Michaels and Amasino 1999; Sheldon *et al.* 1999) that is regulated by both the autonomous and the vernalisation pathways (see next section). Elevated expression of *FLC* leads to suppression of *FT* and *SOC1* and inhibition of flowering (Hepworth *et al.* 2002). Genes in the autonomous pathway include *FCA*, *FY*, *FVE*, *FPA*, and *LUMINIDEPENDENS (LD)* (Koornneef *et al.* 1991). Mutations in these genes cause a late-flowering phenotype in long or short photoperiods, but this is suppressed by a vernalisation treatment (Martínez-Zapater and Somerville 1990; Koornneef *et al.* 1991). These genes independently promote flowering by down-regulating expression of *FLC* (Sheldon *et al.* 1999, 2000), although the complete pathways are not fully understood. Several of the autonomous genes encode proteins that likely function in RNA processing, although direct targets are not known. For example, *FCA* encodes a protein with 2 RNA recognition motifs (Burd and Dreyfuss 1994). Processing of *FCA* transcripts is complex (Macknight *et al.* 1997, 2002) and shows negative autoregulation (Quesada *et al.* 2003). *FY* is a conserved mRNA 3' end processing factor that functions with *FCA* (Simpson *et al.* 2003). *FPA* also encodes an RNA-binding protein (Schomburg *et al.* 2001), whereas *LD* encodes a homeodomain protein (Lee *et al.* 1994) that could be involved in RNA regulation (Dubnau and Struhl 1996). *FVE* and another autonomous pathway gene, *FLD*, may directly act in the repression of *FLC* by promoting modification of the chromatin in the *FLC* gene (Kenziar and Folk 1998; He *et al.* 2003; Ausin *et al.* 2004). In addition, *FVE* may not just be involved in flowering time control, but also during all stages of plant development (Martínez-Zapater *et al.* 1995). It is interesting that *fve* and *fca* mutants are less sensitive to growth temperature than either wild type or other mutants are in the range of 24–16°C (Blazquez *et al.* 2003; Dong 2003; Welch *et al.* 2003), suggesting that *FCA* and *FVE* function is especially responsive to temperature, which may contribute to differential temperature-dependent growth rates, leaf development rates, and flowering time.

### Vernalisation pathway

Cold treatment leads to down-regulation of *FLC* in genotypes that require vernalisation (1–3 months under 4°C), protecting the plant from flowering before spring (Reeves and Coupland 2000). Mutations decreasing *FLC* function have created naturally occurring vernalisation-independent ecotypes (summer annuals) (Michaels *et al.* 2003). *FRIGIDA* (*FRI*), a second major determinant of natural variation in *Arabidopsis* flowering time, enhances *FLC* function (Michaels and Amasino 1999; Johanson *et al.* 2000).

The vernalisation pathway senses low temperatures and down-regulates *FLC* mRNA levels, which promotes flowering. A set of vernalisation genes has been identified in screens for mutants unable to respond to vernalisation (Chandler *et al.* 1996; Sung and Amasino 2004). *VERNALIZATION INSENSITIVE-3* (*VIN3*) is activated by long cold periods, and establishes the vernalisation response by promoting histone deacetylation at the *FLC* locus, effectively repressing *FLC* expression (Sung and Amasino 2004). The autonomous genes *FVE* and *FLD* described previously independently promote histone deacetylation at *FLC* (He *et al.* 2003; Ausin *et al.* 2004). Two other vernalisation genes, *VRN1* and *VRN2*, are involved in the maintenance of *FLC* repression. *VRN1* encodes a DNA-binding protein and *VRN2* encodes a nuclear-localised protein; these both promote histone methylation and silencing at the *FLC* locus (Gendall *et al.* 2001; Levy *et al.* 2002; Bastow *et al.* 2004; Sung and Amasino 2004).

### Gibberellin pathway

The plant hormone, gibberellin (GA), influences flowering time, and is the main promoting pathway for flowering in short days in *Arabidopsis* (Reeves and Coupland 2001). It plays only a minor role during long days, when other pathways are dominant. Several major genes that have been identified in the pathway are repressors of the GA response, including *GIBBERELLIN INSENSITIVE* (*GAI*) and *REPRESSOR OF GA* (*RGA*), which presumably must be inhibited during GA signalling (Peng *et al.* 1997; Silverstone *et al.* 1997, 1998). They are both members of the GRAS family of plant transcription factors (TFs) (Lee *et al.* 2002). The gibberellin pathway stimulates flowering by causing up-regulation of *SOC1* and *LFY*, but not *FT* (Blazquez *et al.* 1998; Moon *et al.* 2003).

### Meristem identity genes

Promotion of flowering from the flowering time genes finally stimulates expression of the floral meristem identity genes *LFY* and *API* as the terminal output from the converging signal pathways (Hempel *et al.* 1997). The floral meristem identity genes then control floral organ identity genes that pattern development of the floral organs (Simpson *et al.* 1999; Espinosa-Soto *et al.* 2004). Over-expression of *LFY*

causes early flowering (Weigel and Nilsson 1995), whereas mutations in the *LFY* gene result in the conversion of early flowers into shoot-like structures (Weigel *et al.* 1992). *API* encodes a TF that also regulates genes controlling floral structures (Alejandra Mandel *et al.* 1992; Ng and Yanofsky 2001; Lamb *et al.* 2002).

*LFY* is activated by *SOC1* (Mouradov *et al.* 2002), possibly via an intermediary factor *AGAMOUS-LIKE 24* (*AGL24*) (Yu *et al.* 2002), and is up-regulated by gibberellin, either directly or via *SOC1* as the activator (Blazquez and Weigel 1999; Moon *et al.* 2003), whereas *API* is activated by *FT* (Ruiz-Garcia *et al.* 1997). *API* and *LFY* have distinct but overlapping functions and positively regulate each other; *LFY* has been shown to directly activate *API* transcription (Bowman *et al.* 1993; Weigel and Meyerowitz 1993; Parcy *et al.* 1998; Liljegren *et al.* 1999; Wagner *et al.* 1999). *TERMINAL FLOWER 1* is an *FT*-like gene that acts antagonistically with *API* and *LFY* to maintain inflorescence meristem indeterminacy and is activated during development by *CO* (Simon *et al.* 1996; Bradley *et al.* 1997; Liljegren *et al.* 1999; Ratcliffe *et al.* 1999; Ferrandiz *et al.* 2000; Parcy *et al.* 2002). A currently tenable view is that floral commitment relies on a 3-gene double negative (i.e. positive) feedback loop ( $API \rightarrow TFL1 \rightarrow LFY \rightarrow API$ ) highly suggestive of a redundant bi-stable switch (Welch *et al.* 2005).

### Other floral repressors and regulation of flowering time by microRNAs

In addition to the floral repressor *FLC*, other genes have been identified that seem to negatively regulate flowering, including the *EMBRYONIC FLOWERING* (*EMF*) genes, *EARLY BOLTING IN SHORT DAYS* (*EBS*), and *FWA* (Reeves and Coupland 2000; Chou *et al.* 2001; Sung *et al.* 2003). Recently reported gene expression profiles of induced shoot apices reveal not only induced genes, but also additional down-regulated genes that may function as floral repressors (Schmid *et al.* 2003). In addition to traditional scenarios of gene repression, microRNAs are the products of a new class of genes that can inhibit the function of other genes at a post-transcriptional step (Carrington and Ambros 2003). Recently, the *miR172* gene has been described that can target a family of AP2-domain containing genes, including some that may be floral repressors (Park *et al.* 2002; Auckerman and Sakai 2003; Schmid *et al.* 2003). So, the plant must coordinate the repression and activation of a diverse set of genes to achieve the switch to flowering.

### Gene network modelling

Quantification of systems as intricate as the one just summarised requires sophisticated descriptive methods. Multiple mathematical formalisms have been used to model genetic and metabolic networks. Examples include (1) Boolean (ON/OFF) networks (Kauffman 1993; Frank 1999; Liang *et al.* 1998; Mendoza and Alvarez-Buylla

1998; Szallasi and Liang 1998; Samsonova and Serov 1999; Akutsu *et al.* 2000; Ideker *et al.* 2000; Mendoza and Alvarez-Buylla 2000; Maki *et al.* 2001; Kauffman *et al.* 2003, 2004), including work on *A. thaliana* floral-organ specification (Espinosa-Soto *et al.* 2004); (2) Petri (concurrent information flow) nets (Goss and Peccoud 1999; Matsuno *et al.* 2000); (3) S-systems (continuous time models motivated by chemical kinetics) (Liang *et al.* 1998; Akutsu *et al.* 1999, 2000; Tominaga *et al.* 1999; Maki *et al.* 2001); (4) differential equation models (Wolf and Eeckman 1998; Chen *et al.* 1999; Baldi and Hatfield 2002; Welch *et al.* 2005); (5) neural network models (Reinitz and Sharp 1995; D'Haeseleer *et al.* 1999; Weaver *et al.* 1999; Marnellos *et al.* 2000); and (6) Bayesian networks (Friedman *et al.* 2000; Barash and Friedman 2001; Hartemink *et al.* 2001).

Welch *et al.* (2005) extended the 'rough network model' described by Chen *et al.* (1999) and Baldi and Hatfield (2002, p. 151) to include thermal effects. They commented that their model did not explicitly treat protein-protein interactions, although they gave examples in which the model functioned well despite this deficiency. Their model also did not encompass the concept of ploidy, being effectively haploid. These issues can be remedied. Let the state vector  $\mathbf{s}$  be a  $(m + p)$  column vector partitioned into  $\mathbf{m}$  and  $\mathbf{p}$ , which contain effective levels of RNA expression and protein, respectively, although other metabolites could easily be added. The rate of change of any  $s_i$  is the environmentally dependent difference between production and degradation rates:

$$\frac{ds_i}{dt} = R_i g_i(\mathbf{s}, \mathbf{e}, \mathbf{a}_i) - \lambda_i h_i(\mathbf{s}, \mathbf{e}, \mathbf{b}_i) s_i \quad (1)$$

where  $\mathbf{a}_i$  and  $\mathbf{b}_i$  are parameter vectors,  $\mathbf{e}$  is a vector of environmental inputs, and  $R_i$  and  $\lambda_i$  are scalars. The  $\mathbf{a}_i$ ,  $\mathbf{b}_i$ ,  $R_i$ , and  $\lambda_i$  parameters are all positive. Common RNA- and protein-level measurement techniques (e.g. northern or western blotting, microarray, and rt-PCR technology) involve normalisation against standards, so the  $s_i$  are dimensionless. The parameter values are assumed to offset the use of different standards for different  $i$  s. At this level of generality, exact forms for  $g$  and  $h$  need not be specified. However, because all biological processes have finite rates, we assume that  $0 \leq g, h \leq 1$ , but not that  $R_i, \lambda_i$  are necessarily reached. Although it is not always true (e.g. Comai and Harada 1990), unless compelled by the exigencies of a particular system, we assume that  $s_i(0) = 0$ .

Eqn 1 allows RNA, protein, and the environment to influence each other in many combinations, thus supporting modelling of transcription control (e.g. *API*, *FLD*, *GAI*, *RGA*, and *VRNI*), RNA processing (e.g. *FCA*, *FPA*, and *FY*), photothermal regulation (e.g. *CO* and *FVE*), and numerous other interactions. The equation can also be modified to incorporate ploidy by assuming that the total RNA produced at each locus is the sum of that produced by each allele.

