BIOINFORMATICS

Vol. 14 no. 7 1998
Pages 632–653

QU-GENE: a simulation platform for quantitative analysis of genetic models

D. W. Podlich and M. Cooper

School of Land and Food, The University of Queensland, Brisbane, Queensland 4072, Australia

Received on March 30, 1998; revised and accepted on May 18, 1998

Abstract

Motivation: *Classical quantitative genetics theory makes a number of simplifying assumptions in order to develop mathematical expressions that describe the mean and variation (genetic and phenotypic) within and among populations, and to predict how these are expected to change under the influence of external forces. These assumptions are often necessary to render the development of many aspects of the theory mathematically tractable. The availability of high-speed computers today provides opportunity for the use of computer simulation methodology to investigate the implications of relaxing many of the assumptions that are commonly made.*

Results: *QU-GENE (QUantitative-GENEtics) was developed as a flexible computer simulation platform for the quantitative analysis of genetic models. Three features of the QU-GENE software that contribute to its flexibility are (i) the core E(N:K) genetic model, where E is the number of types of environment, N is the number of genes, K indicates the level of epistasis and the parentheses indicate that different N:K genetic models can be nested within types of environments, (ii) the use of a two-stage architecture that separates the definition of the genetic model and genotype–environment system from the detail of the individual simulation experiments and (iii) the use of a series of interactive graphical windows that monitor the progress of the simulation experiments. The E(N:K) framework enables the generation of families of genetic models that incorporate the effects of genotype-by-environment ($G \times E$) interactions and epistasis. By the design of appropriate application modules, many different simulation experiments can be conducted for any genotype–environment system. The structure of the QU-GENE simulation software is explained and demonstrated by way of two examples. The first concentrates on some aspects of the influence of $G \times E$ interactions on response to selection in plant breeding, and the second considers the influence of multiple-peak epistasis on the evolution of a four-gene epistatic network.*

Availability: *QU-GENE is available over the Internet at {http://pig.ag.uq.edu.au/qu-gene/}*

Contact: *m.cooper@mailbox.uq.edu.au*

Introduction

Quantitative genetics provides the principal and most widely used theoretical framework that is used to link the genotype and phenotype of individuals and populations of individuals within genotype–environment systems. Among its many uses, this framework has been applied in agriculture to design plant and animal breeding strategies. In evolutionary studies, it is used to study the factors that determine the genetic structure of natural populations and how this can change under the influence of external forces, such as selection, population size, migration patterns among sub-populations and mutation.

Simplifying assumptions are often used in the application of the quantitative genetics framework to theoretical and applied problems. Attitudes to these assumptions are variable, but it is acknowledged that they are often necessary to render many of the problems tractable to some form of algebraic solution. Common assumptions include: Mendelian inheritance, no mutation, infinite populations, Hardy–Weinberg equilibrium, many genes with equal and small effects, no linkage or linkage phase equilibrium, no epistasis, no genotype-by-environment ($G \times E$) interactions, no correlated environmental effects. While these and other assumptions are widely used, there is always interest in the implications of removing the restrictions imposed by making them, particularly when experimental evidence suggests that this is appropriate.

With the increasing availability of high-speed computers and more powerful software, Kempthorne (1988) made a strong case for investigating more general (or less restricted) genetic models by means of computer simulation methodology. An increasing number of applications of computer simulation methodology can be found in the areas of evolutionary genetics (Bürger *et al*., 1989; Kauffman, 1993), plant breeding (Cress, 1967; Cox, 1995), animal breeding (Jeyaruban and Gibson, 1996) and teaching of genetics (St Martin and Skavaril, 1984; Partner *et al*., 1993; Tinker and Mather, 1993). Fraser and Burnell (1970) discussed some general principles involved in the use of computer simulation methods in genetics. Also within the field of evolutionary programming and genetic algorithms, there has been some interest in the use of animal breeding strategies to design algorithms that search for optimal solutions to problems that

generate complex response surfaces (Mühlenbein and Schlierkamp-Voosen, 1993).

A general simulation platform with a level of flexibility that allows investigation of a wide range of issues in quantitative genetics would be useful. To date, most applications of simulation tools have been developed in a way that is specific to the problems under investigation, resulting in a nexus between the software and the problem, and ultimately reducing the flexibility of the software for use in further investigations. The QU-GENE (QUantitative-GENEtics) software (Podlich and Cooper, 1997) was developed to provide a simulation platform with a high degree of flexibility for quantitative analysis of genetic models.

With the case for the use of computer simulation methodology in quantitative genetics already made (Kempthorne, 1988), the objectives of this paper are (i) to explain the basic architecture of the QU-GENE software environment and (ii) to demonstrate the application of QU-GENE by way of two examples. Both examples relax assumptions that are commonly made and compare theoretical expectations with the results of a computer simulation experiment. The first example considers the joint influence of G × E interactions and finite samples of environments in multi-environment trials (METs) on the response to selection that is achieved by a plant breeding programme. The results are compared to theoretical considerations given to this problem by Cooper *et al*. (1996). The second example considers the influence of multiple-peak epistasis on selection response and is based on an example reported by Wright (1963).

Systems and methods: structure and operation of QU-GENE

A schematic representation of the two-stage architecture of QU-GENE is depicted in Figure 1. The QU-GENE software platform consists of two major components: (i) the engine (referred to as QUGENE), which is used to define the genetic model for the genotype–environment system; (ii) the application modules that are used to investigate, analyse or manipulate populations of genotypes within the defined genotype–environment system (Podlich and Cooper, 1997). With this two-stage architecture, there is no unnecessary nexus between the definition of the genetic model for the genotype–environment system and the specific problems investigated by the individual application modules. Only core information on the genotype–environment system is generated by the engine. The detail required for specific simulation experiments is defined or generated in the application modules. Thus, simulation experiments are designed and implemented by developing an application module that interacts with the information generated by the engine. With this architecture, QU-GENE has a high degree of flexibility and enables the conduct of many alternative simulation experiments for a given genotype–environment system.

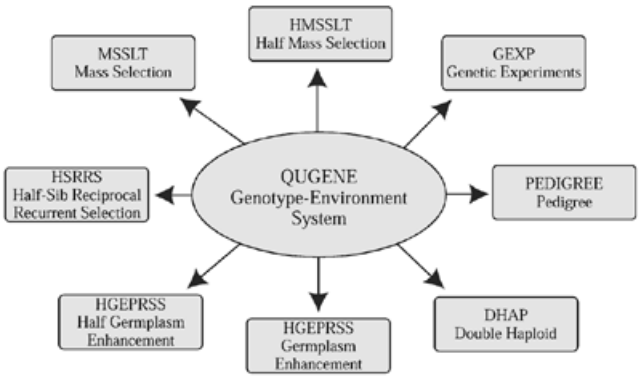

Fig. 1. Schematic outline of the structure of the QU-GENE simulation software. The central ellipse represents the engine (QUGENE) and the surrounding boxes represent the application modules.

The application modules currently available were developed to represent alternative plant breeding strategies. They include:

1. mass selection where selection is for both female and male parents (MSSLT) or only for the female parent (HMSSLT);
2. pedigree and single-seed-descent breeding strategies (PEDIGREE);
3. double-haploid breeding strategies (DHAP);
4. S1 recurrent selection strategies where selection is for both female and male parents (GEPRSS; Fabrizius *et al*., 1996) or for only one parent by use of a dominant male sterile gene (HGEPRSS; Cooper and Podlich, 1997);
5. half-sib reciprocal recurrent selection strategies (HSRRS; Cooper and Podlich, 1998).

The GEXP application module implements a number of the commonly used genetic experiments: bi-parental progenies, North Carolina I, II and III, triple test cross, diallel, generation means and variances analysis, random homozygous lines and a number of molecular marker-facilitated linkage analysis scenarios. Several modules have been developed to assist in the teaching of population and quantitative genetics. These range in complexity from one-gene models (POPGEN) to multiple-gene models (MPOPGEN1 and MPOPGEN2), but they will not be discussed here.

The E(N:K) model for a genotype–environment system

The question of defining an appropriate genetic model to represent the genetic architecture of quantitative traits for a genotype–environment system is a challenging topic. Linear

statistical models have been widely used (e.g. Fisher, 1918; Kempthorne, 1957; Mather and Jinks, 1982; Falconer and Mackay, 1996; Kearsey and Pooni, 1996). Wright (1932) defined the relative performance of genotypes in terms of the concept of population flow on a fitness landscape. A common analogy, though not necessarily sufficient, is a population of individuals exploring a mountain range, where peaks represent genotypes of high fitness and valleys low fitness. More recently, Kauffman (1989, 1993) has used the fitness landscape model for computer simulation studies applied to a wide range of applications in evolutionary genetics. Cooper and Hammer (1996) discussed the use of the landscape model as a framework for investigating plant adaptation and crop improvement in an agricultural context.

The underlying framework used in QU-GENE to define a genotype–environment system combines the deterministic and stochastic features of both linear statistical and landscape models, and is referred to as the $E(N{:}K)$ model. E is the number of different types of environments in the genotype–environment system. The number of different types of environments and their frequency of occurrence define a target population of environments (TPE). N is the total number of genes involved in the genomic network controlling the expression of the traits and K indicates the level of epistasis in the genotype–environment system. It is defined as:

$$K = \sum_{i=1}^{E} e_i \, K_i$$

where E is as defined above, e_i is the frequency of occurrence (expressed as a proportion) of environment type i in the TPE, and K_i indicates the level of epistasis in environment type i. Following the definition of K in Kauffman's (1989, 1993) NK model, K_i is the average number of genes which epistatically affect the fitness contribution of each gene for environment type i. Thus, K in the $E(N{:}K)$ model is the weighted average level of epistasis in the genotype–environment system. Let N_i be defined as the number of genes being expressed in environment type i, while N is the number of distinct genes in the total genotype–environment system. The $E(N{:}K)$ notation identifies that different $N_i{:}K_i$ genetic models can be nested within the different types of environments that make up the TPE. Thus, the $E(N{:}K)$ model is a generalization of the NK model developed by Kauffman (1989, 1993). For consistency with expression of the model in terms of numerals, we denote the $E(N{:}K)$ model with a colon between N and K, as distinct from Kauffman who represents his models without the colon (NK). By including the E term, this form of the NK model can incorporate the influence of G × E interactions.

To demonstrate the application of the $E(N{:}K)$ notation to genotype–environment systems, consider a simple example where the TPE consists of two types of environment ($E = 2$) and 10 genes contribute value to the performance of the individuals in the population ($N = 10$). If we consider that two of the genes are operating in a digenic epistatic network (and the remaining eight genes are additive) in the first environment, the computed level of epistasis for environment 1 is $K_1 = (2 \times 1 + 8 \times 0)/10 = 0.2$. If the epistatic network did not operate in the second environment (all 10 genes additive), the computed level of epistasis for environment 2 is $K_2 = (10 \times 0)/10 = 0.0$. Thus, if we assume that the frequency of occurrence of each of the two environments in the TPE is 0.7 and 0.3, respectively, the computed level of epistasis in the model (K) is $0.7 \times K_1 + 0.3 \times K_2 = 0.14$. Therefore, this example belongs to the family of $E(N{:}K) = 2(10{:}0.14)$ genotype–environment system models.

The $E(N{:}K)$ model provides flexibility for classifying and investigating a wide range of genetic models. At one extreme, the specification of one target environment and no epistasis, i.e. $E(N{:}K) = 1(N{:}0)$, gives the family of classical additive quantitative genetic models for a single TPE that are discussed in most introductory genetics text books. With $E > 1$, G × E interactions can be explicitly introduced into the genotype–environment system and multiple target environments can be incorporated into the TPE. With $K > 0$, epistasis can be explicitly incorporated into the genetic model. Thus, by manipulating the factors E, N and K, families of genetic models ranging from simple to complex can be generated by the engine for investigation within the application modules. Following the discussions by Kauffman (1993), the complexity of each of these genotype–environment systems can be quantified in relation to the ruggedness of the fitness (or adaptation) landscape that they generate. For example, the simple additive model $E(N{:}K) = 1(N{:}0)$ represents a single-peak adaptation landscape model. Inclusion of G × E interaction and epistatic components generates more complex (more rugged) multiple-peaked adaptation landscapes.

Once the basic $E(N{:}K)$ model is selected, the user specifies detailed information on the:

1. chromosomal positions of genes (linkage groups and recombination frequencies);
2. number of traits that the genes regulate;
3. form of intra- and inter-locus gene action;
4. environment types of the TPE in which the genes are expressed and the extent to which they are expressed;
5. form of gene action by environment interactions;
6. frequency of the alternative alleles in the reference genetic population;
7. heritability of the traits.

