Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 65(2) 2011 225-235

Simulation Modeling in Crop Breeding

Jiankang Wang

Institute of Crop Science and CIMMYT China, Chinese Academy of Agricultural Sciences, No. 12, Zhongguancun South Street, Beijing 100081, China

Received 02 December 2010; Revised 28 February 2011 Accepted 28 February 2011

SUMMARY

Along with the fast developments in molecular biology and biotechnology, a large amount of biological data is available from genetic studies of important breeding traits in plants, which in turn provides an opportunity for undertaking genotypic selection in the breeding process. However, gene information has not been effectively used in crop improvement due to the lack of appropriate tools. The simulation approach can utilize the vast and diverse genetic information, predict the cross performance and compare different selection methods. Hence, the best performing crosses and effective breeding strategies can be identified. QuLine and QuHybrid are computer tools capable of defining a range from simple to complex genetic models and simulating breeding processes for developing final advanced lines and hybirds. Based on the results from simulation experiments, breeders can optimize their breeding methodology and greatly improve the breeding efficiency. In this paper, we first introduce the underlying principles of simulation modeling in crop enhancement, and then summarize several applications of QuLine in comparing different selection strategies, precision parental selection using known gene information, and the design approach in breeding. Breeding simulation allows the definition of complicated genetic models consisting of multiple alleles, pleiotropy, epistasis and gene-by-environment interaction, and provides a useful tool to efficiently use the wide spectrum of genetic data and information available to the breeders.

Keywords : Breeding strategy and method, Computer simulation, Crop breeding, Genetic model.

INTRODUCTION

The 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 (Allard 1960, Fehr 1987, Cooper *et al.* 1999). To achieve this objective, breeders face many complex choices in the design of efficient crossing and selection strategies aimed at combining desired alleles into a single target genotype. It is normally 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. Quantitative genetics provides much of the framework for the design and analysis of selection methods used within breeding programs (Allard 1960, Falconer and Mackay 1996, Cooper *et al.* 1999). However, there are usually associated assumptions, some of which can be easily tested or satisfied by experimentation; others can seldom, if ever, be met. Computer simulation provides us with a tool to investigate the implications of relaxing some of the assumptions and the effect this has on the conduct of a breeding program (Kempthone 1988).

QU-GENE is a simulation platform for quantitative analysis of genetic models, which consists of a two-stage architecture (Podlich and Cooper 1998). The first stage is the engine (referred to as QU-GENE), and its role is to: (1) define the genotype-by-environment (GE) system (i.e., all the genetic and environmental information of the simulation experiment), and (2) generate the starting population of individuals (base germplasm). The second stage encompasses the application modules, whose role is to investigate, analyze, or manipulate the starting

E-mail address : wangjk@caas.net.cn

population of individuals within the GE system defined by the engine. The application module will usually represent the operation of a breeding program. A QU-GENE strategic application module, QuLine, has therefore been developed to simulate breeding procedure deriving inbred lines.

Built on QU-GENE, QuLine (previously called QuCim) is a genetics and breeding simulation tool that can integrate various genes with multiple alleles operating within epistatic networks and differentially interacting with the environment interaction, and predict the outcomes from a specific cross following the application of a real selection scheme (Wang *et al.* 2003, 2004, 2007a). Therefore, it has the potential to serve as a bridge between the vast amount of biological data and breeder's queries on optimizing selection gain and efficiency. QuLine has been used to compare two selection strategies (Wang *et al.* 2003), to study the effects on selection of dominance and epistasis (Wang *et al.* 2004), to predict cross performance using known gene information (Wang *et al.* 2005), and to optimize marker assisted selection to efficient pyramid multiple genes (Wang *et al.* 2007b, 2007c). Building on the advanced development of QuLine, QuHybrid was developed to simulate the breeding programs for selecting hybrids. While retaining the most functionalities in QuLine, QuHybrid is able to conduct the testcross and hybrid performance prediction, making it possible to simulate a hybrid breeding program and to optimize the hybrid breeding strategy.

In this paper, we introduce the underlying principles behind genetics and breeding simulation modeling, and summarize some applications on comparing breeding strategies, choosing parents with known gene information, and optimizing marker-assisted selection strategies.