Representing a diploid genotype can be accomplished with the  $(m + p)$  column vector:

$$\mathbf{X} = \begin{bmatrix} \alpha_{1j_{11}} + \alpha_{1j_{22}} \\ \vdots \\ \alpha_{mj_{m1}} + \alpha_{mj_{m2}} \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (2)$$

where the last  $p$  entries are 1.  $\alpha_{ijk}$  is the *effective* relative RNA production rate of allele  $j$  at locus  $i$ , where  $i$  indexes the haploid genome. The  $k = 1, 2$  subscript denotes the chromosomes in each homologous pair; extension to higher ploidy is straightforward. The  $\alpha_{ijk}$  are measured relative to single-copy wild-type alleles, which have unit values. For a null allele,  $\alpha_{ijk} = 0$ . A mutation that increases the copy number increases  $\alpha_{ijk}$  proportionately (or decreases it if copy number increases lead to gene silencing). A partial-loss-of-function mutant allele has an intermediate  $\alpha_{ijk}$ . This approach treats mutation as altering regulation or product activity proportionately across all situations. For conciseness, let  $*$  denote component-wise multiplication; the model is then written:

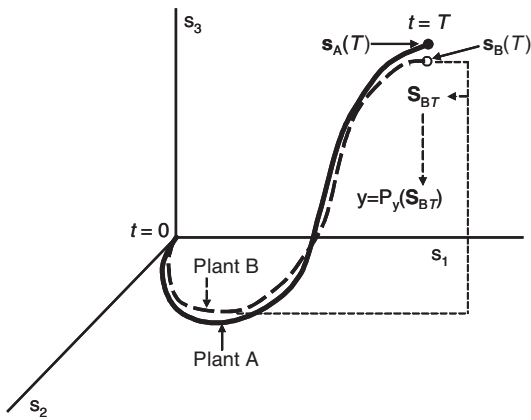
$$\dot{\mathbf{s}} = \mathbf{X} * \mathbf{R} * \mathbf{g} - \Lambda * \mathbf{h} * \mathbf{s} \quad (3)$$

Eqn 3 is biochemical in nature and therefore says nothing about phenotypic traits such as plant height, flowering date, etc. It seems virtually impossible that 2 plants  $A$  and  $B$  would ever be biochemically identical [i.e.  $\mathbf{s}_A(t) = \mathbf{s}_B(t)$ ], except perhaps trivially at  $t = 0$ . Uniqueness of the biochemical state at observation time  $T$  makes it possible, in principle, to express the phenotype solely in terms of  $\mathbf{s}_A(T)$ . But many traits (e.g. yield at harvest) depend on earlier events (e.g. water availability at silking, see Campos *et al.* 2004). It is therefore more convenient to relate trait scores to the entire time series of  $\mathbf{s}$  from  $t = 0$  to  $T$ , which we symbolise by  $\mathbf{S}_T$  (adding an additional plant subscript when needed for clarity). This is illustrated in Fig. 2 for  $(m + p) = 3$ . This notation exhibits massive data compression. For example, if whole-genome microarrays were used to take repeated gene-expression snapshots from 0 to  $T$  then  $\mathbf{S}_T$  could be the entire multi-sample database.

$\mathbf{S}_T$  is a function of time. A *functional* is a mathematical operation that pairs a function with a real value  $y$  (e.g. for some quantitative trait). The notation is:

$$y = P_y(\mathbf{S}_T) \quad (4)$$

where  $P_y$  is the phenotype functional. The combination of Eqns 3 and 4 is a gene-to-phenotype model; to be specific it exemplifies the  $\Gamma_{(g|N)ij}$  term in Eqn 2 of Cooper *et al.* (2005). Although  $P_y$  can take any form, a common functional



**Fig. 2.** Biochemical trajectories for two plants (*A* and *B*). The coordinates of each point in this space represent a simultaneous combination of the levels of three gene products  $S_1$ ,  $S_2$ , and  $S_3$ . A single curve is therefore equivalent to a plot of three individual time series. Because a plant's phenotype ( $P_y$ ) may depend on its complete developmental history, phenotype can be expressed as a functional of entire curve ( $S_T$ ) as shown in dashed lines for *B*.

calculates the cumulative effect of an ongoing process such as grainfill. In this case:

$$P_y(S_T) = \int_0^T p(S_\tau) d\tau \quad (5)$$

Dong (2003) developed an *A. thaliana* flowering time model that specialised the theory just described. The  $h_i$  were either zero or one, and most  $g_i$  operated below an upper bound. The model simulated the temperature- and photoperiod-dependent dynamics of mRNA expression for 9 genes and 4 protein levels. Eqn 4 was a threshold trigger for budding based on *LFY* expression (Fig. 1). Two datasets were used for parameter estimation: (1) bud dates from replicated growth chamber experiments on *fca-6*, *fpa-2*, *fve-2*, *co-6*, *cry2-1*, *gi-6*, and *phyB-1* single-locus homozygous mutants and wild-type line under constant photothermal regimes and (2) gene expression time series from the literature. The model accounted for 85% of the variation in observed bolting time for the 8 calibration genotypes. The model was validated against 121 independent observations, all but 8 of which were from literature. Of the 121, four *fve-1* double mutants and 2 genotypes reared at 6°C were statistical outliers that were discarded. The retained data contained heterozygotes, single and double mutants, and over-expression lines under constant or variable temperatures and alternative photoperiods, including continuous light; in short, a far wider range of G × E combinations than were present in the calibration data. As is common with literature data, bud dates were reported as total leaf numbers (TLN) rather than in days after planting (DAP), which is what the model predicts. Typical correlations between DAP and TLN exceed 90% (Dong 2003). TLN observations ranged from 4.3 to 74.4 and were quite uniformly distributed on this

interval. The model accounted for 73% of the variation in TLN despite having included only 9 out of the 100+ genes known to affect flowering time. This suggests that the true complexity of real genetic networks is not necessarily an impediment to effective GN models.

### Genetic networking and quantitative genetics

The theory of quantitative genetics (Lynch and Walsh 1998) is an immense field of inquiry; indeed, a Google search on the phrase 'quantitative genetics' scores *c.* 72 600 hits. Two major goals of QG are to provide methods for (1) inferring and (2) manipulating genetic properties on the basis of phenotypic data. As stated, these goals almost seem to reverse the objectives of genetic networking evident in the previous section. In fact, at least a dozen contrasts may be drawn between the 2 topics:

- (1) QG is largely algebraic and focusses on changes across generations; GN often applies differential equations or Boolean models on time scales of hours or days;
- (2) QG variables have phenotypic units; GN variables are dimensionless, normalised biochemical levels;
- (3) as a consequence, QG needs no additional mathematical mechanisms to relate 'genotype' to 'phenotype' values; GN does, in the form of Eqn 4;
- (4) QG is a Mendelian theory, explicitly depicting alleles, loci, chromosomes, mutation, crossover, reproduction, selection, etc.; GN focusses on the processes of transcription, translation, their controls, and the interactions of the resulting chemical species, although mechanisms like the  $\mathbf{X}$  vector in Eqn 3 do allow for alternative alleles;
- (5) thus, QG well describes the physical structure of the genetic mechanism (chromosomes, alleles, markers, etc.); GN better represents the biochemical and information-processing functions;
- (6) QG models are linear, with interaction terms whose number can rise rapidly to impractical levels; GN models are nonlinear, with considerably fewer direct interactions that are explicitly defined by directed graphs;
- (7) QG is a population-level theory that addresses relationships between means and (co)variances; GN operates at the individual level or, most often, lower;
- (8) relatedly, QG relies heavily on basic concepts of probability; GN models may include stochastic elements (e.g. due to molecular randomness), but most often do not;
- (9) QG yields useful results even when the genetic basis of a trait is unknown; GN models explicitly reference particular genes;
- (10) QG is a mature theory with an accepted set of core axioms and an expanding set of applications, some of great commercial value; GN is, as yet, none of these;

- (11) QG views phenotypes in a sequence of increasingly complex contexts: single-allele effects at one locus (additivity), multiple-allele effects at one locus (dominance), multiple-locus effects (epistasis), and environmental effects (general/specific/G × E); GN sees all phenotypes as emerging from the interaction of environmentally influenced network elements;
- (12) QG analyses inheritance in terms of alleles or allele combinations whose effects combine by addition; many GN models do not consider inheritance, but do provide a mechanistic, mathematical representation for explaining how traits originate.

The existence of such a range and diversity of differences is surprising given that the 2 approaches supposedly converge on the same set of biological phenomena. These contrasts cry out for the creation of some hybrid framework combining the features of both. It is plausible that an ideal hybrid would have equal contributions from both parents being QG-like for contrasts 4, 8, 9, and 10; GN-like for 1, 6, 11, and 12; and a blend of both for 2, 3, 5, and 7. Although we adopt complementary points of departure, finding such a hybrid is a goal we share with Cooper *et al.* (2005).

**Dominance and additivity as emergent properties**

GN theory is enabled by advances that allow deep probing of genetic mechanisms. This suggests that fundamental concepts in QG such as additivity, dominance, and epistasis may be emergent properties that become explicable at the network level. It is now easy to compare mean trait scores within known genotypes and, by appropriate differencing, construct individual-level estimates of additive, dominance, and epistasis effects. Furthermore, these means can be related to the corresponding population-level values by appropriate computations involving allele frequencies (Cheverud and Routman 1995). However, theory is mute as to how within-genotype means arise, despite the fact that they are sometimes called ‘physiological’ (Cheverud and Routman 1995). GN moves beyond this juncture. Using a model equivalent to Eqn 6, Omholt *et al.* (2000) studied 2 simple gene networks both of which involved feedback (either positive or negative). They found instances of additive, dominance, and overdominance behaviour depending on the model and the parameter values used. Furthermore, although classical genetics would consider such effects as single-locus properties, the behaviours arose through epistatic interactions.

Although feedback systems are rampant in biological systems, other topologies are also important. For example, the *A. thaliana* flowering time control system has a negative feedback loop in the diurnal clock that provides needed oscillatory inputs, and a positive feedback loop in the bistable switch that records floral initiation. But between these 2 loops is a feed-forward network within which the crucial

signal integration occurs that actually determines flowering time. Except for switches, the elaborate developmental genetic network that controls the embryonic differentiation of sea urchin (*Strongylocentrotus purpuratus*) endomesoderm is also remarkably feed-forward (Davidson *et al.* 2002). The next 2 sections present dominance and additivity examples in the context of feed-forward networks. One example entails protein complexing, a feature not touched on by Omholt *et al.* (2000).

