# PLANT BREEDING

Edited by **Ibrokhim Y. Abdurakhmonov**

# Contents

# Modelling and Simulation of Plant Breeding Strategies

Jiankang Wang
*Institute of Crop Science and CIMMYT China,*
*Chinese Academy of Agricultural Sciences (CAAS),*
*China*

## 1. 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 environment or a target population of environments (TPE). To achieve this objective, plant breeders employ a range of selection methods (Allard, 1999; Hallauer et al., 1988). Quantitative genetic theory generally provides much of the framework for the design and analysis of selection methods used within breeding programs, based on various assumptions in order to render mathematically or statistically tractable theories (Hallauer et al., 1988; Falconer and Mackay, 1996; Lynch and Walsh, 1998). Some of these assumptions can be easily tested or satisfied by certain experimental designs; others can seldom be met, such as the assumptions of no linkage and no genotype by environment (G×E) interaction. Still others, such as the presence or absence of epistasis and pleiotropy, are difficult to test. Field experiments have been conducted to compare the efficiencies from different breeding methods. However, due to the time and effort needed to conduct field experiments, the concept of modeling and prediction have always been of interest to plant breeders. Computer simulation gives breeders the opportunity to lessen the impact of these assumptions, thereby establishing more valid genetic models for use in plant breeding (Kempthone, 1988). Simulation as a tool has been applied in many special plant breeding studies that use relatively simple genetic models. A tool capable of simulating the performance of a breeding strategy for a continuum of genetic models ranging from simple to complex, embedded within a large practical breeding program including marker-assisted-selection, had not been available until recently (Wang et al., 2003; Wang and Pfeiffer, 2007). In this chapter, the principles and applications of simulation modeling in plant breeding are introduced.

## 2. Principles of plant breeding simulation

### 2.1 Available simulation tools

QU-GENE is a simulation platform for quantitative analysis of genetic models, which consists of the two-stage architecture (Podlich and Cooper, 1998). The first stage is the engine, and its role is to: (1) define the gene and environment (G×E) interaction 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 population of individuals within the G×E system defined by the engine. The application module usually represents the operation of breeding programs.

Three application modules have been developed. QuLine, a computer program, was firstly designed in 2002-2003 for simulating CIMMYT's wheat breeding, one of the most successful wheat breeding programs in the world. QuLine can integrate enormous amounts of data from different sources, process them in many ways, and produce alternative theoretical but realistic scenarios that the breeder can draw on to make a decision. It can simulate almost all breeding activities in CIMMYT's wheat breeding program, including male master selection, female master selection, parental selection, single cross, backcross, top cross, double cross, doubled haploid, marker-assisted selection, pedigree breeding, selected bulk etc. QuLine can simulate other breeding programs for selecting inbred lines, which means all major food cereals in the world, plus basically all leguminous crops (Wang et al., 2003 and 2004).

Taking advantage of the sophisticated state of QuLine, QuHybrid was developed in 2008-2009. The major development required for QuHybrid is the implementation of test crossing. To make the testcrosses, an additional population defining all the testers was added (Zhang et al. 2011). When the testcross functionality is activated, testcrosses will be made between all families and testers. Among-family selection is conducted based on the mean performance of all testcrosses in each tested family. Breeding methods can be compared by the line *per se* and testcross genetic gains. For hybrid prediction, another population consisting of inbred lined in another heterotic group is also needed to run QyHybrid. At the end of each breeding cycle, performance of all potential $F_1$ hybrids between the final selected inbred lines and lines in the other heterotic group are predicted.

Marker assisted recurrent selection (MARS) was proposed to overcome the disadvantages when using markers in selecting complex traits. It has been commercially used for selecting complex traits in maize, sunflower and soybean breeding programs (Bernardo and Charcosset, 2006). As a result, QuMARS was developed in 2009-2010. Prediction models include best linear unbiased prediction (BLUP), and regression models. Prediction can be built on both line *per se* performance and testcross performance. Therefore, it can simulate both MARS and genomic selection (GS), starting from a single cross between two parental lines. With QuMARS, various issues using MARS or GS can be investigated. For example, how many cycles of recurrent selection are suitable? How many markers should be used? How can the breeding values of lines under development be best predicted?

## 2.2 Selected applications of modeling and simulation in plant breeding

Simulation can be used to investigate both strategic (say comparison of two breeding methods; Wang et al., 2003, 2004, 2009b) and tactical (say identification of optimum crossing and selection schemes given the gene distribution in parents; Wang et al., 2005, 2007a, 2007b, 2009a) issues in plant breeding. Two strategic and two tactical applications using QuLine are summarized in this section.

