

# Evaluating Plant Breeding Strategies by Simulating Gene Action and Dryland Environment Effects

Scott Chapman,\* Mark Cooper, Dean Podlich, and Graeme Hammer

## ABSTRACT

Functional genomics is the systematic study of genome-wide effects of gene expression on organism growth and development with the ultimate aim of understanding how networks of genes influence traits. Here, we use a dynamic biophysical cropping systems model (APSIM-Sorg) to generate a state space of genotype performance based on 15 genes controlling four adaptive traits and then search this space using a quantitative genetics model of a plant breeding program (QU-GENE) to simulate recurrent selection. Complex epistatic and gene  $\times$  environment effects were generated for yield even though gene action at the trait level had been defined as simple additive effects. Given alternative breeding strategies that restricted either the cultivar maturity type or the drought environment type, the positive (+) alleles for 15 genes associated with the four adaptive traits were accumulated at different rates over cycles of selection. While early maturing genotypes were favored in the Severe-Terminal drought environment type, late genotypes were favored in the Mild-Terminal and Midseason drought environment types. In the Severe-Terminal environment, there was an interaction of the stay-green (SG) trait with other traits: Selection for + alleles of the SG genes was delayed until + alleles for genes associated with the transpiration efficiency and osmotic adjustment traits had been fixed. Given limitations in our current understanding of trait interaction and genetic control, the results are not conclusive. However, they demonstrate how the per se complexity of gene  $\times$  gene  $\times$  environment interactions will challenge the application of genomics and marker-assisted selection in crop improvement for dryland adaptation.

WHILE THIS PAPER FOCUSES on the simulation of plant breeding programs from an understanding of gene action, it is useful to begin with a description of how connections between crop modeling, genomics (the understanding of how multiple genes function together), and plant breeding are developing. Crop simulation models have been used to integrate physiological understanding and evaluate alternative strategies of system management to account for the soil, climate, and agronomic technologies available. The principal objective of a plant-breeding program is the generation and selection of new gene combinations to create genotypes with trait performance that is superior to current genotypes, within the target population of environments (TPE) (Comstock, 1977). This objective applies equally to conventional, molecular, and combined approaches.

In quantitative genetics, computer simulation is commonly used to evaluate alternative plant breeding strategies on the basis of stochastic descriptions of gene action and interaction (e.g., Hospital et al., 1997; Podlich and Cooper, 1998). Our aim is to demonstrate how linkages between gene action and crop performance in dryland environments can be investigated by combining the biophysical response simulation of the crops to the moisture environment with the quantitative genetics simulation of plant breeding programs. A review of present research will reveal that substantial resources are being invested into the cellular and molecular basis for adaptation to dry environments while plant breeding companies and public programs continue to make advances in yield through conventional means. Exploitation of the investment in the former requires the integration of knowledge from agronomy and cellular, plant, and crop physiology as well as plant breeding and quantitative genetics. For those less familiar with the breeding and quantitative genetics, there are some useful background texts (e.g., Hallauer et al., 1988; Falconer and Mackay, 1996) in addition to Podlich and Cooper (1998) and Chapman et al. (2002a).

In combining gene sequencing, gene cloning, and plant transformation with biochemical and genome databases, scientists in private and public industry have identified and constructed genes that control relatively linear pathways like herbicide tolerance, disease resistance, and product quality (Somerville and Somerville, 1999; Mazur et al., 1999). Molecular biology is beginning to investigate the role of the other genes that relate to adaptation to the abiotic environment. For these genotype-environment systems, thousands of genes interact in complex ways to generate crop responses to the environment via mediation of responses over both short time scales (e.g., cellular response to environment shocks like frost) and long time scales (e.g., morphological growth responses of crop development and morphology). Some pathways for abiotic adaptation are comparatively straightforward, e.g., direct cellular tolerance of salt stresses (see review by Hasegawa et al., 2000). However, it will be some time (>20 yr?) before we understand how the interactions of developmental and signaling genes control yield of crops as a function of responses at the biochemical, cellular, plant, and canopy or crop levels of organization. Until then, we need to deal with adaptive traits at a more integrative level (i.e., traits observable at the plant or crop level) while beginning

S. Chapman, CSIRO Plant Industry, Long Pocket Lab., 120 Meiers Rd., Indooroopilly, 4068, QLD, Australia; M. Cooper and D. Podlich, School of Land and Food Sci., The University of Queensland, Brisbane, 4072, QLD, Australia; and G. Hammer, Agric. Prod. Syst. Res. Unit, Queensland Dep. of Primary Industries, P.O. Box 102, Toowoomba, QLD, 4350, Australia. M. Cooper and D. Podlich, current address: Pioneer Hi-Bred Int., 7300 N.W. 62nd Ave., P.O. Box 1004, Johnston, IA 50131. Received 1 May 2001. \*Corresponding author (scott.chapman@csiro.au).

**Abbreviations:** G  $\times$  E, genotype  $\times$  environment; MET, multienvironment trial; OA, osmotic adjustment; PH, flowering time; QTLs, quantitative trait loci; SG, stay-green; TE, transpiration efficiency; TPE, target population of environments.

to connect the tools and databases that are developing in all of the research areas (molecular biology, plant breeding, and plant and crop physiology) to understand both the effects of genes on pathways and how these are mediated in the responses of crops to the environment. Complexity per se has become an area of serious research (Gell-Mann, 1994), and the process of plant breeding is an example of the challenges to be faced in understanding the interactions of genes with each other and with environments.

### **Simulation of Crop Response to Environment and Gene Flow through Plant Breeding Programs**

Apart from the ability to accumulate large amounts of gene and phenotype data, another innovation of the information age in agriculture is the simulation of growth processes and gene action (see papers in this issue and Cooper et al., 2002a, 2002b; Hammer et al., 2002). Yin et al. (2003) show that models can be parameterized using quantitative trait loci (QTLs) derived from experiments of real near-isogenic lines to test alternative ideotypes in a wider range of conditions than experienced in the observed trial. Our intention is not to model biochemical pathways per se (see Giersch, 2000, for ideas on the current status of this work in plants) but rather to model the trait effects at the crop level and simulate near-isogenic lines for different combinations of traits. As knowledge improves, there is the opportunity to connect these traits to genes via simulation of the transcriptome, proteome, and biochemical pathways using models like GEPASI (Mendes and Kell, 1998).

Not all crop simulation models are suitable for use in these genetic frameworks. Hammer (1998) revisited the concept of emergent properties (de Wit and Penning de Vries, 1983). This concept suggests that modelers should attempt to define the rules that set the boundary conditions for simulation processes rather than applying a descriptive structure. The model needs to be able to handle perturbations to any process and self-correct, as do plants under hormonal control when growing in the field. This philosophy of parameterization and modeling of the principles of response and feedbacks, cf. description of response, infers that models should be able to express complex behavior of the type observed in the field, even given simple operational rules at a functional crop physiological level. The sorghum [*Sorghum bicolor* (L.) Moench] crop module (APSIM-Sorg) within the APSIM cropping systems model (McCown et al., 1996) contains several deliberate parameterizations to address genetic variation using a *boundary conditions* approach (Hammer and Muchow, 1994; Hammer et al., 1999; Chapman et al., 2002a), e.g., the model employs a switching method to estimate crop growth rate when limited by either radiation or water (Chapman et al., 1993) to utilize our ability to characterize the crop level efficiencies of radiation or water use for different genotypes.

The QU-GENE simulation platform (Podlich and

Cooper, 1998) simulates the stochastic properties of genes, genotypes, and environments in the operation of plant breeding programs. It can model breeding programs as search strategies that seek higher peaks on the adaptation landscape (genetic space) for a given genotype–environment system. Searches progress by creating, identifying, and selecting genotypes with improved adaptation to the TPE. The rate at which a population improves with selection is monitored by the change in grain yield of successive cycles and in the changes in the fixation (gene frequency) of both positive and negative alleles related to this yield improvement. Statistical analyses determine the effectiveness of searches in creating and finding superior combinations of alleles in the simulated populations. The superior methods of recombination and searching genetic space can then be considered for application in conventional plant breeding programs. The methods evaluated might include such things as different methods of recombining genotypes and different levels of selection pressure (the proportion of the population selected for recombination) as well as improved statistical interpretations of adaptation.

In using QU-GENE to define the genetic space to be searched, the actions of genes and their interactions with other genes (epistasis) and with environments (gene  $\times$  environment interactions) are prescribed for different crop traits, as are gene associations with molecular markers. As in other genetic simulation studies (e.g., Hospital et al., 1997; Van Berloo and Stam, 1998), these actions and interactions have normally been derived from field experiments as stochastic parameters (estimates of variance components and heritability) and from direct knowledge of the allelic effects of genes on traits, yield, or both. Until now, in QU-GENE, as in other genetic simulation models, there has been no direct biophysical connection between the gene effects associated with a trait and the yield phenotype of resulting genotypes as modulated by abiotic environmental influences. Establishing this direct connection by linking QU-GENE and APSIM enables direct definition of the actions of genes on traits so that epistatic and genotype  $\times$  environment ( $G \times E$ ) interactions for yield are emergent properties of the dynamics of the APSIM crop simulation model.

### **Improving the Efficiency of Plant Breeding for Dryland Environments**

Three areas suggested by Cooper et al. (2002a) in which crop modeling could assist in assessing in silico the multitude of options to improve the efficiency of plant breeding are: (i) characterizing environments to define the TPE, (ii) assessing the value of specific putative traits in improved plant types, and (iii) enhancing integration of molecular genetic technologies. Hence, plant breeders can pose questions that range from how to better utilize field performance data to how knowledge of gene action and/or function can be utilized for selection in a complex TPE. For example, the sequence of sample environments in Australia is extremely variable among locations and seasons, such that the selec-

tion pressures on germplasm are quite different to that experienced when the sample environments are constant (Podlich and Cooper, 1998), i.e., in the 2 yr of testing across six locations (typical for late-stage breeding trials), the proportion of low-stress environments sampled can range from 0 to 100% of the sample and confound the ranking of cultivars where they have differential adaptation to low- and high-stress environments (Chapman et al., 2002a). Podlich and Cooper (1998) showed that weighting the data from different environments by their expected frequency in the TPE can be used to improve the efficiency of selection. Chapman et al. (2000b) showed how models can assist in achieving Option i, which can then be used in the weighted selection approach of Podlich and Cooper (1998), while other papers in this issue also consider Option ii (Asseng et al., 2003; Yin et al., 2003) and Option iii (Yin et al., 2003).