These variables can be manipulated to generate genotype–environment systems of varying complexity. For a two-allele system, gene action and the genetic value of the genotypes at each locus are specified by use of the familiar midpoint (m), additive (a) and dominance (d) effects, using the completely inbred generation ($F_{n\to\infty}$) as the reference population (e.g.

Falconer and Mackay, 1996). When the user has information on the genetic architecture of traits, the gene action associated with variation for a trait, and the influence of environmental conditions on gene expression, relevant *E*(*N*:*K*) models can be specified. Information of this form is now becoming available from research involving molecular and classical methods for analysis of quantitative traits (e.g. Allard, 1996; Chase *et al*., 1997). Linkage maps involving molecular markers, Quantitative Trait Loci (QTLs) and known genes can be directly entered into the QU-GENE engine to enable analysis of the effectiveness of the alternative direct and indirect selection-based breeding strategies that are represented by the application modules.

Procedures for specifying a genotype–environment system in QU-GENE were explained by Podlich and Cooper (1997). An artificial example from the family of *E*(*N*:*K*) = 3(5:0.20) models is discussed here to demonstrate the main elements (Table 1). In this example, a single trait (defined by column *ATT*) is controlled by genes at five loci (column *GN*) each with two alleles (*N* = 5). There are three types of environment (*E1*, *E2* and *E3*), each occurring with a different frequency in the TPE (*E* = 3). There is heterogeneity in the magnitude of genetic variance among the three environment types, with *E2* generating the most, *E1* intermediate and *E3* the least. Gene 4 may be considered to be a major gene relative to genes 1, 2, 3 and 5 as it has a larger effect on the expression of the trait as specified by the *m*, *a* and *d* columns. When the effects of epistasis are ignored, genes 1 and 5 have a completely dominant gene action (*d* = *a*), gene 4 has a partial dominance gene action (0 < *d* < *a*) and genes 2 and 3 are additive (*d* = 0). Genes 1, 2 and 3 are located on chromosome 1, with a recombination frequency of 0.1 between genes 1 and 2, and a recombination frequency of 0.2 between genes 2 and 3 (column *RF*). Genes 4 and 5 are located on two other chromosomes, referred to as chromosomes 2 and 3, respectively. Recombination is simulated as described by Fraser and Burnell (1970), where the recombination frequency is the probability of a cross-over event occurring during gametogenesis. The form of gene expression for each gene in each type of environment is specified by the coefficients given in the columns *GX*(*E1*), *GX*(*E2*) and *GX*(*E3*) for each combination of gene and environment. Within these columns, a gene expression coefficient of 0 indicates that the gene is not expressed in that environment, a code of 1 indicates that the gene is expressed as defined by the *m*, *a* and *d* coefficients, and a code of –1 indicates that there is a rank reversal (cross-over interaction) of the performance of the genotypes relative to the performance defined by the *m*, *a* and *d* coefficients. Thus, the number of genes being expressed in each of the three environments is $N_1 = 5$, $N_2 = 2$ and $N_3 = 4$, respectively. These are only examples of the possible forms of G × E interaction that can be defined. The performance profiles for the three genotypes generated by gene 1, for gene expression coefficients 0, 1 and –1, and assuming no epistasis, are shown in Figure 2a.

Table 1. An artificial example of the specification of a genotype–environment system based on an *E*(*N*:*K*) = 3(5:0.20) model.[a] There are five genes (*GN*) and three types of environments (*E1*, *E2* and *E3*) and one digenic epistatic family exists

						E1	E2	E3		
Frequencies of environment types in the TPE						0.5	0.3	0.2		
Heterogeneity G × E interaction multipliers for environment types						1.0	1.5	0.7		
GN[b]	*m*[c]	*a*[c]	*d*[c]	*ATT*[d]	*RF*[e]	*GX*(*E1*)[f]	*GX*(*E2*)[f]	*GX*(*E3*)[f]	*EPI*[g]	*p*[h]
1	1.0	0.3	0.3	1	1	1	0	–1	1	0.6
2	0.5	0.4	0.0	1	0.1	1	0	1	0	0.2
3	0.5	0.4	0.0	1	0.2	1	0	1	0	0.2
4	2.0	1.0	0.5	1	2	1	1	1	0	0.3
5	0.3	0.1	0.1	1	3	1	1	0	1	0.5

[a]The *E*(*N*:*K*) = 3(5:0.20) model was defined by three types of environment (*E* = 3), five distinct genes (*N* = 5) and using the level of epistasis in each environment type ($K_1 = 2/5$, $K_2 = 0$, $K_3 = 0$), the epistasis value $K = 0.5 \times 0.4 + 0.3 \times 0 + 0.2 \times 0 = 0.20$ was computed.
[b]For each gene (1–5), the following information is specified.
[c]*m* = midpoint, *a* = additive effect, *d* = dominance effect (Falconer and Mackay, 1996).
[d]*ATT* = the attribute that the gene influences (in this example, all five genes influence one attribute).
[e]*RF* = the recombination frequency between adjacent genes (whole numbers, e.g. 1, 2 and 3, identify the start of new linkage groups).
[f]*GX*(*E1*), *GX*(*E2*) and *GX*(*E3*) are columns of gene expression coefficients for the three environment types and these determine the expression of the genes in the respective environment type (0 = the gene is not expressed, 1 = the gene is expressed as defined by the *m*, *a* and *d* parameters in the absence of epistasis, –1 = a cross-over interaction).
[g]*EPI* is an identifier that indicates which genes are involved in an epistatic network (in this example, genes 1 and 5 form a digenic network, and the remaining genes are not part of an epistatic network).
[h]*p* defines the allele frequencies at each locus (e.g. if the alleles at a locus are represented by *A* and *a*, then *p* is the frequency of the upper-case allele *A* and the frequency of the lower-case allele *a* is 1 – *p*).

Genes 1 and 5 act epistatically upon each other (defined by column *EPI*). The form of epistasis is specified in an input file by the user. Epistasis can be defined in many ways. For example, epistasis can be described statistically in terms of orthogonal partitions into interaction terms based on additive and dominance effects of loci (Kempthorne, 1957; Mather and Jinks, 1982; Kearsey and Pooni, 1996). Kauffman (1993) described epistatic models based on distributions of random effects within genetic networks. Alternatively, particular values can be defined for the genotypes in an epistatic network as for situations where multiple-peak epistasis occurs (Wright, 1963). Within the QU-GENE $E(N{:}K)$ model, epistasis is expressed only in those environments where all of the genes involved in the epistatic network are expressed, as defined by the gene expression coefficients. Therefore, in this example, the epistatic interaction between genes 1 and 5 will only occur in environment type E1, as this is the only environment type in which both genes 1 and 5 are expressed (Table 1). The performance profiles for the nine genotypes from the combination of gene 1 (*AA*, *Aa*, *aa*) and gene 5 (*BB*, *Bb*, *bb*), for the three environment types, are shown in Figure 2b. In environment type 1, an example of multiple-peak epistasis is defined. Note that in environment types 2 and 3, no epistasis is expressed and the different performance profiles result from the combination of *m*, *a* and *d* effects and gene expression coefficients (Table 1). Thus, to obtain the value $K = 0.20$ in the definition of the $E(N{:}K) = 3(5{:}0.20)$ model, the computed levels of epistasis for each environment are K_1 $(2\times1 + 3\times0)/5$, $K_2 = (5\times0)/5$ and $K_3 = (5\times0)/5$, respectively. Based on equation (1), the level of epistasis in the $E(N{:}K)$ genotype–environment system model is $K_1 = 0.5\times0.4 + 0.3\times0 + 0.2\times0 = 0.20$.

The frequency with which any genotype is observed depends on the frequencies of the alleles at each locus (defined by column *p*) (Table 1) and the mating system selected. There is flexibility within QU-GENE to work with mating systems based on self-pollination, cross-pollination, and mixtures of self- and cross-pollination. This example presents only a few of the many possible options for definition of a genetic model. These are discussed more fully by Podlich and Cooper (1997).

Figure 3 gives a schematic outline of the operation of the QU-GENE engine. The user can manipulate the variables described in Table 1 to define the genotype–environment system

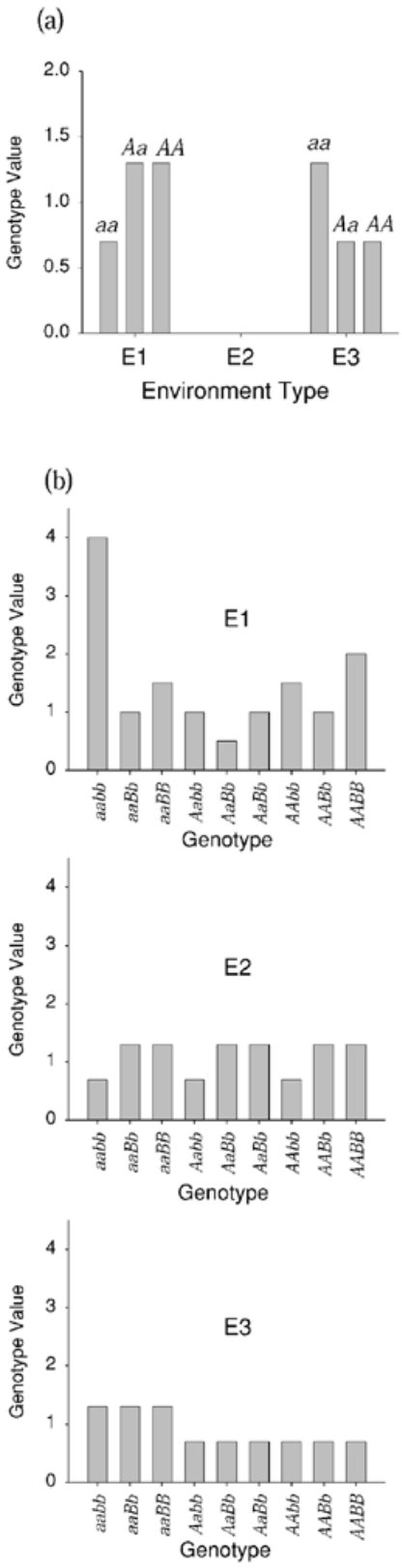

Fig. 2. (**a**) Performance profiles for the three genotypes generated by gene 1 (*A*: *AA Aa aa*) (Table 1), for the three environment types with gene expression coefficients 0, 1 and –1, and assuming no epistasis. (**b**) Performance profiles for the nine genotypes generated by the epistatic interaction between gene 1 (*A*: *AA Aa aa*) and gene 5 (*B*: *BB Bb bb*) for the three environment types.

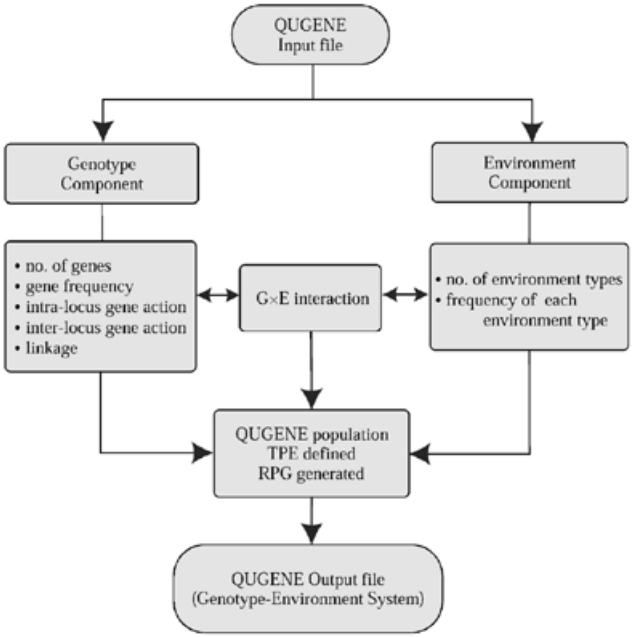

Fig. 3. Schematic outline of the operation of the QU-GENE engine (QUGENE) used to define the target population of environments (TPE) and generate the reference population of genotypes (RPG).

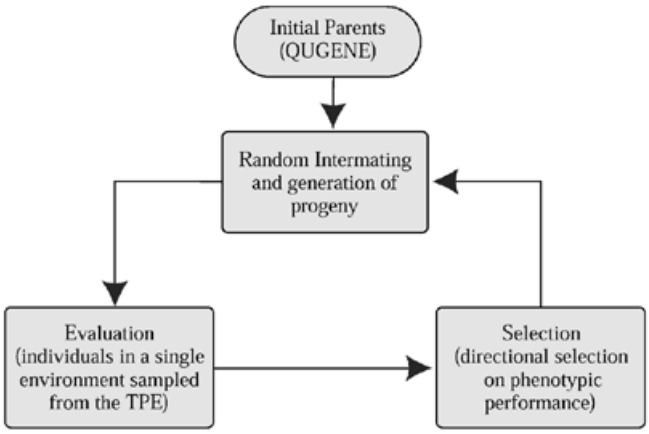

Fig. 4. Schematic outline of the operation of the MSSLT (Mass selection) application module.

that is to be used for a simulation experiment. The engine then generates a reference population of genotypes (RPG) and the target population of environments (TPE). Broad-sense heritability on an individual basis is defined for each trait and an appropriate microenvironmental (within environment error) component of variance is estimated once the base population of individuals is generated within the engine. These environmental effects are assumed to be normally and independently distributed. Any deviations from this basic model of the structure of the environmental variation are coded in the application modules. All of the genotype–environment system information is stored in an output file that is read by the selected application module. The simulation experiments are then conducted by the application modules. The information on the genotype–environment system is stored in a form that can be read by each of the application modules. Thus, any genetic model defined within the engine can be analysed by each of the application modules.

