Simulating the Effects of Dominance and Epistasis on Selection Response in the CIMMYT Wheat Breeding Program Using QuCim

Jiankang Wang,* Maarten van Ginkel, Richard Trethowan, Guoyou Ye, Ian DeLacy, Dean Podlich, and Mark Cooper

ABSTRACT

Plant breeders use many different breeding methods to develop superior cultivars. However, it is difficult, cumbersome, and expensive to evaluate the performance of a breeding method or to compare the efficiencies of different breeding methods within an ongoing breeding program. To facilitate comparisons, we developed a QU-GENE module called QuCim that can simulate a large number of breeding strategies for self-pollinated species. The wheat breeding strategy "Selected Bulk" used by CIMMYT's wheat breeding program was defined in QuCim as an example of how this is done. This selection method was simulated in QuCim to investigate the effects of deviations from the additive genetic model, in the form of dominance and epistasis, on selection outcomes. The simulation results indicate that the partial dominance model does not greatly influence genetic advance compared with the pure additive model. Genetic advance in genetic systems with overdominance and epistasis are slower than when gene effects are purely additive or partially dominant. The additive gene effect is an appropriate indicator of the change in gene frequency following selection when epistasis is absent. In the absence of epistasis, the additive variance decreases rapidly with selection. However, after several cycles of selection it remains relatively fixed when epistasis is present. The variance from partial dominance is relatively small and therefore hard to detect by the covariance among half sibs and the covariance among full sibs. The dominance variance from the overdominance model can be identified successfully, but it does not change significantly, which confirms that overdominance cannot be utilized by an inbred breeding program. QuCim is an effective tool to compare selection strategies and to validate some theories in quantitative genetics.

T HE MAJOR OBJECTIVE of plant breeding programs is to develop new genotypes that are genetically superior to those currently available for a specific target population of environments (Comstock, 1996; Cooper et al., 1999). To achieve this objective, plant breeders employ a range of selection methods (Allard, 1960; Jensen, 1988; Stoskopf, 1993). Many field experiments have been conducted to compare the efficiencies of different breeding methods (for review see Stoskopf, 1993). However, because of the time and effort spent in conducting field experiments, the concept of modeling and prediction has always been of interest to plant breeders.

Quantitative genetic theory generally provides much of the framework for the design and analysis of selection methods used within breeding programs. However, various assumptions are made in quantitative genetics to render theories mathematically or statistically tractable (Falconer, 1989; Comstock, 1996; Kearsey and Pooni, 1996). Some of these assumptions can be easily tested or satisfied by certain experimental designs; others, such as the assumptions of no linkage, no multiple alleles, and no genotype by environment interaction, can seldom, if ever, be met. Other assumptions, like the presence or absence of epistasis and pleiotropy, are statistically difficult to define and test (Holland, 2001). Computer simulation provides an opportunity to lessen the impact of these assumptions by accommodating these factors, thereby improving the validity of genetic models for use in plant breeding (Kempthone, 1988; Comstock, 1996).

Simulation, using relatively simple genetic models, has been used for many special studies in plant breeding (Casali and Tigchelaar, 1975; Reddy and Comstock, 1976; van Oeveren and Stam, 1992; van Berloo and Stam, 1998; Frisch and Melchinger, 2001). Nevertheless, a tool capable of simulating the performance of breeding and selection strategies for a continuum of genetic models ranging from simple to complex, imbedded within an existing practical breeding program, has not been available until recently.

QU-GENE, a simulation platform for quantitative analysis of genetic models, provides the opportunity to develop a general simulation program for actual breeding programs through its two-stage architecture (Podlich and Cooper, 1998). The first stage involves QU-GENE as the central *engine*, the role of which is to define the genotype and environment system and generate the starting population of individuals and the reference population to estimate genetic variances and error variances. In the second stage, external application modules are developed and linked to the QU-GENE engine to manipulate, investigate, and analyze the starting population of individuals according to crossing and selection approaches set by the user within the genotype and environment system defined by the engine.

QuCim is such a QU-GENE application module and was specifically developed to simulate the wheat breeding programs in the International Maize and Wheat Improvement Center (CIMMYT), gaining its name from the contraction of these two names (software available; refer inquiry to the senior author). However, the breeding strategies defined in QuCim represent the operations of most breeding programs for self-pollinated crops, and hence in principle have wide potential application. The breeding methods that can be simulated by QuCim are mass selection, pedigree system (including single seed descent), bulk population system, backcross breeding, top cross (or three-way cross) breeding, doubled haploid breeding, marker-assisted selection, and many com-

J. Wang, M. van Ginkel, and R. Trethowan, CIMMYT, Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico; G. Ye, I. DeLacy, D. Podlich, and M. Cooper, School of Land and Food Sciences, The Univ. of Queensland, Brisbane, Qld 4072, Australia; D. Podlich and M. Cooper, current address: Pioneer Hi-Bred International Inc., 7250 N.W. 62nd Avenue, PO Box 552, Johnston, IA 50131, USA. Received 12 Nov. 2003. *Correspondence author (jkwang@cgiar.org).
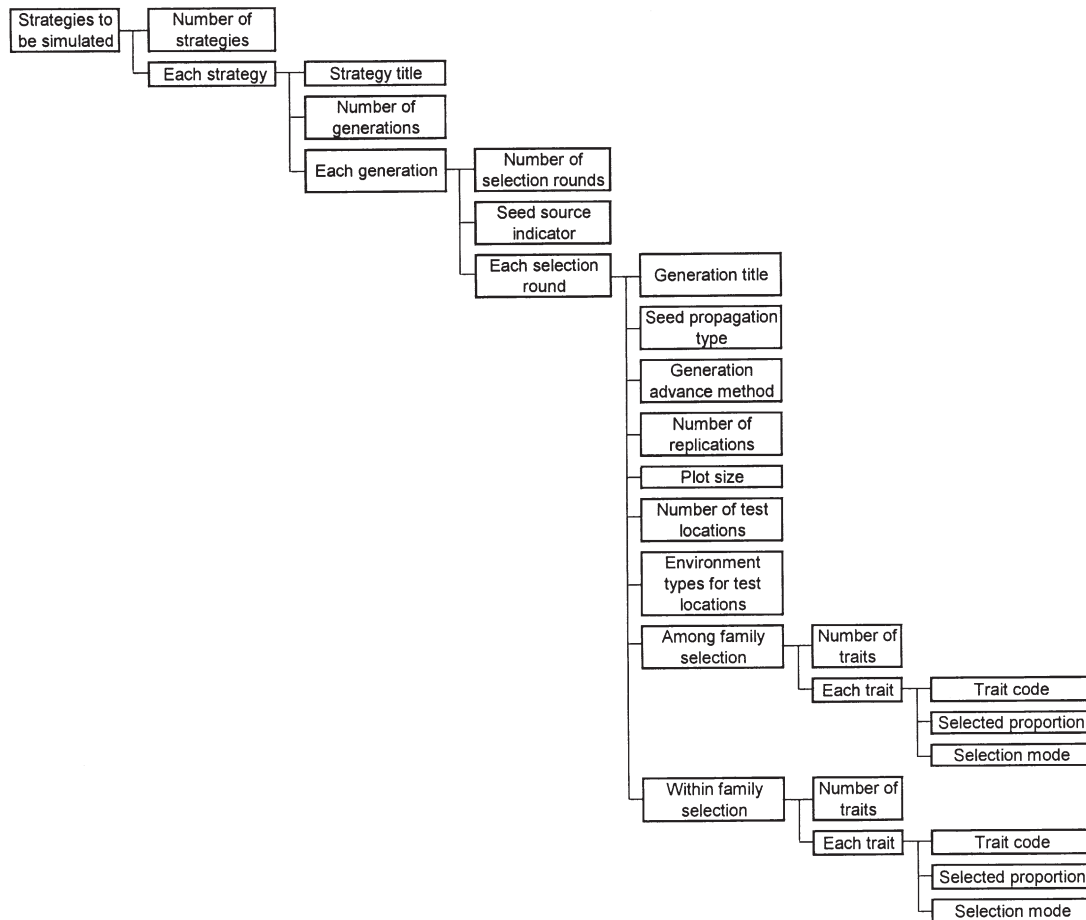
Fig. 1. Data structure to define breeding strategies in QuCim.

binations and modifications of these methods. Simulation experiments can therefore be designed to compare the breeding efficiencies of different selection strategies, or various modifications within a selection strategy, for any self-pollinating crop or line-development breeding program, including that for inbred lines in cross-pollinated crops such as corn (*Zea mays* L.). QuCim has been used to compare two breeding strategies common in CIMMYT's wheat breeding program (Wang et al., 2003a). The objectives of this paper are (i) to explain how a complex crossing and selection strategy is defined in QuCim and (ii) to investigate effects on selection of dominance and epistasis in a breeding program developing inbred lines.

MATERIALS AND METHODS

Definition of Breeding Strategies in QuCim

QuCim allows for several breeding strategies to be defined simultaneously in one input file (Fig. 1). The program then makes the same virtual crosses for all the defined strategies at the first breeding cycle. Hence, all strategies start from the same point (the same initial population, the same crosses, and the same genotype and environment system) allowing appropriate comparisons.