### 2.2.1 Comparison of two breeding strategies in CIMMYT's wheat breeding

The main elements of international wheat improvement program at CIMMYT have been shuttle breeding at two contrasting locations in Mexico, wide adaptation, durable rust and

Septoria resistances, international multi-environment testing, and the appropriate use of genetic variation to enhance yield gains (Rajaram et al., 1994; Rajaram, 1999). Two breeding strategies are commonly used in CIMMYT's wheat breeding programs (van Ginkel et al., 2002; Wang et al., 2003, 2004). The modified pedigree (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 $F_6$. In the selected bulk (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 $F_5$, whereas, pedigree selection is used only in the $F_6$. Assuming that planting intensity is similar, SELBLK uses approximately two thirds of the land allocated to MODPED, and produces smaller number of families. Therefore when SELBLK is used, fewer seed lots need to be handled at both harvest and sowing, resulting in a significant saving in time, labor, and cost. Will the two strategies result in similar genetic gain on yield and other breeding traits?

The genetic models developed accounted for epistasis, pleiotropy, and G×E. For both breeding strategies, the simulation experiment comprised of the same 1000 crosses developed from 200 parents. A total of 258 advanced lines remained following 10 generations of selection. The two strategies were each applied 500 times on 12 GE systems (Wang et al., 2003). The average adjusted genetic gain on yield across all genetic models is 5.83 for MODPED and 6.02 for SELBLK, with a difference of 3.3%. 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 replication (50 models by 10 runs in this experiment) is feasible with simulation and can better account for the stochastic properties from a run of a breeding strategy and 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.

The number of crosses remaining after one breeding cycle is significantly different among models and strategies, but not among runs (Wang et al., 2003). The number of crosses remaining from SELBLK is always higher than that from MODPED, which means that delaying pedigree selection favors diversity. On an average, 30 more crosses were maintained in SELBLK. However, there was a crossover between the two breeding strategies. Prior to $F_5$ the number of crosses in MODPED was higher than that in SELBLK. The number of crosses became smaller in MODPED after $F_5$, when pedigree selection was applied in $F_6$. Among-family selection from $F_1$ to $F_5$ in SELBLK was equal to among-cross selection, and resulted in a greater reduction in the cross numbers for SELBLK compared to MODPED, in the early generations. In general, only a small proportion of crosses remained at the end of a breeding cycle (11.8% for MODPED and 14.8% for SELBLK); therefore, intense among-cross selection in early generations was unlikely to reduce the genetic gain. On the contrary, breeders would tend to concentrate on fewer but "higher probability" crosses. As more crosses remained in SELBLK, the population following selection from SELBLK might have a larger genetic diversity than that from MODPED. In this context also, SELBLK is superior to MODPED.

### 2.2.2 Modeling of the single backcrossing breeding strategy

Regarding the crossing strategies in CIMMYT wheat breeding, top (or three-way) crosses and double (or four-way) crosses were employed to increase the genetic variability of

breeding populations in the early 1970s. By the late 1970s, double crosses were dropped due to their poor results relative to single cross, top crosses and limited backcrosses. From the 1980s onwards, all crosses onto selected $F_1$ generations were single cross, backcrosses or top crosses (van Ginkel et al., 2002). Single and top (or three-way) crosses are commonly used among adapted parental lines, while backcrosses are preferred for transferring a few useful genes from donor parents to adapted lines. In CIMMYT, the single backcrossing approach (one backcross to the adapted parent) was initially aimed at incorporating resistance to rust diseases based on multiple additive genes (Singh and Huerta-Espino, 2004). However, it soon became apparent that the single backcross approach also favored selection of genotypes with higher yield potential. The reason why single backcrossing shifts the progeny mean toward the higher side is that it favors the retention of most of the desired major additive genes from the recurrent, while simultaneously allowing the incorporation and selection of additional useful small-effect genes from the donor parents.