PRINCIPLES OF BREEDING SIMULATION

QuLine allows for several breeding strategies to be defined simultaneously in one input file. 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' in QuLine is defined as all crossing, seed propagation, and selection activities in an entire breeding cycle. A breeding cycle begins with crossing and ends at the generation when the selected advanced lines are returned to the crossing block as new parents. By defining breeding strategy, QuLine translates the complicated breeding process into a way that the computer can understand and simulate. We illustrate the breeding strategy using CIMMYT's wheat breeding program as an example.

Definition of Breeding Strategies in QuLine

The strategies used by CIMMYT breeders have evolved with time. Pedigree selection was used primarily from 1944 until 1985. From 1985 until the second half of the 1990s the main selection method was a modified pedigree/bulk method (MODPED) (van Ginkel *et al.* 2002), which successfully produced many of the widely adapted wheats now being grown in the developing world. This method was replaced in the late 1990's by the selected bulk method (SELBLK) (van Ginkel *et al.* 2002) in an attempt to improve resource-use efficiency. The major differences between MODPED and SELBLK are outlined below.

The MODPED method begins with pedigree selection of individual plants in the F_2 followed by three bulk selections from F_3 to F_5, and pedigree selection in the F_6; hence the name modified pedigree/bulk. In the SELBLK method, spikes of selected F_2 plants within one cross are harvested in bulk and threshed together, resulting in one F_3 seed lot per cross. This selected bulk selection is also used from F_3 to F_5, while pedigree selection is used only in the F_6. A major advantage of SELBLK compared with MODPED is that fewer seed lots need to be harvested, threshed, and visually selected for seed appearance. In addition, significant savings in time, labor, and costs associated with nursery preparation, planting and plot labelling ensue, and potential sources of error are avoided.

Fig. 1 gives the definition of MODPED in QuLine. At the beginning of this file are some general information used in simulation, such as the number of strategies to be simulated, number of simulation runs, number of breeding cycles, number of crosses to make, and some output options. Before the definition of a strategy, a strategy number, name and the number of generations in this strategy has to be specified.

```
!*********************************QMP file for QuLine 1.2************************************
!*******************General information for the simulation experiment************************
!NumStr NumRun NumCyc NumCro CBUpdate OutGES OutPOP OutHIS OutROG OutCOE OutVar Cross
 1      50     1      1000   0        0      0      0      0      0      0      random
!******************Information for selection strategies to be simulated**********************
!StrategyNumber StrategyName  NumGenerations
   1              MODPED       10
   !NR SS  GT       PT         GA       RP  PS       NL  ET...       Row 1
   !                                AT (ID SP  SM)...                          Row 2
   !                                WT (ID SP  SM)...                          Row 3
   1   0   CB       self       bulk     1   10       1   2
                                                     0
                                                     0
   1   0   F1       singlecross bulk    1   20       1   1
                                                     7   2 0.98 B  3 0.99 B  4 0.85 B  6 0.99 M  7 0.90 T  8 0.98 B  9 0.97 T
                                                     0
   1   0   F2       self       pedigree 1   1000     1   2
                                                     7   2 0.99 B  3 0.99 B  5 0.90 B  6 0.99 M  7 0.99 T  8 0.99 B  9 0.99 T
                                                     8   2 0.95 B  4 0.99 B  5 0.40 B  6 0.85 M  7 0.60 T  8 0.90 B  9 0.50 T  10 0.60 T
   1   0   F3       self       bulk     1   70       1   1
                                                     7   2 0.90 B  3 0.99 B  4 0.70 B  6 0.97 M  7 0.75 T  8 0.95 B  9 0.80 T
                                                     5   4 0.90 B  6 0.95 M  8 0.95 B  9 0.30 T  10 0.60 T
   1   0   F4       self       bulk     1   70       1   2
                                                     6   2 0.90 B  5 0.65 B  6 0.97 M  7 0.85 T  8 0.97 B  9 0.85 T
                                                     5   5 0.90 B  6 0.95 M  8 0.95 B  9 0.30 T  10 0.60 T
   1   0   F5       self       bulk     1   70       1   1
                                                     6   2 0.90 B  4 0.75 B  6 0.97 M  7 0.85 T  8 0.95 B  9 0.85 T
                                                     5   4 0.90 B  6 0.95 M  8 0.95 B  9 0.30 T  10 0.60 T
   1   0   F6       self       pedigree 1   140      1   2
                                                     6   2 0.90 B  5 0.75 B  6 0.97 M  7 0.85 T  8 0.97 B  9 0.85 T
                                                     5   5 0.90 B  6 0.90 M  7 0.95 T  8 0.95 B  9 0.10 T
   4   0   F7       self       bulk     1   70       1   1
                                                     7   2 0.90 B  4 0.75 B  6 0.97 M  7 0.90 T  8 0.95 B  9 0.85 T  10 0.75 T
                                                     0
               F8(T)    self       bulk     1   70       1   2
                                                     6   2 0.95 B  5 0.90 B  6 0.99 M  7 0.98 T  8 0.99 B  9 0.85 T
                                                     0
               F8(B)    self       bulk     1   70       1   3
                                                     1   4 0.90 B
                                                     0
               F8(YT)   self       bulk     1   100      1   1
                                                     1   1 0.40 T
                                                     0
   4   0   F8(SP)   self       bulk     1   30       1   1
                                                     0
                                                     0
               F9(T)    self       bulk     1   70       1   2
```