A breeding strategy as defined in this paper includes all information regarding crossing, seed propagation, selection activities, and field experimental designs in an entire breeding cycle. A breeding cycle begins with crossing two parents and

ends at the generation when the selected advanced lines are considered sufficiently homozygous and homogenous to be used as ready products by growers, or are returned to the crossing block as new parents. Most breeding programs of self-pollinated crops can be approximated as described in Fig. 2. This figure demonstrates the germplasm flow in one breeding cycle in CIMMYT's bread wheat breeding program (van Ginkel et al., 2002). The selection method described in Fig. 2 is called "Selected Bulk" in CIMMYT, and will be used as an example throughout this paper to demonstrate how a breeding strategy is defined in QuCim. It should be noted that variations of this selection approach are commonly used in breeding programs for self-pollinated crops. For this study, we simulated 20 complete cycles of selection for each combination of experimental parameters.

There are three key Mexican locations used by CIMMYT's wheat breeding program: Ciudad Obregon [27°N, 39 m above sea level (masl)], Toluca (19°N, 2640 masl), and El Batan (19°N, 2300 masl). Cd. Obregon is an arid, irrigated location, and growing season conditions are similar to many other irrigated environments around the world (Trethowan et al., 2001). The yield trials for materials targeted to low rainfall and irrigated areas are conducted at Cd. Obregon. High yields (8–11 Mg/ha) are obtained under near optimal irrigation, while reduced irrigation can result in yields as low as 1 to 2 Mg/ha, indicative of drought-prone areas. The main diseases are leaf rust [caused by *Puccinia triticina* Eriks. = *P. recondita* Roberge ex Desmaz. f. sp. *tritici* (Eriks. & E. Henn.) D.M. Henderson] and stem rust (caused by *Puccinia graminis* Pers.:Pers. = *P. graminis* Pers.:Pers. f. sp. *tritici* Eriks. & E. Henn.). Precipita-
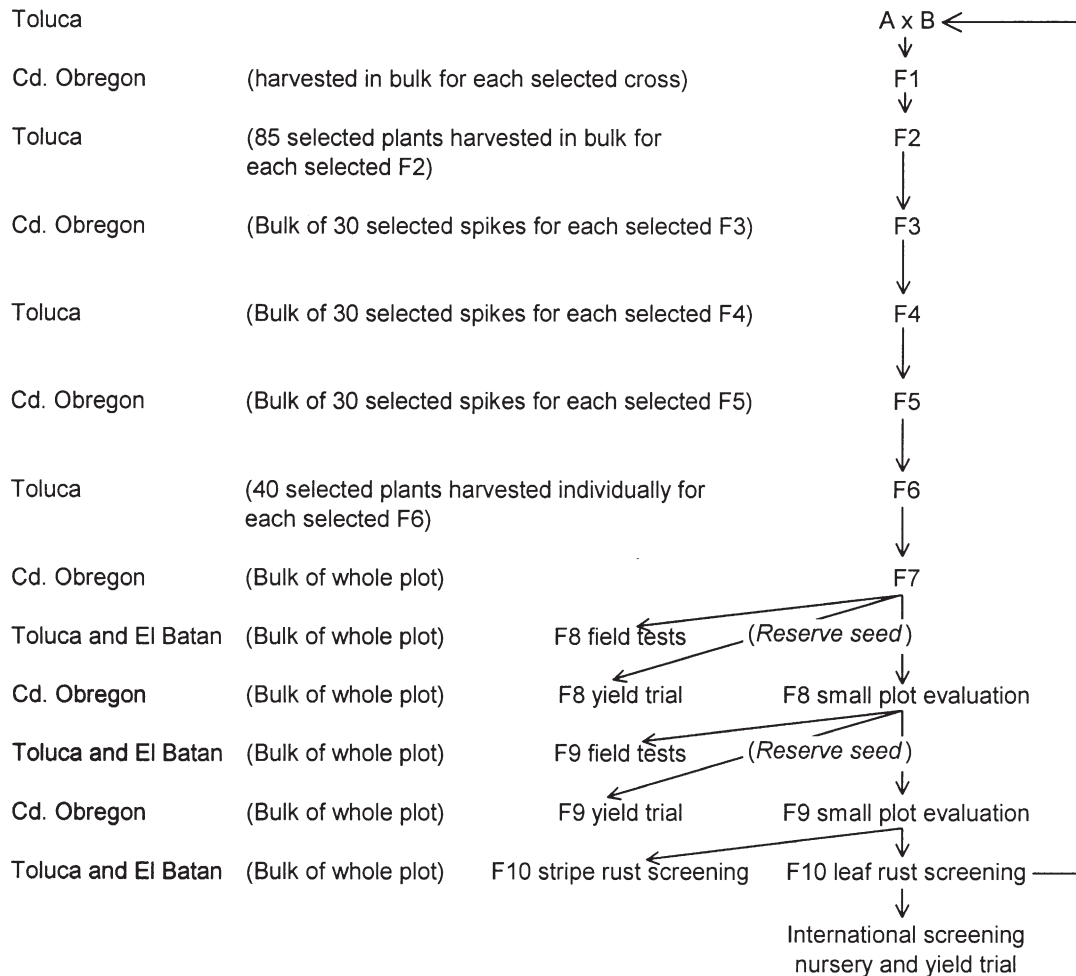
Toluca		A x B
Cd. Obregon	(harvested in bulk for each selected cross)	F1
Toluca	(85 selected plants harvested in bulk for each selected F2)	F2
Cd. Obregon	(Bulk of 30 selected spikes for each selected F3)	F3
Toluca	(Bulk of 30 selected spikes for each selected F4)	F4
Cd. Obregon	(Bulk of 30 selected spikes for each selected F5)	F5
Toluca	(40 selected plants harvested individually for each selected F6)	F6
Cd. Obregon	(Bulk of whole plot)	F7
Toluca and El Batan	(Bulk of whole plot)	F8 field tests (Reserve seed)
Cd. Obregon	(Bulk of whole plot)	F8 yield trial F8 small plot evaluation
Toluca and El Batan	(Bulk of whole plot)	F9 field tests (Reserve seed)
Cd. Obregon	(Bulk of whole plot)	F9 yield trial F9 small plot evaluation
Toluca and El Batan	(Bulk of whole plot)	F10 stripe rust screening F10 leaf rust screening

International screening nursery and yield trial

Fig. 2. Germplasm flow in the Selected Bulk selection method used in CIMMYT's wheat breeding program.

tion in Toluca is high (about 800-900 mm during the summer crop cycle), providing favorable conditions for foliar diseases, in particular stripe rust (caused by *Puccinia striiformis* Westend.) and foliar blights, such as that caused by *Septoria tritici* Roberge in Desmaz. Precipitation at CIMMYT's headquarters, El Batan is more erratic, with an annual average of 600 to 700 mm, and irrigation is available when needed. El Batan is mainly used for leaf rust screening because of its slightly higher temperature profile, and for small-scale seed increases. Two wheat cycles can be grown each year: November through April at Cd. Obregon, and May through October at Toluca and El Batan.

Number of Generations in Selected Bulk and Number of Selection Rounds in Each Generation

In the breeding program in Fig. 2, the best advanced lines developed from the F10 generation will be returned to the crossing block to be used for new crosses; that is to say, a new breeding cycle starts after F10 leaf rust screening at El Batan. Therefore, the number of generations in one breeding cycle is 10 for the Selected Bulk breeding strategy. The crossing block (viewed as F0) and the 10 generations must first be defined in QuCim. The parameters to define a generation consist of the number of selection rounds in the generation, an indicator for seed source (explained later), and the planting and selection details for each selection round (Fig. 1; Tables 1 and 2). Most generations in this breeding program have just

one selection round, e.g., F1 to F6, while some generations have more than one selection round since they are grown simultaneously at different sites or under different conditions, e.g., F7, F8, and F9.

QuCim was developed such that the key components of the field breeding process are retained and described. This is important for the integrity of the simulation, and it allows the breeder to retain confidence in the value of the simulation process.

Seed Propagation Type for Each Selection Round

The seed propagation type describes how the selected plants in a retained family from the previous selection round or generation are propagated to generate the seed for the current selection round or generation. There are seven options for seed propagation, presented here in the order of increasing genetic diversity (the F1 excluded): (i) *clone* (asexual reproduction), (ii) *DH* (doubled haploid), (iii) *self* (self-pollination), (iv) *backcross* (back crossed to one of the two parents), (v) *topcross* (crossed to a third parent, also known as three-way cross), (vi) *random* (random mating among the selected plants in a family), and (vii) *noself* (random mating but self-pollination is eliminated). The seed for the F1 is derived from crossing among the parents in the initial population (or crossing block). QuCim randomly determines the female and the male parents for each cross from a defined initial population, or alternately, one may select some preferred parents from the crossing block.

Table 1. Definition of the selected bulk selection method.

Number of selection rounds	Seed source†	Generation title	Seed propagation type‡	Generation advance method§	Number of replications	Plot size¶	Number of locations	Environment type of test location
1		Crossing block (F0)	self	bulk	1	20	1	Toluca
1		F1	random	bulk	1	20	1	Cd. Obregon
1		F2	self	bulk	1	1000	1	Toluca
1		F3	self	bulk	1	500	1	Cd. Obregon
1		F4	self	bulk	1	625	1	Toluca
1		F5	self	bulk	1	625	1	Cd. Obregon
1		F6	self	pedigree	1	750	1	Toluca
4	0	F7	self	bulk	1	70	1	Cd. Obregon
		F8(T)#	self	bulk	1	70	1	Toluca
		F8(B)#	self	bulk	1	70	1	El Batan
		F8(YT)††	self	bulk	1	100	1	Cd. Obregon
4	0	F8(SP)‡‡	self	bulk	1	30	1	Cd. Obregon
		F9(T)§§	self	bulk	1	70	1	Toluca
		F9(B)§§	self	bulk	1	70	1	El Batan
		F9(YT)††	self	bulk	2	100	1	Cd. Obregon
1		F9(SP)‡‡	self	bulk	1	30	1	Cd. Obregon
2	0	F10(LR)¶¶	self	bulk	1	30	1	El Batan
		F10(YR)¶¶	self	bulk	1	30	1	Toluca

† Seed source, an indicator to identify where the seed for selection round 2 and afterwards comes from for the generation with more than one selection round. This value can only be either 0 or 1. Value 0 means seed for round i ($i \geq 2$) comes from round 1, and value 1 means seed for round i ($i \geq 2$) comes from the previous round i-1.