The breeding efficiency of this strategy compared with other crossing and selection strategies was investigated through computer simulation for many scenarios, such as the number of genes to be transferred, frequency of favorable alleles in donor and recurrent parents etc. Results indicated this breeding strategy has advantages in retaining or overtaking the adaptation of the recurrent parents and at the same time transferring most of the desired donor genes for a wide range of scenarios (Wang et al., 2009). Two times of backcrossing have advantages when the adaptation of donor parents is much lower than that of the adapted parents, and the advantage of three times of backcrossing over two times of backcrossing is minimal. We recommend the use of single backcrossing breeding strategy based on three assumptions: (1) multiple genes governing the phenotypic traits to be transferred from donor parents to adapted parents, (2) donor parents still have some favorable genes that may contribute to the improvement of adaptation in the recipient parents even under low adaptation, and (3) the conventional phenotypic selection is applied or the individual genotypes cannot be precisely identified.

### 2.2.3 Optimization of marker assisted selection (MAS)

Many breeding programs in a range of crops are using molecular markers to screen for one to several alleles of interest. The availability of an increasing number of useful molecular markers is allowing accurate selection at a greater number of loci than has been previously possible (Dekkers and Hospital, 2002; Dubcovsky, 2004). However, larger population sizes are required to ensure with reasonable certainty that an individual with the target genotype is present. Different crossing and selection strategies may require vastly different population sizes to recover a target genotype with the same certainty even when the same parents are used (Bonnett et al., 2005). Determination of the most efficient strategy has the potential to dramatically decrease the amount of resources (plants, plots, marker assays, and labor) required to combine a set of target alleles into a new genotype.

The drought-suitable lines in wheat should be semi-dwarf with long coleoptiles, resistant to multiple diseases, have good dough properties, and have productive tillers. To achieve this, nine target alleles need to be combined into one genotype (Wang et al., 2007a). Three parent lines were used: Sunstate, a commercial Australian line; HM14BS, a germplasm line combining an allele for height reduction and long coleoptiles; and Silverstar+tin, a derivative of Silverstar with a restricted tillering allele. The largest target genotype

frequency was found in the Silverstar+tin/HM14BS//Sunstate topcross. The optimum MAS strategy to combine the nine target alleles from this topcross could be divided into three steps: (i) selection for Rht-B1a and Glu-B1i homozygotes, and enrichment selection of Rht8c, Cre1, and tin in top cross $F_1$, (ii) selection of homozygotes for one target allele, e.g. Rht8c, and enrich the remaining target alleles in top cross $F_2$, and (iii) selection of the target genotype with doubled haploid lines or recombination inbred lines. Enrichment of allelic frequencies in top cross $F_2$ reduced the total number of lines screened from >3500 to <600.

### 2.2.4 Design breeding with known gene information

The concept of design breeding was proposed in recent years as the fast development in molecular marker technology (Peleman and Voort, 2003; Wang et al., 2007b). Three steps are involved in design breeding. The first step is to identify the genes for breeding traits, the second step is to evaluate the allelic variation in parental lines, and the third step is to design and conduct breeding. A permanent mapping population of rice consisting of 65 non-idealized chromosome segment substitution lines (denoted as CSSL1 to CSSL65) and 82 donor parent chromosome segments (denoted as M1 to M82) was used to identify QTL with additive effects for two rice quality traits, area of chalky endosperm (ACE) and amylose content (AC), by a likelihood ratio test based on stepwise regression. These CSS lines were generated from a cross between the japonica rice variety Asominori (the background parent, denoted as P1) and the indica rice variety IR24 (the donor parent, denoted as P2) (Wan et al., 2004, 2005).

Through QTL studies, it is impossible to derive an inbred with the minimum of ACE and the maximum of AC, because QTL on segments M35, M57, and M59 have unfavorable pleiotropic effects on ACE and AC. However, the ideal inbred with relatively low ACE and high AC can be identified through simulation (Wang et al. 2007b). This designed inbred contains four segments from P2, which are, M19, M35, M57, and M60, and another genome is from the background P1. The value of ACE in this inbred is 9.2%, where the theoretical minimum ACE is 0. The value of AC is 17.73%, whereas, the theoretical maximum of AC is 22.3%. Among the 65 CSS lines, the three lines, CSSL15, CSSL29, and CSSL49, have the required target segments, therefore, can be used as the parental lines in breeding. Three possible topcrosses can be made among the three parental lines, Topcross 1: (CSSL15 × CSSL29) × CSSL49, Topcross 2: (CSSL15 × CSSL49) × CSSL29, and Topcross 3: (CSSL29 × CSSL49) × CSSL15. Different MAS schemes can be used to select the target inbred line. Here two schemes are considered, Scheme 1: 200 topcross $F_1$ (TCF$_1$) are first generated. Then 20 doubled haploid (DH) are derived from each TCF$_1$ individual. The target inbred lines are selected from the 4000 DH lines. Scheme 2: 200 TCF1 are first generated. An enhancement selection (Wang et al., 2007a), is conducted among the 200 TCF$_1$ individuals. Then 20 doubled haploid (DH) are derived from each selected TCF$_1$ individual. The target inbred lines are selected from those derived DH lines.