*Dominance*

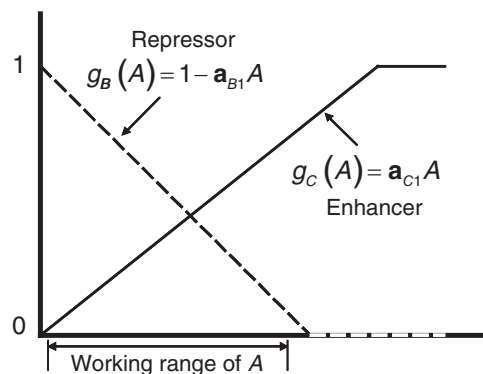
Consider a gene *A* that regulates *B* and *C*, but in opposite directions ( $C \leftarrow A \rightarrow B$ ). Let the dynamics be modelled with a simplified form of Eqn 3:

$$\dot{s} = X_s R_s g_s - \lambda_s s \tag{6}$$

where *s* is either *A*, *B*, or *C*. If *A* is constitutively expressed, then, from the definition of  $X_A$  in Eqn 2 and by setting  $\dot{s}_A = 0$ , the possible equilibrium levels are  $\bar{A} = 2\alpha_A R_A / \lambda_A$ ,  $(\alpha_A + \alpha_a) R_A / \lambda_A$ , and  $2\alpha_a R_A / \lambda_a$ , where the subscripts on  $\alpha$  denote 2 different alleles. Assume that  $g_B$  and  $g_C$  are the simple piecewise linear functions shown in Fig. 3; a rationale for this assumption is presented further below. Finally, suppose that phenotype is determined by the heterodimer  $D \leftarrow B + C$ , whose formation is governed by the Law of Mass Action. Then the steady-state levels of *B*, *C*, and *D* are:

$$\begin{aligned} \bar{B} &= X_B R_B (1 - \mathbf{a}_{B1} \bar{A}) / \lambda_B \\ \bar{C} &= X_C R_C (\mathbf{a}_{C1} \bar{A}) / \lambda_C \\ \bar{D} &= K_{eq} \bar{B} \bar{C} = K_{eq} \mathbf{a}_{C1} (X_B R_B) (X_C R_C) (\bar{A} - \mathbf{a}_{B1} \bar{A}^2) / (\lambda_B \lambda_C) \end{aligned} \tag{7}$$

It can be seen from the  $(\bar{A} - \mathbf{a}_{B1} \bar{A}^2)$  factor that a graph of  $\bar{D}$  (and thus the phenotype) would be a concave-down parabola peaking at  $\bar{A} = 1/(2\mathbf{a}_{B1})$ . This would happen at the steady-state level of the *A* heterozygote if the system parameters were such that  $(\alpha_a + \alpha_A) = \lambda_A / (2\mathbf{a}_{B1} R_A)$ . Such exactitude is unlikely, but there will be finite parameter



**Fig. 3.** Simplified gene product production functions for a repressor ( $g_B$ ) and an enhancer ( $g_C$ ).

ranges for which the condition is well enough satisfied that  $\bar{D}$  for the  $A$  heterozygote exceeds that of either homozygote (i.e. overdominance). Note that dominance would be absent at any time when either  $B$  or  $C$  was present in excess, because the formation of  $D$  would then have first-order kinetics. This illustrates how dominance behaviour can depend on the background, on the environment, or even occur transiently.

*Additivity*

Additivity is mysterious from the GN perspective, especially when present at high levels, as in flowering time. The prerequisite for physiological additivity is linearity, which is far from apparent in Eqn 3, given the kinetic complexities buried in  $g$  and  $h$ . CS models are also highly nonlinear, even in phenology (Yin et al. 1995, 1997). Nevertheless, examples of additivity have been observed in CS outputs (Boote et al. 2003) and in genetic coefficients, which can be viewed as quantitative traits (White and Hoogenboom 1996; Stewart et al. 2003). Finally, although a few genes are sufficient to create epistasis and dominance effects, a single gene might well destroy linearity and, thus, additivity. So, what properties of genetic networks preserve additivity despite the seeming rarity and fragility of linearity? No complete accounting can be given at this time, but some indications and special cases can be described.

Equations 2 and 3 introduce linearity under allelic substitution through the additive  $\mathbf{X}$  factor, a notion suggested by mass balance for transcription products. Suppose (1) that  $C$  is a gene (or set of genes) that proximally affect phenotype, (2) that  $A$  is a gene upstream of  $C$ , and (3) that neither  $C$  nor any gene,  $B$ , between  $A$  and  $C$  feeds back to regulate  $A$  either directly or indirectly. Given these assumptions,  $g_A$  and  $h_A$  are not dependent on  $A$ , and can thus be written as functions of time only. The  $A$  component of Eqn 3 is then:

$$\dot{s}_A + \lambda h_A(t)s_A = X_A R_A g_A(t) \tag{8}$$

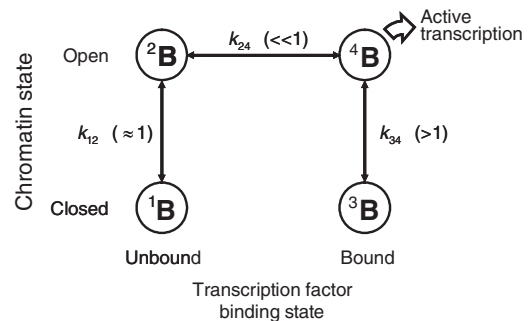
which is a linear differential equation. Thus, if  $s(t)$  is a solution to Eqn 8 when written without  $X_A$ , and if  $\alpha$  and  $\beta$  are the  $X_A$  values for 2 different alleles of  $A$ , then  $s_A(t)$  equals  $2\alpha s(t)$  for the  $\alpha$  homozygote,  $(\alpha + \beta)s(t)$  for the  $\alpha\beta$  heterozygote, and  $2\beta s(t)$  for the  $\beta$  homozygote. Therefore, under this model, a sufficient condition for the expression of a gene to be additive in its own alleles is for it to be part of a feed-forwards signalling cascade.

If additivity in  $A$  is to be seen in the phenotype, however, it must survive modification by any of the downstream  $B$  genes. As a first step in understanding how this might occur, it is useful to consider  $g$  from the perspective of promoter reaction thermodynamics (Kingston and Narlikar 1999). The  $g$  function encodes the dynamics of TF binding and chromatin remodelling that regulate transcription. The rates of these processes are much faster than the dynamics of (1) the RNA polymerase (RNAP) motor responsible for transcription,

which, itself, involves disparate time scales (Wang et al. 1998), and (2) translation activities at the ribosome (Bolouri and Davidson 2003). For this reason, forms for  $g$  can be derived by using equilibrium thermodynamics. In this way, the manifold details of the kinetic convergence to equilibrium may be ignored, reducing both the experimental burden and the intricacy of required simulations. Interest in the latter benefit led Kurata and Taira (2000) to describe a *two-phase partition method*. They partitioned events into *binding* and *reaction* phases, with process rates in the former being sufficiently high that the chemical species involved could be assumed to reside at equilibrium. Calculation speedups as great as  $5000\times$  were achieved without loss of accuracy.

Consider the most elementary case of a single TF that promotes  $s$ . Figure 4 identifies four equilibrium binding states. In  $^1B$  and  $^3B$ , the chromatin is in a closed, condensed state in which there is no space for RNAP to bind and initiate transcription. The other 2 states are open. In  $^1B$  and  $^2B$ , the TF is not bound to the DNA, but in the other 2 it is. Following Kingston and Narlikar (1999), the transition between  $^1B$  and  $^2B$  is assumed to be sufficiently rapid that there is no time for RNAP attachment despite the open state of  $^2B$ . Let  $M_i$  be the molar concentration of promoter binding state  $^iB$ , where the unit volume is such that the  $M_i$ 's sum to 1, allowing their numerical values to be interpreted as  $^iB$  state probabilities. Define  $v$  to be the concentration of the unbound TF, and  $V$  to be the total of bound plus unbound TF (both in the same units as the  $M_i$ ). Thus,  $V$  is the argument of  $g$  and is the solution of an upstream equation of the same form as Eqn 1. The 5 system state variables  $M_1 \dots M_4$  and  $v$  are determined by the nonlinear equations:

$$\begin{aligned} V &= v + 0M_1 + 0M_2 + 1M_3 + 1M_4 \\ 1 &= M_1 + M_2 + M_3 + M_4 \\ M_2 &= k_{12}M_1 \\ M_4 &= k_{24}M_2v \\ M_4 &= k_{34}M_3 \end{aligned} \tag{9}$$



**Fig. 4.** Thermodynamic state diagram. As the TF concentration increases the system is driven towards the transcribing state,  $^3B$ . The equilibrium constant  $k_{24}$  is much less than one because non-target sites, although of low specificity, so outnumber target sites.

The first 2 equations are from mass conservation. From the principle of detailed thermodynamic balance (Hammes 2000) there are also 3 independent mass action equations that complete the set. The  $k_{ij}$  are equilibrium constants.

Some algebra yields:

$$0 = \left(1 + \frac{1}{k_{34}}\right)v^2 + \left(\frac{1}{k_{12}k_{24}} + \frac{1}{k_{24}} + \left(1 + \frac{1}{k_{34}}\right)\right)v - \left(\frac{1}{k_{12}k_{24}} + \frac{1}{k_{24}}\right)V$$

$$g(V) = \frac{M_4}{\left(\sum_{i=1}^4 M_i\right)} = M_4 = v / \left(\left(\frac{1}{k_{12}k_{24}} + \frac{1}{k_{24}}\right) + \left(1 + \frac{1}{k_{34}}\right)v\right) \quad (10)$$

where, for any value of  $V$ ,  $v$  is obtained from the first equation by the quadratic formula and is then inserted into the second to yield  $g$ .

As  $V$  increases, so does the probability of being in one of the bound states  ${}^3B$  or  ${}^4B$ . This will increase the probability of transcription towards an asymptotic upper limit of  $M_4 \approx k_{34}/(1 + k_{34})$ . For any value of  $k_{12}$ , the rate of the approach to the asymptote is determined by  $k_{24}$ . In the neighbourhood of its asymptote,  $g$  is largely unresponsive to  $V$ , but the initial rise portion is surprisingly linear, as can be demonstrated by approximating  $g(V)$  with a 2-piece linear function  $g_r(V) = \min(\alpha V, \beta)$ , where  $\alpha > 0$  is the ramp slope,  $0 < \beta \leq 1$  is the asymptote, and  $V \geq 0$ . We assumed that  ${}^1B$  and  ${}^2B$  are equiprobable ( $k_{12} = 1$ ) and that the ratio of open v. closed bound chromatin states ( $k_{34}$ ) ranges from 1 : 1 to  $10^2$  : 1. Stormo and Fields (1998) gave a nominal specificity of  $10^6$  for the relative likelihood in prokaryotes that a TF binds to its target site rather than to a random site of the same size. Based on this value,  $k_{24}$  was varied from  $10^0$  to  $10^{-4}$ , corresponding, in theory, to genome sizes from 1 Mbp to 10 Gbp, respectively. Several factors probably increase binding specificities in eukaryotes and/or reduce effective genome sizes (G. D. Stormo, pers. comm., 2004), so these values likely subsume the range relevant to higher plants. For several hundred factorial ( $k_{24}, k_{34}$ ) combinations, nonlinear least-squares was used to fit  $g_r(V)$  to  $g(V)$  in the interval  $[0 \dots V_{\max}]$  where  $g(V)$  reached 95% of its asymptote at  $V_{\max}$ . Values of  $r^2$  were tabulated separately for the ramp portion of  $g_r$  (as an index of linearity) and for the function as a whole. Both measures were almost totally insensitive to  $k_{ij}$  values within the ranges tested;  $r^2$ s for the linear ramp were 73–82% and total  $r^2$ s were 85–90%.