‡ Seed propagation type, ways to propagate seed from the selected plants, which can only be clone, DH (doubled haploid), self, backcross, topcross, random, or noself. For the F1 generation only random is available, indicating that the parents for each cross are randomly selected from the crossing block.

§ Generation advance method, ways to process the selected individuals in a family, which can only be *bulk* or *pedigree*.

¶ The plot size (number of plants in a plot or replication) in yield trials can be 1000 or more in practice. Since we only conducted among-family selection based on yield data, 100 plants in a plot is large enough to make a good estimate of the mean yield of a family in simulation. A smaller plot size also reduces computational requirements.

F8(T) and F8(B), the F8 field tests grown at Toluca (T) and El Batan (B) for stripe rust and leaf rust selections, respectively.

†† F8(YT) and F9(YT), yield trials (YT) in F8 and F9 grown at Cd. Obregon.

‡‡ F8(SP) and F9(SP), small plot (SP) evaluation in F8 and F9 grown at Cd. Obregon.

§§ F9(T) and F9(B), the F9 field tests grown at Toluca (T) and El Batan (B) for stripe rust and leaf rust selections, respectively.

¶¶ F10(LR) and F10(YR), the F10 leaf rust (LR) screening at El Batan and F10 stripe rust (YR = yellow rust) screening at Toluca.

The selection criteria used to identify such preferred parents (grouped here as the male and female master lists) can be defined in terms of among-family and within-family selection descriptors (see below for details) within the crossing block (referred to as F0 generation).

By using the parameter of seed propagation type, most if not all methods of seed propagation in self-pollinated crops can be simulated by QuCim.

Generation Advance Method for Each Selection Round

The generation advance method describes how the selected plants within a family are harvested. There are two options for this parameter: *pedigree* (the selected plants within a family are harvested individually, and therefore each selected plant will result in a distinct family in the next generation) and *bulk* (the selected plants in a family are harvested in bulk, resulting in just one family in the next generation). This parameter and the seed propagation type allow QuCim to simulate not only the traditional breeding methods, such as pedigree breeding and bulk population breeding, but also many combinations of different breeding methods (e.g., pedigree selection until the F4 and then doubled haploid production on selected F4 plants). The bulk generation advance method will not change the number of families in the following generation if no among-family selection is applied in the current generation, while the pedigree method increases the number of families rapidly if there is weak among-family selection intensity, and several plants are selected within each retained family. For a generation with more than one selection round, the generation advance method for the first selection round can be either pedigree or bulk. The subsequent selection rounds are used to determine which families derived from the first selection round will be advanced to the next generation. In the majority of cases, bulk generation advance is the preferred option for the subsequent selection rounds.

Field Experiment Design for Each Selection Round

The parameters used to define the virtual field experiment design in each selection round include the number of replications for each family, the number of individual plants in each replication, the number of test locations, and the environment type for each test location (Fig. 1 and Table 1). The concept of an environment type within the target population of environments was defined by Podlich and Cooper (1998) to distinguish sets of environmental conditions (e.g., CIMMYT's megaenvironments; Rajaram, 1999) that conditioned different genetic effects, and thus different genetic requirements for adaptation. Each environment type defined in the genotype and environment system has its own gene action and gene interaction, which provides the framework for defining the genotype × environment interaction. Therefore, by defining the target population of environments as a mixture of environment types, genotype × environment interactions are defined as a component of the genetic architecture of a trait (Cooper and Podlich, 2002). An integer number represents the environment type for each test location, and whenever possible, it should be consistent with known features that are defined for the target population of environments of the genotype and environment system. For those locations where the environment types are little understood, QuCim will randomly assign environment types to them with a likelihood based on the frequencies of environment types in the target population of environments. Additional examples demonstrating the application of these procedures to define genotype × environment interactions were given by Podlich et al. (1999) and Cooper et al. (2002).

Table 2. Among-family and within-family selected proportions for each trait in the 10 generations in one breeding cycle.

Generation	Yield	Lodging	Stem rust	Leaf rust	Stripe rust	Height	Heading	Tillering	Grains per spike	Kernel weight	Total selected proportion	Number of families after selection
Selection mode†	top	bottom	bottom	bottom	bottom	middle	middle	top	top	top		400‡
F1, among§	¶	0.98	0.99	0.85		0.99	0.98	0.90	0.97		0.70	279
F2, among		0.99	0.99		0.90	0.99	0.99	0.99	0.99		0.85	238
F2, within		0.95	0.99		0.40	0.85	0.90	0.60	0.50		0.08	
F3, among		0.99		0.90				0.95			0.85	201
F3, within		0.90		0.70		0.90	0.90	0.80	0.25	0.60	0.06	
F4, among		0.99			0.96			0.95			0.90	181
F4, within		0.90		0.65		0.95	0.90	0.80	0.20	0.60	0.05	
F5, among		0.99		0.96				0.95			0.90	163
F5, within		0.90		0.70		0.90	0.90	0.80	0.20	0.60	0.05	
F6, among		0.99			0.96			0.95			0.90	5880
F6, within		0.90		0.70		0.90	0.98	0.95	0.10		0.05	
F7, among¶		0.85		0.70	0.98	0.96	0.85	0.70		0.75	0.25	1469
F8(T), among		0.55			0.70	0.99	0.99	0.98	0.90		0.55	861
F8(B), among				0.90							0.90	775
F8(YT), among	0.40										0.40	310
F8(SP), among											1.00	310
F9(T), among		0.97			0.95			0.99	0.99		0.90	280
F9(B), among				0.95							0.95	266
F9(YT), among	0.40										0.40	106
F9(SP), among											1.00	106
F10(YR), among					0.98						0.98	104
F10(LR), among				0.98							0.98	102

† There are four possible selection modes for each trait: *top*, *bottom*, *middle*, and *random*.
‡ 400 is the number of crosses made at the beginning of each breeding cycle.
§ There is no within-family selection in F1 and the F7 generation onwards.
¶ Blank means no selection for that trait or the selected proportion can be viewed as 1.00.

Among-Family Selection and Within-Family Selection for Each Selection Round

Ten traits have been included as relevant (van Ginkel et al., 2002) for the selection process in the breeding program described in Fig. 2. Among-family selection and within-family selection are distinct processes in a breeding strategy. However, the definition of these two types of selection is essentially the same: the number of traits to be selected is followed by the definition of each trait (Fig. 1).

Apart from the trait code (defined in the genotype and environment system) there are two parameters that define a trait used in selection: selected proportion and selection mode. For among-family selection, the selected proportion is the percentage of families to be retained; for within-family selection, it is the percentage of individual plants to be selected in each retained family. There are four options for trait selection mode: (i) *top* (the individuals or families with highest phenotypic values for the trait of interest will be selected, e.g., yield, tillering, grains per spike, and kernel weight), (ii) *bottom* (the individuals or families with the lowest phenotypic values will be selected, e.g., lodging, stem rust, leaf rust, and stripe rust), (iii) *middle* (individuals or families with medium trait phenotypic values will be selected, e.g., height and heading), and (iv) *random* (individuals or families will be randomly selected). Independent culling is used if multiple traits are considered for within-family or among-family selection. If there is no among-family or within-family selection for a specific selection round, the number of selected traits is noted as 0 (Table 2). In generations F1 to F6 in the Selected Bulk method, each family is derived from one distinct cross since the method of bulk generation advance is applied from the F2 onwards. Among-family selection from F1 to F6 is, in fact, among cross selection. The traits for both among-family and within-family selections can be the same or different, as is the case for selected proportions (Table 2). The traits for selection may also differ from generation to generation, as may the selected proportions for traits.

Genetic Models Used to Investigate the Effects of Dominance and Epistasis on Selection

Seven agronomic traits and three rust resistances have been used in the simulation of the Selected Bulk breeding method. The gene number and genetic values were derived from discussions with breeders and from analyses of past unpublished experiments. In total, we postulated that 59 independently segregating genes control these traits (Table 3). The genetic effects for traits other than yield were considered fixed. Pleiotropic effects were included to account for trait correlations, and they were also considered fixed. Two kinds of pleiotropic effects were included (Fig. 3), although more complicated pleiotropic interaction can also be defined within the QU-GENE engine. The first kind is positive pleiotropy, such as the pleiotropic effects on lodging from genes for grains per spike (Fig. 3-a). The second kind is the negative pleiotropy, such as the pleiotropic effects on kernel weight from genes for grains per spike (Fig. 3-b). As shown in Table 3, at Cd. Obregon the three lodging genes, the five stem rust genes, and the five leaf rust genes have some degree of negative effect on yield, and the five kernel weight genes have a positive pleiotropic effect. Stem rust, leaf rust, heading, tillering, and grains per spike genes all have a negative pleiotropic effect on kernel weight (Table 2 in Wang, 2003a). Stripe rust rarely occurs at Cd. Obregon, so there is no selection for stripe rust when the nursery is grown there (Table 2) and the genetic effects of stripe rust genes are considered to be zero in this environment.