Information management and simulation experiments

A common problem encountered with simulation experiments is the large volume of data that can be rapidly generated. The results from a QU-GENE simulation experiment can be investigated either by considering graphical windows that interactively monitor the simulation experiment as it progresses, or by analysis of the results stored in output files, or by a combination of both approaches. Well-designed graphical windows that summarize the information relevant to the objectives of the simulation experiment can overcome many of the problems associated with dealing with the large volume of results that can be generated. They enable results from the experiments to be scanned while the experiment is in progress. This has proved to be more efficient than attempting to scan the large output files on completion of a simulation experiment.

Implementation of QU-GENE

Following the details given above, the engine and MSSLT (Mass selection) application module (Figure 1) were used to demonstrate how QU-GENE can be used to conduct a simulation experiment. Using the genotype–environment information summarized in Table 1 and Figure 2, and defining a broad-sense heritability of 0.7, the QU-GENE engine was used to generate a reference population of genotypes of 1000 individuals. These individuals were the base population for a mass selection simulation experiment using the MSSLT application module (selection for both male and female parents; Figure 4). Fifteen cycles of mass selection were conducted with a selection pressure of 20% for higher mean performance applied at each cycle of selection. A population size of 1000 individuals was maintained as the base population for each cycle of selection after intermating the selected individuals from the previous cycle. An example of typical results from 10 runs of the experiment are presented in the form of one of many possible graphical windows (Figure 5).

The graphical window used in this example has six subfigures: environment frequency (top left), genes fixed for an allele (middle left), gene frequency (bottom left), genetic

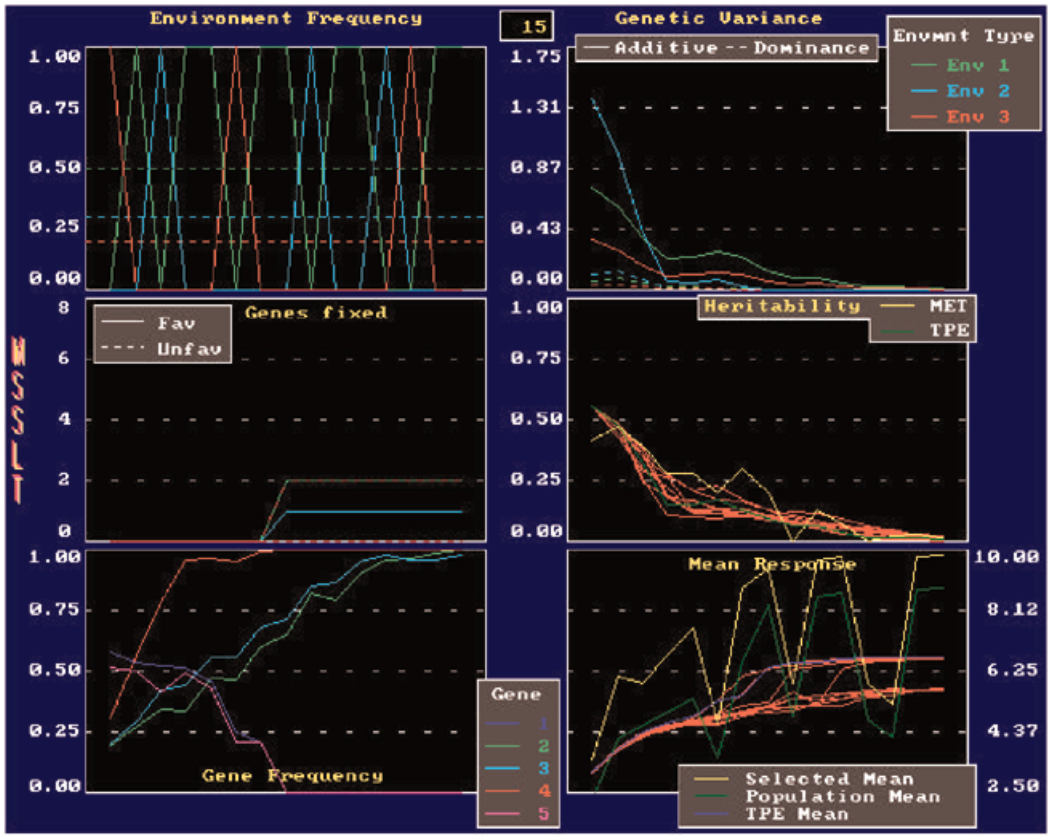

Fig. 5. An example of the basic graphical screen showing results from 10 runs of the MSSLT (Mass selection) application module for the $E(N{:}K) = 3(5{:}0.20)$ model described in Table 1 and Figure 2.

variance (top right), heritability (middle right), and population mean response (bottom right). The sub-figures are each described briefly to emphasize their main features. For all sub-figures, the horizontal axis is time, measured in cycles of the breeding programme. A cycle counter is located in the top centre of the figure. The vertical axis of each of the sub-figures depends on the information presented. For three of the sub-figures (environment frequency, genes fixed and genetic variance), the results for the different types of environments are distinguished to take into account the different $N_i{:}K_i$ models within each environment type. Different colours are used to distinguish the information for each type of environment over cycles of selection. A legend is used to distinguish the environment types and is given in the top right corner of the graphical window.

The environment frequency sub-figure (top left) presents the expected frequency (expressed as a proportion) of the three types of environments in the TPE as the dashed horizontal lines (E1 = 0.5, E2 = 0.3 and E3 = 0.2 for this example; see Table 1) and the frequency of the three environment types as they were sampled in each cycle to evaluate the individuals. Since only a single environment is sampled in each cycle of the MSSLT module, these frequencies are either 0 or 1. In the example shown, the sequence of 15 environments sampled in the final of the 10 runs is presented. In cycle 1, environment type 3 was sampled. For the first four cycles, the environment type sampled changed between cycles. For cycles four and five, the same environment type was sampled (i.e. environment type 1). Over a large number of cycles, the environments are expected to be represented in proportion to their frequency of occurrence in the TPE, when the environments are sampled at random. In the example depicted, environment type 1 was sampled nine times, environment type 2 three times, and environment type 3 was sampled three times. Thus, in relation to the TPE (frequencies E1 = 0.5, E2 = 0.3 and E3 = 0.2), environment type 1 was oversampled, environment type 2 undersampled and environment type 3 was sampled with a frequency equivalent to the expectation for the TPE.

The genes fixed sub-figure (middle left) monitors the number of genes fixed for both the unfavourable and favourable alleles in the genotype population. In this example, the definition of an allele as favourable or unfavourable ignores the influence of epistasis. An alternative graphical screen is

used to investigate the influence of epistasis and this is discussed below. Counts of genes fixed are made for each environment type. The gene frequency sub-figure (bottom left) monitors the frequency of the allele considered to be the most favourable for the TPE, again ignoring the influence of epistasis. In this example, after the 15 cycles of mass selection, the favourable allele for gene 4 was fixed; for genes 2 and 3 the favourable alleles increased in frequency, but were not fixed. For the two genes in the digenic epistatic network (genes 1 and 5), the alleles considered to be unfavourable in the absence of epistasis were fixed, thus the desirable epistatic combination *aabb* (Figure 2b) was fixed. This is discussed further below.

The genetic variance sub-figure (top right) indicated that the additive and dominance genetic variance declined over the 15 cycles of selection. In this example, the components of genetic variance were computed assuming no epistasis and the population was in Hardy–Weinberg equilibrium after random mating, following each cycle of selection. The heritability sub-figure (middle right) indicated that there was a corresponding reduction in narrow-sense heritability in the TPE and in the sequence of environments sampled across the cycles of selection (MET). In this example, the heritability in the TPE for each of the 10 runs was superimposed to facilitate comparisons among the 10 runs. The mean response sub-figure (bottom right) monitored the change in genetic mean of the population in the TPE (TPE Mean), the phenotypic mean of the population (Population Mean) and the selected group of individuals (Selected Mean) in the environment sampled in each cycle of selection. The mean of the population in the TPE was superimposed for each of the 10 runs. In this example, the population mean was computed taking into account the effects of epistasis. Two general forms of response for the population mean in the TPE were observed. These differed for the genotypes fixed at the loci involved in the digenic epistatic network (genes 1 and 5). The lower endpoint mean was associated with those runs of the simulation experiment where there was fixation of the *AABB* genotype and the higher endpoint mean with runs where there was fixation of the *aabb* genotype (Figure 2b). All of the information displayed on the graphical window is also stored in an output file to enable analysis of the results on completion of the simulation experiment.

As discussed above, the definition of what is the favourable allele for a gene becomes complicated and in many cases is meaningless when epistasis is incorporated into the genetic model. Therefore, it becomes difficult to follow the progress from selection in terms of the increase in frequency of the favourable alleles at loci. The example shown in Figure 5 is used to demonstrate this point. In the gene frequency sub-figure (bottom left), the two genes in the epistatic network (genes 1 and 5) are fixed for what is considered to be the unfavourable alleles at these loci when epistasis is not considered (Table 1). However, this is the best allelic combination for these two genes when the specific nature of the epistatic network is considered (Figure 2b). To overcome this problem, an alternative graphical screen, referred to as the spatial games graphical screen, is used (Figure 6). For this screen, the user defines a target genotype in terms of a combination of alleles at all loci. The performance of the genotypes within a population is then monitored in relation to genetic distance from the target genotype. The target genotype can be defined to take into account the effects of epistasis. The genetic distance from the target genotype is measured as a Hamming distance, which is a measure of the number of alleles that are different from the target genotype for all loci. This procedure was used by Fontana and Schuster (1987) in their analysis of the evolution of genetic networks. The example shown in Figure 6 is a run of the MSSLT application module for a 30-gene model with five digenic epistatic networks and three types of environments in the TPE [$E(N{:}K) = 3(30{:}0.23)$, data not shown]. The mean of the population and its distribution are depicted in the TPE (top left) and for the three types of environments (top right, middle left and right) after 15 generations (cycles) of selection. The horizontal axes measure the Hamming distance from the target genotype and the vertical axes measure the genetic value of the individuals. The distribution of the individuals is depicted, with the legend (bottom left) relating colour to frequency (expressed as a proportion) of individuals for a particular combination of Hamming distance and genetic value. A yellow trace maps out the flow of the population towards the target genotype over the 15 generations of selection and a generation counter is located with the legend (bottom left). For this example, no individuals had reached the target genotype after 15 generations. Progress from selection was more consistent in environment type 1 (top right) than in environment types 2 (middle left) and 3 (middle right). The bottom right sub-figure shows the distribution of individuals in each generation in terms of their Hamming distance from the target genotype in the TPE. The horizontal axis is generation and the colours mapped along the base of the sub-figure show the environment type (defined in the legend) sampled in each generation. The vertical axis is the Hamming distance from the target genotype. Therefore, within each generation, the vertical spread depicts the distribution of individuals within the population, with the legend relating colour to frequency of individuals with a given Hamming distance from the target genotype. The yellow trace monitors the mean Hamming distance of the population from the target genotype. In this example (Figure 6), the population progresses towards the target genotype for generations 1–6. Then, for generations 7–10, it remains at a similar Hamming distance from the target genotype. From generations 11 to 15, there is further progress towards the target genotype.