Genomic type projects being initiated in the area of adaptation to drought or other abiotic stresses are largely focused on traits observed at the molecular and cellular levels, such as membrane stability or modified ion exchange and/or exclusion, e.g., Hasegawa et al. (2000). While these cellular traits may be essential in plant survival and, in some cases, contribute to economic yield, adaptation to variable rainfall environments is greatly mediated by traits observed at the crop level that influence the seasonal pattern and total water use of the crop, such as flowering time (PH), canopy transpiration efficiency (TE), leaf development and senescence, and repartitioning of dry matter. Richards and Belhassen (1996) have discussed examples of these effects for the adaptation of wheat (*Triticum aestivum* L.) to dryland environments. To a great extent, adaptation to drought as exploited through plant breeding has resulted from modifications of the normal process of growth and development (e.g., to change the pattern of water use), rather than the introduction of strong localized reactions of novel genes to a stress.

The objectives of this paper are to first demonstrate the simulation of yield resulting from definition of gene action for four physiological crop traits (additive effects with several genes and interacting states of expression) acting in different drought environment types. Second, we demonstrate how breeding progress is influenced by selection for yield given two common constraints experienced by plant breeders: restricted sampling of representative environments and the need to select within different maturity groups. The results are presented for long sequences ( $>10$  cycles of evaluation, selection, and intermating) of recurrent S1 selection. As shall be evident, we do not (and may never) have a complete understanding of the genetic controls and physiological interactions among the targeted traits. Our objective is not to provide all of the answers, but to demonstrate tools to begin to integrate quantitative genetics with dynamic crop simulation in the investigation of the complexity of gene  $\times$  gene  $\times$  environment interaction.

## MATERIALS AND METHODS

To undertake this study, we established a linkage between the definition of gene action by the QU-GENE software (Pod-

lich and Cooper, 1998; <http://pig.ag.uq.edu.au/qu-gene/>; verified 21 Aug. 2002), which simulates the change in the genotype population between successive steps within a plant breeding program, and the sorghum module of the APSIM cropping systems model (McCown et al., 1996; <http://www.apsru.gov.au/Products/apsim.htm>; verified 21 Aug. 2002), which simulates the effect of environment and gene action to generate crop yield (Fig. 1). The following sections describe the processes in detail, for which a summary (Fig. 1) is given here:

1. Use APSIM-Sorg to simulate a reference genotype and characterize the degree of stress and frequency of occurrence of three drought stress environment types across a sample of locations and years (Chapman et al., 2000a, 2002a).
2. For four crop traits [TE, PH, osmotic adjustment (OA), and SG], determine the yield value of all combinations of different gene expression levels for each trait (genotypes as near-isogenic lines) in all location–season combinations. For each genotype, calculate the mean yield across all of the location and year combinations that comprise each of the drought environment types. These data represent the yield genotype–environment space for the entire possible population of genotype–environment type combinations.
3. Using QU-GENE, model the processes of an example breeding program:
  - a. Initially sample the genotype–environment space where there was a low to moderate frequency (0.2) of favorable alleles for each gene to choose parents for intermating.
  - b. Evaluate S1 offspring in multienvironment trials (METs) (from a fixed or random sample of environment types) and select offspring based on mean yield.
  - c. Repeat the evaluation, selection, and intermating process for 12 cycles of S1 recurrent selection.

The results were then interpreted in terms of changes in the grain yield of the offspring and in the frequency of favorable alleles for the genes associated with each trait.

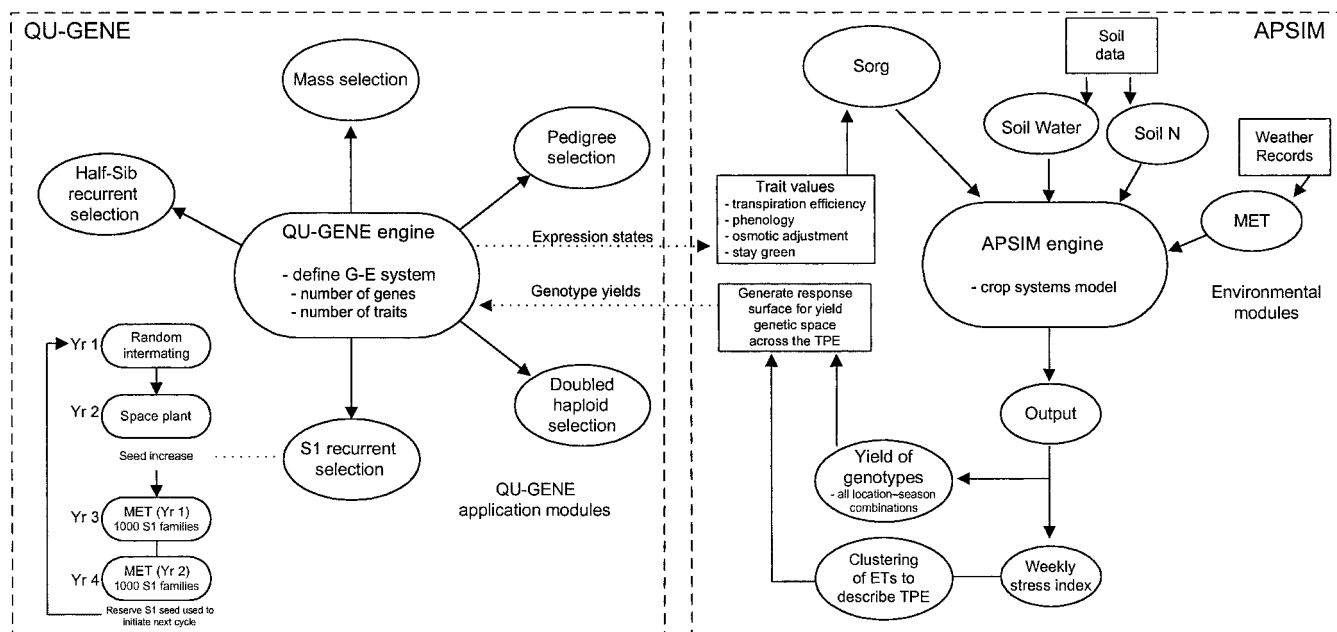
### 1. Characterizing the Target Population of Environments

Chapman et al. (2002a) have described the environments (locations–season combinations) used in this experiment and their effects on the performance of 54 genotypes with the extreme (low and high) and standard levels of gene expression for the four traits considered.

Version 1.5 of APSIM was used to run the SORG (sorghum) module using weather data on a daily time step to interact with a specified soil profile and simulate the soil and plant processes associated with water and N during fallow and in-crop states (Fig. 1). We simulated an opportunity cropping system, i.e., winter or extended fallow, followed by sorghum during a summer planting window whenever minimum rainfall (25 mm in 4 d) and soil water conditions (80 mm) were achieved. Nitrogen was nonlimiting, and the crop was grown at 50 000 plants ha<sup>-1</sup>.

A genotype with all parameters set to the standard value for each of the traits (Table 1) was run using 108 yr of daily weather data at six locations (648 potential trials) across the sorghum production region of northeastern Australia. The weather data for these locations, obtained from the SILO data set (<http://www.dnr.qld.gov.au/silo/>; verified 21 Aug. 2002), contains actual temperature and rainfall, with actual solar radiation or, prior to 1956, solar radiation estimated from functions of temperature and cloud cover. The reference genotype was used to simulate, for each trial, the final grain yields and the average level of a water stress index (water supply/





**Fig. 1.** Schematic of the modular structures and linkages between QU-GENE and APSIM used to simulate S1 recurrent selection of sorghum for adaptation to dryland environments. Several other selection strategies that can be simulated as QU-GENE application modules are indicated (e.g., pedigree selection) although the multiple crop and systems modules of APSIM are not. Gene action is defined as expression states that become trait value inputs to APSIM-Sorgh, together with soil and weather data. Output from APSIM is processed to define both the yield of all possible genotypes (expression state combinations) and the frequency of drought environment types (ETs) encountered in the target population of environments (TPE). This output comprises the genotype-environment space to which QU-GENE applied S1 recurrent selection to search for superior genotypes. MET, multienvironment trial.

demand ratio) for each successive 100 degree days (thermal-time weeks) from emergence (Chapman et al., 2000a). This is simply to provide an objective basis (i.e., compared to sets of locations and years) for the classification of different stress environments that might be encountered in the breeding program. Individual simulations of genotypes in each environment (see next section) generate independent levels and timing of stress and therefore flow-on effects to other traits.

Using hierarchical cluster analysis, each of the 547 runs for the reference genotype was classified into one of three groups of drought stress environment types (Mild-Terminal, Severe-Terminal, and Midseason) on the basis of similarities in the seasonal sequence of the stress index over thermal-time weeks. The classification process has been described by Muchow et al. (1996) and Chapman et al. (2000a) and, for the data set used in this study, by Chapman et al. (2002a). Given that the sites used are representative of >80% of the sorghum area, the frequency of occurrence of drought environment types across the entire set of simulations is taken as an estimate

of their distribution in the TPE for the Australian sorghum production region (Table 2). As shown by Chapman et al. (2002a), the frequencies of occurrence as might be experienced by a six-location breeding program that samples only two to three successive years of the weather record can differ substantially from these 108-yr frequencies.

## 2. Simulation of Gene-Environment Effects

### Specification of Trait Action

The current sorghum module (APSIM version 1.5) has undergone development to enhance its capacity to simulate the interactions among physiological processes that influence growth and yield (Hammer, 1998). The model has been validated across a range of agronomic (density, irrigation, and N rate) and genetic (varying maturity and SG) treatments in Australia and India, with predicted yields well correlated ( $r^2 = 0.89$ ,  $n = 23$ ) with observed yields from 0.5 to 8 t/ha (Hammer

**Table 1.** Values and number of additive genes (and expression states of equal size from the lower to upper range values) for each of four physiological traits. For trait OA, two model variables were modified by the same two genes; the five expression states ranged from the standard to upper values only; and the trait only operated when the crop demand for water exceeded the available water supply from the roots. See Chapman et al. (2002a) for more detail on traits.