Apart from the pleiotropic effects of genes affecting other traits, we postulated that there are 20 genes for yield per se, even though their very existence has been debated (Grafius, 1959). Four gene effect models were considered for yield, i.e., pure additive [AD0, $Aa = (AA+aa)/2$, where A and a represent the two alleles at each locus affecting yield], partial dominance [AD1, genetic value of $Aa \neq (AA+aa)/2$, but is between AA and aa], a combination of partial, complete and

Table 3. Number of segregating genes and their genetic effects in the Cd. Obregon environment type.

Gene classification†	Number of genes	Traits effected	Individual gene effect		
			AA	Aa	aa
Yield	20	yield (Mg/ha)	Four models for yield genes AD0: pure additive AD1: partial dominance AD2: partial or overdominance ADE: digenic epistasis		
Lodging	3	lodging (%)	0.00	5.00	10.00
		yield (Mg/ha)	0.00	−0.40	−0.80
Stem rust	5	stem rust (%)	0.00	0.50	1.00
		yield (Mg/ha)	0.00	−0.25	−0.50
		kernel weight (g)	0.00	−0.75	−1.50
Leaf rust	5	leaf rust (%)	0.00	5.00	10.00
		yield (Mg/ha)	0.00	−0.25	−0.50
		kernel weight (g)	0.00	−0.75	−1.50
Height	3	height (cm)	40.00	30.00	20.00
		lodging (%)	5.00	2.50	0.00
Heading	5	heading (days)	20.00	16.00	12.00
		kernel weight (g)	−1.00	−0.50	0.00
Tillering	3	tillering (no.)	5.00	3.00	1.00
		lodging (%)	2.00	1.00	0.00
		heading (days)	1.00	0.50	0.00
		grains per spike (no.)	−1.00	−0.50	0.00
		kernel weight (g)	−1.50	−0.75	0.00
Grains per spike	5	grains per spike (no.)	14.00	10.00	6.00
		lodging (%)	2.00	1.00	0.00
		kernel weight (g)	−1.00	−0.50	0.00
Kernel weight	5	kernel weight (g)	12.00	8.50	5.00
		yield (Mg/ha)	1.00	0.50	0.00
		lodging (%)	2.00	1.00	0.00

† Five stripe rust genes were not present in this table because of the absence of stripe rust at Cd. Obregon.

overdominance (AD2, the genetic values of *AA*, *Aa*, and *aa* are independent), and digenic interaction (ADE). Following the procedures described by Cooper and Podlich (2002), the genetic effects of 20 yield genes in each environment type were sampled from the uniform distribution before the simulation was run. These sampled gene effects are approximations of the distribution of real gene effects (Cooper et al., 2002). Some genes have relatively large effects and can be viewed as major genes, and others have relatively small effects and can be viewed as minor genes (Fig. 4). For Model AD0 (Fig. 4-1a and -1b), gene 19 has the largest additive effect in the Cd. Obregon environment type and the first allele is the favorable allele (*AA* = 0.91, *Aa* = 0.47, and *aa* = 0.03). It explains more than 20% of genetic variance (variation from pleiotropic genes excluded) in the Cd. Obregon environment in the reference population where all gene frequencies were set at 0.5, but only 5 to 8% in the other two environment types. Gene 19 can be viewed as a major gene at Cd. Obregon, but probably not at Toluca and El Batan. Gene 8 has the second largest additive effect in the Cd. Obregon environment type, with the first allele also being favorable (*AA* = 0.83, *Aa* = 0.51, and *aa* = 0.18). This gene explains 12% of the total genetic variation

and therefore functions as a major gene in the Cd. Obregon environment. However, it explains only a very small amount of genetic variation (less than 1%) in the Toluca and El Batan environment types and is considered a minor gene in these environments. For Model AD1 (Fig. 4-2a and -2b), genes 19, 9, 2, and 3 each contribute more than 12% to genetic variance, and are therefore major genes in the Cd. Obregon environment. Genes 19 and 3 have a positive dominance effect (*a* = 0.34, *d* = 0.22 and *d*/*a* = 0.64 for gene 19, and *a* = 0.23, *d* = 0.01 and *d*/*a* = 0.02 for gene 3, where *a* and *d* are the additive and dominance gene effects, and *d*/*a* is the degree of dominance). Genes 9 and 2 have a negative dominance effect (*a* = 0.26, *d* = −0.22 and *d*/*a* = −0.85 for gene 9, and *a* = 0.28, *d* = −0.13 and *d*/*a* = −0.45 for gene 2). For Model AD2 (Fig. 4-3a and -3b), genes 11, 7, 15, and 17 each contribute more than 10% to the genetic variance in the Cd. Obregon environment. For Model ADE (Fig. 4-4a and -4b), the interaction between genes 1 and 2 resulted in two local peaks and one global peak in the distribution of genotypic values. The importance of each epistasis network can also be represented by the proportion of genetic variance explained by the network. Among the 10 epistasis networks, the interaction be-
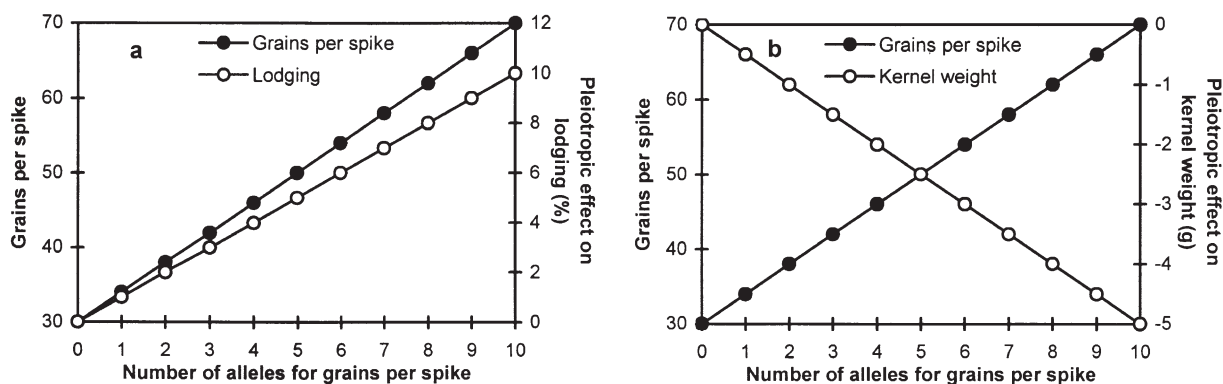

Fig. 3. Two kinds of pleiotropic effects exemplified by the pleiotropic effects of grains per spike on lodging and kernel weight.
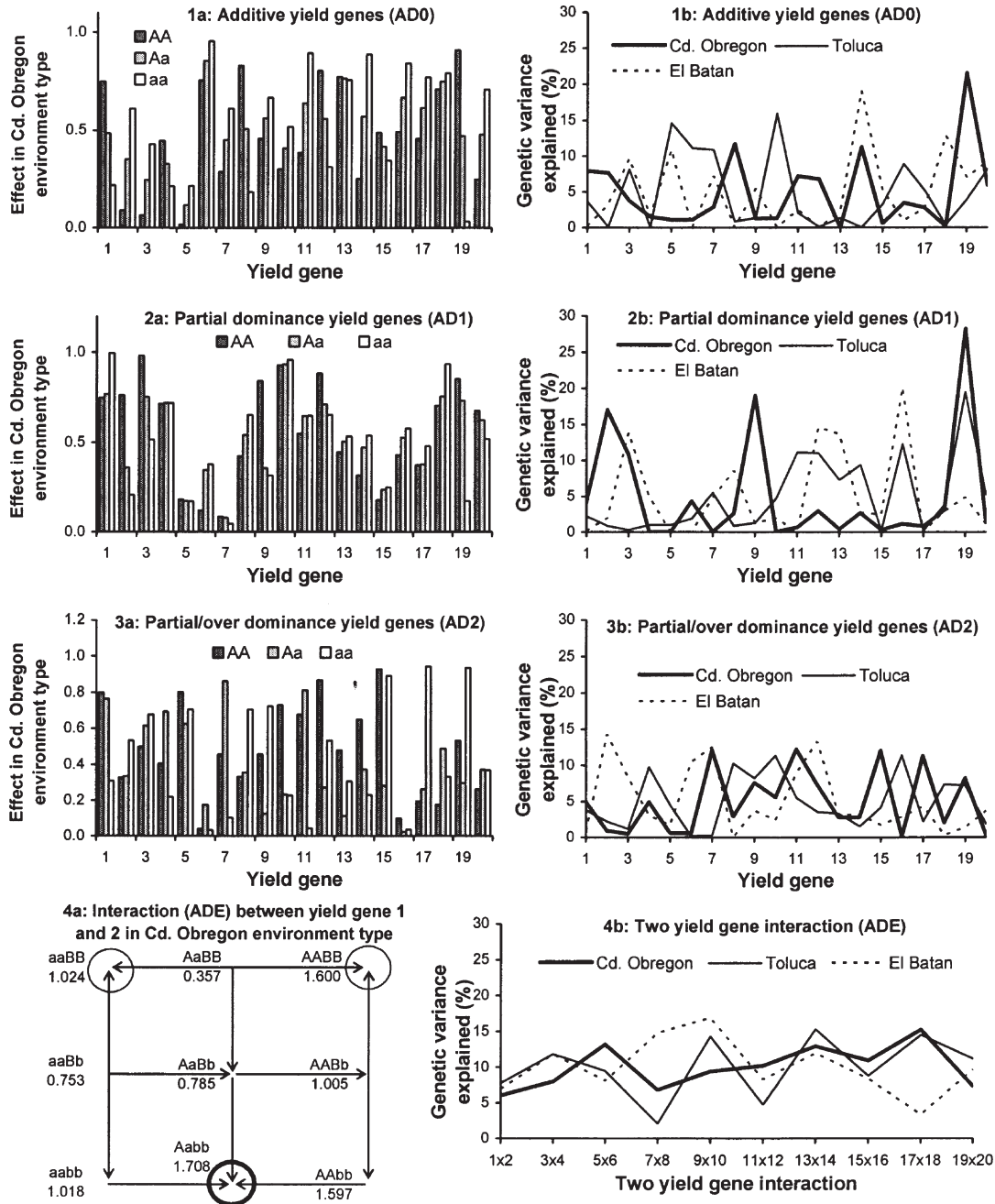
Fig. 4. Four genetic models used to investigate the effects of dominance and digenic epistasis on selection. 1, 2, 3, and 4 represent the four genetic models, i.e., pure additive yield genes (AD0), partial dominance yield genes (AD1), partial or overdominance yield genes (AD2), and digenic epistasis yield genes (ADE); Column a shows the three genotypic values on each of 20 yield genes for Models AD0, AD1, and AD2, and the gene 1 and gene 2 interaction for Model ADE; Column b shows the percentage of genetic variance explained by each gene or each epistasis network. All genes have the same frequency, 0.5, in the reference population used to estimate the genetic variation. Variation from pleiotropic genes was excluded. A and a are the two alleles on each locus. A and a, and B and b are the alleles on the two interacting loci.