From 100 simulation runs, it was found that by using Scheme 1, 27 target inbred lines were selected from Topcross 1, 13 from Topcross 2, and 8 from Topcross 3. Therefore Topcross 1 had the largest probability to select the target inbred line, and should be used in breeding low ACE and AC inbred lines. The two MAS schemes resulted in significant difference in cost when genotyping for MAS. Scheme 1 required 4000 DNA samples for each topcross. On the contrary, Scheme 2 required 462 DNA samples for Topcross 1, 324 for Topcross 2, and

691 for Topcross 3. Topcross 1 combined with Scheme 2 resulted in the least DNA samples per selected line, and therefore was the best crossing and selection scheme.

### 2.3 Definition of a gene and environment system in QU-GENE

G×E system underlies the genetic and environmental model for simulation experiments. In general, information about a G×E system consists of some general information, the target population of environments (TPE) for the breeding program, traits to be selected during the breeding procedure, random environmental deviations for these traits, genes for traits, their locations on chromosomes, and their effects on traits in different environment types. Information about the population consists of the number of parents and their genotypes.

**Genes distributed on chromosomes 1 to 5**

Recombination frequency 0.05

Gene of maturity, additive effect is 3 days on maturity, and 0.1 t/ha on yield

Gene of TKW, additive effect is 2 g on TKW, and 0.1 t/ha on yield

Recombination frequency 0.05

Gene of yield per se, additive effect is 0.1 t/ha on yield

**Genes distributed on chromosomes 6 to 20**

Gene of yield per se, additive effect is 0.1 t/ha on yield

Fig. 1. A putative genetic model consisting of five genes for maturity, five genes for TKW (thousand kernel weight), and 20 gene for yield.

As a simplified example, we assume TPE of a plant breeding program only contain one environment type, and three traits are used in selection in, i.e., maturity, thousand kernel weight (TKW), and yield. A putative genetic model consists of 5 genes for maturity, five genes for TKW, and 20 genes for yield (Fig. 1). Each maturity gene has an additive effect of 3 days on maturity, and 0.1 t/ha on yield (Fig. 1). Each TKW gene has an additive effect of 2 g on TKW, and 0.1 t/ha on yield. Each yield gene has an additive effect of 0.1 t/ha on yield (Fig. 1). One maturity gene, one TKW gene and one yield gene are linked on each of the first 5 chromosomes, and one yield gene is located on chromosomes 6 and 20 (Fig. 1). These information needs to be organized in certain formats in QU-GENE.

### 2.3.1 General information about a G×E system

The first part is the general information about the G×E system (Fig. 2). Number of models is specifically designed for a G×E system with random gene effects. For a G×E system with all gene effects (additive, dominance, epistasis, and pleiotropy) fixed, this parameter should be set at 1. The random effects model in a G×E system will most likely mimic the real genetic effects of a large number of genes, such as the genes for yield. With this model some genes will have relatively larger effects and others, smaller effects. The large number of G×E systems, different yield gene effects in each G×E system, and replications feasible within the

simulation allow many potential realities to be compared. If one breeding methodology is superior to another for all, or most, permutations, the breeder can be confident that a superior breeding methodology has been identified that is also robust to the complexities and perturbations that may emerge, regardless of the G×E system. Random seed of random gene effects will ensure that the same gene effects will be assigned whenever the G×E system is used, so that all random gene effects are repeatable.

```
! **************************************************************************
! *   QUGENE engine input file
! *
! **************************************************************************
! *** General information on the G-E system ***

! Engine G-E output filename prefix (*.ges)
WheatModel

1                   ! Number of models
0                   ! Random seed of random gene effects
30                  ! Number of genes (includes markers and qtls)
1                   ! Number of environment types
3                   ! Number of traits (not including markers)
1 1 1 0 0 0 0   ! Specify names (ETs, Trts, Genes, Alls, EPN, GPM, pop)
```

Fig. 2. General information about a gene and environment system in QU-GENE

### 2.3.2 Environment information

The TPE for a breeding program consists of a set of distinct, relatively homogenous environment types, each with a frequency of occurrence. Each environment type has its own gene action and interaction, providing the framework for defining G×E interactions (Fig. 3). Each environment type takes three rows. Row 1 is an ID number to distinguish each environment type (arranged in order and starting from 1). Row 2 gives the name of the environment type (if defined). If the indicator for environment type names is 1 (Fig. 3), a valid name must be specified for each environment type. If the indicator for environment type names is 0, the place is left blank. Row 3 specifies the frequency of occurrence in the TPE. Each frequency should be equal to or greater than 0.0, and the sum of all frequencies should be equal to 1.0. In Fig. 3, the one environment type is given the name "Obregon", with a frequency of 1.0 in TPE.