Physiological trait	Model variable description and units	Value in allele combination			No. of additive genes (expression states)
		Lower	Standard	Upper	
Transpiration efficiency coefficient (TE)	Transpiration efficiency coefficient (MPa)	0.008	0.009	0.010	5 (11)
Flowering time (PH)	Thermal time from the end of juvenile stage to panicle initiation (°Cd)	90	115	140	3 (7)
Osmotic adjustment (OA)	Growth amount required per grain set (g grain <sup>-1</sup> )	No down-regulation from standard			2 (5)
	Fraction of stem biomass available for retranslocation (%)				
Stay-green (SG)	Target specific leaf N for new leaf (g N m <sup>-2</sup> leaf)	1.35	1.5	1.65	5 (11)

**Table 2.** From the data set of 4235 genotypes, the mean yield within each of the three drought environment types, and for each drought environment type and trait, the difference between the mean of all genotypes containing the highest expression state for the trait and the mean of all genotypes containing the lowest expression state for the trait. See Table 1 for trait definitions and expression states.

Trait	Drought environment type (frequency of occurrence across the target population of environments)		
	Severe-Terminal (35%)	Midseason (28%)	Mild-Terminal (37%)
	t/ha		
TE	0.82	0.61	0.46
PH	-0.32	0.89	1.36
OA	0.42	0.22	0.18
SG	0.10	0.49	1.03
Mean yield	2.58	3.55	4.99

et al., 2001). It has been particularly designed to account for the genetic variation for the crop traits of TE coefficient, PH, OA, and SG (Table 1). Traits are defined as heritable characteristics for which genetic variation is known to exist. As in any plant breeding context (apart from disease resistance, perhaps), they are rarely associated directly with a specific cellular-level characteristic, but rather with a network of gene controls, the effects of which can be observed by measurements on segregating populations. The mechanisms of operation of the traits and the evidence for the chosen ranges were detailed by Chapman et al. (2002a) and are outlined briefly here, with the gene expression effects described in the next section.

The TE trait and traits correlated with it (isotopic C discrimination ratio) are considered heritable as they have been selected for in plant breeding programs and backcrossed successfully into new germplasm to increase yield, e.g., wheat (Richards et al., 2002). The TE coefficient is here referred to as a trait that depends on the balance of the exchange of CO<sub>2</sub> and water vapor at the cellular level of the leaf and is expressed at the level of the crop as the crop growth rate per unit of water transpired (absorbed by roots). The actual or realized TE depends on the vapor pressure deficit (VPD) such that:  $TE_{actual} = TE/VPD$ . The value of TE has been shown in glasshouse and field trials of sorghum to vary across cultivars by approximately 10%, independent of vapor pressure deficit (Mortlock and Hammer, 1999). The effect of raising TE is to increase the efficiency with which water is utilized by the crop to meet the demand for potential (radiation limited) growth. Hence, when water supply is sufficient to meet demand, less soil water is extracted, leaving a larger late-season soil water reserve in seasons when rainfall is low. Alternatively, when water is insufficient to meet demand, more dry matter growth can be produced with that supply. Hammer et al. (1996) considered that cultivars with an increased TE also had a reduced radiation use efficiency. This effect was not implemented here as the data of Mortlock and Hammer (1999) did not support a clear negative association between these traits at the plant or crop level.

For trait PH, the thermal time requirement for the completion of the development stage end-juvenile to floral initiation was varied to simulate the genetic range of flowering dates (ca. 14 d) observed in the local germplasm pool. A longer thermal time requirement allowed more leaves to be initiated. Hence, increasing the trait PH increases the final number of leaves produced and therefore delays flowering because flowering only begins once all of the initiated leaves have appeared and expanded. To simplify this example, no modification of photoperiod response was introduced. Though pho-

toperiod is believed to be a major control of PH in sorghum (Rooney and Aydin, 1999), the genotype effect can be satisfactorily mimicked using the method outlined above as the planting date and latitude variation in our example is not extreme.

Trait OA was implemented in terms of the observed crop-level effects in near-isogenic lines that differed in the level of OA under drought (P. Snell, unpublished data as summarized by Hammer et al., 1999). The crop-level effect of OA under drought conditions was to, first, reduce the amount of crop growth in the period from panicle initiation to flowering that is required to set a given number of grains and, second, increase the potential remobilization pool for the filling of grains. Increasing the genetic value of OA only has an effect under drought (Table 1), and its effect is to increase both the number of grains (sink size) and the retranslocation potential should drought continue into grain filling (Table 1).

Genetic variation in SG was simulated by modifying the target specific leaf N (g N m<sup>-2</sup> leaf area) of new leaf. Borrell et al. (2000) reported on physiological studies of the mode of action of SG in hybrids from a cross of parents with high and low levels of SG. Increasing the target specific leaf N for new leaf allows increased N uptake during canopy development as found by Borrell et al. (2000) in SG type germplasm. Subsequently, during grain filling, depletion of N from leaves is delayed, causing the SG effect although other forms of SG are also known to exist (Thomas and Howarth, 2000).

In general, higher values for these different traits result in higher yields but not under all environmental conditions. This is particularly the case for PH where lower values result in early maturity, which may be an advantage to escape the effects of severe terminal drought. The background studies and references relating to the ranges of the traits described in Table 1 are given by Chapman et al. (2002a).

### Defining Gene Action and Expression States for Physiological Traits

In QU-GENE, we modeled the multilocus effects for the states of expression of each trait as unlinked cumulative additive alleles across loci, with two alleles per locus, e.g., for the trait OA, we described two genes that are located at two loci, which can result in the five evenly distributed levels of expression (see below). As a convention for referring to genes and alleles, we use uppercase bold letters to refer to the gene and upper- and lowercase italic letters to refer to their alleles (e.g., gene **A** with alleles *A* and *a*). For each locus (position of a gene), one allele was considered to result in increased trait expression relative to the other allele. The expression state of a trait for a genotype was then determined by the total number of alleles for increased trait expression possessed by the genotype across all loci for the trait. The uppercase alleles were defined as the alleles that increased the level of a trait and are referred to as + alleles. Conversely, the lowercase alleles were defined as the alleles that decreased trait expression and are referred to as - alleles. Importantly, the + or - designation refers to their influence on the expression of a trait but not necessarily their effect on expressed traits such as leaf area, biomass, or yield. For a trait (OA) regulated by two genes [gene **A** with alleles *A* (+) and *a* (-) and gene **B** with alleles *B* (+) and *b* (-)], there were five possible expression states for the trait, based on a genotype possessing either zero, one, two, three, or four of the + alleles across genes **A** and **B**. With this gene expression model, different genotypes can have the same expression state. For example, genotypes *Aabb*, *aAbb*, *aaBb*, and *aabB* all have one + allele and therefore have the same expression state. Similarly, for the five expression states:

1. *aabb* (zero + alleles)
2. *Aabb*, *aAbb*, *aaBb*, *aabB* (one + allele)
3. *AAbb*, *AaBb*, *Aabb*, *aABb*, *aaBB* (two + alleles)
4. *AABb*, *AABB*, *AAbB*, *ABbB*, *aABb*, *aABB* (three + alleles)
5. *AABB* (four + alleles)

For the hypothesized genetic models for the four traits (Table 1), there are  $3^{15} = 14\,348\,907$  different genotypes but a much smaller number of different expression states ( $11 \times 7 \times 5 \times 11 = 4235$ ). Alternative gene expression models could have been considered and would have created different relationships between the genotypes and phenotypes. The number of expression states (and genes) chosen for the different traits was determined either from some knowledge of gene action and recombination in breeding experiments (for PH and OA) or was suggested from the approximate number of strong molecular markers that have been found (for TE and SG). With respect to the number of gene expression states used in Table 1, we note that:

- No QTLs or genes have yet been identified as being related to TE in sorghum. We used five genes in Table 1 with the expectation that it may be reasonably complex and potentially associated with SG (see below) as suggested by Borrell et al. (2000).
- Hart et al. (2001) found three major QTLs for maturity in sorghum though breeding studies suggest that there are six major genes controlling flowering date and that four of these are common in sorghum germplasm (Rooney and Aydin, 1999).
- Basnayake et al. (1995) suggested a two-gene model for OA in the germplasm utilized by Snell (unpublished, 1999) and summarized by Hammer et al. (1999).
- Tao et al. (2000) identified five QTLs associated with SG in local sorghum germplasm.

APSIM-Sorg was used as described above to generate yields for all 4235 expression states (equivalent to genotypes for our discussion) from QU-Gene for all 547 location-season combinations (Fig. 1). These yields represent a nonlinear interaction of the expression states of each trait with each other and with the growing environments. For each genotype, the mean yield in each of the three drought environment types was calculated. This was a simplification of the simulation experiment though we intend in future studies to sample environments from the three drought environment types rather than using the genotype mean across each drought environment type.

### 3. Modeling the Breeding Program

The structure of the adaptation landscape for the genetic model was defined by the APSIM-Sorg crop model estimate of yield for different combinations of genes (genotypes) based on different expression levels of the four traits (Table 1). The QU-GENE software managed the creation, evaluation, and selection of genotypes within a breeding program (Fig. 1). The first stage of QU-GENE is the engine, which specifies the properties of the genetic models under investigation and hence the structure of the adaptation landscapes corresponding to the germplasm pool available. The engine creates a starting-point reference population of genotypes for investigation by second-stage application modules that simulate the structure of different plant breeding programs. In the configuration used here, grain yields were generated by APSIM for all of the genotypes and were averaged for each genotype over the three environment types.

A QU-GENE application module representing an S1 family recurrent selection breeding program was used to conduct the

experiment (Fig. 1). The breeding program operated on a 4-yr cycle, with the first 2 yr used for random intermating, space plant selection, and seed increase of S1 families that are the direct offspring of the randomly intermated parents (Podlich et al., 1999). The S1 families were then evaluated in a MET over five locations and in 2 yr. Superior S1 families were selected on mean yield performance in the MET, determined from the database of APSIM runs. Reserve S1 family seed from the selected families was intercrossed to initiate the next cycle. The program was conducted for 12 cycles (equivalent to 48 yr), using a spaced plant population of 5000 individuals, with 1000 S1 families evaluated in both years of the MET. The top 100 S1 families were selected on superior performance in the MET. To simplify the analysis, plot heritability of each gene was assumed to be 1.0, i.e., there was no experimental error.