Fig. 1. Definition of MODPED in QuLine. F8(T), F8 field test in Toluca; F8(B), F8 field test in El Batan; F8(YT), F8 yield trial in Cd. Obregon; F8(SP), F8 small plot evaluation in Cd. Obregon; F9(T), F9 field test in Toluca; F9(B), F9 field test in El Batan; F9(YT), F9 yield trial in Cd. Obregon; F9(SP), F9 small plot evaluation in Cd. Obregon; F10(YR), F10 stripe rust screening in Toluca; F10(LR), F10 leaf rust screening in El Batan.

Number of Generations in MODPED and Number of Selection Rounds in Each Generation

In the breeding program depicted in Fig. 1, the best advanced lines developed from the F_{10} generation will be returned to the crossing block to be used for new crosses; that is to say a new breeding cycle starts after F_{10} leaf rust screening at El Batan. Therefore the number of generations in one breeding cycle is 10 for MODPED. The crossing block (viewed as F_0) and the 10 generations need to be defined in MODPED. 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). Most generations in this breeding program have just one selection round, e.g., F_1 to F_6, while some generations have more than one selection round since they are grown simultaneously at different sites or under different conditions, e.g., F_7, F_8, and F_9 (Column NR in Fig. 1).

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 F_1 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 F_1 is derived from crossing among the parents in the initial population (or crossing block). QuLine 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. 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 F_0 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 QuLine.

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 QuLine 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 F_4 and then doubled haploid production on selected F_4 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.

It can be seen from Fig. 1 that *pedigree* is used in F_2 and F_6 and bulk is used in other generations in MODPED.

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). 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 by environment interaction. Therefore, by defining the target population of environments as a mixture of environment types, genotype by environment interactions are defined as a component of the genetic architecture of a trait. 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, QuLine will randomly assign environment types to them with a likelihood based on the frequencies of environment types in the target population of environments.

For column ET in Fig. 1, 1 is used for the Cd. Obregon environment type, 2 for Toluca, and 3 for El Batan. We can see, for example, that F_7 is grown in the Cd. Obregon environment, F8(T) in Toluca, F8(B) in El Batan, and F8(YT) in Cd. Obregon.

Among Family Selection and Within Family Selection for Each Selection Round

Ten traits have been included as relevant (Table 1) for the selection process in the breeding program described in Fig. 1. 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 Table 1) 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 (Fig. 1). The traits for both among family and within family selections can be the same or different, as is the case for selected proportions. The traits for selection may also differ from generation to generation, as may the selected proportions for traits.

Taking F_6 as an example, 6 traits are used in among family selection, and they are traits 2 (lodging), 5 (leaf rust), 6 (height), 7 (heading), 8 (tillering), and 9 (grains per spike). Five traits are used in within family selection, and they are traits 5 (leaf rust), 6 (height), 7 (heading), 8 (tillering), and 9 (grains per spike). It should be noted that some new functionalities have just been added to QuLine to select families or individuals with trait values above or below some pre-assigned values.

COMPARISON OF BREEDING EFFICIENCIES OF DIFFERENT SELECTION STRATEGIES

The genetic models developed accounted for epistasis, pleiotropy, and genotype by environment (GE) interaction (Table 1). The simulation experiment comprised the same 1000 crosses, developed from 200 parents, for both breeding strategies. A total of 258 advanced lines remained following 10 generations of selection. The two strategies were each applied 500 times on 12 GE systems.