### 2.3.3 Trait information

For the purpose of simulation, the genotypic value of an individual can be calculated from the definition of gene actions in the G×E system and from its genotypic combination. However, breeders select based on the phenotypic value in the field. Therefore, the phenotypic value of a genotype in a specific environment needs to be defined from its genotypic value and associated environmental errors. The trait information will allow QuLine to define the phenotypic value from the genotypic value. Major trait information required by the QU-GENE engine is the environmental effects on traits (within-plot variance and among-plot variance) in each environment type. Either the variance or individual plant-level heritability in the broad sense needs to be specified. For heritability, the QU-GENE

engine will convert the specified heritability into an estimate of environmental variance based on the provided reference population. This environmental variance is used throughout the simulation. The population structure differs from generation to generation; hence, population heritability also varies with changes in the genetic variation within the population.

Each trait takes four rows in Fig. 3. Row 1 is an ID number to distinguish each trait. Row 2 gives the name of the trait (if defined). Row 3 specifies that heritability will be defined for within plot, among-plot error will be defined as a proportion of within-plot error variance. Row 4 specified heritability or error variance in "Obregon", depending on indicators of Row 3. In Fig. 3, the within-plot heritability is 0.4 for maturity, 0.3 for TKW, and 0.2 for yield. The engine will use a reference population to calculate with-in plot error variance of each trait. The among-plot error is defined as 1.0 of the within-plot error. If indicator 2 means the error variance will be given. If more environment types are defined, same information as shown in Row 3 is needed for each environment type.

```
! ************************************************************************
! *** Environment Type Information ***
!     Row 1: Number
!     Row 2: Name (if defined)
!     Row 3: Frequency of occurrence in TPE
! ************************************************************************

1
Obregon
1.000

! ************************************************************************
! *** Trait Information ***
!     Row 1: Number
!     Row 2: Name (if defined)
!     Row 3: Error Specification Type (for within,among,mixture)
!             1=heritability (spb); 2=error
!     Row 4+: Within, Among, Mixture error [each ET]
! ************************************************************************

1
Maturity
1      1      2
0.400  1.000  0.000

2
TKW
1      1      2
0.300  1.000  0.000

3
Yield
1      1      2
0.200  1.000  0.000
```

Fig. 3. Environment and trait definitions in QU-GENE

### 2.3.4 Gene information

Gene information is the most fundamental and complicated part in defining a G×E system. It is used to generate progeny genotypes from any crossing or propagation type, and to define the genotypic value of any genotype for each trait. It consists of the location of genes on wheat chromosomes, the number of alleles for each gene locus, the number of traits

affected by each gene, the genotypic effects in each defined environment type, etc. Linkage, multiple alleles, pleiotropy, epistasis, and G×E interaction are all defined in this part.