The timing of flowering is an important consideration with respect to the occurrence of drought, e.g., earlier-flowering genotypes are able to escape the effects of a terminal drought. Further, to fit in with other operations in a cropping system, breeding programs frequently provide a suite of cultivars differing in maturity. For summer crops like sorghum, a farmer may prefer to plant a longer-season variety if an early planting opportunity arises and a shorter-season variety when planting late so that the crop does not mature into cool conditions and interfere with future crop rotations. To consider these practical issues, several selection scenarios were evaluated:

1. No constraints selection for average yield in the MET, with the 10 testing environments randomly sampled at their occurrence rates in the TPE (Table 2).

2. Selection for average yield in the MET, with the 10 environments sampled from only one drought environment type. This restricted sampling process, repeated for each drought environment type, demonstrates the effects on selection of sampling only one type of drought stress pattern, as can happen when breeding programs experience a series of wet or dry years, e.g., Chapman et al. (2000b), or repeatedly sample a particular drought environment type through managed irrigation or drought treatments, e.g., Edmeades et al. (1999).

3. Selection for yield in the MET as in each of the two scenarios above but with phenology constrained to be within a particular class. Three phenology constraints on the selection process were implemented by retaining in selection only those genotypes (and hence S1 families) that possessed allelic combinations for the PH trait genes that fell into one of the sets of expression states 1 to 3 (early maturity), 3 to 5 (medium maturity), or 5 to 7 (late maturity) (Table 1). This simulates the common process utilized when breeders are selecting for adaptation within different maturity classes.

Given that the genetic composition of the initial parents can vary, the performance of each of the defined selection scenarios was evaluated as the average of 2000 runs (10 independent sets of starting parents  $\times$  200 independent runs of each set of parents) over 12 cycles of S1 recurrent selection. The initial parents were selected such that the population had a fixed gene frequency of 0.2 for the + alleles of each of the 15 genes that described the four traits. At each cycle, the mean of the S1 families in the MET and the average gene frequency of the genes regulating expression in the four traits were tabulated for interpretation. The genotypic and  $G \times E$  interaction components of variance were also calculated for each cycle of selection and selection scenario.

Running the APSIM simulations was a substantial task. APSIM was installed on the QCC (QU-GENE Computer Cluster; Micallef et al., 2001), which comprises forty-eight 400 MHz (or greater) computers, reducing the simulation time to

5 d instead of 250 d on a single computer. Report files from APSIM were assembled from the computer cluster using customized scripts written in the Tcl/Tk control language and were stored in a database before production of data files for input to QU-GENE (Micallef et al., 2001). The QU-GENE simulations (see below) were completed in about 10 h using all 48 computers in the QCC.

## RESULTS

As reported by Chapman et al. (2002a), the overall frequency of the three drought environment types was similar across the 547 simulated trials though slightly lower for the Midseason drought environment type compared with the Mild- and Severe-Terminal types (Table 2). The mean yields of the genotype data set differed by 2.5 t/ha across the three drought environment types. For the trait TE, the difference in yield between the genotypes with the highest (11) and lowest (1) number of expression states (averaged over all expression states for the other traits) was greatest in the Severe-Terminal drought environment type, i.e., the advantage of TE was greater in the Severe-Terminal drought environment type than in the other types

(Table 2). The direction of this effect was similar for the trait OA but with about half the magnitude of that for trait TE. For the trait PH in the Severe-Terminal drought environment type, yields were greater for genotypes with the lowest expression states (i.e., early maturing genotypes) while in the other drought environment types, yields were greater for later-maturing genotypes with the highest expression states. The increase in the size of the difference between the high and low expression state yields was also present for the trait SG. High expression of SG was most advantageous in the Mild-Terminal drought environment type although it was positive in all types.

For each cycle of selection, the proportion of genes fixed (gene frequency) for the + alleles was averaged over the genes for each trait (Table 1) and over the 2000 QU-GENE runs (Fig. 2). When the MET was conducted using only the Severe-Terminal drought environment type, OA and TE were the first traits fixed for + alleles (Fig. 2a). Once these genes had been fixed, the SG genes, which had changed relatively little over the first four cycles, began to be fixed rapidly. The PH genes were fixed gradually to – alleles (early maturity)

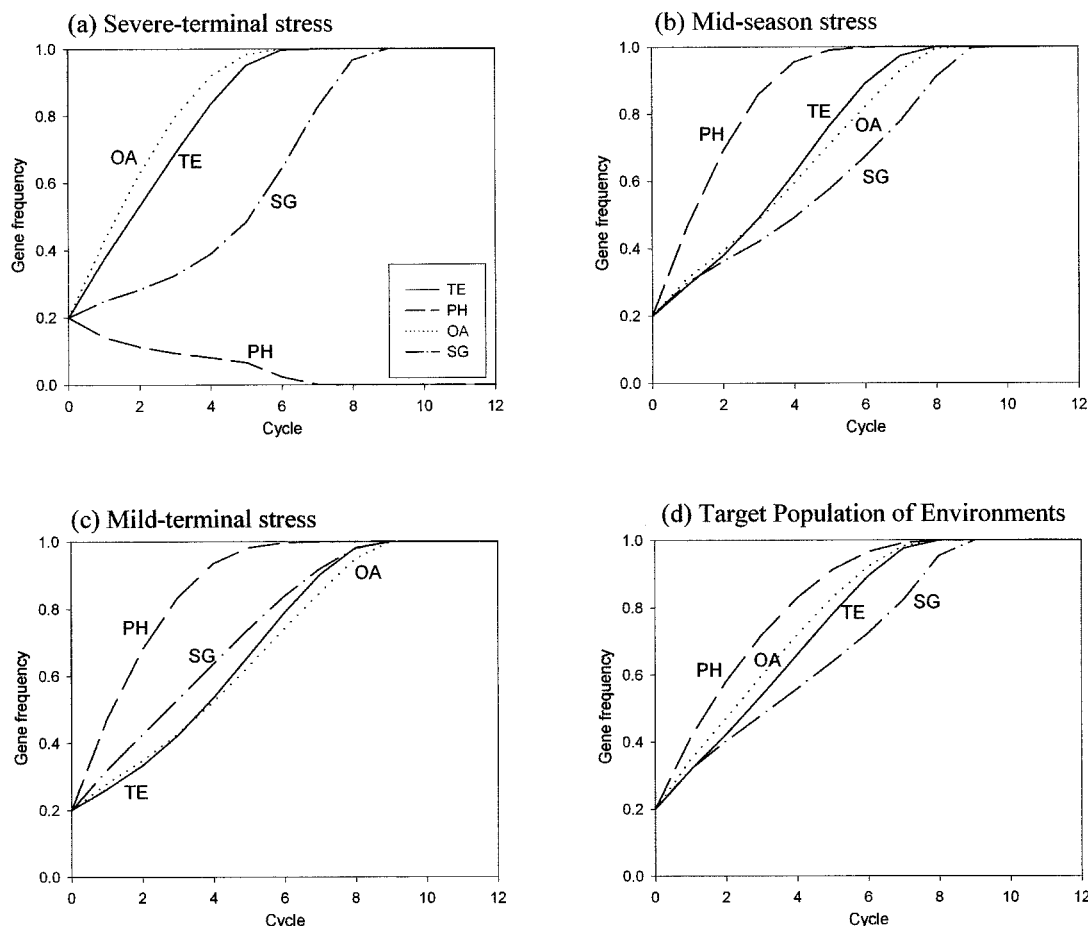


Fig. 2. For the 2000 simulations of cycles of S1 recurrent selection, mean changes in the gene frequency for + alleles associated with four physiological traits—transpiration efficiency (TE, average of five genes), flowering time (PH, three genes), osmotic adjustment (OA, two genes), and stay-green (SG, 5 genes)—given four different selection environments. The selection environments were applied as a multi-environment screen (5 locations by 2 yr) of S1 families consisting of (a, b, and c) the same environment type or (d) the target population of environments in which the three environment types (a, b, and c) were sampled in proportion to represent the target population of environments in the sorghum region of northeastern Australia (Table 2).



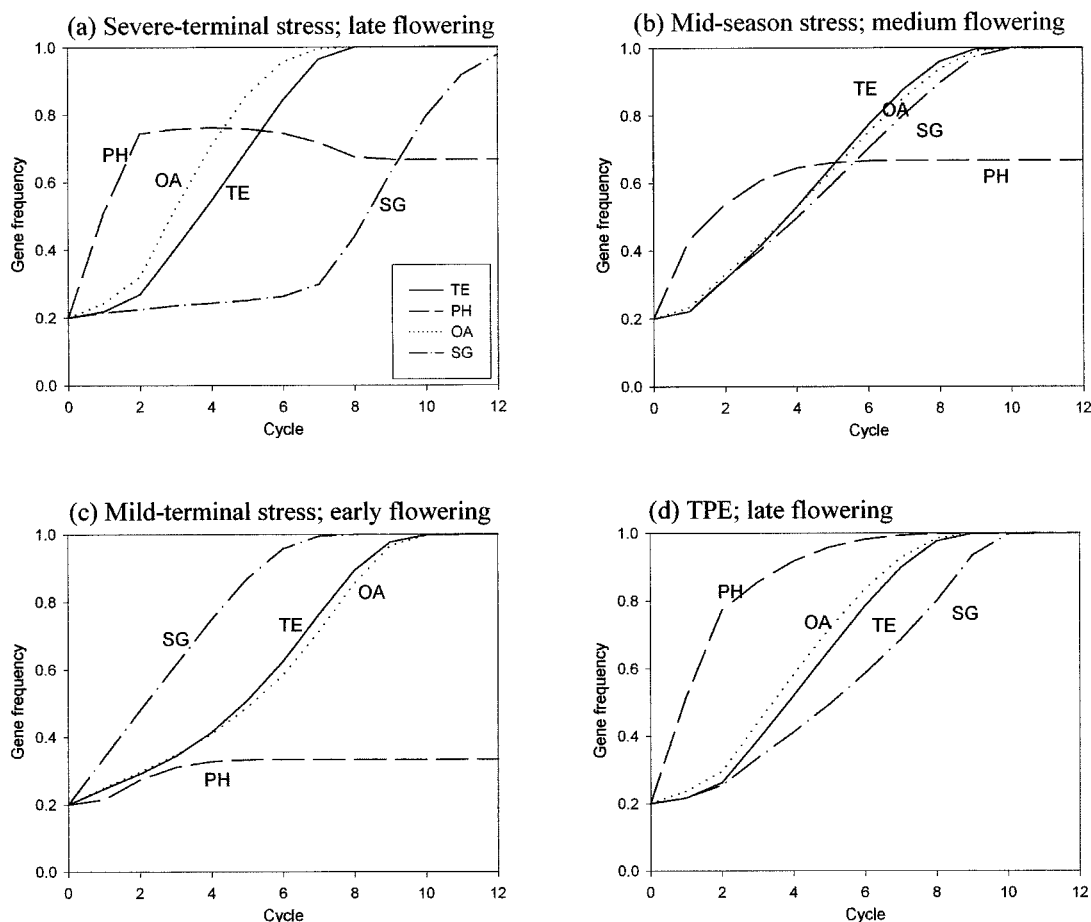


Fig. 3. For the same traits and the selection environments given in Fig. 2, mean changes in the gene frequency for + alleles with the constraint that selection was only made among genotypes within one of three maturity types (early, medium, or late). Within each maturity type, genotypes were restricted to three of seven possible expression states in trait PH (Table 1): early (expression states 1–3), intermediate (3–5), or late (5–7). Simulations (captioned in figure) are given for selection environment–constraint combinations that contrast particularly with Fig. 2 (see text).