tween genes 17 and 18 explains the largest proportion of genetic variance (Fig. 4-4b) in the Cd. Obregon environment type.

Three populations with gene frequencies 0.1, 0.5, and 0.9 were used as initial populations. Here the gene frequency refers to the frequency of the first allele at each locus. Each initial population consisted of 100 homozygous individuals among which 400 crosses were made at the beginning of each breeding cycle. After each cycle of selection, 106 lines were retained in the final selected population and these lines were used as parents for the next cycle. The Selected Bulk selection method was applied to each of the 12 models (AD0, AD1, AD2, and ADE) by population (represented by three gene frequencies 0.1, 0.5, and 0.9) combinations for 20 breeding cycles. This process was repeated 10 times.

RESULTS AND DISCUSSION

Genetic Advance Due to Selection

As there are different scales applied to different genotype and environment systems, the range transformed
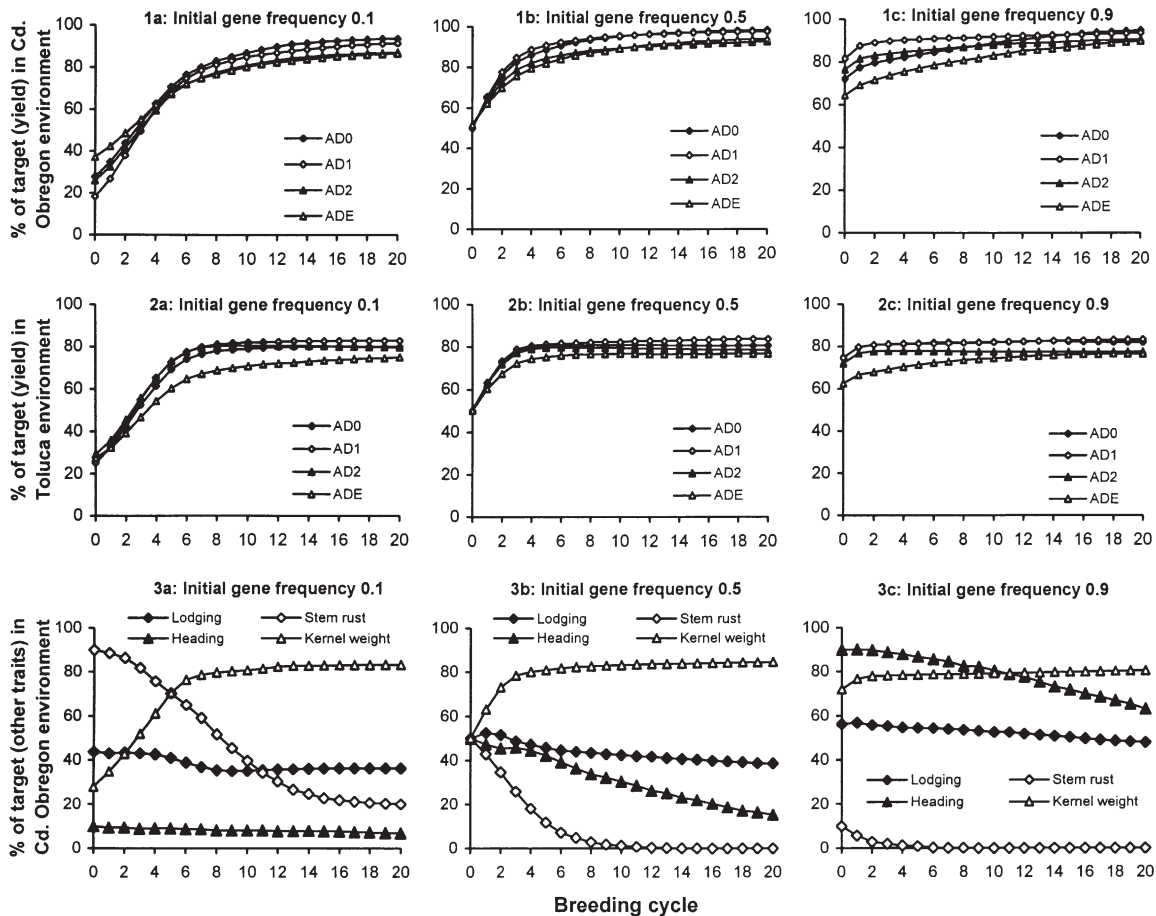
Fig. 5. Effect on genetic advance. 1a, 1b, and 1c: Advances on yield in the Cd. Obregon environment type for the three initial populations identified by gene frequencies 0.1, 0.5, and 0.9, respectively; 2a, 2b, and 2c: Advances in yield in the Toluca environment type for the three initial populations; 3a, 3b, and 3c: Advances on lodging, stem rust, heading and kernel weight in the Cd. Obregon environment type for the three initial populations.

trait value (i.e., genotype value expressed relative to the difference between the worst and best target genotypes) will be used to show the changes that occurred within the selected populations (Wang et al., 2003a). The four models show a similar trend in yield advance. When gene frequency for the first allele at each locus is low in the initial population, the response to selection for yield is greater in the first five to six cycles, after which the response slows down (Fig. 5-1a). When gene frequency is initially set at 0.5, the selection response is greater in the first three cycles (Fig. 5-1b). When the gene frequency is high, the selection response may already start to slow down after just one cycle of selection (Fig. 5-1c). This result is not coincident to the actual data from CIMMYT's wheat breeding program, where the average increase in yield potential per year is estimated at 0.9%, and there is no evidence that a yield plateau has been reached (Rajaram, 1999). One reason may be that the genetic phenomena in the actual breeding program are much more complicated than those considered in this simulation study. Another reason could be the continuous introduction of germplasm from outside breeding programs along with the pathology and wide cross programs within CIMMYT. As new germplasm is introduced into the breeding program, the genotype and environment system may change, such as the gene number and the target genotype. This characteristic was not captured in the current simulation study.

When gene frequency is 0.5 in the initial population, all models have the same starting point, and that being the case, the effects of different models on selection can be more conveniently compared. The purely additive model (AD0) and the partial dominance model (AD1) give a very similar yield advance. The yield advance is slower when overdominance (AD2) and epistasis (ADE) are present (Fig. 5-1b). This is because the most desired genotype could be heterozygous, as determined by overdominance or epistasis (Fig. 4-4a), a condition that cannot be fixed in an inbred breeding program.

As yield in the various environments in this particular genotype and environment system are positively correlated, selection for yield in one environment may also improve yield in the other environments. For example, the genetic correlation of yield at Cd. Obregon with that at Toluca, in the initial population when all gene frequencies were set at 0.5, was $r = 0.56$ for Model AD0, 0.61 for Model AD1, 0.56 for Model AD2, and 0.35 for Model ADE. The genetic correlation coefficient between two environment types j_1 and j_2 for a specific trait is calculated by

$$r_{j_1 j_2} = \frac{COV(g_{ij_1}, g_{ij_2})}{\sqrt{VAR(g_{ij_1})VAR(g_{ij_2})}}$$

where g_{ij_1} and g_{ij_2} are the genotypic values of the trait of interest for individual i in the initial population in the two environment types. In the Selected Bulk method, selection for yield was only conducted at Cd. Obregon. Nevertheless, as a result of this correlation yield in the Toluca environment was also improved (Fig. 5-2a, -2b, and -2c).