These thermodynamic results not only suggest that quasi-linear ramps in  $g$  may contribute to physiological additivity but also support the use of 2-piece linear functions for simplified theoretical analyses such as the

overdominance example previously given. Such functions not only encode the ON/OFF features of genes, but also provide a graded intermediate response. Genetic networks have been extensively studied within the ON/OFF Boolean framework that has inspired the  $E(NK)$  networks described by Cooper *et al.* (2005). Perhaps multivariate  $g_r$ -like functions can be exploited to increment the realism of  $E(NK)$  models, thus enhancing their ability to correct for network interaction effects. Equilibrium analysis also generalises beyond the simple, single-site example given here. One of us has extended Eqn system 9 to arbitrary combinations of TFs, binding sites, and  $k_{ij}$  values and derived a closed-form solution for the rate of change of the equilibrium state (S. M. Welch, unpublished). We are currently examining the properties of this solution in the context of the 2-phase partition method.

Whatever method is used for calculation, expression levels are inputs to the phenotype functional, which must also be linear if the additivity of upstream genes is to be visible. Because phenotype functionals are, by definition, *ad hoc*, few generalisations are possible. Because integration is a linear operation, however, additivity is enhanced when phenotype functionals are accumulative as in Eqn 5. The next section gives an extended analysis of a model introduced by Welch *et al.* (2005) relating flowering development rates in rice (*Oryza sativa*) to the time average (i.e. to the integral) of a gene expression level. The phenotype functional is proportional to  $\bar{y}(f)$  in Eqn 16. Any upstream additivity propagating through the  $g$  functions in that equation would be reflected in the phenotype.

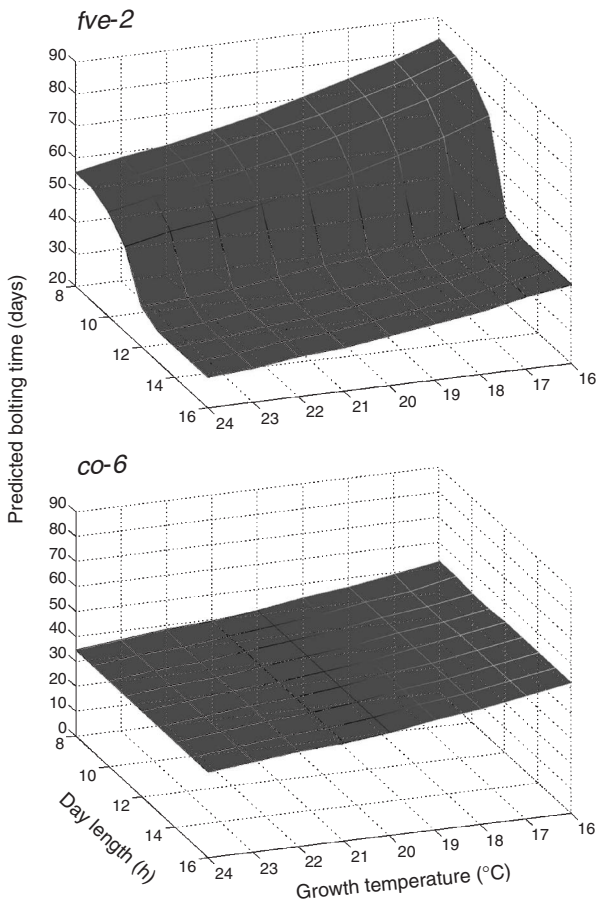
These examples, and those of Omholt *et al.* (2000) show that physiological additivity, far from being an allelic property independent of other genes, depends on the structure and operation of entire gene subnets. The importance of heritability in breeding programs certainly indicates a high priority for understanding its origin. Omholt *et al.* (2000, p. 978) link the success of breeding programs based on heritability as conceptualised by QG to the existence of ‘broad parameter domains where regulatory networks will display additive gene action’. The wide parameter ranges used in our thermodynamic analysis reinforce this conclusion at a very mechanistic level. We have also related physiological additivity to linearity at higher emergent levels such as the one we analyse next.

### Higher-level emergent properties

Existing CS models integrate a wealth of physiological concepts, many of which derive from physical and/or chemical first principles, whereas others reflect empirical observations not yet understood mechanistically. As genetic networks are elucidated, they may provide explanations for the latter. An example is critical short daylength (CSDL), the photoperiod above which developmental rates begin to decline (increase) in SD (LD) plants.

Because accurate flowering time simulation is important, CSDL is a key parameter in some crop models (Tsuji *et al.* 1994), and methods have been developed for its efficient estimation from variety performance trial data (Irmak *et al.* 2000; Welch *et al.* 2002). But how is it determined biologically?

*CO* (Fig. 1) is strongly tied to daylength measurement in *A. thaliana* and its expression profiles have been studied under LD and SD conditions, but the resulting qualitative inferences cannot suggest what patterns might occur at transitional photoperiods. In contrast, Fig. 5 compares *CO* loss of function to an autonomous pathway mutant as simulated by the Dong (2003) model. The latter clearly shows a photoperiod response transition with a mildly temperature-dependent CSDL. The pattern is absent in *co-6*, which has lost the ability to measure daylength. Unfortunately, molecular geneticists, unlike plant physiologists, seldom take data at intermediate daylengths so it remains to evaluate the model in that range.



**Fig. 5.** The *fve-2* autonomous path mutant (*top*) shows a CSDL below which development responds only slightly to photoperiod. In contrast, the *co-6* loss-of-function mutation removes the ability to measure daylength and cancels the response (*bottom*).

Welch *et al.* (2005) modelled genes as Hopfield neurons, a special case of Eqn 1. Among their examples is *HEADING DATE 1* (*Hd1*), the rice homologue of *CO*. The form of the model was suggested by *Hd1* time-series data for Nipponbare, a *japonica* cultivar, collected under SD (9 h) and LD (15 h) by Kojima *et al.* (2002). Specifically:

$$\frac{d}{dt}(Hd1) = \left. \begin{matrix} R_L \\ R_D \end{matrix} \right\} g_{NN}(C(t)) - (Hd1) \left\{ \begin{matrix} \lambda_L \\ \lambda_D \end{matrix} \right. \quad (11)$$

The model assumes (Suárez-López *et al.* 2001) that rates of maximum production ( $R_L$ ,  $R_D$ ) or specific degradation ( $\lambda_L$ ,  $\lambda_D$ ) may differ under light and dark conditions ( $L$  and  $D$ , respectively). Production is governed by  $g_{NN}(u) = [1 + \exp(-u)]^{-1}$ , a sigmoid function used in neural networks, and is driven by a sine-wave clock input,  $C(t)$ . Parameters minimise the sum of absolute errors against the Kojima *et al.* (2002) data.

Welch *et al.* (2005) assumed that daylength is encoded as the time-averaged *Hd1* level and that higher values slow development (Kojima *et al.* 2002), making rice a SD plant. Although per-observation errors average 22% due to model simplicity and the variability of gene expression data, errors in the time averages are 10.3% and 1.3% for SD and LD, respectively. They next compared time averages, plotted as a function of photoperiod, with developmental rates from the rice phenology model of Yin *et al.* (1997) parameterised to mimic Nipponbare. The photoperiod above which development rates start to decline differs by *c.* 15 min between the 2 models, representing the first time that a CSDL has been estimated from gene expression data alone.

The results in Fig. 5 and Welch *et al.* (2005) were obtained by simulation. Both demonstrate the emergent property of CSDL, but neither explains the origin of the effect. We do so now. No assumptions are made about the clock waveform,  $C(t)$ , beyond periodicity, or about  $g$ , which controls transcription and translation.

Because Welch *et al.* (2005) used Eqn 11 only as an example, they did not report parameter values, which are  $R_L = 1.71$ ,  $R_D = 1.03$ ,  $\lambda_L = 0.090$ , and  $\lambda_D = 0.084$ . The 66% difference in  $R$  values *v.* 7% for  $\lambda$  suggests using a single  $\lambda$  for both light and dark photophases. Refitting the data gives  $\lambda = 0.086$  and a slight increase in the goodness-of-fit. Thus, Eqn 11 can be rewritten as:

$$\frac{dy}{dt} - ay = G(t) \quad (12)$$

where  $a = -\lambda$  and  $G(t) = L(t)g[C(t)]$ , with  $L(t) = R_L$  when the lights are on and  $R_D$  otherwise. Define  $t$  to be Zeitgeber (ZT) time (dawn at  $t = 0$ ) and assume that the lights are on for a fixed fraction,  $f$ , of the  $p = 24$  h. diurnal cycle.  $G$  is

periodic because  $C$  and  $L$  are the Fourier series expansion for  $G$  is:

$$G(t) = \sum_{n=0} a_n(f) \cos(s_n t) + b_n(f) \sin(s_n t) \quad (13)$$

where  $s_n = 2n\pi/p$  and

$$\begin{aligned} a_n(f) &= \left(\frac{2}{p}\right) \int_0^p G(\tau) \cos(s_n \tau) d\tau, \\ b_n(f) &= \left(\frac{2}{p}\right) \int_0^p G(\tau) \sin(s_n \tau) d\tau \end{aligned} \quad (14)$$

Ignoring short-term transients, the solution (CRC 1996, p. 405) of Eqn 12 is:

$$\begin{aligned} y(t) &= - \sum_{n=0} a_n(f) [(a \cos(s_n t) - s_n \sin(s_n t)] / (a^2 + s_n^2) \\ &\quad - \sum_{n=0} b_n(f) [(a \sin(s_n t) + s_n \cos(s_n t)] / (a^2 + s_n^2) \end{aligned} \quad (15)$$

Next, average  $y(t)$  over one period. Because  $\int_0^p \cos(s_n \tau) d\tau = \int_0^p \sin(s_n \tau) d\tau = 0$  for  $n > 0$  and  $b_0(f) = 0$ ,

$$\begin{aligned} \bar{y}(f) &= p^{-1} \int_0^p y(\tau) d\tau = a_0(f) / 2\lambda \\ &= \left(\frac{1}{p}\right) \left[ \frac{R_L}{\lambda} \int_0^{fp} g(C(\tau)) d\tau + \frac{R_D}{\lambda} \int_{fp}^p g(C(\tau)) d\tau \right] \end{aligned} \quad (16)$$

Call the 2 integrals on the right-hand side of Eqn 16 the *light* and *dark* integrals, respectively. *Hdl* production begins in the afternoon. The light integral thus contributes little for earlier  $fp$ , so  $\bar{y}(f)$  is constant at the average of  $R_D g(C(t)) / \lambda$ . Because  $R_L > R_D$ ,  $\bar{y}(f)$  rises with further increases in  $fp$  to plateau at the average of  $R_L g(C(t)) / \lambda$ . Thus, CSDL in rice seems to represent the onset of *Hdl* relative to dawn. This linkage adds functionality to what has previously been a purely descriptive CS model input parameter.

Equation 16 is unaltered if, in addition to the clock,  $g$  also responds to  $f$ . This is relevant to *A. thaliana* in which *CO* regulation is more complex. Under SD, *CO* expression increases in the afternoon, exhibits a single peak during the night, and drops to a low level before dawn. Under LD, however, 2 peaks appear in the dark, with the second one extending past dawn (Suárez-López *et al.* 2001). Although this pattern differs from *Hdl*, the result is similar. Under short but increasing daylengths, CSDL occurs, as in rice, when sunset begins to fall later than the rise in *CO*. However,  $\bar{y}(f)$  increases accelerate as the second peak develops and affects the light integral. Although the second peak affects the dark integral, this is more than offset by a shrinking interval of integration. Thus, both species operate consistently with the external coincidence model (Bunning

1936; Pittendrigh 1972) in that daylight during the expression period determines progress towards flowering.