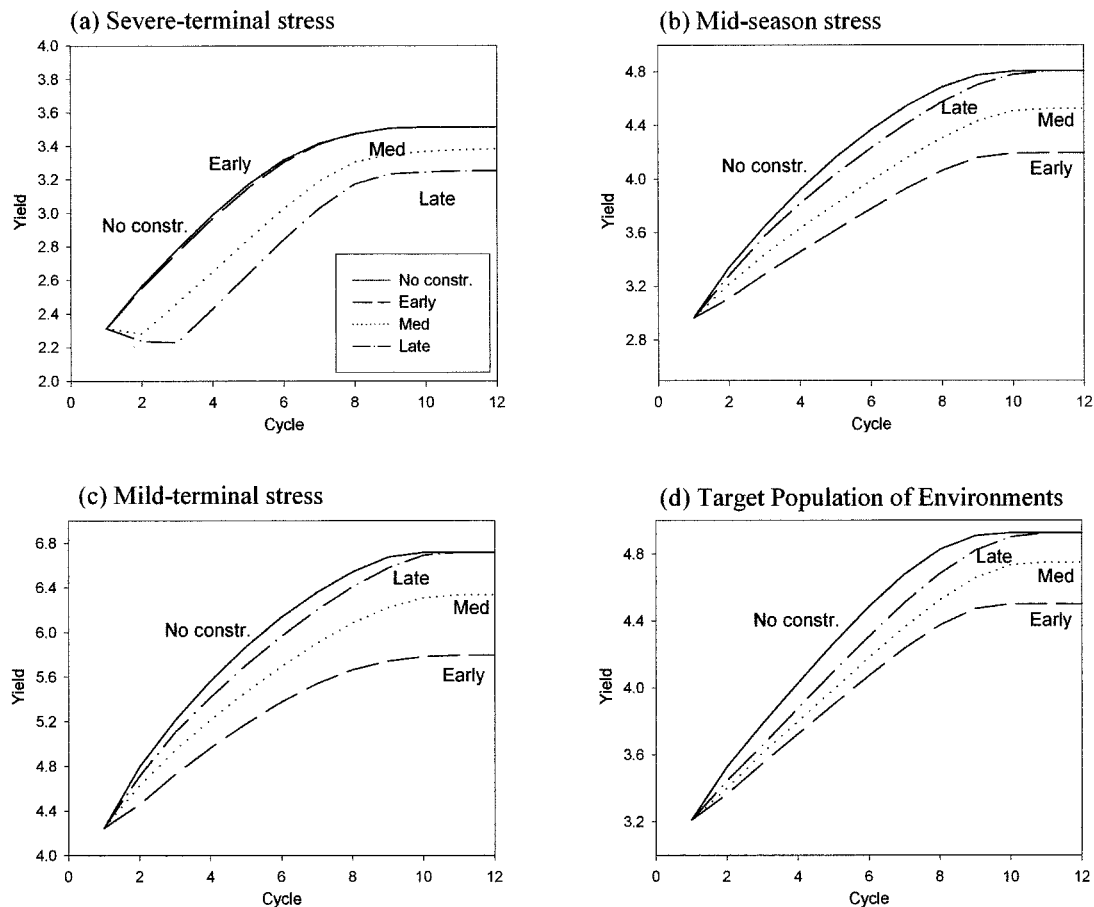
over seven cycles. By Cycle 9, all genes were fixed to either + (OA, TE, and SG) or – (PH) alleles. These effects on PH genes were in contrast to selection under environments that were solely Midseason drought environment type (Fig. 2b) or Mild-Terminal drought environment type (Fig. 2c). In these two environment types, all trait genes were fixed to + alleles, with the PH trait fixed the most quickly, followed by a constant, slower rate of fixing of the + alleles for the TE and OA genes and then SG (Midseason) or the SG and then TE and OA (Mild-Terminal). When the MET evaluation was conducted using environment types sampled in proportions to mimic the TPE (Fig. 2d), the patterns of gene fixing most closely resembled those observed when the Midseason drought environment type alone was the selection environment.

In Fig. 3, several of the more obvious contrasts to the no constraints scenario are shown. The greatest contrast of the early maturing constraint with the no constraints scenario was under selection in a Mild-Terminal stress where the + alleles for the SG genes were fixed first, followed by the PH genes and then by the TE and OA genes (Fig. 3c). The order of fixation of genes was almost reversed, cf. Fig. 2c, where maturity was not constrained.

The changes in gene frequency for selecting early genotypes in a Severe-Terminal drought environment type were similar to those for the no constraints scenario (data not shown). When selection for early genotypes was done by sampling the TPE or the Midseason drought environment types (data not shown), PH was slowly fixed toward the highest value of the three possible expression states for an early genotype, and the other traits were simultaneously fixed for + alleles at a greater rate than the PH genes were but slower than they had been in no constraints selection (Fig. 2).

The other relationships in Fig. 3 illustrate the effect of constraining selection to late genotypes when using a Severe-Terminal drought environment type (Fig. 3a) or the TPE (Fig. 3d) as the selection environment or constraining selection to medium-maturing genotypes in a Midseason drought environment type (Fig. 3b). In Fig. 3a, selection quickly swept out the rejected early and medium types and moved the population toward the earliest (no. 5) of the three expression states (5, 6, or 7) allowed in a late-maturing genotype, i.e., when all selected lines were fixed to that phenotype, the gene frequency was  $5/7 = 0.71$ . In this scenario, the progress in fixing OA and TE traits to all + alleles was slower





**Fig. 4.** For the same traits and selection environments given in Fig. 2, mean yields of successive cycles selected (a, b, and c) within different environment types or (d) by randomly sampling the target population of environments with a specified frequency combination of environments, given no selection constraints or constraining selection to be within one of three genotype maturity groups (see Fig. 3). Scales are varied to emphasize differences within selection environments.

than in the no constraints case (Fig. 2a) while the fixing of + alleles for SG genes was greatly delayed. As was the case for constraining selection to early genotypes (data not shown), selection for medium-maturing genotypes in the Midseason drought environment type quickly moved toward the latest (no. 5) of the three states (3, 4, or 5) for this class of maturity (Fig. 3b). It happens that this maturity (latest medium-maturing genotype) is the same as the earliest of late-maturing genotypes that became fixed in Fig. 3a. In Fig. 3b, the remaining traits were fixed a little slower but more simultaneously toward + alleles compared with the no constraints scenario (Fig. 2b).

The associations between the fixing of genes for + or – alleles and the mean yield of the population in the selection environment can be assessed by comparing Fig. 2 and 3 with Fig. 4. Fig. 4a shows that the maximum yield ceiling of about 3.5 t/ha in the Severe-Terminal drought environment type was reached after about nine cycles of recurrent selection. Constraining the PH genes to only the early maturing genotypes had minimal influence on the rate or end point of yield improvement as these are indeed the best-adapted genotypes. There were two phases of yield improvement in both selection scenarios: first, a linear increase from Cycles 1 to 6

(associated with the complete fixing of OA and TE genes to the + alleles, PH to – alleles, and about 0.6 of the SG genes to + alleles Fig. 2a) and second, a plateauing of yield improvement in Cycles 6 to 9 associated with the fixing of the + alleles for the remaining segregating SG genes and one PH gene. The gene fixation patterns were similar for the early-constraint scenario in this environment (data not shown).

When there were no phenology constraints under selection environments of either Midseason or Mild-Terminal drought environment types or the TPE, the yield improvement had a similar form to that under Severe-Terminal drought environment type, with a plateauing in Cycles 6 to 10 (Fig. 4b, 4c, and 4d) that occurred once most PH genes had been fixed to + alleles (Fig. 2). If selection was restricted to the early maturing genotypes, the rate of yield improvement was slower, and the ultimate yield reached was lower than for no constraints selection.

Figure 4 also shows the yield improvements associated with constraining selection to either medium- or late-maturing genotypes. When selecting in a Severe-Terminal drought environment type, the yield decreased for the first two cycles of evaluation (Fig. 4a) while the PH genes were being fixed to the earliest of the late-