Because of different selection modes (e.g., bottom for lodging and stem rust, middle for heading, and top for kernel weight; Table 2), the selection responses of other traits may be distinct. However, similar trends can be identified regardless of the yield genetic models used. In Fig. 5-3a, -3b, and -3c only the results for a purely additive yield gene model (AD0) are presented. In the Selected Bulk, the individuals and families with the lowest lodging severity (percentage) were selected in each generation. However, because of the related pleiotropic effect of genes for grains per spike and kernel weight, lodging in the selected population hardly decreases, no matter which initial population was used. In contrast, stem rust resistance could be improved as demonstrated by decreasing severity values over cycles, as there was no counteracting pleiotropic effect from other genes. Kernel weight increases with selection, as the individuals and families with highest kernel weights were selected. However, it cannot reach the level of the most desired target genotype because of the negative pleiotropic effects from the other two yield components, tillering and fects from the other two yield components, tillering and grains per spike, that when actively selected for will reduce kernel weight.

Gene Frequency

The change of gene frequency after one cycle of selection is proportional to the relative size of the additive gene effect (Falconer, 1989). This must then also be true for several cycles of selection and for various genetic models without epistasis. In Model AD0, only additive gene effects influence yield, and hence changes in yield must coincide directly with changes in gene frequency. In response to selection, genes with large additive effects change their gene frequencies faster, while those with small additive effects change more slowly. For partial dominance yield genes (AD1), gene 19 has the largest additive gene effect, followed by genes 9, 2, and 3. The changing rate in allele frequency after selection followed this order when all gene frequencies in the initial population were set at 0.5 (Fig. 6-1b). Gene 12 is an exception when gene frequencies in the initial population are set at 0.1. It has a smaller additive effect than gene 9, but responds with a faster increase in first-allele frequency. The size of the additive gene effect also reflects a change in the gene frequency. For example, genes 6 and 8 in Model AD1 and genes 8 and 17 in Model AD2 have negative additive effects for the nominated first-allele. In accordance with these negative additive values, the first-allele frequencies for these loci decrease compared to their frequency in the initial population (Fig. 6-1a, -1b, -1c, -2a, -2b, and -2c).
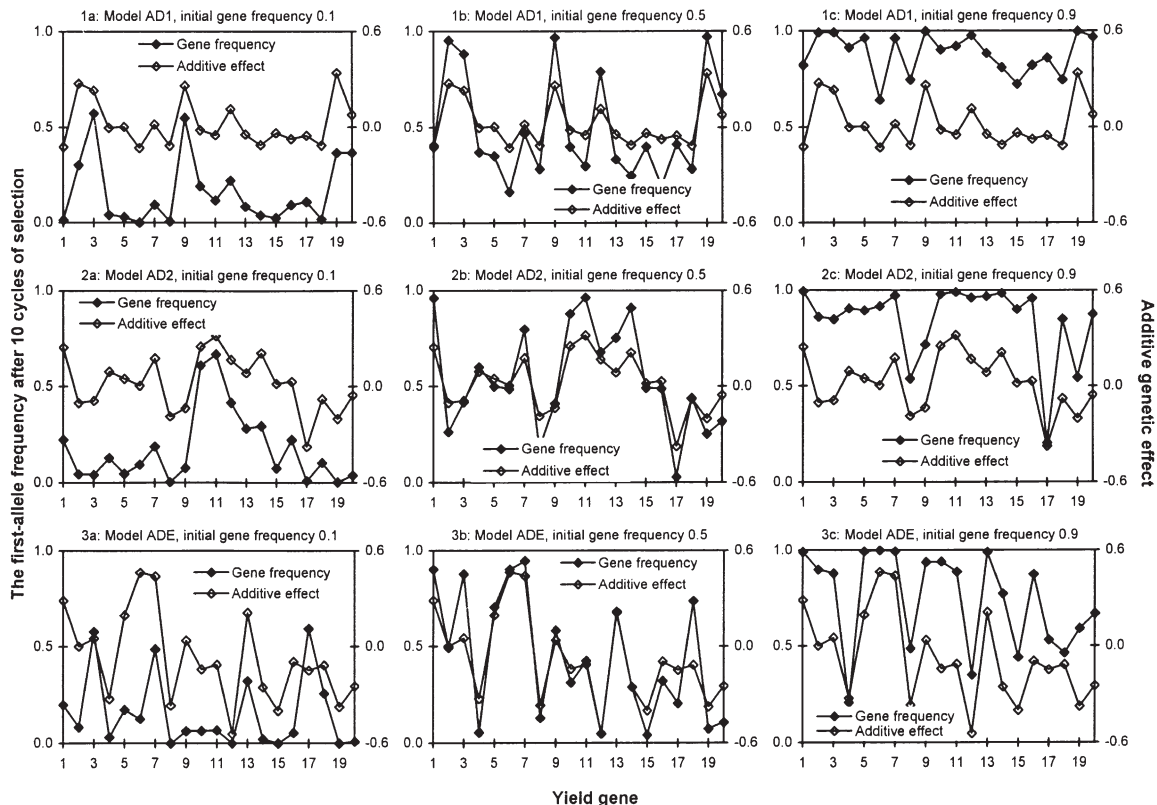

Fig. 6. Gene frequency in the final selected population following 10 cycles of selection. 1, 2, and 3 represent the three genetic models, i.e., partial dominance yield genes (AD1), partial or overdominance yield genes (AD2), and digenic epistasis yield genes (ADE); a, b, and c correspond to the three initial populations identified by gene frequencies 0.1, 0.5 and 0.9, respectively.

Table 4. Genetic values on yield of the four homozygous genotypes in each epistasis network in the Cd. Obregon environment type.

Epistasis network	Genotype				Additive effect	
	AABB†	AAbb	aaBB	aabb	First gene	Second gene
1 × 2	*1.600*‡	1.597	1.024	1.018	0.289	0.002
3 × 4	0.194	*1.636*	0.879	0.739	0.053	−0.326
5 × 6	*1.473*	0.263	0.799	0.152	0.196	0.464
7 × 8	0.956	*1.701*	0.089	0.806	0.441	−0.366
9 × 10	0.821	0.793	0.435	*1.026*	0.038	−0.141
11 × 12	0.648	1.483	0.621	*1.957*	−0.112	−0.543
13 × 14	1.511	*1.759*	0.831	1.594	0.211	−0.253
15 × 16	0.567	0.773	1.381	*1.551*	−0.398	−0.094
17 × 18	0.082	1.682	*1.746*	0.608	−0.148	−0.112
19 × 20	0.481	0.435	0.693	*1.725*	−0.376	−0.247

† *A* and *a* are the two alleles at the first locus, and *B* and *b* are the two alleles at the second locus in each epistasis network.
‡ The largest value among the four genetic values in each epistasis network is presented in italic type.

In the case of epistasis between yield genes (ADE), the additive effects can also be estimated (Cheverud and Routman, 1995; Kearsey and Pooni, 1996; Holland, 2001), and these estimates also provided suitable indicators for the change in gene frequency (Fig. 6-3a, -3b, and -3c). However, this is not generally the rule. For example, in the epistasis network 17 × 18, both genes have negative additive effects (Table 4). Nevertheless, the frequency of gene 17 went up and the frequency of gene 18 went down. This is not surprising when the genetic values of these four homozygous genotypes are examined (Table 4). Genotype *aaBB* has the largest value, which was favored by selection. So the frequency of *aaBB* in the selected population increases, and as a result, the frequency of allele *A* at gene 17 decreased and the frequency of allele *B* at gene 18 increased (Fig. 6-3a, -3b, and -3c).

Genetic Variance Components

The allele combination of any individual in a population can be identified when simulating, from which the genotypic value is calculated based on the defined gene effects in the genotype and environment system. Then the phenotypic value in any specific environment is determined from its genotypic value and estimates of associated random errors (i.e., within plot error and among plot error). Within-family selection is made based on the phenotypes in the selected family, and among-family selection is made based on phenotypic family means. The genetic variance and different variance components can therefore be calculated from the genotypic values. In an inbred breeding program, the final retained population consists of only homozygous or nearly homozygous lines after many generations of self-pollination. So the selected population after each breeding cycle was randomly mated a few times before the genetic variance and variance components were estimated. A total of 500 half sibs and 500 full sibs were generated from the randomly mated population, and the covariance among half sibs and the covariance among full sibs were estimated. The additive variance was estimated by $V_A = 4COV_{(HS)}$, and the dominance variance was estimated by $V_D = 4[COV_{(FS)} − 2COV_{(HS)}]$, where $COV_{(HS)}$ is the covariance among half sibs, and $COV_{(FS)}$ is the covariance among full sibs. The interaction variance was then estimated by $V_I = V_G − V_A − V_D$ for all models (Falconer, 1989), where V_G is the total genetic variance. As expected, additive variance is the only component noted

for model AD0 (Fig. 7-1a, -1b, and -1c). When gene frequency in the initial population is 0.1, the additive variance increases in the first two or three breeding cycles, and then decreases rapidly to a low level. When gene frequencies are set at 0.5 and 0.9, the additive variance decreases rapidly in the first few breeding cycles, and slows down afterwards. This change in additive variance can also be seen in other models (Fig. 7-2a, -2b, -2c, -3a, -3b, and -3c). For the partial dominance model (AD1), there should be some dominance variance. However, it was estimated as null from the covariance among half sibs and the covariance among full sibs (Fig. 7-2a, -2b, and -2c), indicating that the dominance variance from partial dominance is either small or hard to detect. The dominance variance can be observed in the partial or overdominance model (AD2), and very probably it is the overdominance among yield genes that contributed the largest portion to this variance component. However, it remains at almost the same level from cycle to cycle (Fig. 7-3a, -3b, and -3c). The reason for this is that the overdominance results in stabilizing selection, maintaining heterozygosity in the population rather than driving one allele to fixation.