Genetic Gain in Yield from MODPED and SELBLK

The average adjusted gains were 6.70 for no epistasis, 5.36 for di-genic epistasis, and 5.71 for tri-genic epistasis, which indicates that epistasis will reduce the adjusted gain. The adjusted gain associated with the absence of pleiotropy is also higher than that for the presence of pleiotropy. These results show that the increase in gene number and the presence of epistasis and pleiotropy make it more difficult for a breeding strategy to identify the trait performance level of the best genotype in the defined GE system. When the experimental factors are considered individually, the adjusted gain from SELBLK is always significantly higher than that from MODPED, except in the absence of pleiotropy, indicating SELBLK is at least equivalent to or better than MODPED.

The average adjusted genetic gain on yield is 5.83 for MODPED and 6.02 for SELBLK a difference of 3.3% (Fig. 2a). This difference is not large and therefore unlikely to be detected using field experiments (Singh *et al.* 1998). However, it can be detected through simulation, which indicates that the high level of

Table 1. Number of segregating genes and their genetic effects in the Cd. Obregon environment type in CIMMYT's Wheat Breeding Program. (adapted from Table 1, Wang *et al*. (2003))

Gene classification†	Number of genes	Traits effected	Individual gene effect		
			AA	Aa	aa
Yield	20	Yield (t/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 (t/ha)	0.00	– 0.40	– 0.80
Stem rust	5	Stem rust (%)	0.00	0.50	1.00
		Yield (t/ha)	0.00	– 0.25	– 0.50
		Kernel weight (g)	0.00	– 0.75	– 1.50
Stripe rust	5	Stripe rust (%)	0.00	0.00	0.00
Leaf rust	5	Leaf rust (%)	0.00	5.00	10.00
		Yield (t/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 (t/ha)	1.00	0.50	0.00
		Lodging (%)	2.00	1.00	0.00

replication (50 models by 10 runs in this experiment) feasible with simulation can better account for the stochastic properties of a run of a breeding strategy and for the sources of experimental errors. The average adjusted gains for the two yield gene numbers 20 and 40 are 6.83 and 5.02, respectively, suggesting that genetic gain decreases with increasing yield gene number.

Number of Crosses Remaining after Selection

The same 1000 crosses were made for both breeding strategies and 258 advanced lines were selected after a breeding cycle, regardless of the GE system used. The number of crosses remaining after one breeding cycle is significantly different among models and strategies, but not among runs. The number of

crosses remaining from SELBLK is always higher than that from MODPED, which means that delaying pedigree selection favors diversity. On average, 30 more crosses were maintained in SELBLK (Fig. 2b). However, there is a crossover between the two breeding strategies (Fig. 2b). Prior to F_5 the number of crosses in MODPED is higher than that in SELBLK. The number of crosses becomes smaller in MODPED after F_5 when pedigree selection is applied in F_6. Among-family selection from F_1 to F_5 in SELBLK is equal to among-cross selection, and results in a greater reduction in cross number for SELBLK compared to MODPED in the early generations. In general, only a small proportion of crosses remains at the end of a breeding cycle (11.8% for MODPED and 14.8% for SELBLK); therefore, intense among-cross selection in early generations is unlikely to reduce the genetic gain. On the contrary, breeders will tend to concentrate on fewer but "higher probability" crosses (Simmonds 1996). That just a few crosses of the many generated remain after the final yield trial stage is common in most breeding programs. Since more crosses remain in SELBLK, the population following selection from SELBLK may have larger genetic diversity than that from MODPED. In this context, SELBLK is also superior to MODPED.

Resource Allocation

Since the number of families and selection methods after F_8 are basically the same for both MODPED and SELBLK, only the resources allocated from F_1 to F_8 are compared. The total number of individual plants from F_1 to F_8 was calculated to be 5,155,090 for MODPED and 3,358,255 for SELBLK (Fig. 2c). Assuming that planting intensity is similar, SELBLK will use approximately two thirds of the land allocated to MODPED. Furthermore, SELBLK produces a smaller number of families compared to MODPED (Fig. 2d). From F_1 to F_8, there are 63,188 families for MODPED but only 24,260 for SELBLK, approximately 40% of the number for MODPED. Therefore when SELBLK is used, fewer seed lots need to be handled at both harvest and sowing, resulting in significant savings in time, labor and cost.