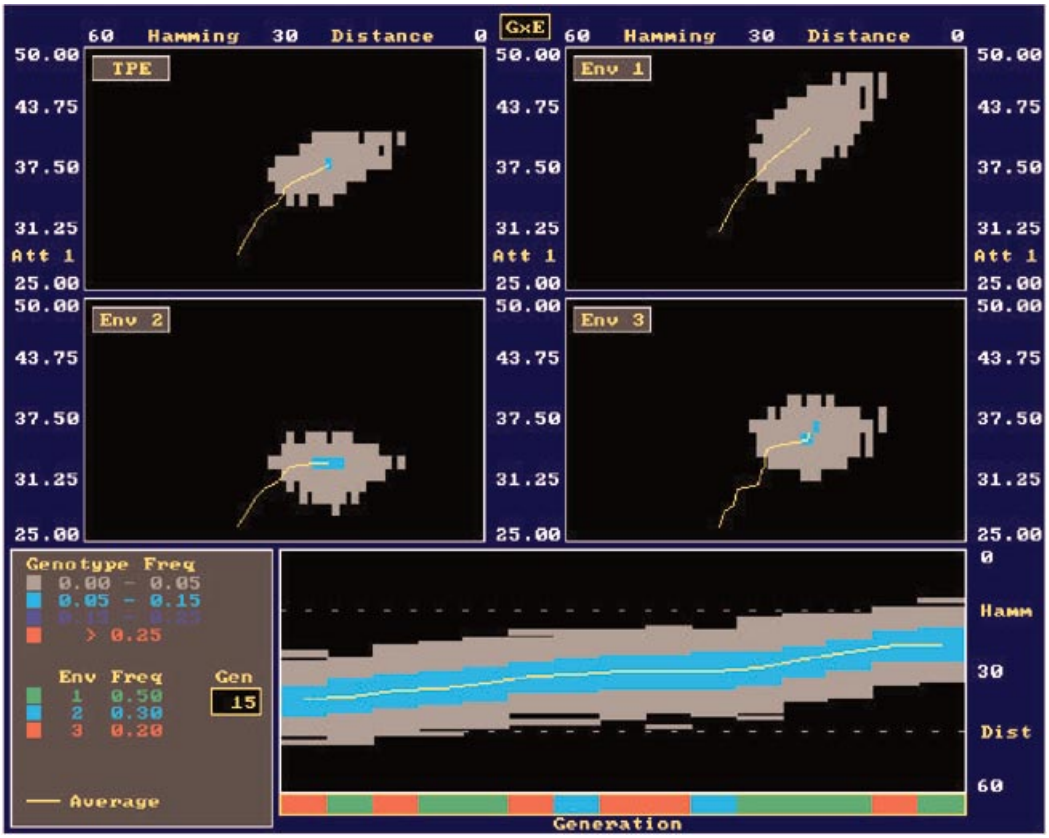

Fig. 6. An example of the spatial games graphical screen showing the results from one run of the MSSLT (Mass selection) application module for an $E(N{:}K) = 3(30{:}0.23)$ model.

All of the application modules have graphical windows to enable a representation of the results of the simulation experiments. As with mass selection, alternative graphical windows can be devised for the rest of the application modules in accordance with the nature and objectives of the simulation experiment. Comparisons of multiple runs of the simulation experiments in the graphical windows provide the user with a methodology to visualize and quantify the expected variability for the change in the genetic parameters. For example, in Figure 5, the variability in the heritability and population mean among the 10 runs was depicted graphically. The capacity to generate and analyse multiple runs can be used to compute bootstrap estimates for any feature of the simulation experiment.

QU-GENE examples

Two examples are presented to demonstrate the use of QU-GENE for the investigation of topical problems found in the quantitative genetics literature. In both cases, theoretical investigations have been undertaken. This enables a comparison between the expectations from the theoretical treatment of the problems and the results of the simulation experiments.

Example 1: Genotype-by-environment interactions

Background. Genotype-by-environment interactions complicate selection for improved genotype performance in a TPE. Plant breeders evaluate candidate breeding lines in METs to predict their expected relative performance in the TPE. The success of this strategy is influenced by the presence and form of G × E interactions, and how well the sample of environments in a MET represents the TPE. Since METs are small samples relative to the size and complexity of the TPE, they are subject to the effects of sampling variance. Therefore, some samples may represent the TPE more accurately than others. Where a particular sample deviates from the TPE, progress from selection would be expected to decrease. Thus, within the plant breeding literature, there is an interest in understanding the influence of this sampling variance and devising procedures for matching the composition of METs with the TPE (e.g. Horner and Frey, 1957; Brennan

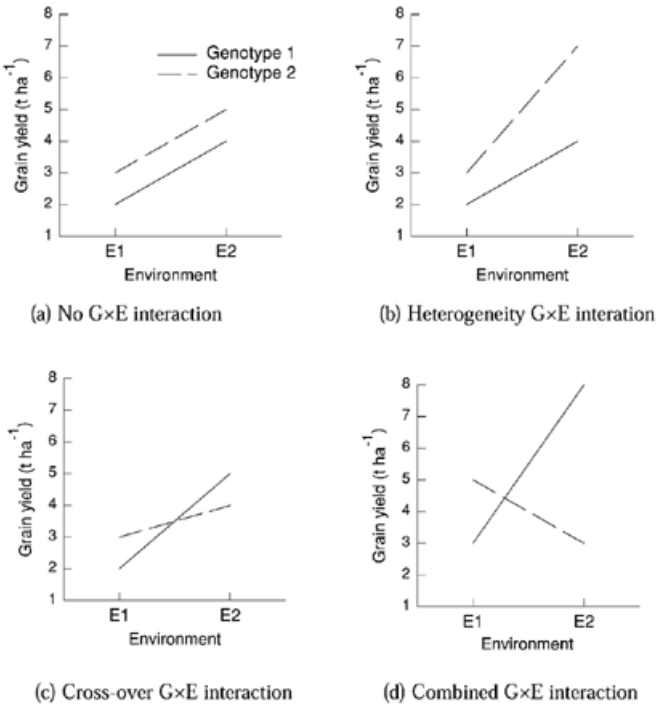

Fig. 7. Schematic representation of the performance profiles of two genotypes evaluated in two types of environment for four G × E interaction models. Reproduced with the permission of CAB International.

et al., 1981; Pederson and Rathjen, 1981; Mirzawan *et al*., 1994).

Cooper *et al*. (1996) considered four possible performance profiles for two genotypes (G1 and G2) in two types of environments (E1 and E2) (Figure 7). The four profiles differed in the presence and form of G × E interaction: no G × E interaction (Figure 7a), G × E interaction due to heterogeneity of genetic variance (Figure 7b), cross-over G × E interactions where there is a change in the rank of the genotypes between the two environments (Figure 7c), and a combination of G × E interactions due to heterogeneity of variance and cross-over (Figure 7d). Using direct and indirect selection theory, Cooper *et al.* (1996) theoretically quantified the joint influence of the G × E interactions for these profiles and the sampling variation associated with METs by way of the genetic covariance between the performance of the genotypes in the METs and their expected performance in a TPE as the frequency (expressed as a percentage) of the two types of environments (E1 and E2) changed in the MET and TPE (Figure 8). When there were no G × E interactions (Figure 7a) or G × E interactions due only to heterogeneity of genetic variance (Figure 7b), the genetic covariance between the MET and TPE was always positive, regardless of the frequencies of the two environment types (Figure 8a and b). However, where cross-over G × E interactions were involved in the performance profiles (Figure 7c and d), the genetic covariance was positive when the frequencies of the two environment types in the MET matched the TPE, but it was negative when they were not well matched (Figure 8c and d). Based on these theoretical relationships, it is expected that for those comparisons between a MET and the TPE where there is a positive genetic covariance, a positive response to selection will be observed in the TPE. Conversely, where there is a negative genetic covariance, a negative response is expected.

Experimental evaluation of these theoretical expectations is impractical. However, further analysis of their implications for a breeding programme is possible within a simulation experiment. The influence of the genetic covariance relationships between a MET and TPE (Figure 8) were examined in terms of the response to selection in the TPE using QU-GENE to conduct a comparable simulation experiment.

Simulation experiment. The QU-GENE engine was used to define four genotype–environment systems that represented each of the four sets of performance profiles shown in Figure 7. In each case, the genetic variation was based on 20 genes

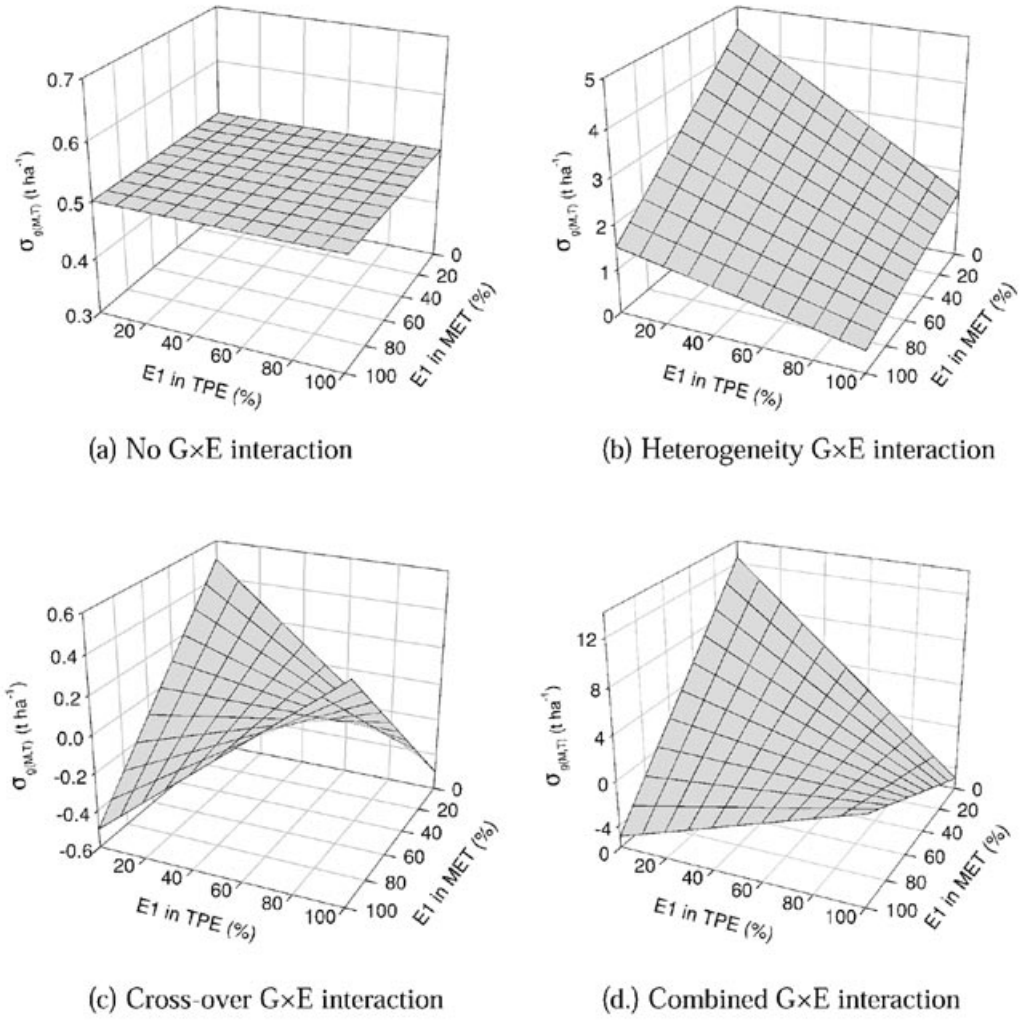

Fig. 8. Genetic covariance ($\sigma_{g(M,T)}$) between genotype performance in the multi-environment trials (METs) and the target population of environments (TPE) for the four G × E interaction models. Reproduced with the permission of CAB International.

of equal effect, each with additive gene action. The alternative alleles at each locus commenced with a frequency of 0.5. The extreme homozygous genotypes represented the two contrasting performance profiles shown for each of the G × E interaction scenarios depicted in Figure 7. Since there were $E = 2$ types of environments, $N = 20$ genes and $K = 0$ no epistasis for each case, all four genotype–environment systems were members of the $E(N:K) = 2(20:0)$ family of genetic models. The G × E interaction aspects of the performance profiles were introduced by use of the heterogeneity of variance and G × E interaction coefficients available in the engine (e.g. Table 1). The engine was used to generate four reference populations of genotypes, one for each of the sets of performance profiles (Figure 7). All four populations comprised 20 homozygous lines.