## Synthesis

Wu *et al.* (2002) (1) fit rice leaf phenological data from 111 lines to a multi-parameter growth curve, (2) analysed the resulting line-specific parameter vectors with multi-trait composite interval mapping, and (3) QTL mapped the parameters individually. They isolated significant QTLs in step (2), but not in step (3). Yin *et al.* (1999a, 1999b) associated QTLs with specific parameters in a barley CS model. Although the QTL were stable for flowering date, they were not so for specific leaf area (SLA) which, along with carbon input to the leaf, was used to compute leaf expansion rate (LER). QTL stability of the pre-flowering interval (Yin *et al.* 1999b) aligns with the thermal behaviour of the multi-genotype GN flowering time model of Welch *et al.* (2003). By re-dissecting the LER trait in terms of vapour pressure deficit, soil water potential, and a linear temperature factor, Reymond *et al.* (2003) were able to achieve stable QTLs for the coefficients of a simple ecophysiological model. When stable QTLs are obtained, they can be used in combination with the model to predict the behaviour of novel genotypes (Yin *et al.* 2000; Reymond *et al.* 2003). As discussed by Tardieu (2003) and Tardieu *et al.* (2005), this finding opens the way to the simulation of *virtual genotypes* by exploitation of their marker-coefficient relationships.

The importance of these studies, however, may extend well beyond virtual genotypes. Previously, researchers have verified biological process submodels by statistically comparing their outputs with independent data. Although this seems a powerful test, it has not been able to choose definitively among the multiple, differently formulated models that exist for most crops. Crop modelling would thus benefit greatly from additional, formal procedures that could be routinely applied in multi-criterion verification. Contrasting those instances in the last paragraph where QTLs were found v. where they were not suggests such a test: is the model realistic enough that its constants seem to result from specific gene actions as inferred from stable QTL maps?

In this light, an interpretation of the results of Wu *et al.* (2002) would be (1) that overall patterns of rice leaf phenology can be associated with specific QTLs, (2) but genetic control is not exercised through any mechanism described directly by the particular mathematical form of the growth curve used.

Cooper *et al.* (2005) lay out a program for building GN models that includes detecting QTL interactions to identify edges in an undirected network graph. The test just proposed may offer a complementary extension. Consider a hypothesised network structure derived by any means (possibly including automated search) and parameterised against a suitable mapping population. One might expect a

positive correlation between model realism and the existence of stable QTLs for network parameters. As Cooper *et al.* (2005) point out, breeders were improving crops long before any understanding of genes existed. Thus, perfect knowledge is not a prerequisite for success in crop improvement. In this context, minimal desiderata for a GN would seem to be (1) consistency with phenotypic data (especially the results of crosses) and with any physiological or genetic features that may be known, and (2) coefficients that map stably, thus identifying markers for use in selection. Beyond this, detailed congruence with actual GN structure, however desirable, would seem to be of secondary importance, and could be allowed to develop over time.

Dynamic models like those in this paper, and in Tardieu *et al.* (2005) can also integrate directly with QTL models in a procedure by Ma *et al.* (2002) called *functional mapping*. Van Eeuwijk *et al.* (2005) discuss this method, opining that a way forwards may be an analogous integration between ecophysiological models and association mapping methods. Although functional mapping is mathematically intricate, efficient algorithms and software tools are becoming available (Zhao *et al.* 2004a; Ma *et al.* 2004). Interestingly, Zhao *et al.* (2004b) present a functional mapping example that is an apt simile for the relationship between GN models and the meta-mechanism models discussed by Tardieu *et al.* (2005). Zhao *et al.* (2004b) isolated QTLs for plant height in rice by mapping the coefficients of a growth model derived from the analysis of von Bertalanffy (1957). This model describes growth in terms of the difference between tissue production (metabolic) and destruction (catabolic) rates. As a result, when cast in differential equation form (Zhao *et al.* 2004b, eqn 1, p. 1752), the model is similar to our Eqn 6. The actual mathematical form used in mapping, however, emerged by (1) solving the differential equation, and (2) reparameterising the results (Richards 1959) to yield an algebraic equivalent. In principle, the QTL analysis could have been carried out using the original differential equation, but it was not, as a matter of convenience. By analogy, GN and meta-mechanism models are related in ways that, although not as close as a mathematical derivation, will become closer in time (e.g. the previous CSDL example). Coefficient mapping of both model types will be important for verification during model development and for the subsequent creation of virtual genotypes, but which form is used in any one instance will be determined by convenience.

Whatever quantitative approaches are used within process submodels, variety development entails understanding how individual processes integrate to produce phenotypes in particular environments. Providing this integration is the role of CS models. Such models are the nearest things to 'virtual plants' (*sensu* Salk Institute 2000) currently in existence, incorporating physiological process submodels for crop phenological development, dry matter production,

and biomass partitioning among plant tissues, including economic yield (usually grain). CS modellers are seeking ways to adapt and exploit these models in the genomic era (White *et al.* 2004). The results in Hammer *et al.* (2005) provide examples of ways to enhance breeding strategies enabled by molecular genetics.

It would be helpful if existing CS models could be screened rapidly to identify current submodels that are problematic in terms of the QTL-based validity test proposed above. Mathematically, complete CS models are just functions that are not intrinsically different from the simple growth curves and meta-mechanisms whose coefficients have been mapped previously. The barrier is merely the enormous amount of computation that screening would require. Welch *et al.* (2000) reported a method for estimating CS model coefficients for large numbers of varieties simultaneously that requires far less effort than if equal numbers of varieties are processed one-by-one. The approach should be adaptable to CS/QTL mapping given the large number of line  $\times$  marker  $\times$  environment combinations. The practical constraints would be that (1) submodels (i.e. 5- to 6-member parameter sets) would have to be screened individually, and (2) accurate environmental data would be required, especially for soils (Welch *et al.* 2002).

As research on the GP problem progresses, GN, QG, and CS models will become increasingly more realistic individually, will synergise with each other, and will support a wider range of applications. This section has presented a unifying framework in which models containing different mathematical forms (GN, meta-mechanisms, and full CS models) are developed or improved; verified in ways that include genome-aware, QG-based tests that produce useful QTL as byproducts; integrated into updated CS models; and used to support on-going breeding programs in the evaluation of virtual genotypes and alternative selection strategies.

### Final words

In 2001, an NSF reviewer told the authors that models like those above are simplistic and 'unlikely to have relevance to real genetic systems', a view oblivious to the 80 years of utility demonstrated by much simpler QG models (Fisher 1918; Wright 1921a, 1921b, 1921c, 1921d). Perhaps because most of differential equations lack closed-form solutions, little attention has been given to GN models in which some analysis is feasible. Yet, in ecology, a domain whose complexity rivals genomics, simple mathematical models have generated insight and practical uses for nearly as long (Lotka 1925; Kot 2001). Lacking computers, early ecologists were compelled to develop quantitative insights by other means. These insights now guide ecologists even as they use computers, today of ubiquitous importance. Bioinformaticists currently seek to develop software that can extract meaning from large masses of diverse genomic data.

Perhaps by studying models that have some realism but great simplicity, we can find ways to better guide the machines in their efforts.

### Acknowledgments

An earlier version of this work is in the published Proceedings of the 4th International Crop Science Congress (26 September–4 October 2004, Brisbane, Australia). The Congress organisers kindly gave their permission to publish this updated version, for which we are grateful. We also benefited from the inputs of Mark Cooper and Graeme Hammer; open discussions with other authors in this Special Issue, especially Francois Tardieu; and comments from anonymous reviewers. This work was supported in part by NSF Projects 32115 and 0425759, USDA Project 2003–35304–13217, and Hatch Project KAES 0507, all at Kansas State University. This is Contribution No. 05-191-J from the Kansas Agricultural Experiment Station.

### References

- Akutsu T, Miyano S, Kuhara S (1999) Identification of genetic networks from a small number of gene expression patterns under the boolean network model. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 4, pp. 17–28. (World Publishing Co.: Singapore)
- Akutsu T, Miyano S, Kuhara S (2000) Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* **16**, 727–734. doi: 10.1093/bioinformatics/16.8.727
- Alabadi D, Oyama T, Yanovsky M, Harmon F, Mas P, Kay S (2001) Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. *Science* **293**, 880–883. doi: 10.1126/science.1061320
- Auckerman M, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *The Plant Cell* **15**, 2730–2741. doi: 10.1105/tpc.016238
- Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nature Genetics* **36**, 162–166. doi: 10.1038/ng1295
- Bagnall D, King R, Whitelam G, Boylan M, Wagner D, Quail P (1995) Flowering responses to altered expression of phytochrome in mutants and transgenic lines of *Arabidopsis thaliana* (L.) Heynh. *Plant Physiology* **108**, 1495–1503. doi: 10.1104/pp.108.4.1495
- Baldi P, Hatfield G (2002) 'DNA microarrays and gene expression.' (Cambridge University Press: Cambridge, UK)
- Barash Y, Friedman N (2001) Context specific bayesian clustering for gene expression data. In 'Proceedings of the 5th Annual International Conference on Computational Molecular Biology RECOMB-2001'. 2–11. ACM.SIGACT, New York.
- Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C (2004) Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* **427**, 164–167. doi: 10.1038/nature02269
- Blazquez M (2000) Flower development pathways. *Journal of Cell Science* **113**, 3547–3548.
- Blazquez MA, Ahn J, Weigel D (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nature Genetics* **33**, 168–171. doi: 10.1038/ng1085
- Blazquez M, Green T, Nilsson O, Sussman M, Weigel D (1998) Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. *The Plant Cell* **10**, 791–800. doi: 10.1105/tpc.10.5.791
- Blazquez M, Koornneef M, Putterill J (2001) Flowering on time: genes that regulate the floral transition. *EMBO Reports* **2**, 1078–1082. doi: 10.1093/embo-reports/kve254
- Blazquez M, Trenor M, Weigel D (2002) Independent control of gibberellin biosynthesis and flowering time by the circadian clock in *Arabidopsis*. *Plant Physiology* **130**, 1770–1775. doi: 10.1104/pp.007625
- Blazquez M, Weigel D (1999) Independent regulation of flowering by phytochrome B and gibberellins in *Arabidopsis*. *Plant Physiology* **120**, 1025–1032. doi: 10.1104/pp.120.4.1025
- Bolouri H, Davidson EH (2003) Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9371–9376. doi: 10.1073/pnas.1533293100
- Boote K, Jones J, Batchelor W, Nafziger E, Myers O (2003) Genetic coefficients in the CROPGRO-Soybean model: links to field performance and genomics. *Agronomy Journal* **95**, 32–51.
- Boss PK, Bastow RM, Mylne JS, Dean C (2004) Multiple pathways in the decision to flower: enabling, promoting, and resetting. *The Plant Cell* **16**, S18–S31. doi: 10.1105/tpc.015958
- Bowman J, Alvarez J, Weigel D, Meyerowitz E, Smyth D (1993) Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119**, 721–743.
- Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in *Arabidopsis*. *Science* **275**, 80–83. doi: 10.1126/science.275.5296.80
- Bunning E (1936) Die endonome tagesrhythmik als grundlage der photoperiodischen reaktion. *Berichte der Deutschen Botanischen Gesellschaft* **54**, 590–607.
- Burd C, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621.
- Campos H, Cooper M, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. *Field Crops Research* **90**, 19–34. doi: 10.1016/j.fcr.2004.07.003
- Carre I (2001) Daylength perception and the photoperiodic regulation of flowering in *Arabidopsis*. *Journal of Biological Rhythms* **16**, 415–423. doi: 10.1177/074873001129002006
- Carrington J, Ambros V (2003) Role of microRNAs in plant and animal development. *Science* **301**, 336–338. doi: 10.1126/science.1085242
- Chandler J, Wilson A, Dean C (1996) *Arabidopsis* mutants showing an altered response to vernalization. *The Plant Journal* **10**, 637–644. doi: 10.1046/j.1365-313X.1996.10040637.x
- Chen T, He H, Church G (1999) Modeling gene expressions with differential equations. In 'Proceedings of Pacific Symposium on Biocomputing'. Vol 4, pp. 17–28. (World Publishing Co.: Singapore)
- Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* **139**, 1455–1461.
- Chou M, Haung M, Yang C (2001) *EMF* genes interact with late-flowering genes in regulating floral initiation genes during shoot development in *Arabidopsis thaliana*. *Plant and Cell Physiology* **42**, 499–507. doi: 10.1093/pcp/pce062
- Comai L, Harada JJ (1990) Transcriptional activities in dry seed nuclei indicate the timing of the transition from embryogeny to germination. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2671–2674.
- Cooper M, Chapman S, Podlich DW, Hammer GL (2002) The GP problem: quantifying gene-to-phenotype relationships. *In Silico Biology* **2**, 151–164.
- Cooper M, Podlich DW, Smith OS (2005) Gene-to-phenotype models and complex trait genetics. *Australian Journal of Agricultural Research* **56**, 895–918.
- CRC (1996) 'Standard mathematical tables and formula.' (CRC Press: Boca Raton, FL)