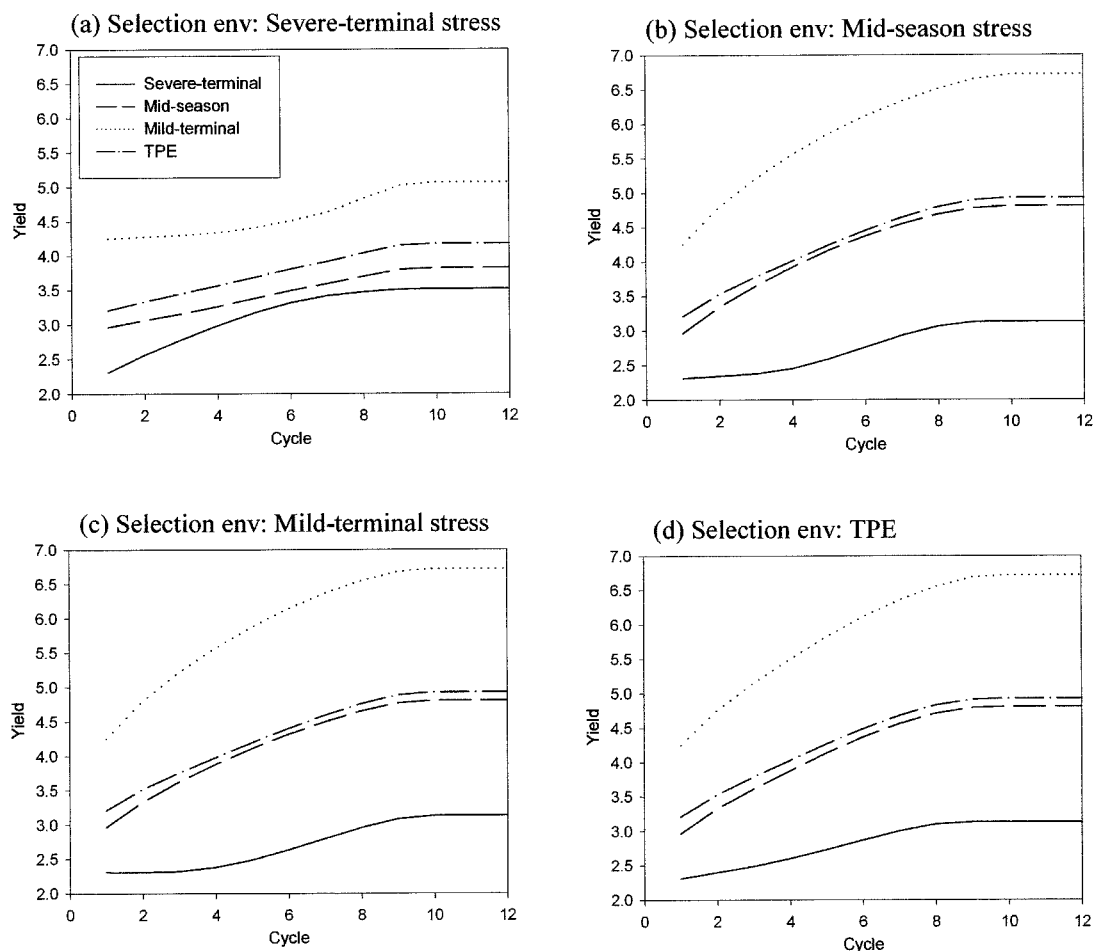


Fig. 5. For the no constraints selection scenario (Fig. 2), the yield of the population when selected under one of four selection environments (env) and then evaluated for the effects of indirect selection in each of the alternate selection environments. TPE, target population of environments.

genotype range (Fig. 3a). After this point, the + alleles for the OA and TE genes were completely fixed around Cycle 7, followed by the SG genes from Cycles 7 to 12 in a manner similar to that seen in Fig. 2a. When constrained to the late genotypes, yield progress in either the Midseason stress, Mild-Terminal stress, or the TPE was slightly slower than in the unconstrained case although the same end point was eventually achieved (Fig. 4b, 4c, and 4d). The more rapid fixing of the PH genes was associated with a delay of between one and two cycles in the fixation of the genes for the remaining traits (e.g., in Fig. 3d cf. Fig. 2d). As may be expected, both rate of progress and ultimate yield of the medium-maturing genotypes was intermediate to that of the early- and late-maturity selection methods.

The results in Fig. 4 show what happens to adaptation for specific drought environment types or the TPE when genotypes are evaluated only in the same drought environment type or TPE in which selection was conducted. For the selected proportions of genotypes under any scenario, it is possible to determine their corresponding value in other drought environment types and in the TPE, i.e., even though the selection has been done in a particular environment combination, we can evaluate the resulting changes in any other combination (Fig. 5).

This demonstrates the principle of indirect selection where selection within an environment leads to improved performance in another environment. For example, when the TPE was used as the selection environment, the yield for the TPE and for each of the drought environment types (i.e., indirectly selected for) increased at similar rates in percentage terms (data not shown) although there were differences in absolute terms (Fig. 5d). If the selection environment was restricted to only the Severe-Terminal drought environment type (Fig. 5a), then the rate of yield improvement was most rapid in that drought environment type but was much lower when the selected population was tested separately in the other drought environment types or the TPE. Notice that the final yield for the TPE was lower than in Fig. 5d (i.e., broad adaptation was lower than if the TPE had been used for selection) but that the final yield for a Severe-Terminal drought environment type was greater than for selection in either of the other drought environment types (Fig. 5b and 5c) or in the TPE (Fig. 5d).

When either the Midseason or Mild-Terminal drought environment type was used as the selection environment, performance in the TPE and in the alternate of these stress environments was relatively rapid (Fig. 5b

and 5c). However, yield improvement for the Severe-Terminal drought environment type was minimal until about Cycle 5. Inspection of Fig. 2 shows that this was the point when the PH genes had been mostly fixed with about half of the genes fixed for the remaining traits.

Under the maturity-constrained scenarios for selection in the TPE, the computed genotypic variance component had decreased by about 30% at Cycle 3, with little further change until Cycle 6, and then decreased more gradually compared with the no constraints scenario (Fig. 6a). The pattern was similar for the  $G \times E$  interaction variance components (Fig. 6b), such that the ratio of  $G \times E$  interaction and genotype effects was initially highest in the no constraints scenario (Fig. 6c). The  $G \times E$  variance for the early maturing constraint was greater than for the other PH constraint scenarios. In the last three cycles, the ratio of  $G \times E$  interaction and genotype effects increased greatly and was variable across the last few cycles of selection as both components of variance had become relatively small by this time.

## DISCUSSION

Hammer et al. (1996) and Chapman et al. (2002a) showed that the methodology of using a sorghum simulation model (APSIM-Sorg) to generate yields for a set of genotypes was able to reproduce genotypic and  $G \times E$  interaction effects over locations and seasons that were similar to those observed in the sorghum breeding program trials in northeastern Australia, e.g., Chapman et al. (2000a). In this paper, the genotypes were extended to a complete near-isogenic set with a defined gene action associated with different expression states for the traits. The connection of biophysical simulation of  $G \times E$  effects (APSIM-Sorg) to the simulation of a plant breeding program (QU-GENE) provided a test framework to evaluate polygenic (genomic) effects on selection for sorghum yield. In this case, it demonstrated that when the genotype population was sampled and recombined in a recurrent selection program, the rate of gene fixation and the alleles that were fixed varied with the traits in association with their value in improving yield for particular combinations of gene effects and drought environments (Fig 2).

Apart from the trait PH in the Severe-Terminal drought environment type, increasing the level of trait expression (by increasing number of + alleles for a trait) resulted in greater grain yield (Table 2). Higher values for the PH and TE traits had a greater effect on yield as the environments changed from Severe-Terminal stress to Midseason stress to Mild-Terminal stress. Conversely, greater values of TE and OA had a greater absolute effect on yield in the Severe-Terminal drought environment type. When the 4235 genotypes were sampled in the recurrent selection breeding program, the relative effects of the traits on yield in different drought environment types determined the rate at which genes were fixed by selection for grain yield.

Under the no constraints on maturity scenario, clear differences existed in the rates at which genes were fixed

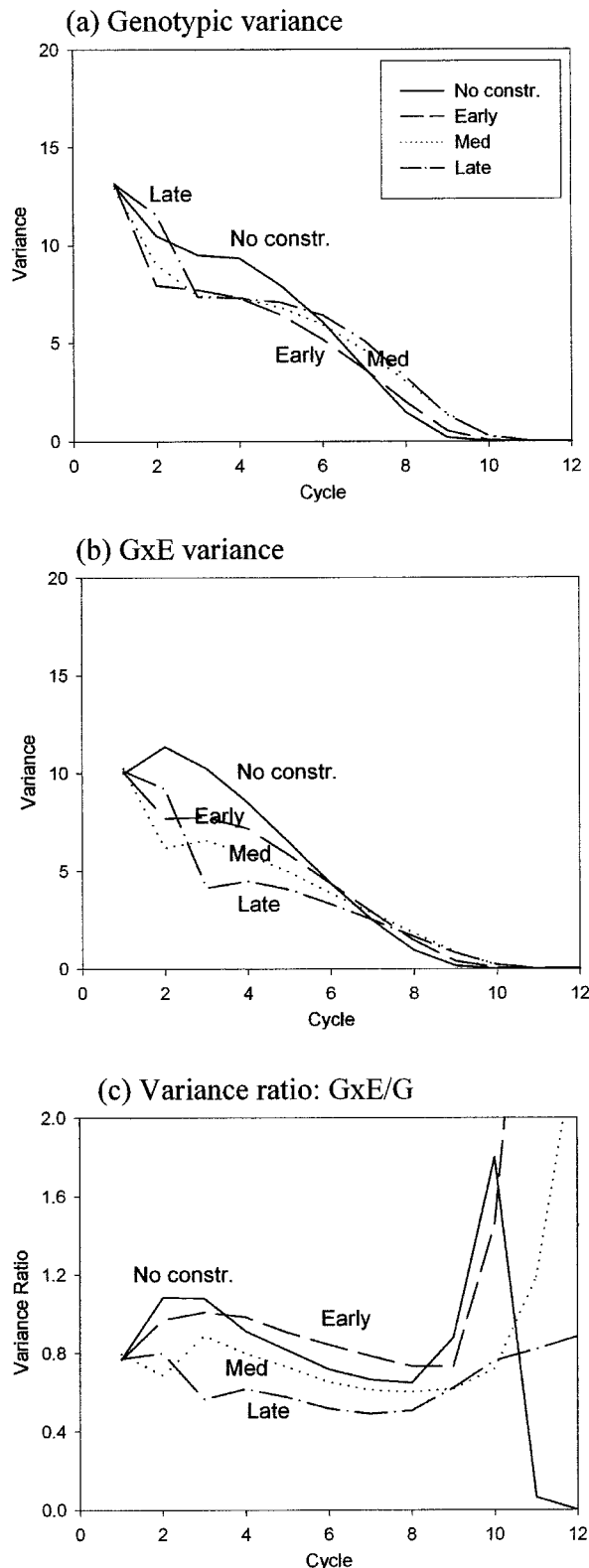


Fig. 6. For the simulations of S1 recurrent selection, the changes in (a) genotypic, (b) genotype  $\times$  environment ( $G \times E$ ) interaction variance components, and (c) their ratio for different cycles of selection under the target population of environments (TPE) (Fig. 2) when evaluated in the TPE.

during selection under different drought environment types or under the sampled mixture of drought environment types representing the TPE. When selected in the TPE, late-maturing genotypes were favored due to their higher yields in 65% of the component environments, i.e., in both Midseason and Mild-Terminal drought environment types (Fig. 4). The pattern of fixation of + alleles in the TPE was similar to that observed in the Midseason drought environment type. Hence, this was the best surrogate of the three drought environment types to represent the variation in the real-world TPE. However, in the TPE, the rate at which the PH genes were fixed to the latest-maturing genotypes was clearly slower than that observed in the Midseason and Mild-Terminal drought environment types.