The reference populations of genotypes were then subjected to five cycles of recurrent selection using the single-seed-descent (SSD) breeding strategy. This was implemented by the PEDIGREE application module (Figure 9). All four populations were subjected to the same evaluation and selection procedures. The 20 homozygous lines in the reference genetic population were allocated into 10 random pairs and each pair was used to generate an F_2 population of size 50. The 50 F_2 individuals were advanced to the F_6 generation by SSD. The lines of descent were then multiplied and entered into a MET consisting of 10 environments. On comple-

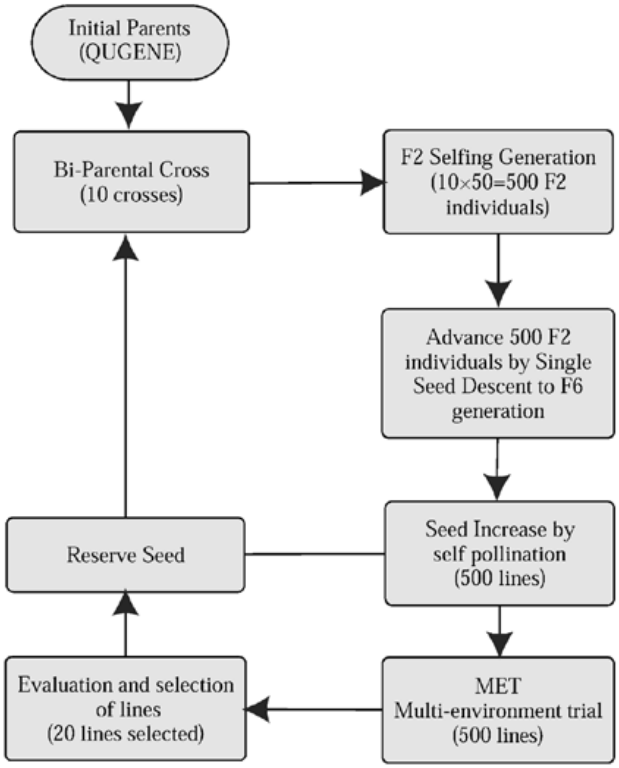

Fig. 9. Schematic outline of the operation of the PEDIGREE (Pedigree) application module used to implement the single-seed-descent (SSD) breeding strategy for the simulation experiment. The information in parentheses relates to the numbers used in the simulation experiment.

tion of the MET, the phenotypic performance of the 500 lines across the 10 environments was determined and the 20 lines with the highest mean performance were selected. These 20 lines were then used as the base population to conduct another cycle of selection as described above. This process was continued for five cycles. At the end of the five cycles, the mean performance of the final population of genotypes in the TPE was determined. This process was repeated 10 times to estimate a mean population performance in the TPE based on 10 runs of the SSD breeding strategy. Each of these means represents one point on the theoretical response surfaces shown in Figure 8.

To generate response to selection figures comparable in form to those shown in Figure 8, the frequencies (expressed as a percentage) of the two environment types were changed in steps of 10% from 0 to 100% for both the MET and the TPE. This gave a total of 121 combinations between the MET and the TPE for each of the four genotype–environment systems that represent the performance profiles (Figure 7). Therefore, to complete the simulation experiment, the five cycles of the SSD breeding strategy were run $121 \times 10 \times 4 = 2420$ times (121 MET–TPE combinations × 10 runs × 4 genotype–environment systems). The mean performance of the population of genotypes was then plotted against the frequency of environment type (E1) in the MET and TPE as a three-dimensional figure comparable to Figure 8. To take into account the effect of the change in the expected mean of the genotypes in the TPE as the frequencies of the two types of environments changed, the mean response from selection was expressed as a percentage deviation from the target genotype. The target genotype was the genotype that gave the highest mean in the TPE based on the genotype performance profiles (Figure 7) and the defined frequencies of the two environment types in the TPE. With this graphical representation of response to selection, the MET and TPE combinations that resulted in a good response to selection in the TPE would have no or small negative deviations from the performance of the target genotype. However, combinations that resulted in a poor response to selection in the TPE would have larger negative deviations. The simulated response for the population mean in the TPE was compared with the expectations based on the genetic covariance relationships (Figure 8).

Results and discussion. The mean performance expressed as a deviation from the target genotype after five cycles of the SSD breeding strategy for the four genotype–environment systems is shown in Figure 10. When there were no G × E interactions (Figure 10a) or G × E interactions due only to heterogeneity of genetic variance (Figure 10b), the response to selection was maximized (no deviation from the target genotype), regardless of the frequency of the two environment types in the MET and TPE. However, when there were cross-over G × E interactions, there was no deviation from the target genotype when the frequencies of the two environment types in the MET matched the TPE, but there was a 100% deviation from the target genotype when they were not well matched (Figure 10c and d). These results correspond with the theoretical expectations based on the derived values of genetic covariance between the MET and TPE (Figure 8) for each of the four genotype–environment systems. For MET and TPE combinations that generate a positive genetic covariance (Figure 8), the target genotype mean performance was realized (Figure 10), and for combinations that generated a negative covariance (Figure 8) the target genotype mean performance was not realized.

Figure 11 illustrates the mean performance expressed as a deviation from the target genotype for the genotype–environment system with cross-over G × E interactions (Figure 7c) from cycle 0 through to cycle 3. Cycle 0 (Figure 11a) depicts the mean performance of the base population (QUGENE population) before entering the first cycle of the SSD breeding strategy. Since all genes started with a frequency of 0.5, the mean performance of the base population (cycle 0) had a deviation of 50% from the target genotype for the 121 MET–TPE combinations. At cycle 1 (Figure 11b), there was approximately a 30% deviation from the target ge-

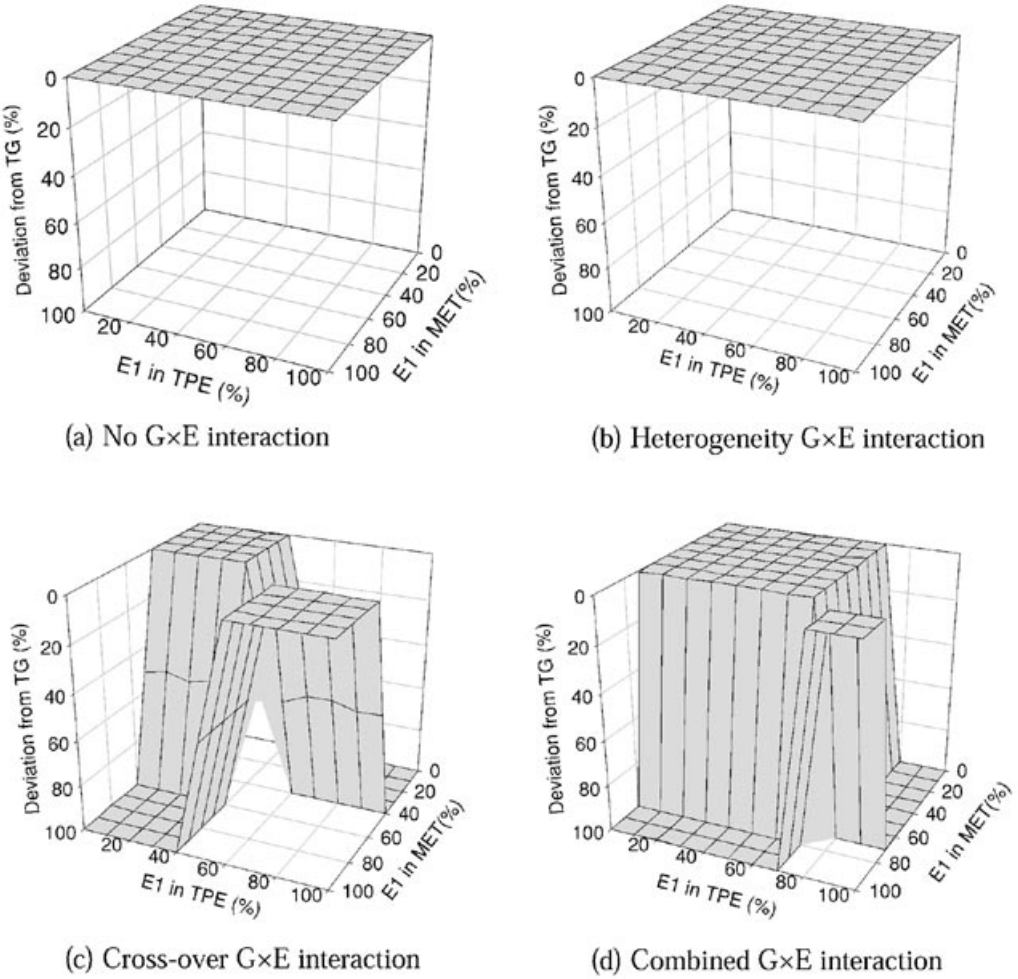

Fig. 10. Mean performance expressed as a deviation from the target genotype after five cycles of the single-seed-descent breeding strat egy for the G × E interaction models.

notype when the frequencies of the two environment types in the MET matched the TPE (i.e. regions of positive genetic covariance between MET and TPE on Figure 8c), but approximately a 70% deviation from the target genotype when they were not well matched (i.e. regions of negative genetic covariance between MET and TPE on Figure 8c). In cycle 2 (Figure 11c) and cycle 3 (Figure 11d), the 121 combinations continued to move towards the extremes of 0 or 100% deviation from the target genotype as in cycle 5 (Figure 10c), depending on whether or not the frequencies of the two environment types in the MET matched those in the TPE.

Considering the results of the simulation experiment as a whole, they are in agreement with the predictions based on the theoretical response surfaces (Figure 8) generated by Cooper *et al*. (1996). When cross-over G × E interactions occur within a genotype–environment system, a greater response to selection will be realized when the environmental composition of the MET matches that of the TPE. Therefore, in plant breeding, the strong emphasis on developing a reliable MET for testing breeding lines becomes more justifiable as cross-over G × E interactions account for a larger proportion of the total G × E interaction and as the variation due to G × E interactions increases relative to that for genetic variation for average performance across environments. The results of both theory and the simulation experiment strongly emphasize the importance of understanding the frequency of occurrence in the TPE of the types of environments that generate cross-over G × E interaction when designing effective METs for breeding programmes.

Example 2: Epistasis

Background. Wright (1963) gave a theoretical treatment of the expected response to selection for a four-gene model with

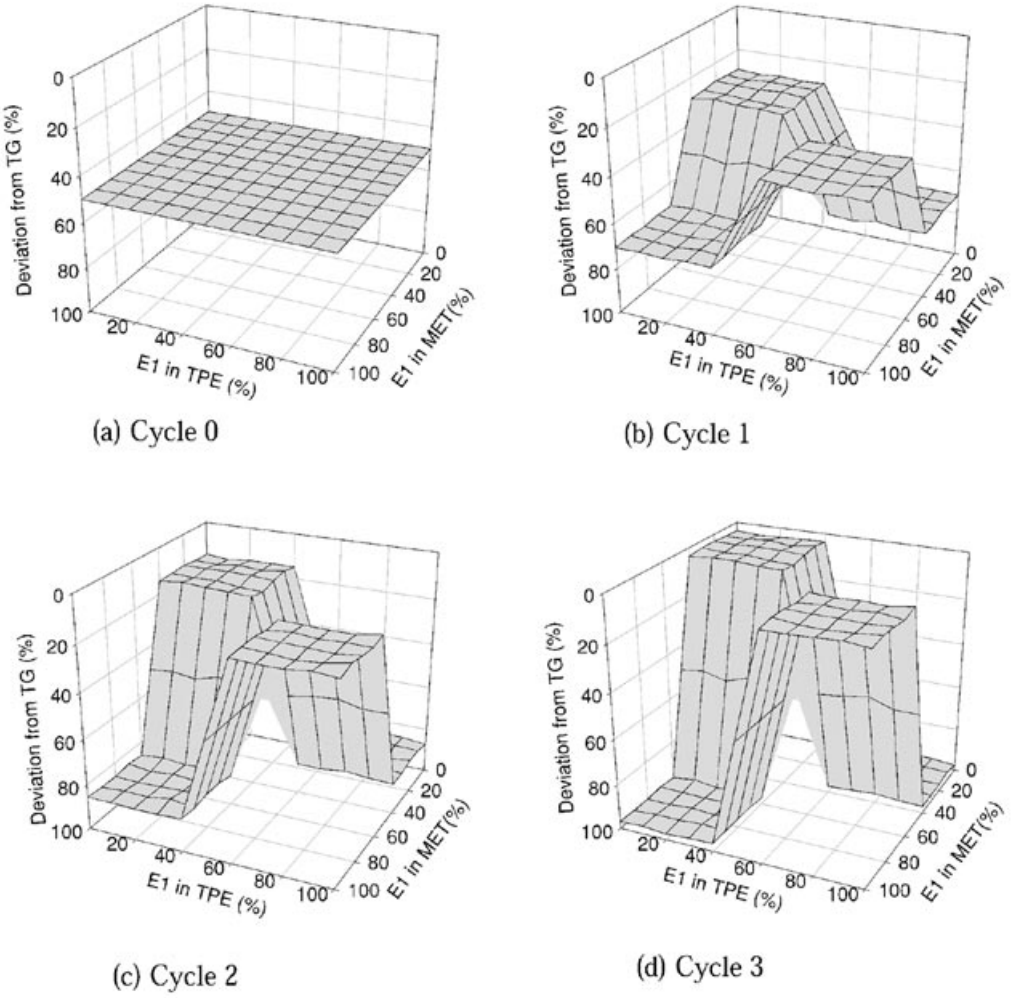

Fig. 11. Mean performance expressed as a deviation from the target genotype at cycles 0, 1, 2 and 3 of the single-seed-descent breeding strategy for the cross-over G × E interaction model.

multiple-peak epistasis and no G × E interactions. This model belongs to the family of $E(N{:}K) = 1(4{:}3)$ genotype–environment system models. The four gene model generates an adaptation landscape that consists of six selective peaks, each separated by regions of lower selective value. Theoretically, selection will move the population to one of the selective peaks, but the peak to which the population is moved is not necessarily the highest and depends on the relative influences of initial gene frequencies, selection pressure and random drift. When the population is localized on a sub-optimal peak and there is no mutation, further improvement would require the population to move down the current peak to explore other regions of the adaptation landscape and provide the chance of finding a higher peak. Wright suggested that the effects of random drift due to small population size may move populations away from one peak and into regions where selection towards an alternate peak takes over.