Both dominance variance and interaction variance can be observed with digenic epistasis yield genes (ADE) (Fig. 7-4a, -4b, and -4c). In the case of epistasis,

$$COV_{(HS)} = 1/4V_A + 1/16V_{AA}$$
$$COV_{(FS)} = 1/2V_A + 1/4V_D + 1/4V_{AA} +$$
$$1/8V_{AD} + 1/16V_{DD}$$

where V_{AA} is the additive by additive variance, V_{AD} is the additive by dominance variance, and V_{DD} is the dominance by dominance variance (Falconer 1989). So we have

$$4COV_{(HS)} = V_A + 1/4V_{AA}$$
$$4[COV_{(FS)} − 2COV_{(HS)}] = V_D + 1/2V_{AA} +$$
$$1/2V_{AD} + 1/4V_{DD}$$
$$V_I = 1/4V_{AA} + 1/2V_{AD} + 3/4V_{DD}$$

Consequently, a quarter of the additive × additive variance is contained within the estimate of interactive variance when epistasis is present, and all three kinds of interaction variance are included in the estimate of dominance variance. Therefore the actual interaction variance (V_I) is always underestimated. As shown in Fig. 7-1, -2, and -3, the pure additive variance decreases to a low
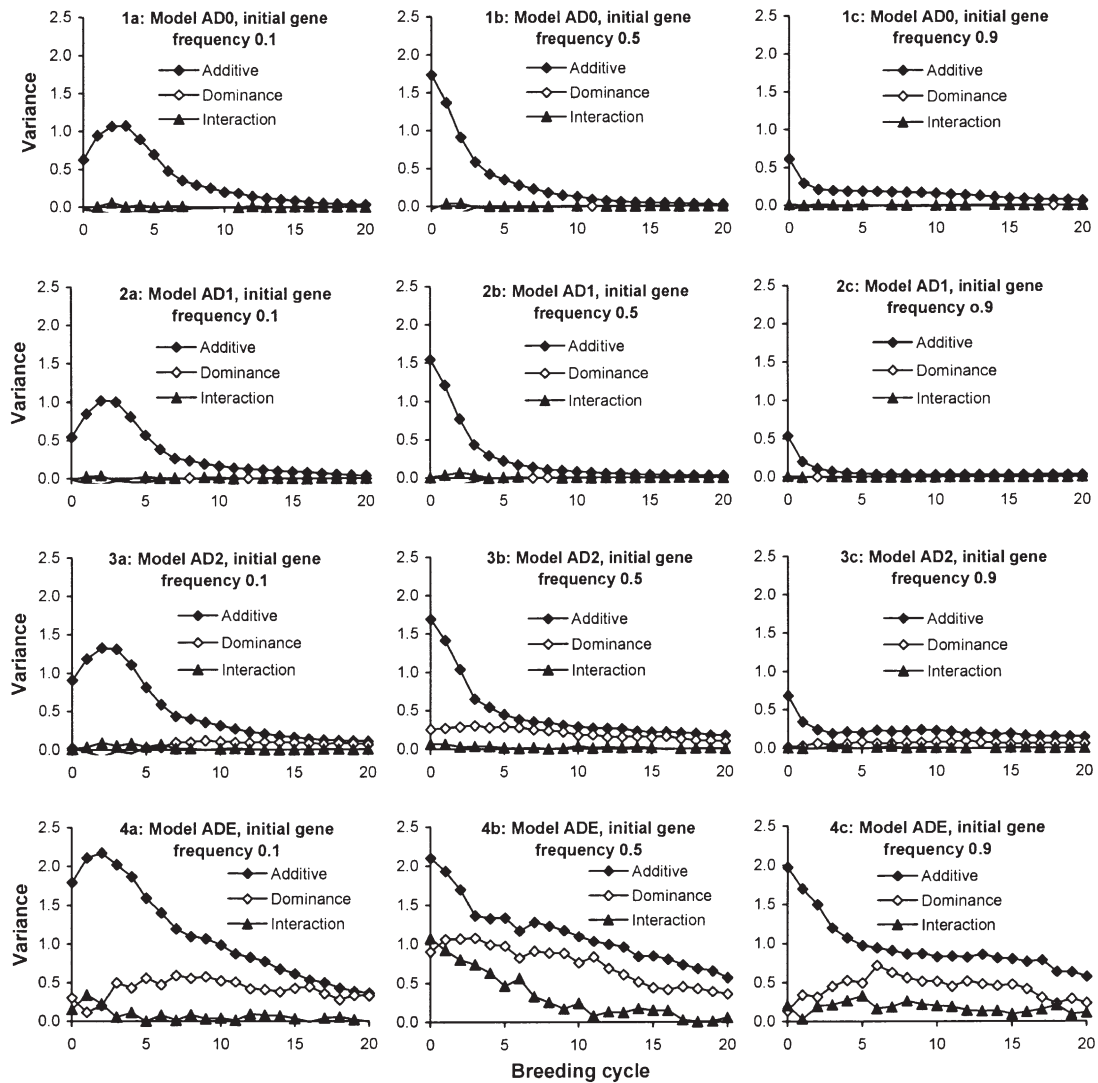
Fig. 7. Changes in variance components following selection. 1, 2, 3, and 4 represent the four genetic models, i.e., pure additive yield genes (AD0), partial dominance yield genes (AD1), partial or overdominance yield genes (AD2), and digenic epistasis yield genes (ADE); a, b, and c correspond to the three initial populations identified by gene frequencies 0.1, 0.5, and 0.9, respectively.

level after 10 cycles of selection. So we may suppose that for Model ADE, the additive × additive variance is the largest portion of the additive variance estimate after a few cycles of selection. The higher level of additive variance with epistasis (Fig. 7-4a, -4b, and -4c) means that selection can still be effective; it just takes more cycles to fix the additive × additive interaction. Figure 5-1b indicates that yield under the Model ADE can be further improved with more cycles of selection.

The Reality of Genotype and Environment Systems in Simulation

As in field breeding, QuCim conducts within-family selection from individual phenotypic values in each family and among-family selection from family means. The genotypic value of an individual was calculated from the definition of gene actions in the genotype and environment system. The phenotypic value and family mean was estimated from the genotypic value and its associated error (random environmental deviation). A sensi-

ble definition of the genotype and environment system is thus essential to any such simulation, since it determines the phenotypic value of a genotype and then the phenotypic mean of a population to which selection is applied. However, given the current state of our knowledge of gene-to-phenotype relationships for complex traits, it is difficult to comprehensively define a real genotype and environment system. It is therefore not possible to ensure that the genotype and environment systems used in this simulation experiment match the biophysical systems within which CIMMYT's wheat breeding program operates.

However, a large amount of data from CIMMYT's historical breeding records has been used to define the genotype and environment systems as realistically as possible. The systems used in the current simulation research result in similar trait correlations, environment correlations, and trait heritabilities as derived from the field data. From a previous simulation experiment (Wang et al., 2003b), we found that the linkage (e.g., recombina-
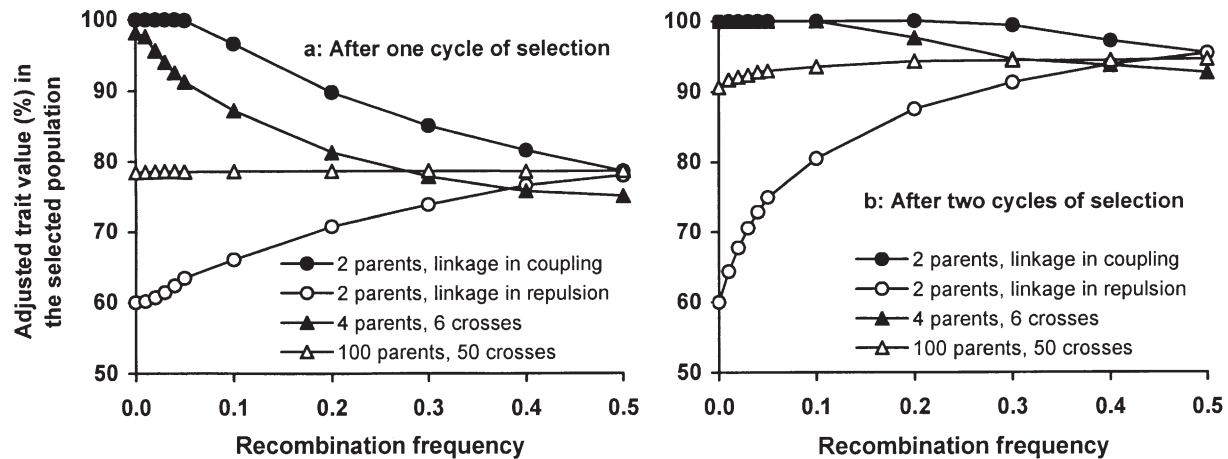
Fig. 8. The effect of linkage phase and degree on response to selection.

tion frequency 0.05) has only a small effect on selection for a large breeding program, in which a large number of crosses have been made. Therefore, linkage was not included in this study.