The utility of + alleles for different traits in improving adaptation depended on the constraints set by both the maturity times of the genotypes and the selection environment. Early genotypes were favored in the Severe-Terminal drought environment type, with most of the + alleles for the SG trait not being fixed until the - alleles of the PH genes had first been fixed (Fig. 2). When the phenology was constrained to early-maturity types, the value of the + alleles for SG was increased in all environments as the SG alleles began to be fixed earlier in the selection process. Where the Severe-Terminal drought environment type was part of the selection environment (Fig. 2a and 2d), + alleles for OA genes were fixed more rapidly than those for TE. This reflects that fact that the OA genes were implemented in the crop simulation model to only have an effect under conditions of severe stress around flowering and during grain filling. As these conditions were less frequent in the Midseason and Mild-Terminal drought environment types than in the Severe-Terminal drought environment type, OA genes were not particularly favored over TE genes, which provide a yield advantage in all environment types.

The selection criteria (maturity) and the selection environment also interacted to influence the effectiveness of selection for yield. For example, in the Severe-Terminal drought environment type, selection with a constraint to early maturity resulted in the same rate of gain as no constraints. In contrast, although late maturity was favored in the other two drought environment types, constraining the selection to late-maturing types in these drought environment types reduced the rate of improvement in grain yield. It seems that restricting the selection in these environments effectively reduced the selection pressure for yield, i.e., if only 300 of the 1000 families met the late-phenology requirement and 100 of the 1000 have to be selected, then selection pressure is lessened to the degree that some superior-yielding genotypes (of early or medium maturity but with other favorable genes) did not meet the maturity criteria. This equates to a *selection bottleneck* in the breeding program where favorable-yield genes can be lost from the breeding population as their value is masked by the use of a rigorous culling criteria, in this case, maturity. Another type of bottleneck arises when breeding programs ex-

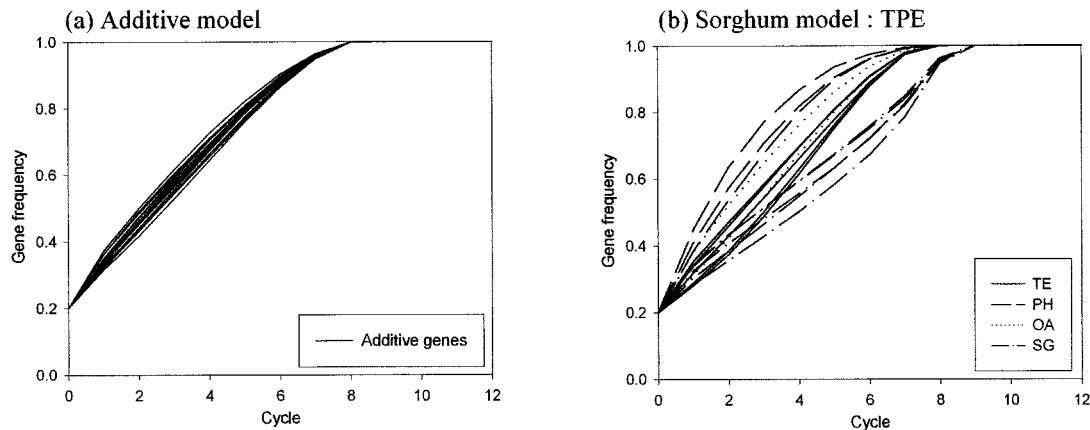
perience a particular sequence of environment types though we do not attempt to evaluate that here.

When evaluated in the TPE, the no constraints selection scenario retained the greatest degree of genotypic variance for yield for a longer period (until approximately Cycle 6) of the selection process (Fig. 6a). Constraining the phenology during selection quickly eroded the genotypic variance for yield although the later-maturing genotypes actually had lower  $G \times E$  interaction effects in the TPE, again due to the dominance of favorable environments (Midseason and Mild-Terminal drought environment types) for them in the TPE compared with the early maturing genotypes.

It is clear from Fig. 2 that, in an unconstrained scenario selecting in the TPE, it is difficult to retain segregation for maturity because of its strong association with yield in the better environments. If we were employing this methodology in practice to deliver a suite of cultivar maturities to the industry, the results suggest that it would be useful to set up a separate early maturing population for adaptation to Severe-Terminal stress drought environment type as this drought environment type tends to be associated with locations in the shallow soils of Central Queensland and the poorer rainfall zones of northern New South Wales (Chapman et al., 2000b, 2002a). There seems less justification to have a medium-maturing population as the Midseason drought environment type where its quality of performance is not particularly biased in any location. Nevertheless, medium-maturing cultivars would provide greater stability (i.e., less  $G \times E$  interaction for yield) in these poorer water environments than late-maturing cultivars.

The process of defining cumulative additive genes for traits and then expressing them via crop simulation generated both pleiotropic (where genes affect multiple traits, e.g., TE and yield) and epistatic (where different gene combinations interact with each other) effects for yield. These effects can be illustrated if we examine the results of a QU-GENE simulation for the same breeding program employing a conventional quantitative genetic model, i.e., many genes of small effect (see Podlich and Cooper, 1998, for examples of this simulation). The model was specified with 15 genes (as we used), each having small, equal additive effects on yield in different environments but with no specified epistatic or  $G \times E$  effects and no elaboration of the trait effects via a biophysical crop simulation model. Figure 7a shows that with selection, each of these genes for yield becomes fixed at a similar rate. This similarity in gene fixation contrasts with the case for the QU-GENE results when using the sorghum crop model to determine crop yield as controlled by 15 genes directly affecting four traits (Fig. 7b). The use of the crop model has modified the relative yield value of the different genes in the environment types sampled and hence has influenced the rate and timing of fixation of the + alleles for the different traits. The difference between these two approaches also indicates the importance of the state of the current germplasm as well as the selection environments in the potential for progress in the breeding program, i.e., while the relative value of each of the genes in Fig. 7a,





**Fig. 7.** For the simulations of S1 recurrent selection, the rate of fixation of + alleles for (a) a 15-gene additive model and (b) the 15 additive genes (across four traits) processed through the sorghum crop simulation model. Data in (b) are for the separate gene effects that had been averaged for each trait in Fig. 2d. TPE, target population of environments.

indicated by its rate of fixation, is more or less similar, this is not the case for Fig. 7b where the genes clearly have different values in terms of yield.

When set up to use information from analysis of quantitative genetics, applications in QU-GENE account explicitly for gene effects,  $G \times E$  interactions, and gene  $\times$  gene interactions (epistasis) (Podlich and Cooper, 1998). In the configuration that we have used here, only additive gene expression models were considered. However, the filtering of these expression states through a crop simulation model to determine the effect on yield has introduced epistasis for the yield trait. The most dramatic epistatic effects are evidenced by the necessity to fix favorable alleles for different genes in some sort of order that varied with environment type. Changing the order of the selection for major genes, by restricting selection for PH genes, could slow progress by reducing the opportunities for recombination of other non-PH genes that were present at low frequencies within the maturity class chosen. By the same inference, where plant breeders have observed epistatic effects for yield, they may really be seeing a complex combination of simple additive effects for traits (perhaps not measured) that determine yield.

Consideration of the effects of epistasis when modeling selection scenarios is vitally important. It suggests that for different combinations of traits being tested in particular environments, the fixation of some traits is unlikely to proceed until one or more other traits have been improved and in some cases, partially fixed, e.g., SG in Fig. 2a and TE and OA in Fig. 3c. If we conducted a molecular marker experiment to identify molecular markers for SG using the Cycle 3 population from Fig. 2a, we would find a low frequency of + alleles for SG genes and possibly discover some useful molecular markers. However, even if we had perfect markers, we would not achieve any advance in yield until the genes for other traits had also become fixed, i.e., the marker experiment may appear to fail even though the newly fixed alleles would be coming into play following further phenotypic selection for yield. Similarly, if we were evaluating the same population in different environ-

ments, our assessment of the existence of markers would change. These scenario results have significant implications for the manner in which strategies such as marker-assisted selection are introduced to improve the efficiency of the breeding program for yield.

The QU-GENE software can be used to investigate the inclusion of markers and degrees of recombination and linkage among the genes in the model to answer some of the questions posed in the previous paragraph. In an extension of our current analysis, for example, we examined the effects of the precision of molecular markers on the efficiency of selection (Chapman et al., 2002b). A sophisticated enhancement would examine the issue of the difficulty of finding molecular markers for yield, per se. This difficulty is not unexpected (although many marker projects continue to attempt it) because yield is a pleiotropic effect of multiple genes controlling subprocesses of crop growth over the season. By outputting the values of other key attributes from the crop model (e.g., crop growth rate at flowering), it is feasible for a researcher using this model to find sets of attributes that control yield but are simpler in their genetic control. Investigating the interaction between the biophysical simulation and the breeding program simulation should also allow derivation of key traits and genetic networks controlling yield in different environments. Additional issues that could be addressed are the effects of variation in the genotype response within an environment type, the effects of selection within fixed sequences of environment types, and the effects of indirect selection for traits other than yield.

With the availability of high-throughput capacity to sequence genomes, it has been widely claimed (e.g., Bassett et al., 1999) that biologists have not determined how to cope with the massive amounts of DNA sequence and gene expression data that are being generated. Worse, there are few examples of linking this information to crop phenotypes in the field—most are laboratory-based phenotypes in simple controlled-environment screens, which may bear little correlation with field performance. As we expand our ability to describe and understand the structure and function of plant ge-

nomes, there is an urgent need to investigate how the effects of genes are integrated at the different organizational levels of the organism and how these organizational structures (e.g., pathways, traits, and trait combinations) interact with the biophysical properties of environments to determine the crop phenotype. While the task of understanding the relationship between gene and phenotype is a major undertaking, even for many simply inherited traits, the integration of genetic models, that can simulate the properties of gene-environment systems, with dynamic biophysical crop models, that simulate plant growth and development processes, provides a quantitative framework to support these ambitious investigations. Within the generic crop template design in APSIM, developments in improved modeling of the traits are quickly captured (Hammer, 1998). Improvements in computer speed and the use of computer clusters, as outlined here, are essential to process the large numbers of scenarios to be investigated. More importantly, it is necessary to continue the field experimentation to improve the ability of crop models to capture the interaction of crop traits with the environment.