The movement of a population of genotypes over the adaptation landscape that is generated by Wright's four-gene epistatic model can be evaluated in terms of the distribution of points in a four-dimensional space, where each dimension represents the gene frequency from each of the four genes (**A**, **B**, **C** and **D**). Here we simplify the presentation of results by considering population flow on part of the adaptation landscape generated by the four-dimensional space. By assuming that two of the genes are fixed for one of their two alternate alleles in the population, one plane in the four-dimensional space can be examined. Based on Wright's model, Figure 12a depicts the expected trajectories of the gene frequency system on one of the planes where alleles *a* and *C* are

fixed, but genes **B** and **D** still vary in the population. The contour plot (Figure 12b) depicts the change in relative fitness of the population for varying combinations of frequencies of the alleles for genes **B** and **D** under the assumption that the population is in Hardy–Weinberg equilibrium. On this plane, there are four possible homozygous combinations: *aaBBCCDD*, *aaBBCCdd*, *aabbCCDD* and *aabbCCdd*. Genotypes *aaBBCCDD* and *aabbCCdd* are low points or non-adaptive pits in the adaptation landscape with fitness values of 0.875 and 0.750, respectively. Genotype *aaBBCCdd* is the highest or optimal adaptive peak on this plane with a fitness value of 1.125. Genotype *aabbCCDD* can be viewed as a sub-optimal or intermediate peak with a fitness value of 1.000. The two peaks are separated by a shallow saddle point with a fitness value of 0.990. The movement of the population on the adaptation landscape can be described in terms of the change in frequency of the alleles at the **B** and **D** loci within the population. Here the gene frequencies of the *B* allele at the **B** locus and the *D* allele at the **D** locus are followed. The population becomes fixed on a peak or pit in the adaptation landscape when the gene frequency of the *B* and *D* alleles is 0 or 1, and all individuals in the population have the same homozygous genotypic combination. Based on the theoretical treatment of this problem by Wright, the directions that populations are expected to move under the influence of selection are indicated by the arrows (Figure 12a).

Simulation experiment. The theoretical expectations for Wright's model were investigated in a QU-GENE simulation experiment. Of particular interest was the movement of the population under the influences of random drift and selection with different population sizes and selection pressures. Three population sizes were considered: 50, 100 and 500 individuals. For each of the population sizes, five levels of selection pressure were considered and were defined as selection proportions (*SP*): 0.2, 0.5, 0.9, 0.95 and 1.00. Here, 0.2 indicates that the 20% of the individuals in the population with the highest fitness values were selected and 1.00 indicates that 100% of the individuals in the population were selected, corresponding to no selection. There were 15 combinations of population size and selection pressure generated. For each of the combinations, the MSSLT (Mass selection) module (Figure 4) was run 5000 times for a maximum of 200 cycles. The initial population for all runs started with gene frequencies of 0.1 for both alleles *B* and *D* with alleles *a* and *C* fixed as indicated by the asterisk on Figure 12a. For each of the comparisons, the number of times the population reached the highest peak (*aaBBCCdd*), intermediate peak (*aabbCCDD*), or the pits (*aaBBCCDD*, *aabbCCdd*), from the 5000 runs after a maximum of 200 cycles was recorded. The gene frequencies of alleles *B* and *D* were recorded for the 200 cycles for each of the 5000 runs. By collating the gene frequencies over all 5000 runs, it was possible to construct a figure depicting the movement of the populations. This figure is comparable to that given by Wright (Figure 12a).

Results and discussion. Table 2 indicates the frequency (expressed as a proportion) and average number of cycles taken for the population to become fixed on the peaks (*aaBBCCdd*, *aabbCCDD*) or pits (*aaBBCCDD*, *aabbCCdd*) for each of the 15 combinations of selection pressure and population size from 5000 runs with a maximum of 200 cycles.

With a relatively strong level of selection (*SP* = 0.2), the population reached the optimal peak (*aaBBCCdd*) for all 5000 runs for population sizes 50, 100 and 500 (Table 2). Random drift from the smaller population size of 50 individuals slightly increased the average number of cycles required to reach the optimum peak in comparison to the larger population sizes of 100 and 500 individuals. Similarly with a selection pressure of *SP* = 0.5, a strong movement of the population across the adaptation landscape towards the optimal peak (*aaBBCCdd*) was observed. Population sizes of 100 and 500 individuals achieved fixation on the optimal peak for all 5000 runs. With the smaller population size of 50 individuals, random drift resulted in three of the 5000 runs moving towards the selection pull of the intermediate peak (*aabbCCDD*), where eventually they became fixed. Figure 13a depicts the general movement of the population across the adaptation landscape for 5000 runs with a population size of 50 individuals and a selection pressure of *SP* = 0.5. Two of the axes of the figure indicate the gene frequencies (expressed as proportions) of the alleles *B* and *D* in the population. The third axis indicates the frequencies (expressed as counts) for all 5000 runs of the pairwise gene frequency combinations (*B* and *D*) in the movement of the population from the gene frequency starting position (0.1, 0.1) at cycle 0 through to cycle 200 or fixation. For example: a frequency of one on the third axis on the figure indicates that the particular gene frequency combination occurred only once in any cycle in any of the 5000 runs. High regions on the figure indicate where the majority of the populations flowed across the adaptation landscape from the 5000 starts. The high-frequency region around approximately (0.1, 0.1) depicted the localized movement of the 5000 populations in the initial cycles. By slightly increasing the frequency of alleles *B* and *D* through selection, the populations were generally increased in average fitness value on the adaptation landscape in the next few cycles. However, once the populations reached the ridge (indicated by the R on Figure 12b), the selective pull of the optimal peak resulted in a decrease in the frequency of allele *D* as it moved towards fixation for genotype *aaBBCCdd* (Figure 13a). With the strong selection pressures of *SP* = 0.2 and *SP* = 0.5, a relatively higher proportion of the superior genotype (*aaBBCCdd*) was selected (Table 2). This

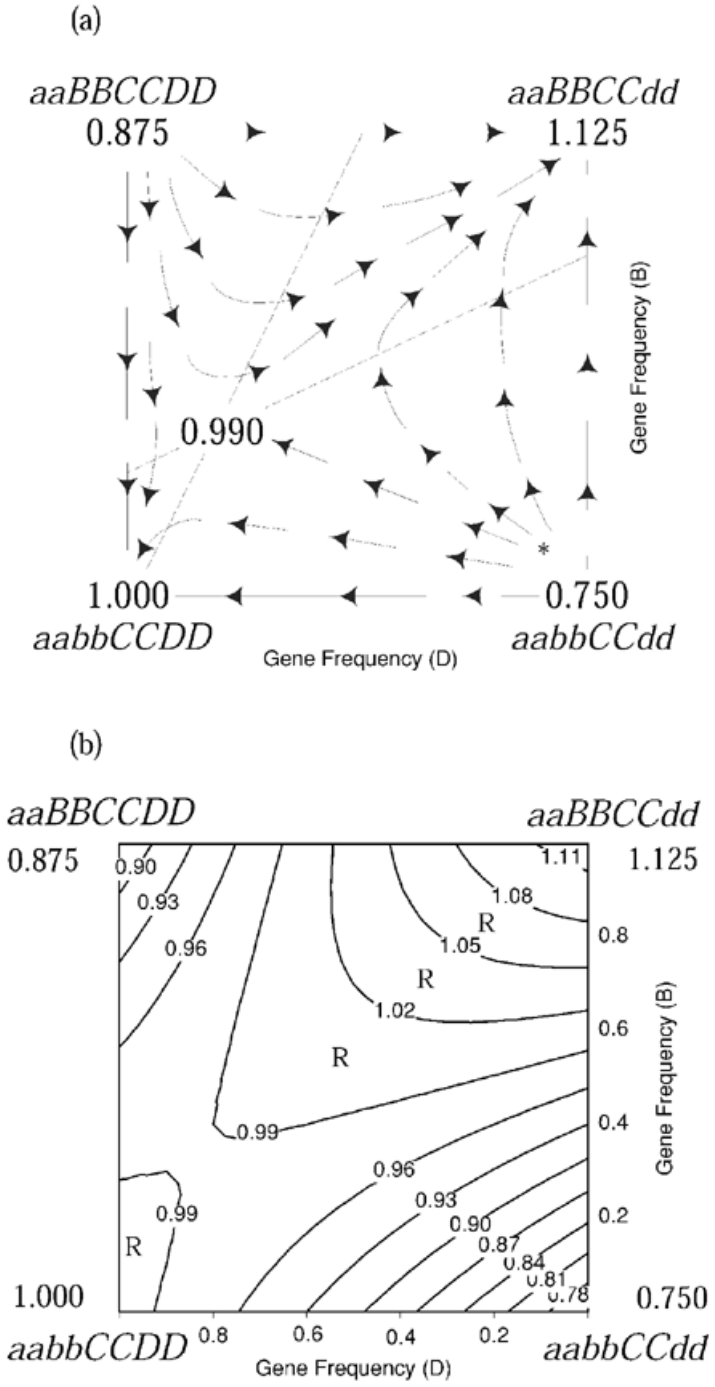

Fig. 12. (**a**) Expected trajectories of the gene frequency system on one plane of the four-dimensional space based on Wright's model. The asterisk indicates the starting position of the populations in the simulation experiment. Original reprinted with the permission of the National Academy of Sciences, Washington, DC. (**b**) Relative average fitness of populations for one plane of the four-dimensional space based on Wright's model. The letter R denotes a ridge in fitness values.

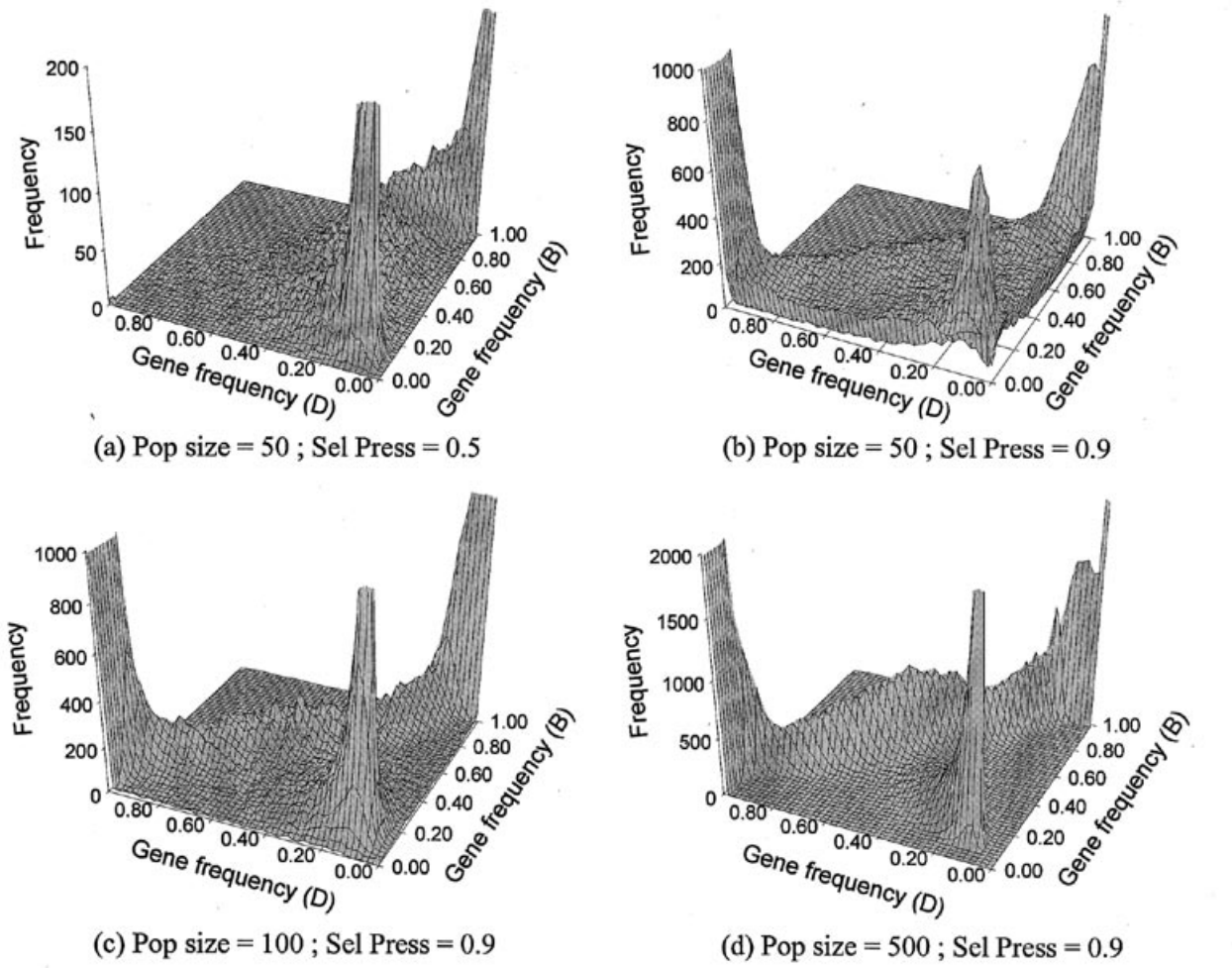

Fig. 13. General movement of the population on one plane of the four-dimensional space based on Wright's model (Figure 12) for 5000 runs with a population size of (**a**) 50 individuals and a selection pressure of 0.5, (**b**) 50 individuals and a selection pressure of 0.9, (**c**) 100 individuals and a selection pressure of 0.9 and (**d**) 500 individuals and a selection pressure of 0.9. Two of the axes indicate the gene frequencies (expressed as proportions) of the alleles *B* and *D* in the population. The third axis indicates the frequencies (expressed as counts) of the pairwise gene frequency combinations (*B* and *D*) in the movement of the populations from the gene frequency starting position (0.1, 0.1) at cycle 0 through to cycle 200 or fixation for all 5000 runs. Note that the figures are presented on different scales to highlight their features.

resulted in the populations being quickly pulled towards the optimal peak as predicted on Wright's diagram (Figure 12a).