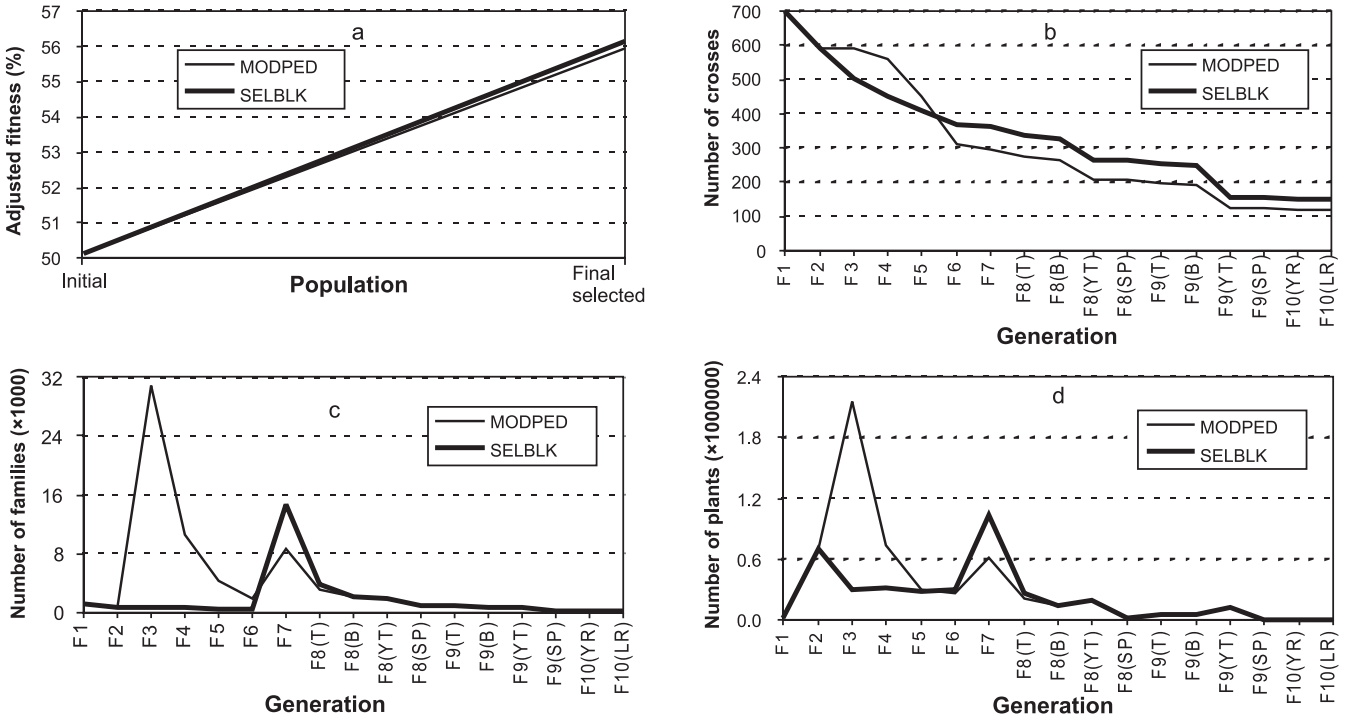

Fig. 2. Results from the simulation experiment (a) Adjusted genetic gain after one breeding cycle across all experimental sets. (b) Number of crosses after each generation's selection across all experimental sets. (c) Number of individual plants in each generation in one breeding cycle. (d) Number of families in each generation in one breeding cycle. F8(T), F8 field test in Toluca; F8(B), F8 field test in El Batan; F8(YT), F8 yield trial in Cd. Obregon; F8(SP), F8 small plot evaluation in Cd. Obregon; F9(T), F9 field test in Toluca; F9(B), F9 field test in El Batan; F9(YT), F9 yield trial in Cd. Obregon; F9(SP), F9 small plot evaluation in Cd. Obregon; F10(YR), F10 stripe rust screening in Toluca; F10(LR), F10 leaf rust screening in El Batan. (adapted from Fig. 2, Wang *et al.* (2003)).