While much basic research investment is in the cellular and molecular physiology of traits, the interaction effects and levels higher in the hierarchy of growth (organ, plant, and crop) greatly complicate the expression of these traits, and many complex crop-level traits are yet to be dissected. In most situations, experimental evaluation of conventional and molecular breeding strategies for manipulating complex yield adaptation traits will be impractical or beyond the resource base of breeding programs. Therefore, simulation tools will have an important role in the design and testing of breeding strategies. This role will become increasingly feasible as our understanding of the genetic architecture of quantitative traits improves. A critical step in developing these simulation tools is to establish the link between gene- and genome-based information and the biophysical processes that determine plant growth and development and adaptation to the biotic and abiotic stresses. In this paper, we have demonstrated the linkages that have been achieved between the QU-GENE and APSIM-Sorg modeling platforms to begin to address the issue of the complexity of interactions among gene, trait, and environment effects.

## ACKNOWLEDGMENTS

The valuable assistance of technical staff in QDPI (Greg McLean and Peter de Voil) and UQ (Kevin Micallef) is gratefully acknowledged in the completion of the APSIM and QU-GENE simulations on the computer cluster and processing and summarizing of output.

## REFERENCES

Asseng, S., N.C. Turner, T. Botwright, and A.G. Condon. 2003. Evaluating the impact of a trait for increased specific leaf area on wheat yields using a crop simulation model. *Agron. J.* 95:10-19 (this issue).  
 Basnayake, J., M. Cooper, M.M. Ludlow, R.G. Henzell, and P.J. Snell. 1995. Inheritance of osmotic adjustment in three grain sorghum crosses. *Theor. Appl. Genet.* 90:675-682.  
 Bassett, D.E., Jr., M.B. Eisen, and M.S. Boguski. 1999. Gene expression informatics—it's all in your mine. *Nat. Genet.* 21(suppl.):51-55.

Borrell, A.K., G.L. Hammer, and R.G. Henzell. 2000. Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. *Crop Sci.* 40:1037-1048.  
 Chapman, S.C., M. Cooper, and G.L. Hammer. 2002a. Using crop simulation to interpret broad adaptation and genotype by environment effects for sorghum in water-limited environments. *Aust. J. Agric. Res.* 53:1-11.  
 Chapman, S.C., M. Cooper, G.L. Hammer, and D. Butler. 2000a. Genotype by environment interactions affecting yield of grain sorghum: II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Aust. J. Agric. Res.* 51:209-222.  
 Chapman, S.C., G.L. Hammer, D. Butler, and M. Cooper. 2000b. Genotype by environment interactions affecting yield of grain sorghum: III. Temporal sequences and spatial patterns in the target population of environments. *Aust. J. Agric. Res.* 51:223-233.  
 Chapman, S.C., G.L. Hammer, and H.M. Meinke. 1993. A sunflower simulation model: I. Model development. *Agron. J.* 85:725-734.  
 Chapman, S.C., G.L. Hammer, D.W. Podlich, and M. Cooper. 2002b. Linking bio-physical and genetic models to integrate physiology, molecular biology and plant breeding. p. 167-188. *In* M. Kang (ed.) Quantitative genetics, genomics, and plant breeding. Proc. Symp. on Quantitative Genet. for the 21st Century, Baton Rouge, LA. March 2001. CABI, Wallingford, UK.  
 Comstock, R.E. (1977). Quantitative genetics and the design of breeding programs. p. 705-18. *In* Proc. of the International Conference on Quantitative Genetics, August 16-21, 1976. Iowa State University Press, Ames, USA.  
 Cooper, M., S.C. Chapman, D.W. Podlich, G.L. Hammer 2002a. The GP problem: Quantifying gene-to-phenotype relationships. *In* Silico Biol. 2:151-164. (Available online at <http://www.bioinfo.de/isb/2002/02/0013/>) (Verified 5 Sept. 2002.)  
 Cooper, M., D.W. Podlich, K.P. Micallef, O.S. Smith, N.M. Jensen, S.C. Chapman, and N.L. Kruger. 2002b. Complexity, quantitative traits and plant breeding: A role for simulation modeling in the genetic improvement of crops. p. 143-166. *In* M. Kang (ed.) Quantitative genetics, genomics, and plant breeding. Proc. Symp. on Quantitative Genet. for the 21st Century, Baton Rouge, LA. March 2001. CABI, Wallingford, UK.  
 de Wit, C.T., and F.W.T. Penning de Vries. 1983. Crop growth models without hormones. *Neth. J. Agric. Sci.* 31:313-323.  
 Edmeades, G.O., J. Bolanos, S.C. Chapman, H.R. Lafitte, and M. Banziger. 1999. Selection improves drought tolerance in tropical maize populations: I. Gains in biomass, grain yield and harvest index. *Crop Sci.* 39:1306-1315.  
 Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Harlow, UK.  
 Gell-Mann, M. 1994. The quark and the jaguar—adventures in the simple and the complex. Abacus, London.  
 Giersch, C. 2000. Mathematical modelling of metabolism. *Curr. Opin. Plant Biol.* 3:249-253.  
 Hallauer, A.R., W.A. Russell, and K.R. Lamkey. 1988. Corn breeding. p. 463-564. *In* G.F. Sprague and J.W. Dudley (ed.) Corn and corn improvement. Agron. Monogr. 18. ASA, CSSA, SSSA, Madison, WI.  
 Hammer, G.L. 1998. Crop modelling: Current status and opportunities to advance. *Acta Hort.* 456:27-36.  
 Hammer, G.L., D. Butler, R.C. Muchow, and H. Meinke. 1996. Integrating physiological understanding and plant breeding via crop modelling and optimization. p. 419-441. *In* M. Cooper and G.L. Hammer (ed.) Plant adaptation and crop improvement. CAB Int., ICRISAT, and IRRI, Wallingford, UK.  
 Hammer, G.L., S.C. Chapman, and P. Snell. 1999. Crop simulation modelling to improve selection efficiency in plant breeding programs. p. 79-85. *In* P. Williamson et al. (ed.) Proc. Assembly Wheat Breeding Soc. of Australia, 9th, Toowoomba. 27 Sept.-1 Oct. 1999. Wheat Breeding Soc. of Australia, Toowoomba, Australia.  
 Hammer, G.L., M.J. Kropff, T.R. Sinclair, and J.R. Porter. 2002. Future contributions of crop modelling—from heuristics and supporting decision-making to understanding genetic regulation and aiding crop improvement. *Eur. J. Agron.* (in press)  
 Hammer, G.L., and R.C. Muchow. 1994. Assessing climatic risk to sorghum production in water-limited subtropical environments: I.

- Development and testing of a simulation model. *Field Crops Res.* 36:221–234.
- Hammer, G.L., E. van Oosterom, S.C. Chapman, and G. Mclean. 2001. Supply and demand economics applied to crop growth. *In* A.K. Borrell and R.G. Henzell (ed.) *Proc. Australian Sorghum Conf.*, 4th, Kooralbyn. 5–8 Feb. 2001 [CD-ROM]. Range Media, Toowoomba, Australia.
- Hart, G.E., K.F. Schertz, Y. Peng, and N.H. Syed. 2001. Genetic mapping of *sorghum bicolor* (L.) Moench QTLs that control variation in tillering and other morphological characters. *Theor. Appl. Genet.* 103:1232–1242.
- Hasegawa, P.M., R.A. Bressan, J.K. Zhu, and H.J. Bohnert. 2000. Plant cellular and molecular responses to high salinity. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51:463–499.
- Hospital, F., L. Moreau, F. Lacoudre, A. Charcosset, and A. Galtais. 1997. More on the efficiency of marker assisted selection. *Theor. Appl. Genet.* 95:1181–1189.
- Mazur, B., E. Krebbers, and S. Tingey. 1999. Gene discovery and product development for grain quality traits. *Science* 285:372–375.
- McCown, R.L., G.L. Hammer, J.N.G. Hargreaves, D.P. Holzworth, and D.M. Freebairn. 1996. APSIM: A novel software system for model development, model testing, and simulation in agricultural systems research. *Agric. Syst.* 50:255–271.
- Mendes, P., and D.B. Kell. 1998. Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* 14:869–883.
- Micallef, K., M. Cooper, and D.W. Podlich. 2001. Using clusters of computers for large QU-GENE simulation experiments. *Bioinformatics* 17:194–195.
- Mortlock, M.Y., and G.L. Hammer. 1999. Genotype and water limitation effects on transpiration efficiency in sorghum. *J. Crop Prod.* 2:265–286.
- Muchow, R.C., M. Cooper, and G.L. Hammer. 1996. Characterizing environmental challenges using models. p. 349–364. *In* M. Cooper and G.L. Hammer (ed.) *Plant adaptation and crop improvement*. CAB Int., ICRISAT, and IRRI, Wallingford, UK.
- Podlich, D., and M. Cooper. 1998. QU-GENE: A simulation platform for quantitative analysis of genetic models. *Bioinformatics* 14:632–653.
- Podlich, D.W., M. Cooper, and K.E. Basford. 1999. Computer simulation of a selection strategy to accommodate genotype–environment interactions in a wheat recurrent selection programme. *Plant Breed.* 118:17–28.
- Richards, R.A., and E. Belhassen. 1996. Defining criteria to improve yield under drought. *Plant Growth Regul.* 20:157–166.
- Richards, R.A., G.J. Rebetzke, A.G. Condon, and A.F. van Herwaarden. 2002. Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. *Crop Sci.* 42:111–121.
- Rooney, W.L., and S. Aydin. 1999. Genetic control of a photoperiod-sensitive response in *Sorghum bicolor* (L.) Moench. *Crop Sci.* 39:397–400.
- Somerville, C., and S. Somerville. 1999. Plant functional genomics. *Science* 285:380–383.
- Tao, Y.Z., R.G. Henzell, D.R. Jordan, D.G. Butler, A.M. Kelly, and C.L. McIntyre. 2000. Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. *Theor. Appl. Genet.* 100:1225–1232.
- Thomas, H., and C.J. Howarth. 2000. Five ways to stay green. *J. Exp. Bot.* 51:329–337.
- Van Berloo, R., and P. Stam. 1998. Marker-assisted selection in autogamous RIL populations: A simulation study. *Theor. Appl. Genet.* 96:147–154.
- Yin, X., P. Stam, M.J. Kropff, and A.H.C.M. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agron. J.* 95:90–98 (this issue).