As far as a specific cross is concerned, linkage does affect the response to selection. Linkage in repulsion delays the response while linkage in coupling favors the response (Allard, 1960). The following example illustrates the effect of linkage on selection for single crosses. Twenty genes are supposed to be evenly distributed on 4 chromosomes, each having 5 genes. The recombination frequency between two neighboring genes is set at 0.00, 0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.20, 0.30, 0.40, and 0.50 (Fig. 8). Two crosses are made from two pairs of parents with genotypes

$$\frac{11111,22222,11111,22222}{11111,22222,11111,22222} \text{ and } \frac{22222,11111,22222,11111}{22222,11111,22222,11111} \text{ and}$$

$$\frac{12121,21212,12121,21212}{12121,21212,12121,21212} \text{ and } \frac{21212,12121,21212,12121}{21212,12121,21212,12121},$$

respectively, where 1 represents the favorable allele at each locus and 2 the nonfavorable allele, and the upper and lower sequences represent the four pairs of homologous chromosomes. For the first cross, the genes are linked in coupling, but for the second cross the genes are linked in repulsion. One thousand individual plants were generated in each of the two F2 populations, and single seed descent was used from F3 to F6 to derive pure lines. Ten lines were selected in F6 on the basis of the performance of the trait of interest. Environmental effects were not included to minimize the effects from other factors. As expected, the selected populations from the two crosses have the same trait performance, given that all genes are unlinked (i.e., recombination frequency is 0.50). As the degree of linkage increases, the population from the first cross (coupling phase linkage) displays increasingly superior performance compared to that of the second cross (repulsion phase linkage). It is obvious that the linkage has a tendency to keep the coupling genes together in the first cross, so that the target genotype can be achieved more easily than in the second cross in which the genes are in repulsion (Fig. 8). When six distinct single crosses were made from all four parents, linkage increased the selection

response after one or two cycles of selection. In contrast, when 100 parents are involved in crossing, the linkage distance does not make any difference in genetic gain after the first cycle of selection. Only the very close linkage (recombination frequency less than 0.02) shows a little difference after two cycles of selection. It delays the genetic advance in the specific population used (Fig. 8-a and -b).

In the future, it will be possible to build more realistic genotype and environment systems if advances in genomics improve our understanding of the genotype-to-phenotype relationship and genotype × environment interactions. This information will be useful in determining gene number and gene effects on phenotype. Conclusions on the relative merits of breeding strategies based on simple gene-to-phenotype models may have to be reevaluated in the context of an exponentially growing knowledge base.

CONCLUSIONS

The QU-GENE engine provides a practical way to define a complicated genotype and environment system, which contains linkage, epistasis, multiple alleles, pleiotropy, molecular markers, and genotype × environment interaction (Podlich and Cooper, 1998). Meanwhile, QuCim provides a flexible way to define complex selection strategies such as the pedigree system, bulk population system, doubled haploid breeding, backcross breeding, and within-population recurrent selection. By using the QU-GENE engine and the QuCim application module, different breeding strategies can be simulated on the basis of various genotype and environment systems. Two distinctly different applications may be considered when using QU-GENE and QuCim. One application allows different selection strategies to be compared among a large number of genetic models, from which the most efficient strategy may be identified. This leads to practical modifications in breeding programs (Podlich et al., 1999; Wang et al., 2003a). The other application allows the effects on selection of different genetic models to be investigated, as presented in this paper. As a result of the second application, some classical quantitative genetics theories based on simplified assumptions

may be tested or improved. Kempthone (1988) and Comstock (1996) advocated this approach to simulation.

We used QuCim to investigate some of the possible implications of increasing the genetic complexity of traits on the outcomes from the CIMMYT breeding program. This raises the question of how to introduce a range of simple to complex genetic models in simulation experiments. This paper reports a preliminary investigation in which our objective was to compare some specific models relative to the classical additive model (AD0). The models of interest, selected in consultation with the CIMMYT breeders, were to introduce partial dominance (AD1), partial or overdominance (AD2), and digenic epistasis (ADE). Analyses were then conducted to make specific comparisons among more complex models and AD0, to measure impact on genetic advance, gene frequency, and components of genetic variance when we challenge the assumption of additivity.

The simulation results indicate that the fully additive model provides similar outcomes regarding genetic advance as the partial dominance model, when all genes have the same frequency of 0.5 in the initial population. When gene frequency is not 0.5, the initial population is located at different points for different models (Fig. 5), making it more complicated to compare the effects of various models. However, genetic advance in genetic systems consisting of pure additive or partial dominance gene effects are generally faster in reaching the target genotype than systems with considerable overdominance and epistasis. The additive gene effect is an appropriate indicator of the change in gene frequency following selection for genetic models without epistatic effects. The gene with the largest additive effect changes its frequency fastest in response to selection. The positive or negative sign of the additive gene effect indicates the direction in which allele frequency will change. When epistasis is absent, the additive variance decreases rapidly following selection. Nevertheless, additive variance remains at a certain level for many breeding cycles when epistasis effects are active. This suggests that it could take a considerably longer time to fix useful additive $\times$ additive interaction in an inbred breeding program. This realization may motive breeders to take other, swifter measures than further selection, such as producing large doubled haploid populations to identify desired recombinants. The variance from partial dominance is small and therefore hard to detect by using the covariance among half sibs and the covariance among full sibs. The variance from overdominance can be properly identified, but it does not change significantly following selection. This is because overdominance cannot be effectively utilized by an inbred breeding program.

REFERENCES

Allard, R.W. 1960. Principles of plant breeding. John Wiley and Sons, New York.

Casali, V.W.D., and E.C. Tigchelaar. 1975. Computer simulation studies comparing pedigree, bulk, and single seed descent selection in self pollinated populations. J. Am. Soc. Hortic. Sci. 100:364–367.

Cheverud, J.M., and E.J. Routman. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461.

Comstock, R.E. 1996. Quantitative genetics with special reference to plant and animal breeding. Iowa State University Press, Ames.

Cooper, M., and D.W. Podlich. 2002. The E(NK) model: Extending the NK model to incorporate gene-by-environment interactions and epistasis for diploid genomes. Complexity 7:31–47.

Cooper, M., D.W. Podlich, N.W. Jensen, S.C. Chapman, and G.L. Hammer. 1999. Modelling plant breeding programs. Trends Agron. 2:33–64.

Cooper, M., D.W. Podlich, K.P. Micallef, O.S. Smith, N.M. Jensen, S.C. Chapman, and N.L. Kruger. 2002. Complexity, quantitative traits and plant breeding: A role for simulation modelling in the genetic improvement of crops. p. 143–166. *In* M.S. Kang (ed.) Quantitative genetics, genomics and plant breeding. CAB International, Wallingford, UK.

Falconer, D.S. 1989. Introduction to quantitative genetics. 3rd ed. Longman, Essex, UK.

Frisch, M., and A.E. Melchinger. 2001. Marker-assisted backcrossing for simultaneous introgression of two genes. Crop Sci. 41:1716–1725.

Grafius, J.E. 1959. Heterosis in barley. Agron. J. 51:551–554.

Holland, J.B. 2001. Epistasis and plant breeding. Plant Breed. Rev. 21:27–92.

Jensen, N.F. 1988. Plant breeding methodology. John Wiley and Sons, New York.

Kearsey, M.J., and H.S. Pooni. 1996. The genetical analysis of quantitative traits. Chapman and Hall, London.

Kempthone, O. 1988. An overview of the field of quantitative genetics. p. 47–56. *In* B.S. Weir et al. (ed.) Proceedings of the second international conference on quantitative genetics. Sinauer Associate Inc., Sunderland, MA.

Podlich, D.W., and M. Cooper. 1998. QU-GENE: A platform for quantitative analysis of genetic models. Bioinformatics14:632–653.

Podlich, D.W., M. Cooper, and K.E. Basford. 1999. Computer simulation of a selection strategy to accommodate genotype-environment interaction in a wheat recurrent selection programme. Plant Breed. 118:17–28.

Rajaram, S. 1999. Historical aspects and future challenges of an international wheat program. p. 1–17. *In* M. van Ginkel et al. (ed.) *Septoria* and *Stagonospora* diseases of cereals: A compilation of global research. CIMMYT, Mexico, D.F., Mexico.

Reddy, B.V.S., and R.E. Comstock. 1976. Simulation of the backcross breeding method. I. Effect of heritability and gene number on fixation of desired alleles. Crop Sci. 16:825–830.

Stoskopf, N.C. 1993. Plant breeding. Westview Press, Boulder, CO.

Trethowan, R.M., J. Crossa, M. van Ginkel, and S. Rajaram. 2001. Relationships among bread wheat international yield testing locations in dry areas. Crop Sci. 41:1461–1469.

van Berloo, R., and P. Stam. 1998. Marker-assisted selection in autogamous RIL populations: A simulation study. Theor. Appl. Genet. 96:147–154.

van Ginkel, M., R. Trethowan, K. Ammar, J. Wang, and M. Lillemo. 2002. Guide to bread wheat breeding at CIMMYT (rev.). Wheat special report no. 5. CIMMYT, Mexico, D.F., Mexico.

van Oeveren, A.J., and P. Stam. 1992. Comparative simulation studies on the effects of selection for quantitative traits in autogamous crops: Early selection versus single seed decent. Heredity 69:342–351.

Wang, J., M. van Ginkel, D. Podlich, G. Ye, R. Trethowan, W. Pfeiffer, I.H. DeLacy, M. Cooper, and S. Rajaram. 2003a. Comparison of two breeding strategies by computer simulation. Crop Sci. 43:1764–1773.

Wang, J., M. van Ginkel, R. Trethowan, G. Ye, I.H. DeLacy, D. Podlich, and M. Cooper. 2003b. QuCim: A software program that simulates inbreeding programs is exemplified by the role of pleiotropy, dominance, epistasis and linkage on selection. p. 252–253. *In* CIMMYT 2003. Book of abstracts: Arnel H. Hallauer International Symposium on Plant Breeding, Mexico City, Mexico. 17–22 August 2003. CIMMYT, Mexico, D.F., Mexico.