OPTIMIZATION OF THE MARKER-ASSISTED SELECTION STRATEGIES

Three Wheat Parental Lines with Nine Major Genes

Sunstate is a commercial Australian line, HM14BS is a source of the 'long coleoptile' trait that utilizes the *Rht8* allele for reduced height, and Silverstar+tin is a modified Australian variety that is a source of the tin 'reduced-tillering' trait (Table 2). The target genotype given in Table 2 was determined by semi-dwarfing with long coleoptile length, multiple resistances, good grain quality, and reduced tillering. Any of the three semi-dwarfing alleles, i.e. *Rht-B1b*, *Rht-D1b*, and *Rht8*, will be able to produce the required plant height, and multiple dwarfing alleles make the plant too short to be useful. However, *Rht-B1b* and *Rht-D1b* also reduce the coleoptile length as well as plant height, contributing to reduced drought-resistance. *Rht8* reduces the plant height without affecting the coleoptile length. Therefore *Rht8* is the favorable dwarfing allele, and should be present in our target genotype. Other alleles in the target genotype are easily understood as they increase the resistance to some diseases, increase the grain quality, or reduce the number of tillers.

Target alleles are distributed unequally between the 3 parents with HM14BS carrying three target alleles, Sunstate carrying five target alleles, and Silverstar+tin carrying four target alleles. A topcross between lines HM14BS, Sunstate and Silverstar+tin (Table 2) was simulated to determine the minimum population sizes required to recover a target genotype, given selection among DH lines with and without prior enrichment in the F_2 generation. The frequency of the target genotype will be maximized if Sunstate is used as the third parent in topcrossing, so the other two topcrosses were not considered.

Selection in the F_1 Generation of the Topcross

In the F_1 generation of the topcross (TCF1), *Rht-B1*, *Rht8*, *Cre1*, *Glu-B1*, and *tin* are segregating. The target genotypes of *Rht-B1aRht-B1a* and *Glu-B1iGlu-B1i* have a frequency of 0.5 in TCF1, and all other target alleles exist in heterozygous form at frequencies of 0.5. Therefore selection of *Rht-B1a* and *Glu-B1i* homozygotes and allele enrichment for *Rht8*, *Cre1*, and tin can be applied in TCF1, and the theoretical selected proportion in TCF1 is 0.55=0.0313. Considering this high proportion and for simplicity, no other selection option was applied in TCF1.

Table 2. Nine major genes, their locations on chromosomes, and the genotypes in three wheat parental lines (adapted from Table 1, Wang *et al.* (2007)).

Gene (locus) symbol	*Rht-B1*	*Rht-D1*	*Rht8*	*Sr2*	*Cre1*	*VPM*	*Glu-B1*	*Glu-A3*	*tin*
Chromosome	4BS	4DS	2DL	3BS	2BL	7DL	1BL	1AS	1AS
Marker type	Codom	Codom	Codom	Codom	Dom	Dom	Codom	Codom	Codom
Distance between marker and gene (cM)	0.0	0.0	0.6	1.1	0.0	0.0	0.0	0.0	0.8
HM14BS	*Rht-B1a*	***Rht-D1a***	***Rht8***	*sr2*	*cre1*	*vpm*	*Glu-B1a*	*Glu-A3e*	*Tin*
Sunstate	***Rht-B1a***	*Rht-D1b*	*rht8*	***Sr2***	*cre1*	***VPM***	***Glu-B1i***	***Glu-A3b***	*Tin*
Silverstar+*tin*	*Rht-B1b*	***Rht-D1a***	*rht8*	*sr2*	***Cre1***	*vpm*	***Glu-B1i***	*Glu-A3c*	***tin***
Target genotype	***Rht-B1a***	***Rht-D1a***	***Rht8***	***Sr2***	***Cre1***	***VPM***	***Glu-B1i***	***Glu-A3b***	***tin***

Note : Alleles *Rht-B1b*, *Rht-D1b*, and *Rht8* reduce plant height. Allele *Sr2* confers resistance to stem rust, and alleles *Cre1* and *VPM* confer resistance to cereal cyst nematode. Alleles *Glu-B1i* and *Glu-A3b* improve dough quality, and allele *tin* reduces the tiller number. The genes are all unlinked, except for *Glu-A3* and *tin* which are 3.8 cM apart on chromosome 1A. The target genotype is determined when all the 9 genes are considered together. Alleles in the target genotype contribute to semi-dwarfing with long coleoptile length, multiple disease resistances, good grain quality, and less tillering. The three semi-dwarfing alleles can all produce the required plant height. However, *Rht-B1b* and *Rht-D1b* also reduce the coleoptile length, which is unfavorable for breeding drought-resistant wheat cultivars. *Rht8* reduces the plant height without affecting the coleoptile length, and therefore is the favorable dwarfing allele. Other alleles in the target genotype are easily understood as they increase the resistance to some diseases, increase the grain quality, or reduce the number of tillers.