With a selection pressure of *SP* = 0.9, random drift and the selective pull of the intermediate peak (*aabbCCDD*) influenced the ability of the populations to reach the optimal peak. Approximately 42, 40 and 28% of the populations from the 5000 runs reached the sub-optimal peak (*aabbCCDD*) for populations sizes of 50, 100 and 500 individuals, respectively (Table 2). Figure 14 illustrates the general movement of the populations on the adaptation landscape at various stages (cycles 10, 20, 40 and 60) on the path to fixation from 5000 runs with a population size of 100 individuals and selection pressure of *SP* = 0.9. Owing to the relatively higher fitness value of the gene frequency combinations in the centre of the landscape, the majority of the populations had moved away from the starting position of (0.1, 0.1) as observed at cycle 10 (Figure 14a). As well as the movement towards the centre of the landscape, random drift in the first few cycles had moved many of the populations into regions where selection towards the peaks (*aaBBCCdd*, *aabbCCDD*) had taken over. By cycle 20 (Figure 14b), several of the populations had become fixed; however, the majority had reached or were in the process of reaching the fitness ridge (as indicated by the *R* in Figure 12b). The selective advantage of neighbouring gene frequency combinations on the ridge is relatively small (Figure 12b), greatly reducing the

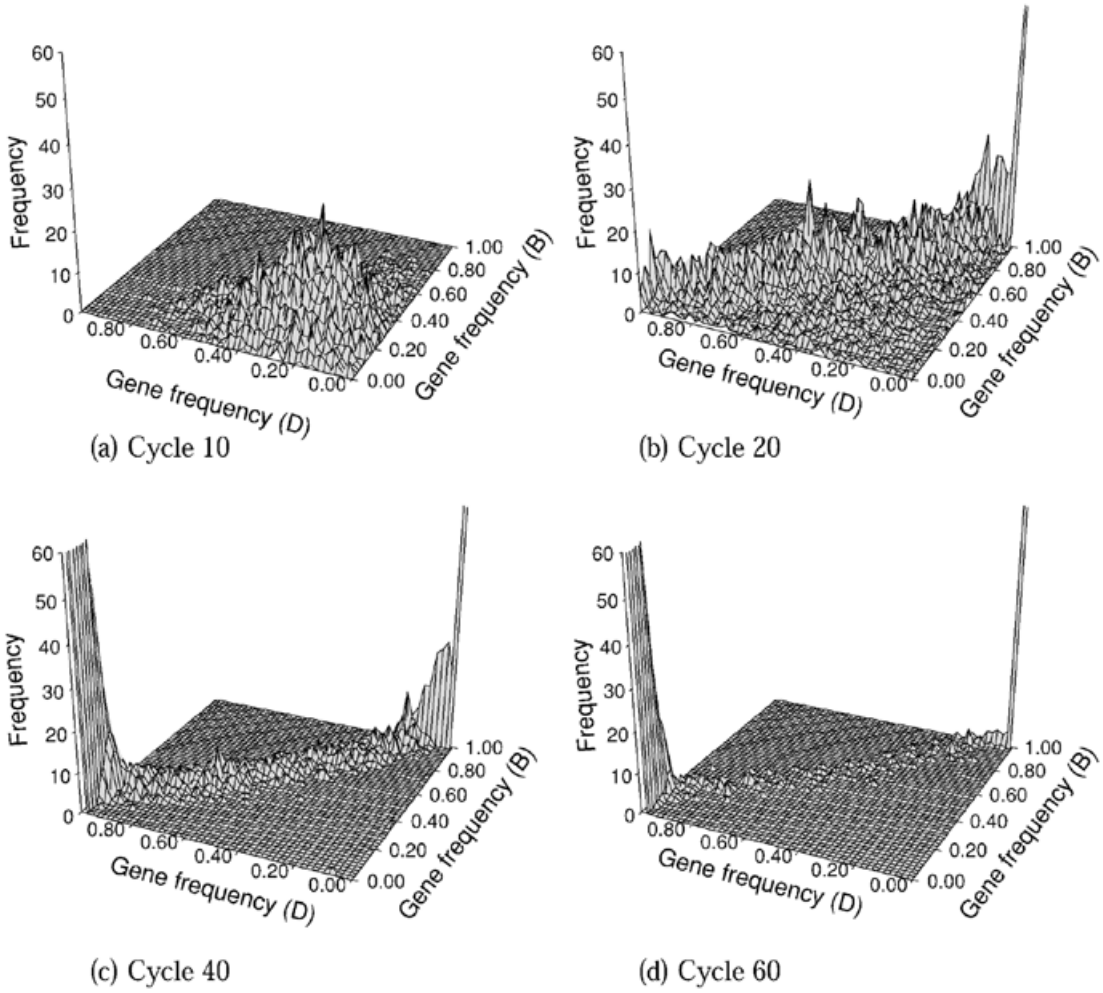

Fig. 14. General movement of the populations on one plane of the four-dimensional space based on Wright's model (Figure 12) at various stages in the movement to fixation (cycles 10, 20, 40 and 60) for 5000 runs with a population size of 100 individuals and a selection pressure of 0.9.

rate at which the populations were moved towards a peak. Several hundred of the 5000 populations were still on the ridge at cycle 40 (Figure 14c). The movement of the populations between the saddle point (0.990) and the sub-optimal peak (*aabbCCDD*) was extremely slow due to the slight fitness advantage on the ridge. By cycle 60 (Figure 14d), most of the populations had become fixed on the two peaks determined by the genotypes *aaBBCCdd* and *aabbCCDD*, with a few populations still moving along the ridge. The movement of the populations under the joint influences of selection and drift (Figure 14) on the adaptation landscape is comparable to Wright's theorized movement indicated by the arrows in Figure 12a.

Figure 13b–d illustrates the general movement of the populations for 5000 runs over all cycles for populations of size 50, 100 and 500 individuals and a selection pressure of $SP = 0.9$. Random drift moved many of the populations of 50 individuals towards regions in the landscape where selection towards one of the peaks took over. With the larger populations of size 100 and 500, random drift had a smaller influence; consequently, in general, the populations moved along a path to fixation that followed the ridge. The larger populations reached the optimal peak (*aaBBCCdd*) in 61 and 70% of the runs for populations of size 100 and 500, respectively. In comparison, of the 5000 runs with a population size of 50 individuals, 57% became fixed on the optimal peak. The influence of selection pressure on the movement of the populations over the adaptation landscape at a population size of 50 can be observed by comparing Figure 13a and b. With the higher selection pressure ($SP = 0.50$), a relatively higher frequency of the superior genotype was selected, resulting in a greater percentage of the populations becoming fixed on the optimal peak.

Table 2. The frequency (expressed as a proportion) and average number of cycles taken for the population to become fixed on the adaptive peaks (*aaBBCCdd*, *aabbCCDD*) or non-adaptive pits (*aaBBCCDD*, *aabbCCdd*) for each of the 15 combinations of selection pressure and population size from 5000 runs with a maximum of 200 cycles. The final column labelled 'No fixation' indicates the frequency (as a proportion) of the 5000 runs which were not fixed after 200 cycles

Pop	Sel	*aaBBCCdd*		*aabbCCDD*		*aaBBCCDD*		*aabbCCdd*		No Fixation
size	pres	1.125		1.000		0.875		0.750		
		Freq.	Cycles	Freq.	Cycles	Freq.	Cycles	Freq.	Cycles	Freq.
50	0.20	1.0000	3.73	0.0000	—[a]	0.0000	—	0.0000	—	0.0000
	0.50	0.9994	6.93	0.0006	7.67	0.0000	—	0.0000	—	0.0000
	0.90	0.5686	30.59	0.4156	44.27	0.0000	—	0.0158	18.32	0.0000
	0.95	0.4286	59.43	0.4136	66.32	0.0002	99.00	0.1570	31.51	0.0006
	1.00	0.0634	131.13	0.0508	131.80	0.0062	152.81	0.7452	56.84	0.1344
100	0.20	1.0000	3.63	0.0000	—	0.0000	—	0.0000	—	0.0000
	0.50	1.0000	6.80	0.0000	—	0.0000	—	0.0000	—	0.0000
	0.90	0.6050	33.90	0.3946	62.69	0.0000	—	0.0000	—	0.0004
	0.95	0.4986	58.84	0.4852	79.56	0.0000	—	0.0152	38.43	0.0010
	1.00	0.0186	155.80	0.0140	160.29	0.0004	150.00	0.6074	83.41	0.3596
500	0.20	1.0000	3.62	0.0000	—	0.0000	—	0.0000	—	0.0000
	0.50	1.0000	6.77	0.0000	—	0.0000	—	0.0000	—	0.0000
	0.90	0.6976	39.95	0.2840	114.17	0.0000	—	0.0000	—	0.0184
	0.95	0.4940	73.52	0.4438	134.80	0.0000	—	0.0000	—	0.0622
	1.00	0.0000	—	0.0000	—	0.0000	—	0.1126	144.74	0.8874

[a]The average number of cycles to fixation was not defined since no runs of the simulation resulted in fixation for the genotype.

Random drift had a larger influence on the movement of the population with a selection pressure of $SP = 0.95$ (Table 2). Only one run with a population size of 50 individuals drifted across the ridge on the adaptation landscape and became fixed in the opposite pit (*aaBBCCDD*). With no selection pressure applied ($SP = 1.00$), random drift alone controlled the fate of the population. Owing to the initial position of the starting population, ~75 and 61% of the runs obtained the nearest fixation pit (*aabbCCdd*) for population sizes of 50 and 100 individuals, respectively. As the size of the population increased, a greater percentage of populations did not obtain fixation after 200 cycles.

Considering the simulation experiment as a whole, the movement of the population under the joint influences of random drift and selection was as Wright suggested. There was an interaction between population size and selection pressure on the frequency of fixation on the four adaptation peaks and pits considered, and the average number of cycles to reach those peaks. Relatively small amounts of random drift frequently moved the populations into regions on the adaptation landscape at which the strong selection towards one of the peaks took over.

Discussion

Quantitative genetic theory provides the major framework for linking genotype to phenotype in genotype–environment systems. Much of this work has involved the development of mathematical expressions that are based on the assumption of a large number of Mendelian loci that each make a small contribution to the expression of genetic variation which combines with environmental variation to give the observed phenotypic variation. While this framework has served us well, and will continue to do so, the expanding body of work on the molecular, biochemical and physiological–genetic architecture of quantitative traits is stimulating considerable demand for a more general treatment of these traits than is currently possible by classical quantitative genetic theory.

Kempthorne (1988) made a strong case for the use of high-speed computers to investigate the influence of features of quantitative traits that are otherwise accommodated in the classical theory by the use of restrictive assumptions, e.g. assuming no epistasis, no $G \times E$ interactions, a large number of genes each making a similar contribution to variation, no mutation. Many agree that these assumptions are unrealistic.