- D'Haeseleer P, Wen X, Fuhrman S, Somogyi R (1999) Linear modeling of mRNA expression levels during CNS development and injury. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 4, pp. 41–52. (World Publishing Co.: Singapore)
- Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, et al. (2002) A genomic regulatory network for development. *Science* **295**, 1669–1678. doi: 10.1126/science.1069883
- Davis S (2002) Photoperiodism: the coincidental perception of the season. *Current Biology* **12**, R841–R843. doi: 10.1016/S0960-9822(02)01348-9
- Devlin P (2002) Signs of the time: environmental input to the circadian clock. *Journal of Experimental Botany* **53**, 1535–1550. doi: 10.1093/jxb/erf024
- Dong Z (2003) Incorporation of genomic information into the simulation of flowering time in *Arabidopsis thaliana*. PhD dissertation, Kansas State University, USA.
- Dubnau J, Struhl G (1996) RNA recognition and translational regulation by a homeodomain protein. *Nature* **379**, 694–699. doi: 10.1038/379694a0
- Erikssohn M, Millar A (2003) The circadian clock. A plant's best friend in a spinning world. *Plant Physiology* **132**, 732–738. doi: 10.1104/pp.103.022343
- Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *The Plant Cell* **16**, 2923–2939. doi: 10.1105/tpc.104.021725
- Ferrandiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* **127**, 725–734.
- Fisher R (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society Edinburgh* **52**, 399–433.
- Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J (1999) *GIGANTEA*: a circadian clock-controlled gene that regulated photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *The EMBO Journal* **18**, 4679–4688. doi: 10.1093/emboj/18.17.4679
- Frank S (1999) Population and quantitative genetics of regulatory networks. *Journal of Theoretical Biology* **197**, 281–294. doi: 10.1006/jtbi.1998.0872
- Friedman N, Linial M, Nachman I, Pe'er D (2000) Using Bayesian networks to analyze expression data. *Journal of Computational Biology* **7**, 601–620. doi: 10.1089/106652700750050961
- Gendall A, Levy Y, Wilson A, Dean C (2001) The *VERNALIZATION2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525–535. doi: 10.1016/S0092-8674(01)00573-6
- Goff SA, Ricke D, Lan T, Presting G, Wang R, et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100. doi: 10.1126/science.1068275
- Goss P, Peccoud J (1999) Analysis of the stabilizing effect of Rom on the genetic network controlling ColE1 plasmid replication. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 4, pp. 65–76. (World Publishing Co.: Singapore)
- Goto N, Kumagai T, Koornneef M (1991) Flowering responses to lightbreaks in photomorphogenic mutants of *Arabidopsis thaliana*, a long-day plant. *Physiologia Plantarum* **83**, 209–215. doi: 10.1034/j.1399-3054.1991.830202.x
- Guo H, Yang H, Mocker T, Lin C (1998) Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279**, 1360–1363. doi: 10.1126/science.279.5355.1360
- Halliday K, Salter M, Thingnaes E, Whitelam G (2003) Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator *FT*. *The Plant Journal* **33**, 875–885. doi: 10.1046/j.1365-3113X.2003.01674.x
- Hammer GL, Chapman S, van Oosterom E, Podlich DW (2005) Trait physiology and crop modelling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research* **56**, 947–960.
- Hammes GG (2000) 'Thermodynamics and kinetics for the biological sciences.' (Wiley-Interscience: New York)
- Hanks J, Ritchie J (Eds) (1991) 'Modeling plant and soil systems.' Agronomy Monographs 31. (ASA, CSSA, and SSSA: Madison, WI)
- Hartemink A, Gifford D, Jaakkola T, Young R (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 6, pp. 422–433. (World Publishing Co.: Singapore)
- Haughn G, Schultz E, Martinez-Zapater T (1995) The regulation of flowering in *Arabidopsis thaliana*: meristems, morphogenesis, and mutants. *Canadian Journal of Botany* **73**, 959–981.
- Hayama R, Coupland G (2003) Shedding light on the circadian clock and the photoperiodic control of flowering. *Current Opinion in Plant Biology* **6**, 13–19. doi: 10.1016/S1369-5266(02)00011-0
- He Y, Michaels SD, Amasino RM (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science* **302**, 1751–1754. doi: 10.1126/science.1091109
- Hempel F, Weigel D, Mandel M, Ditta G, Zambryski P, Feldman L, Yanofsky M (1997) Floral determination and expression of floral regulatory genes in *Arabidopsis*. *Development* **124**, 3845–3853.
- Hepworth S, Valverde F, Ravenscroft D, Mouradov A, Coupland G (2002) Antagonistic regulation of flowering-time gene *SOCI* by *CONSTANS* and *FLC* via separate promoter motifs. *The EMBO Journal* **21**, 4327–4337. doi: 10.1093/emboj/cdf432
- Hudson M (2000) The genetics of phytochrome signaling in *Arabidopsis*. *Seminars in Cell and Developmental Biology* **11**, 475–483. doi: 10.1006/scdb.2000.0204
- Huq E, Tepperman J, Quail P (2000) *GIGANTEA* is a nuclear protein involved in phytochrome signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9789–9794. doi: 10.1073/pnas.170283997
- Ideker T, Thorsson V, Karp R (2000) Discovery of regulatory interactions through perturbation: inference and experimental design. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 5, pp. 302–313. (World Publishing Co.: Singapore)
- Irmak A, Jones J, Mavromatis T, Welch S, Boote K, Wilkerson G (2000) Evaluating methods for simulating soybean cultivar responses using cross validation. *Agronomy Journal* **92**, 1140–1149.
- Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347. doi: 10.1126/science.290.5490.344
- Johnson C (2001) Endogenous timekeepers in photosynthetic organisms. *Annual Review of Physiology* **63**, 695–728. doi: 10.1146/annurev.physiol.63.1.695
- Johnson E, Bradley M, Harberd NP, Whitelam GC (1994) Photoresponses of light-grown *phyA* mutants of *Arabidopsis*. *Plant Physiology* **105**, 141–149. doi: 10.1104/pp.105.3.1023
- Kardailsky I, Shakla V, Ahn J, Dagenais N, Christensen S, Nguyen J, Chory J, Harrison M, Weigel D (1999) Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965. doi: 10.1126/science.286.5446.1962