Selection in the F_2 and F_2-derived DH Generation of the Topcross

The target genotype lacks *Rht-B1b* and *Rht-D1b* and is homozygous for *Rht8*, *Sr2*, *Cre1*, *VPM*, *Glu-B1i*, *Glu-A3b*, and *tin* (last row in Table 2). We considered three options for selection in TCF2: (1) no selection in TCF2, (2) F_2 enrichment for all genes except *Rht-B1* and *Glu-B1* (as *Rht-B1a* and *Glu-B1i* have been fixed after selection of the homozygotes in TCF1 at the two loci), and (3) selection of *Rht8* homozygotes and F_2 enrichment of all remaining alleles. Selection of homozygotes at two loci in TCF_2 was also simulated but a much larger minimum population size in TCF_2 was required (results not shown).

For the three options considered, selection of target homozygotes was conducted in DHs, i.e. the first option (no selection in TCF_2) consists of two selection stages, one in TCF_1, the other in DHs. The simulation shows the proportion selected in TCF_1 is close to the theoretical upper limit of 0.0313 (Table 5). The selected proportion in DHs is about 0.0009, requiring quite a large DH population to select the target genotype. The second and the third options both consist of three selection stages, one in TCF_1, one in TCF_2, and one in DHs. For the second option, the selected proportion is 0.1190 in TCF_2, and 0.0071 in DHs. The third option has a more evenly-distributed selected proportion over stages and requires the smallest number of lines overall (Table 3). In practice, if multi-stage selection is applied, the general rule to minimize population size would be to minimize differences in selection intensity at the different stages, which will minimize cost if markers are equal in cost. Multiplexing appropriate sets of markers provides further cost savings.

Final Target Allele Frequencies Following Marker-assisted Selection

Due to the complete linkage of genes *Rht-B1*, *Rht-D1*, *Cre1*, *VPM*, *Glu-B1* and *Glu-A3* with their markers (Table 2), the frequencies of alleles *Rht-B1a*, *Rht-D1a*, *Cre1*, *VPM*, *Glu-B1i* and *Blu-A3b* are 1.0 after marker-assisted selection in the final selected population. *Rht8* has a distance of 0.6 cM to its marker, and *Sr2* 1.1 cM to its marker. Through simulation, we found the allele frequency is near 0.99 for *Rht8*, and 0.98 for *Sr2* after marker-assisted selection, which should be acceptable in practical breeding.

Given that *tin* and its microsatellite marker are 0.8 cM apart, the estimated allele frequency of tin is at 0.77 in the final selected population. The reason for the lower than expected frequency is due to its linkage in repulsion with the important glutenin allele, *Glu-A3b*, in parents Sunstate and Silverstar+tin (Table 3). The haplotype frequency from the biparental cross between

Table 3. Selected proportion and number of individuals (or DH lines) selected in each marker selection scheme (adapted from Table 3, Wang *et al.* (2007)).

Breeding population	No enrichment selection in TCF_2		Enrichment selection for all target genes in TCF_2		Homozygous selected for *Rht8*, and enrichment selection for others in TCF_2	
	Selected proportion	Minimum population size	Selected proportion	Minimum population size	Selected proportion	Minimum population size
TCF_1†	0.0313	145	0.0316	144	0.0313	145
TCF_2			0.1190	37	0.0397	114
DHs derived from TCF_2	0.0013	3440	0.0112	408	0.0160	286
Total population size required		3585		589		545

Note : In TCF_1, homozygous selection is conducted for *Rht-B1a* and *Glu-B1i*, and enrichment selection for *Rht8*, *Cre1*, and *Tin*. The other loci are not segregating in TCF_1. The homozygous frequency for *Rht-B1a* and *Glu-B1i*, and the heterozygous frequencies for *Rht8*, *Cre1*, and *tin* are all equal to 0.5. So the theoretical selected proportion in TCF_1 is 0.5^5=0.0313.

Sunstate and Silverstar+tin illustrates the effect of repulsive linkage on allele frequency. When three linked loci *Glu-A3*, *tin* and marker for *tin* (denoted as *Mtin*) are considered, there are eight haplotypes. When no crossover interference is assumed, the frequency of each haplotype can be calculated from the recombination frequency between *Glu-A3* and *tin*, and between tin and its marker. After marker-assisted selection for *Glu-A3b* and *tin*, only haplotypes 2 and 3 are retained, with a frequency for tin of 0.01488 / (0.01488 + 0.00388) = 0.79318, which in turn confirms our simulation results. The frequency of tin may not be sufficient, and therefore the presence of the tin allele following marker-assisted selection must be confirmed by other methods.