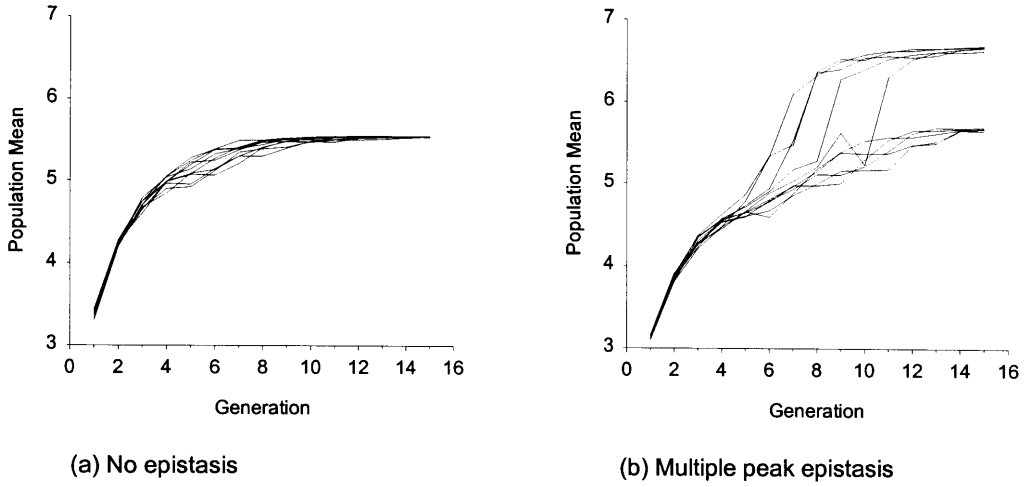

Fig. 15. Change in the population mean response to directional selection for 10 runs of the MSSLT (Mass selection) module, using the model described in Table 1 and Figure 2, with (**a**) no epistasis and (**b**) multiple-peak epistasis.

However, their use has been advocated on two grounds: (i) without the assumptions, much of the algebraic treatment of the quantitative theory would be intractable; (ii) more importantly, with the use of the theory, as developed with the assumptions, useful predictions can be made on some general features of the traits and how they are expected to change, or can be manipulated, through the influence of external forces, e.g. natural and artificial selection. Nevertheless, many observations are not easily explained by the theory with the assumptions in place. An interesting example is the punctuation of linear rates of improvement of quantitative traits by larger stepwise improvements. The classical theory predicts the linear improvement, but not the stepwise jumps. Kauffman (1993) used computer simulation methods to present a convincing demonstration of how the inclusion of epistasis in a quantitative genetic model can explain the presence of the steps of improvement interspersed among phases of stasis or linear improvement.

QU-GENE can be used to provide a simple demonstration of this point (Figure 15). Using the genetic model given in Table 1 and the mass selection (MSSLT) application module, response to selection was simulated with and without the epistatic effects shown in Figure 2b. Comparing the simulated response to selection for the population mean without (Figure 15a) and with (Figure 15b) the presence of multiple-peak epistasis, two contrasting response patterns are observed. Without epistasis, there was a relatively smooth curvilinear response towards the expected limit. With epistasis present, there were a number of paths observed, with two possible major limits. These limits differed for the epistatic combinations of genes that were selected during the course of the cycles of selection. Where the most favourable epistatic combination was selected (i.e. *aabb*), there was a significant step in the profile of the response to selection. Where the less favourable epistatic combination was selected (i.e. *AABB*), the step was not observed. A further point of significance is that the absence of this step in the mean response profile does not correspond with the absence of epistasis. For the selection scenario with epistasis, a number of the runs of the simulation experiment did not select the desirable epistatic combination and had a mean response profile comparable in form with that for the genetic model without epistasis. Clearly, the inherent variability in response to selection for the genetic models investigated was greater for the scenario with epistasis (Figure 15b) than that without it (Figure 15a).

Computer simulation has been used to investigate many features of genetic models. However, much of the software that has been developed to date has been linked to the specific research topics under investigation. QU-GENE was designed to be a simulation platform that enables the user to investigate the implications of a wide range of alternative quantitative genetic models in context with the interests of the user. The initial applications have been in the area of applied plant breeding (e.g. Fabrizius *et al*., 1996; Cooper and Podlich, 1997, 1998; Podlich and Cooper, 1997). However, the architecture of QU-GENE renders it useful for a wider array of applications requiring the investigation of quantitative traits. The two examples considered here, G × E interactions and epistasis, serve only as examples of the flexibility of QU-GENE and demonstrations of how investigations can be conducted to complement the theoretical treatment of problems. In both examples, the results of the simulation experiments were in accordance with the theoretical expecta-

tions. The important point is that these two diverse issues were examined by the same software environment.

The two-stage architecture of QU-GENE, i.e. the engine and the application modules, provides a high degree of flexibility in the design and implementation of simulation experiments. The separation of the specification of the base information on a genotype–environment system from the application modules means that any genotype–environment system can be subjected to many forms of investigation without unnecessary duplication of information and coding across the application modules. New simulation experiments are conducted by designing and coding an appropriate application module that can interact with the information that specifies the genotype–environment system generated by the engine. The current version of the software is designed for investigation of diploid or amphidiploid genomic structures with two alleles per locus. However, the software can be extended to accommodate haploids and autopolyploids, multiple alleles per locus and mutation. These improvements will enable investigation of a wide range of combinatorial features of adaptation landscapes for genotype–environment systems.

The *E*(*N*:*K*) model provides a useful framework for developing and characterizing the genetic models used to define a genotype–environment system. A further advantage is that it provides a link between the experience from extensive work on epistasis that resides in the evolutionary genetics literature (Wright, 1932; Fontana and Schuster, 1987; Kauffman, 1989, 1993) and that on G × E interactions which have been extensively investigated in the plant breeding literature (Comstock and Moll, 1963; Kang, 1990; Cooper and Hammer, 1996; Kang and Gauch, 1996).

Our experience is that computer simulation methodology has much to offer current research programmes investigating the architecture of quantitative traits and the implications of the architecture of these traits in applications such as plant and animal breeding in agriculture. QU-GENE is considered to be a first step in developing a general simulation platform for quantitative analysis of genetic models. It will continue to undergo revision in order to enhance user capability to investigate realistic genetic models as experimental information on the genetic architecture of quantitative traits is accumulated.

Acknowledgements

We thank the National Academy of Sciences, Washington, DC, for permission to reproduce Figure 12a, and CAB International for permission to reproduce Figures 7 and 8.

References

Allard,R.W. (1996) Genetic basis of the evolution of adaptedness in plants. *Euphytica*, **92**, 1–11.

Brennan,P.S., Byth,D.E., Drake,D.W., DeLacy,I.H. and Butler,D.G. (1981) Determination of the location and number of test environments for a wheat cultivar evaluation program. *Aust. J. Agric. Res.*, **32**, 189–201.

Bürger,R., Wagner,G.P. and Stettinger,F. (1989) How much heritable variation can be maintained in finite populations by mutation-selection balance? *Evolution*, **43**, 1748–1766.

Chase,K., Adler,F.R. and Lark,K.G. (1997) Epistat: a computer program for identifying and testing interactions between pairs of quantitative trait loci. *Theor. Appl. Genet.*, **94**, 724–730.

Comstock,R.E. and Moll,R.H. (1963) Genotype-environment interactions. In Hanson,W.D. and Robinson,H.F. (eds), *Statistical Genetics and Plant Breeding. Publication 982*. National Academy of Sciences–National Research Council, Washington, DC, pp. 164–196.

Cooper,M. and Hammer,G.L. (1996) Synthesis of strategies for crop improvement. In Cooper,M. and Hammer,G.L. (eds), *Plant Adaptation and Crop Improvement*. CAB International in association with IRRI and ICRISAT, Wallingford, pp. 591–623.

Cooper,M. and Podlich,D.W. (1997) Genotype-by-environment interactions and selection response. CIMMYT. Book of Abstracts. *The Genetics and Exploitation of Heterosis in Crops*; *An International Symposium*. Mexico, D.F., Mexico, pp. 14–15.

Cooper,M. and Podlich,D.W. (1998) Genotype-by-environment interactions, selection response and heterosis. In *The Genetics and Exploitation of Heterosis in Crops*, in press.

Cooper,M., DeLacy,I.H. and Basford,K.E. (1996) Relationships among analytical methods used to analyse genotypic adaptation in multi-environment trials. In Cooper,M. and Hammer,G.L. (eds), *Plant Adaptation and Crop Improvement*. CAB International in association with IRRI and ICRISAT, Wallingford, pp. 193–224.

Cox,T.S. (1995) Simultaneous selection for major and minor resistance genes. *Crop Sci.*, **35**, 1337–1346.

Cress,C.E. (1967) Reciprocal recurrent selection and modifications in simulated populations. *Crop Sci.*, **7**, 562–567.

Fabrizius,M.A., Cooper,M., Podlich,D.W., Brennan,P.S., Ellison,F.W. and DeLacy,I.H. (1996) Design and simulation of a recurrent selection program to improve yield and protein in spring wheat. In Richards,R.A., Wrigley,C.W., Rawson,H.M., Rebetzke,G.J., Davidson,J.L. and Brettell,R.I.S. (eds), *Proceedings of the Eighth Assembly, Wheat Breeding Society of Australia*. The Australian National University, Canberra, ACT, P8–P11.

Falconer,D.S. and Mackay,T,F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Essex.

Fisher,R.A. (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh*, **52**, 399–433.

Fontana,W. and Schuster,P. (1987) A computer model of evolutionary optimization. *Biophys. Chem.*, **26**, 123–147.

Fraser,A.S. and Burnell,D.G. (1970) *Computer Models in Genetics*. McGraw-Hill, San Francisco, CA.

Horner,T.W. and Frey,K.J. (1957) Methods for determining natural areas for oat varietal recommendations. *Agron. J.*, **49**, 313–315.

Jeyaruban,M.G. and Gibson,J.P. (1996) Estimation of additive genetic variance in commercial layer poultry and simulated populations under selection. *Theor. Appl. Genet*., **92**, 483–491.

Kang,M.S. (ed.) (1990) *Genotype-by-Environment Interaction and Plant Breeding*. Louisiana State University, Baton Rouge, LA.

Kang,M.S. and Gauch,H.G. (eds) (1996) *Genotype-by-Environment Interaction: New Perspectives*. CRC Press, Boca Raton, FL.

Kauffman,S.A. (1989) The *NK* model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol*., **141**, 211–245.

Kauffman,S.A. (1993) *The Origins of Order—Self Organization and Selection in Evolution*. Oxford University Press, New York.

Kearsey,M.J. and Pooni,H.S. (1996) *The Genetical Analysis of Quantitative Traits*. Chapman and Hall, London.

Kempthorne,O. (1957) *An Introduction to Genetic Statistics*. Wiley, New York.

Kempthorne,O. (1988) An overview of the field of quantitative genetics. In Weir,B.S., Eisen,E.J., Goodman,M.M. and Namkoong,G. (eds), *Proceedings of the Second International Conference on Quantitative Genetics*. Sinauer Associates Inc., Sunderland, MA, pp. 47–56.

Mather,K. and Jinks,J.L. (1982) *Biometrical Genetics*, 3rd edn. Chapman and Hall, London.

Mirzawan,P.D.N., Cooper,M., DeLacy,I.H. and Hogarth,D.M. (1994) Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding programme. *Theor. Appl. Genet*., **88**, 707–716.

Mühlenbein,H. and Schlierkamp-Voosen,D. (1993) Predictive models for the breeder genetic algorithm. *Evol. Comput*., **1**, 25–49.

Partner,P.L.R., Smith,M.L., Spoor,W. and Clarkson,M.I. (1993) Computer simulation of selection in a hypothetical crop species. *Comput. Applic. Biosci*., **9**, 597–605.

Pederson,D.G. and Rathjen,A.J. (1981) Choosing trial sites to maximize selection response for grain yield in spring wheat. *Aust. J. Agric. Res*., **32**, 411–424.

Podlich,D.W. and Cooper,M. (1997) QU-GENE: A platform for quantitative analysis of genetic models. Centre for Statistics Research Report 83, The University of Queensland, Brisbane, Queensland.

St Martin,S.K. and Skavaril,R.V. (1984) Computer simulation as a tool in teaching introductory plant breeding. *J. Agron. Educ*., **13**, 43–47.

Tinker,N.A. and Mather,D.E. (1993) GREGOR: Software for genetic simulation. *J. Hered*., **84**, 237.

Wright,S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In Jones,D.F. (ed.), *Proceedings of the Sixth International Congress of Genetics*. Ithaca, New York.

Wright,S. (1963) Discussion: plant and animal improvement in the presence of multiple selective peaks. In Hanson,W.D. and Robinson,H.F. (eds), *Statistical Genetics and Plant Breeding*. *Publication 982*. National Academy of Sciences–National Research Council, Washington, DC, pp. 116–122.