- Kauffman S (1993) 'The origins of order: self-organization and selection in evolution.' (Oxford University Press: Oxford, UK)
- Kauffman S, Peterson C, Samuelsson B, Troein C (2003) Random Boolean network models and the yeast transcriptional network. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14796–14799. doi: 10.1073/pnas.2036429100
- Kauffman S, Peterson C, Samuelsson B, Troein C (2004) Genetic networks with canalizing Boolean rules are always stable. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17102–17107. doi: 10.1073/pnas.0407783101
- Kenzior AL, Folk WR (1998) AtMSI4 and RbAp48 WD-40 repeat proteins bind metal ions. *FEBS Letters* **440**, 425–429. doi: 10.1016/S0014-5793(98)01500-2
- Kingston R, Narlikar G (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes and Development* **13**, 2339–2352. doi: 10.1101/gad.13.18.2339
- Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962. doi: 10.1126/science.286.5446.1960
- Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araaki T, Yano M (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant and Cell Physiology* **43**, 1096–1105. doi: 10.1093/pcp/pcf156
- Koornneef M, Hanhart C, van der Veen J (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Molecular and General Genetics* **229**, 57–66.
- Kot M (2001) 'Elements of mathematical ecology.' (Cambridge University Press: Cambridge, UK)
- Kurata H, Taira K (2000) Two-phase partition method for simulating a biological system at extremely high speed. *General Bioinformatics* **11**, 185–195.
- Lamb R, Hill T, Tan Q, Irish V (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. *Development* **129**, 2079–2086.
- Lee I, Aukerman M, Gore S, Lohman K, Michaels S, Weaver L, John M, Feldmann K, Amasino R (1994) Isolation of *LUMINIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. *The Plant Cell* **6**, 75–83. doi: 10.1105/tpc.6.1.75
- Lee S, Cheng H, King K, Wang W, He Y, Hussain A, Lo J, Harberd N, Peng J (2002) Gibberellin regulates *Arabidopsis* seed germination via *RGL2*, a *GAI/RGA*-like gene whose expression is up-regulated following imbibition. *Genes and Development* **16**, 646–658. doi: 10.1101/gad.969002
- Levy Y, Dean C (1998) The transition to flowering. *The Plant Cell* **10**, 1973–1989. doi: 10.1105/tpc.10.12.1973
- Levy Y, Mesnage S, Mynne J, Gendall R, Dean C (2002) Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. *Science* **297**, 243–247. doi: 10.1126/science.1072147
- Liang S, Fuhrman S, Somogyi R (1998) REVEAL: A general reverse engineering algorithm for inference of genetic network architecture. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 3, pp. 18–29. (World Publishing Co.: Singapore)
- Liljegren SJ, Gustafson-Brown C, Pinyopich A, Ditta G, Yanofsky M (1999) Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *The Plant Cell* **11**, 1007–1018. doi: 10.1105/tpc.11.6.1007
- Lin C (2000) Photoreceptors and regulation of flowering time. *Plant Physiology* **123**, 39–50. doi: 10.1104/pp.123.1.39
- Lotka A (1925) 'Elements of physical biology.' (Williams and Wilkins Co.: Baltimore, MD)
- Lynch M, Walsh B (1998) 'Genetics and analysis of quantitative traits.' (Sinauer Associates, Inc.: Sunderland, MA)
- Ma CX, Casella G, Wu RL (2002) Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* **161**, 1751–1762.
- Ma CX, Wu R, Casella G (2004) FunMap: functional mapping of complex traits. *Bioinformatics* **20**, 1808–1811. doi: 10.1093/bioinformatics/bth156
- Macknight R, Bancroft I, Page T, Lister C, Schmidt R, et al. (1997) *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* **89**, 737–745. doi: 10.1016/S0092-8674(00)80256-1
- Macknight R, Duroux M, Laurie R, Dijkwel P, Simpson G, Dean C (2002) Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter *FCA*. *The Plant Cell* **14**, 877–888. doi: 10.1105/tpc.010456
- Maki Y, Tominaga D, Okamoto M, Watanabe S, Eguchi Y (2001) Development of a system for the inference of large-scale genetic networks. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 6, pp. 446–458. (World Publishing Co.: Singapore)
- Makino S, Kiba T, Imamura A, Hanaki N, Nakamura A, Taniguchi M, Ueguchi C, Sugiyama T, Mizuno T (2000) Genes encoding pseudo-response regulators: insight into *His-to-Asp* phosphorelay and circadian rhythm in *Arabidopsis thaliana*. *Plant and Cell Physiology* **41**, 791–803.
- Makino S, Matsushika A, Kojima M, Oda Y, Mizuno T (2001) Light response of the circadian waves of the *APRR1/TOC1* quintet: When does the quintet start singing rhythmically in *Arabidopsis*? *Plant and Cell Physiology* **42**, 334–339. doi: 10.1093/pcp/pce036
- Makino S, Matsushika A, Kojima M, Yamashino T, Mizuno T (2002) The *APRR1/TOC1* quintet implicated in circadian rhythms of *Arabidopsis thaliana*: I. characterization with *APRR1*-overexpressing plants. *Plant and Cell Physiology* **43**, 58–69. doi: 10.1093/pcp/pcf005
- Alejandra Mandel M, Gustafson-Brown C, Savidge B, Yanofsky M (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**, 273–277. doi: 10.1038/360273a0
- Marnellos G, Deblandre G, Mjolsness E, Kintner C (2000) Delta-notch lateral inhibitory patterning in the emergence of ciliated cells in *Xenopus*: experimental observations and a gene network model. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 5, pp. 326–337. (World Publishing Co.: Singapore)
- Martínez-Zapater JM, Coupland G, Dean C, Koornneef M (1994) The transition to flowering in *Arabidopsis*. In '*Arabidopsis*'. (Eds E Meyerowitz, C Somerville) pp. 403–433. (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY)
- Martínez-Zapater JM, Jarillo J, Cruz-Alvarez M, Roldán M, Salinas J (1995) *Arabidopsis* late-flowering *five* mutants are affected in both vegetative and reproductive development. *The Plant Journal* **7**, 543–551. doi: 10.1046/j.1365-313X.1995.7040543.x
- Martínez-Zapater JM, Somerville C (1990) Effect of light quality and vernalization on late-flowering mutants of *Arabidopsis thaliana*. *Plant Physiology* **92**, 770–776.
- Matsuno H, Doi A, Nagasaki M, Miyano S (2000) Hybrid Petri net representation of gene regulatory network. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 5, pp. 338–349. (World Publishing Co.: Singapore)
- Matsushika A, Imamura A, Yamashino T, Mizuno T (2002a) Aberrant expression of the light-inducible and circadian-regulated *APRR9* gene belonging to the circadian-associated *APRR1/TOC1* quintet results in the phenotype of early flowering in *Arabidopsis thaliana*. *Plant and Cell Physiology* **43**, 833–843. doi: 10.1093/pcp/pcf118

- Matsushika A, Makino S, Kojima M, Mizuno T (2000) Circadian waves of expression of the *APRR1/TOC1* family of pseudo-response regulators in *Arabidopsis*. *Plant and Cell Physiology* **41**, 1002–1012. doi: 10.1093/pcp/pcd043
- Matsushika A, Makino S, Kojima M, Yamashino T, Mizuno T (2002b) The *APRR1/TOC1* quintet implicated in circadian rhythms of *Arabidopsis thaliana*: II. Characterization with *CCA1*-overexpressing plants. *Plant and Cell Physiology* **43**, 118–122. doi: 10.1093/pcp/pcf006
- McClung C (2001) Circadian rhythms in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**, 139–162. doi: 10.1146/annurev.arplant.52.1.139
- McCown RL, Hochman Z, Carberry P (2002) Probing the enigma of decision support systems for farmers: learning from experience and from theory. *Agricultural Systems* **74**, 1–10. doi: 10.1016/S0308-521X(02)00021-5
- Mendoza L, Alvarez-Buylla E (1998) Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. *Journal of Theoretical Biology* **193**, 307–319. doi: 10.1006/jtbi.1998.0701
- Mendoza L, Alvarez-Buylla E (2000) Genetic regulation of root hair development in *Arabidopsis thaliana*: a network model. *Journal of Theoretical Biology* **204**, 311–326. doi: 10.1006/jtbi.2000.2014
- Michaels S, Amasino R (1999) *FLOWERING LOCUS C* encodes a novel *MADS* domain protein that acts as a repressor of flowering. *The Plant Cell* **11**, 949–956. doi: 10.1105/tpc.11.5.949
- Michaels S, He Y, Scortecci K, Amasino R (2003) Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10102–10107. doi: 10.1073/pnas.1531467100
- Millar A (1999) Biological clocks in *Arabidopsis thaliana*. *New Phytologist* **141**, 175–197. doi: 10.1046/j.1469-8137.1999.00349.x
- Mockler T, Guo H, Yang H, Duong H, Lin C (1999) Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. *Development* **126**, 2073–2082.
- Moon J, Suh SS, Lee H, Choi KR, Hong C, Paek NC, Kim SG, Lee I (2003) The *SOC1 MADS*-box gene integrates vernalization and gibberellin signals for flowering in *Arabidopsis*. *The Plant Journal* **35**, 613–623. doi: 10.1046/j.1365-3113X.2003.01833.x
- Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: integrating pathways as a basis for diversity. *The Plant Cell* **14**, S111–S130.
- Murakami-Kojima M, Nakamichi N, Yamashino T, Mizuno T (2002) The *APRR3* component of the clock-associated *APRR1/TOC1* quintet is phosphorylated by a novel protein kinase belonging to the WNK family, the gene for which is also transcribed rhythmically in *Arabidopsis thaliana*. *Plant and Cell Physiology* **43**, 675–683. doi: 10.1093/pcp/pcf084
- Ng M, Yanofsky M (2001) Activation of the *Arabidopsis* B class homeotic genes by *APETALA1*. *The Plant Cell* **13**, 739–753. doi: 10.1105/tpc.13.4.739
- Omholt SW, Plahte E, Oyehaug L, Xiang K (2000) Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* **155**, 969–980.
- Parcy F, Bernablies K, Weigel D (2002) Interaction of *LEAFY*, *AGAMOUS* and *TERMINAL FLOWER1* in maintaining floral meristem identity in *Arabidopsis*. *Development* **129**, 2519–2527.
- Parcy F, Nilsson O, Busch MA, Lee I, Weigel D (1998) A genetic framework for floral patterning. *Nature* **395**, 561–566. doi: 10.1038/26903
- Park D, Somers D, Kim Y, Choy Y, Lim H, Soh M, Kim H, Kay S, Nam H (1999) Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis GIGANTEA* gene. *Science* **285**, 1579–1582. doi: 10.1126/science.285.5433.1579
- Park W, Li J, Song R, Messing J, Chen X (2002) *CARPEL FACTORY*, a *Dicer* homolog, and *HEN1*, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology* **12**, 1484–1495. doi: 10.1016/S0960-9822(02)01017-5
- Peng J, Carol P, Richards D, King K, Cowling R, Murphy G, Harberd N (1997) The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes and Development* **113**, 194–207.
- Pittendrigh C (1972) Circadian surfaces and the diversity of possible roles of circadian organization in photoperiodic induction. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 2734–2737.
- Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857. doi: 10.1016/0092-8674(95)90288-0
- Quesada V, Macknight R, Dean C, Simpson G (2003) Autoregulation of *FCA* pre-mRNA processing controls *Arabidopsis* flowering time. *The EMBO Journal* **22**, 3142–3152. doi: 10.1093/emboj/cdg305
- Ratcliffe OJ, Bradley DJ, Coen ES (1999) Separation of shoot and floral identity in *Arabidopsis*. *Development* **126**, 1109–1120.
- Reed J, Nagpal P, Poole D, Furuya M, Chory J (1993) Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *The Plant Cell* **5**, 147–157. doi: 10.1105/tpc.5.2.147
- Reeves P, Coupland G (2000) Response of plant development to environment: control of flowering by daylength and temperature. *Current Opinion in Plant Biology* **3**, 37–42. doi: 10.1016/S1369-5266(99)00041-2
- Reeves P, Coupland G (2001) Analysis of flowering time control in *Arabidopsis* by comparison of double and triple mutants. *Plant Physiology* **126**, 1085–1091. doi: 10.1104/pp.126.3.1085
- Reinitz J, Sharp D (1995) Mechanism of eve stripe formation. *Mechanisms of Development* **49**, 133–158. doi: 10.1016/0925-4773(94)00310-J
- Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F (2003) Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiology* **131**, 664–675. doi: 10.1104/pp.013839
- Richards FJ (1959) A flexible growth function for empirical use. *Journal of Experimental Botany* **10**, 290–300.
- Roden L, Song H, Jackson S, Morris K, Carre I (2002) Floral responses to photoperiod are correlated with the timing of rhythmic expression relative to dawn and dusk in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13313–13318. doi: 10.1073/pnas.192365599
- Rosenzweig C, Phillips J, Goldberg R, Carroll J, Hodges T (1996) Potential impacts of climate change on citrus and potato production in the US. *Agricultural Systems* **52**, 455–479. doi: 10.1016/0308-521X(95)00059-E
- Rouse DT, Sheldon C, Bagnall D, Peacock W, Dennis E (2002) *FLC*, a repressor of flowering, is regulated by genes in different inductive pathways. *The Plant Journal* **29**, 183–191. doi: 10.1046/j.0960-7412.2001.01210.x
- Ruiz-Garcia L, Madueno F, Wilkinson M, Haughn G, Salinas J, Martinez-Zapater J (1997) Different roles of flowering-time genes in the activation of floral initiation genes in *Arabidopsis*. *The Plant Cell* **9**, 1921–1934. doi: 10.1105/tpc.9.11.1921