Optimum Strategy to Combine Nine Genes from a Topcross

In summary, the optimum strategy to combine the nine target alleles in the topcross Silverstar+tin/ HM14BS//Sunstate can be divided into four steps: Step 1 – selection of Sunstate as the final parent (having largest number of favorable alleles) in the topcross; Step 2 - selection for *Rht-B1a* and *Glu-B1i* homozygotes, and enrichment of *Rht8*, *Cre1*, and tin in TCF1; Step 3 – selection of homozygotes for one target allele, e.g. *Rht8*, and enrich remaining target alleles in TCF2; Step 4 - selection of the target genotype (last row in Table 1) in DHs/RILs. The selected proportion in Table 5 can be used to determine the minimum population size for each selection stage. At this point, the presence of the tin gene needs to be reconfirmed by phenotyping. Currently, laboratory progeny marker screening and field selection experiments are underway with these populations so that we can validate the simulation results.

CONCLUSIONS

The breeding methods that can be simulated in QuLine include mass selection, pedigree breeding (including single seed descent), bulk population breeding, backcross breeding, topcross (or three-way cross) breeding, doubled haploid breeding, marker assisted selection, and combinations and modifications of these methods. The chromosomal locations of genes and markers, and their occurrence in specific parents can be explicitly and precisely defined. Simulation experiments can therefore be designed to compare the breeding efficiencies of different selection strategies under a series of pre-determined genetic models. A great amount of studies on QTL mapping have been conducted for various traits in plants and animals in the recent ten years. How QTL mapping results can be used to pyramid desired alleles at various loci has only rarely been addressed in the literature. As the number of published genes and QTLs for various traits continues to increase, the challenge for plant breeders is to determine how to best utilize this multitude of information in the improvement of crop performance. QuLine provide an appropriate tool that can combine different types and levels of biological data such that the complex and voluminous data is turned into knowledge that can be applied in breeding.

ACKNOWLEDGEMENTS

Development of the QU-GENE simulation tools was supported by the GRDC (Grain Research and Development Corporation, Australia), and the Generation Challenge Program.

REFERENCES

Allard, R.W. (1960). *Principles of Plant Breeding*. John Wiley and Sons, New York.

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

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

Fehr, W.R. (1987). Principles of Cultivar Improvement. Vol. 1. Theory and technique. Macmillian Publishing Company, New York.

Kempthone, O. (1988). An overview of the field of quantitative genetics. p. 47-56. In: B.S. Weir, E.J. Eisen, M.M. Goodman and G. Namkoong (ed.) *Proceedings of the second international conference on quantitative genetics. Sinauer Associate Inc., Sunderland, MA*.

Podlich, D.W. and Cooper, M. (1998). QU-GENE: A platform for quantitative analysis of genetic models. *Bioinformatics*, **14**, 632-653.

Simmonds, N.W. (1996). Family selection in plant breeding. *Euphytica*, **90**, 201-208.

Singh, R.P., Rajaram, S., Miranda, A., Huerta-Espino, J. and Autrique, E. (1998). Comparison of two crossing and four selection schemes for yield, yield traits, and slow rusting resistance to leaf rust in wheat. *Euphytica*, **100**, 35-43.

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

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

Wang, J., van Ginkel, M., Trethowan, R., Ye, G., DeLacy, I., Podlich, D. and Cooper, M. (2004). Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. *Crop Sci.*, **44**, 2006-2018.

Wang, J., Eagles, H.A., Trethowan, R. and van Ginkel, M. (2005). Using computer simulation of the selection process and known gene information to assist in parental selection in wheat quality breeding. *Aust. J. Agric. Res*., **56**, 465-473.

Wang, J. and Pfeiffer, W.H. (2007a). Simulation modeling in plant breeding: principle and applications. *Agric. Sci. China*, 908-921.

Wang, J., Li, H., Wan, X., Pfeiffer, W., Crouch, J., Wan, J. (2007b). Application of identified QTL-marker associations in rice quality improvement through a design breeding approach. *Theo. Appl. Genet*., **115**, 87-100.

Wang, J., Chapman, S.C., Bonnett, D.B., Rebetzke, G.J. and Crouch, J. (2007c). Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. *Crop Sci.*, **47**, 580-588.