- Salk Institute (2000) Functional genomics and the virtual plant: a blueprint for understanding how plants are built and how to improve them. NSF Workshop Report, The Arabidopsis Information Resource (TAIR).
- Samach A, Coupland G (2000) Time measurement and the control of flowering in plants. *BioEssays* **22**, 38–47. doi: 10.1002/(SICI)1521-1878(200001)22:1<38::AID-BIES8>3.0.CO;2-L
- Samach A, Gover A (2001) Photoperiodism: the consistent use of *CONSTANS*. *Current Biology* **11**, R651–R654. doi: 10.1016/S0960-9822(01)00384-0
- Samach A, Onouchi H, Gold S, Ditta S, Schwarz-Sommer Z, Yanofsky F, Coupland G (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616. doi: 10.1126/science.288.5471.1613
- Samsonova M, Serov V (1999) NetWork: An interactive interface to the tools for analysis of genetic network structure and dynamics. In 'Proceedings of Pacific Symposium on Biocomputing'. Vol 4, pp. 102–111. (World Publishing Co.: Singapore)
- Schmid M, Uhlenhaut N, Godard F, Demar M, Bressan R, Weigel D, Lohmann J (2003) Dissection of floral induction pathways using global expression analysis. *Development* **130**, 6001–6012. doi: 10.1242/dev.00842
- Schomburg F, Patton D, Meinke D, Amasino R (2001) *FPA*, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. *The Plant Cell* **13**, 1427–1436. doi: 10.1105/tpc.13.6.1427
- Sheldon C, Burn J, Perez P, Metzger J, Edwards J, Peacock J, Dennis E (1999) The *FLF MADS* box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *The Plant Cell* **11**, 445–458. doi: 10.1105/tpc.11.3.445
- Sheldon C, Rouse D, Finnegan E, Peacock W, Dennis E (2000) The molecular basis of vernalization: the central role of *FLOWERING LOCUS C (FLC)*. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3753–3758. doi: 10.1073/pnas.060023597
- Shimamoto K, Kyozuka J (2002) Rice as a model for comparative genomics of plants. *Annual Review of Plant Biology* **53**, 399–419. doi: 10.1146/annurev.arplant.53.092401.134447
- Silverstone A, Ciampaglio C, Sun T (1998) The Arabidopsis *RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *The Plant Cell* **10**, 155–169. doi: 10.1105/tpc.10.2.155
- Silverstone A, Mak P, Martinez E, Sun T (1997) The *RGA* locus encodes a negative regulator of gibberellin response in *Arabidopsis thaliana*. *Genetics* **146**, 1087–1099.
- Simon R, Igeno MI, Coupland G (1996) Activation of floral meristem identity genes in *Arabidopsis*. *Nature* **384**, 59–62. doi: 10.1038/384059a0
- Simpson GG, Dean C (2002) *Arabidopsis*, the rosetta stone of flowering time? *Science* **296**, 285–289.
- Simpson G, Dijkwel P, Quesada V, Henderson I, Dean C (2003) *FY* is an RNA 3' end-processing factor that interacts with *FCA* to control the Arabidopsis floral transition. *Cell* **113**, 777–787. doi: 10.1016/S0092-8674(03)00425-2
- Simpson G, Gendall A, Dean C (1999) When to switch to flowering. *Annual Review of Cell and Developmental Biology* **15**, 519–550. doi: 10.1146/annurev.cellbio.15.1.519
- Smith H (2000) Phytochromes and light signal perception by plants — an emerging synthesis. *Nature* **407**, 585–591. doi: 10.1038/35036500
- Somers D (1999) The physiology and molecular bases of the plant circadian clock. *Plant Physiology* **121**, 9–20. doi: 10.1104/pp.121.1.9
- Stewart D, Cober E, Bernard R (2003) Modeling genetic effects on the photothermal response of soybean phenological development. *Agronomy Journal* **95**, 65–70.
- Stormo GD, Fields DS (1998) Specificity, free energy, and information content in protein-DNA interactions. *Trends in Biochemical Sciences* **23**, 109–113. doi: 10.1016/S0968-0004(98)01187-6
- Suárez-López P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**, 1116–1120. doi: 10.1038/35074138
- Sung S, Amasino RM (2004) Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* **427**, 159–164. doi: 10.1038/nature02195
- Sung Z, Chen L, Moon YH, Lertpiriyapong K (2003) Mechanisms of floral repression in *Arabidopsis*. *Current Opinion in Plant Biology* **6**, 29–35. doi: 10.1016/S1369-5266(02)00014-6
- Szallasi Z, Liang S (1998) Modeling the normal and neoplastic cell cycle with realistic boolean genetic networks: their application for understanding carcinogenesis and assessing therapeutic strategies. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 3, pp. 54–65. (World Publishing Co.: Singapore)
- Tardieu F (2003) Virtual plants: modeling as a tool for the genomics of tolerance to water deficit. *Trends in Plant Science* **8**, 9–14. doi: 10.1016/S1360-1385(02)00008-0
- Tardieu F, Reymond M, Muller B, Granier C, Simonneau T, Sadok W, Welcker C (2005) Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. *Australian Journal of Agricultural Research* **56**, 937–946.
- Tominaga D, Okamoto M, Maki Y, Watanabe S, Eguchi Y (1999) Nonlinear numerical optimization technique based on genetic algorithm for inverse problem: towards the inference of genetic networks. In 'Proceedings of the German Conference on Bioinformatics. GCB '99'. 4–6 October 1999, Vol 4. (Eds R Giegerich, R Hofestädt, T Lengauer, W Mewes, D Schomburg, M Vingron, E Wingender) pp. 127–140. (GBF-Braunschweig and University of Bielefeld: Hanover, Germany)
- Tsuji G, Uehara G, Balas S (Eds) (1994) 'DSSAT version 3: A decision support system for agrotechnology transfer (3 volumes).' (University of Hawaii: Honolulu, HI)
- Tubiello F, Rosenzweig C, Kimball B, Pinter P Jr, Wall G, Hunsaker D, LaMorte R, Garcia R (1999) Testing CERES-wheat with free-air carbon dioxide enrichment (FACE) experiment data: CO<sub>2</sub> and water interactions. *Agronomy Journal* **91**, 247–255.
- van Eeuwijk FA, Malosetti M, Yin X, Struik PC, Stam P (2005) Statistical models for genotype by environment data: From conventional ANOVA to eco-physiological QTL models. *Australian Journal of Agricultural Research* **56**, 883–894.
- von Bertalanffy L (1957) Quantitative laws for metabolism and growth. *The Quarterly Review of Biology* **32**, 217–231. doi: 10.1086/401873
- Wagner D, Sablowski R, Meyerowitz E (1999) Transcriptional activation of *APETALAI* by *LEAFY*. *Science* **285**, 582–584. doi: 10.1126/science.285.5427.582
- Walsh B (2001) Quantitative genetics, genomics, and the future of plant breeding. In 'Quantitative genetics, genomics, and plant breeding'. (Ed. M Kang) pp. 23–32. (CAB International)
- Walsh B (2005) The struggle to exploit non-additive variation. *Australian Journal of Agricultural Research* **56**, 873–881.
- Wang H, Elston T, Mogliner A, Oster G (1998) Force generation in RNA polymerase. *Biophysical Journal* **74**, 1186–1202.
- Weaver D, Workman C, Stormo G (1999) Modeling regulatory network with weight matrices. In 'Proceedings of the Pacific Symposium on Biocomputing'. Vol 4, pp. 112–123. (World Publishing Co.: Singapore)
- Weigel D, Alvarez J, Smyth D, Yanofsky M, Meyerowitz E (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859. doi: 10.1016/0092-8674(92)90295-N

- Weigel D, Nilsson O (1995) A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377**, 495–500. doi: 10.1038/377495a0
- Weigel D, Meyerowitz EM (1993) Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**, 1723–1726.
- Welch S, Roe J, Das S, Dong Z, He R, Kirkham M (2005) Merging genomic control networks and soil-plant-atmosphere-continuum (SPAC) models. *Agricultural Systems* (In press).
- Welch SM, Roe J, Dong Z (2003) A genetic neural network model of flowering time control in *Arabidopsis thaliana*. *Agronomy Journal* **95**, 71–81.
- Welch SM, Wilkerson G, Whiting K, Sun N, Vagts T, Buol G, Mavromatis T (2002) Estimating soybean model genetic coefficients from private-sector variety performance trial data. *Transactions of the American Society of Agricultural Engineers* **45**, 1163–1175.
- Welch SM, Zhang J, Sun N, Mak TY (2000) Efficient estimation of genetic coefficients for crop models. In 'Proceedings of the Third International Symposium Systems Approaches for Agricultural Development'. CD-ROM computer file. (Ed. W Bowen) (International Potato Center (CIP): Lima, Peru)
- White J, Hoogenboom G (1996) Simulating effects of genes for physiological traits in a process-oriented crop model. *Agronomy Journal* **88**, 416–422.
- White JW, McMaster GS, Edmeades GO (Eds) (2004) Linking functional genomics and physiology for global change research. *Field Crops Research* **90**, 1–170. doi: 10.1016/j.fcr.2004.07.001
- Wolf D, Eeckman F (1998) On the relationship between genomic regulatory element organization and gene regulatory dynamics. *Journal of Theoretical Biology* **195**, 167–186. doi: 10.1006/jtbi.1998.0790
- Wright S (1921a) Correlation and causation. *Journal of Agricultural Research* **20**, 557–585.
- Wright S (1921b) Systems of mating. I. The biometric relations between parents and offspring. *Genetics* **6**, 111–123.
- Wright S (1921c) Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6**, 124–143.
- Wright S (1921d) Systems of mating. III. The biometric relations between parents and offspring. *Genetics* **6**, 144–161.
- Wu W, Zhou Y, Li W, Mao D, Chen Q (2002) Mapping of quantitative trait loci based on growth models. *Theoretical and Applied Genetics* **105**, 1043–1049. doi: 10.1007/s00122-002-1052-8
- Yano M, Kojima S, Takahashi Y, Lin H, Sasaki T (2001) Genetic control of flowering time in rice, a short-day plant. *Plant Physiology* **127**, 1425–1429. doi: 10.1104/pp.127.4.1425
- Yanovsky M, Kay S (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* **419**, 308–312. doi: 10.1038/nature00996
- Yin X, Chasalow S, Dourleijn C, Stam P, Kropff M (2000) Coupling estimated effects of QTLs for physiological traits to a crop growth model: predicting yield variation among recombinant inbred lines in barley. *Heredity* **85**, 539–549. doi: 10.1046/j.1365-2540.2000.00790.x
- Yin X, Kropff M, McLaren G, Visperas R (1995) A nonlinear model for crop development as a function of temperature. *Agricultural and Forest Meteorology* **77**, 1–16. doi: 10.1016/0168-1923(95)02236-Q
- Yin X, Kropff M, Nakagawa H, Horie T, Goudriaan J (1997) A model for photothermal responses of flowering in rice I. Model evaluation. *Field Crops Research* **51**, 201–211. doi: 10.1016/S0378-4290(96)03457-0
- Yin X, Kropff M, Stam P (1999a) The role of ecophysiological models in QTL analysis: the example of specific leaf area in barley. *Heredity* **82**, 415–421. doi: 10.1038/sj.hdy.6885030
- Yin X, Stam P, Dourleijn C, Kropff M (1999b) AFLP mapping of quantitative trait loci for yield determining physiological characters in spring barley. *Theoretical and Applied Genetics* **99**, 244–253. doi: 10.1007/s001220051230
- Yu H, Xu Y, Tan E, Kumar P (2002) AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16336–16341. doi: 10.1073/pnas.212624599
- Zhao W, Wu R, Ma C, Casella G (2004a) A fast algorithm for functional mapping of complex traits. *Genetics* **167**, 2133–2137. doi: 10.1534/genetics.103.024844
- Zhao W, Zhu J, Gallo-Meagher M, Wu R (2004b) A unified statistical model for functional mapping of environment-dependent genetic expression and genotype × environment interactions for ontogenetic development. *Genetics* **168**, 1751–1762. doi: 10.1534/genetics.104.031484

Manuscript received 9 May 2005, accepted 20 June 2005