Definition of three genes located on chromosome 5, and one on chromosome 6 was shown in Fig. 4. The following parameters define each gene (including markers) (Fig. 4). Row 1 is the locus ID number to distinguish each gene (arranged in order, and starting from 1). Please note all genes should be arranged in order starting from the first chromosome or linkage group. Genes in one chromosome or linkage group should also be arranged as they appear on the chromosome. Row 2 gives the name of the gene (if defined). If the indicator for gene name is 1, a valid name must be specified, or the place is left blank. Row 3 specifies chromosome, recombination frequency, number of alleles, and number of traits the gene affects. All the genes, including markers, in the GE system are supposed to be arranged in order on the chromosomes. Recombination frequency of a gene is the crossover rate between the gene and the gene located just before it (two flanking genes). If a gene is located at the beginning of a chromosome, its recombination frequency should be set at 0.5. Row 4 specifies name of each allele (if defined). If the indicator for allele names is 1, a valid name must be specified for each allele, or the place left blank. The number of rows used to define genetic effects of the gene depends on number of traits affected, and number of environments (Fig. 4). For each affecting trait and each environment, Column 5 specifies the trait ID the gene affects. Column 6 specifies the environment ID. Column 7 specifies the three genotype to phenotype (or gene effect) types, i.e., additive (including dominance), epistasis, and QU-GENE plug-in.

```
!  ************************************************************************
!               Columns
!  CH     RF    NA NT      WT  ET  GP  EF   Gene effects
!  1      2     3  4       5   6   7   8    9+
!  ************************************************************************
        Locus ID number
   13
   Mat5          Locus name        Chromosome ID number, recombination with previous locus,
   5   0.5000  2   2                number of alleles at the locus, and number of traits affecting
                         1   1   1   -1    0.0 3.000 0.000    Genetic effects for all
                         3   1   1   -1    0.0 0.100 0.000    affected traits in all
   14                                                         defined environments
   TKW5
   5   0.0500  2   2
                         2   1   1   -1    0.0 2.000 0.000
                         3   1   1   -1    0.0 0.100 0.000
   15
   Yld5
   5   0.2300  2   1
                         3   1   1   -1    0.0 0.100 0.000
   16
   Yld6
   6   0.5000  2   1
                         3   1   1   -1    0.0 0.100 0.000
```

Fig. 4. Gene definition in QU-GENE.

Column 8 specifies how gene effects are stored. For additive genes, value –1 means that midpoint (m), additive (a) and dominance (d) will be specified later. This option is only available for genes with two alleles. For a gene with multiple alleles, value 0 should be used.

Value 0 means that genotypic values in the order of AA, Aa, and aa, where A-a are the two alternative alleles on the gene locus. In case of three alleles, e.g. $A_1$, $A_2$ and $A_3$ at locus A, the genotypic values are arranged in the order of $A_1A_1$, $A_1A_2$, $A_1A_3$, $A_2A_2$, $A_2A_3$, and $A_3A_3$. The order is similar for more than three alleles at a gene locus. Value 1 means that random gene effects with no dominance. In the case of two alleles A and a, genotypes AA and aa have random values AA and aa ranged from 0.0 to 1.0, but the value (Aa) of genotype Aa is at the mid–point between AA and aa. Value 2 means that random gene effects with no over-dominance. In the case of two alleles A and a, genotypes AA, Aa and aa have random values ranged from 0.0 to 1.0, but Aa is between AA and aa. Value 3 means that random gene effects with partial/over-dominance. In the case of two alleles A and a, genotypes AA, Aa and aa have independent random values ranged from 0.0 to 1.0, which will result in either partial dominance or over-dominance, depending on chance.

For epistatic genes, a number is given for the epistatic network the gene is included. Genotypic values of all possible combination in an epistatic network will be defined at a later stage, once all genes in the network have been determined. For QU-GENE plugin genes, a number is given for the plugin the gene is included. If a gene is only a marker, the trait number has to be set at 0. Trait number 0 is reserved to identify which gene locus is a marker.

### 2.3.5 Definition of starting populations

In QU-GENE, a population can be defined by gene frequency, or by genotypes. Four populations are defined in Fig. 5, and the first population "Poperror" will be used as reference to translate heritability to error variance. The other three, i.e. Pop02, Pop05, and

```
4     ! Number of populations to create
1     ! Which population to use for error estimates

1
Poperror
100
1
   0   1   1   2   1   0    0.5000

2
Pop02
20
1
   0   1   1   2   1   0    0.2000

3
Pop05
20
1
   0   1   1   2   1   0    0.5000

4
Pop08
20
1
   0   1   1   2   1   0    0.8000
```

Fig. 5. Four populations defined in QU-GENE.

Pop08, have a size of 20, and allele frequencies 0.2, 0.5, and 0.8, respectively. Each population takes 5 rows. "0" at the beginning of row 5 represents frequencies of alleles at all loci are identical. Otherwise, each locus will take a row. Pop02, Pop05, and Pop08 will be used as the starting population in breeding simulation.

## 2.4 Definition of breeding strategies in QuLine

By defining breeding strategy, QuLine translates the complicated breeding process in a way that the computer can understand and simulate. QuLine allows for several breeding strategies, which were contained in one input file, to be defined simultaneously. 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 comparison. A breeding strategy in QuLine is defined as all the crossing, seed propagation, and selection activities in an entire breeding cycle. For illustration, two breeding strategies, denoted by I-M, and II-M, are described in Fig. 6. Strategy I-M is similar to modified pedigree and bulk, where pedigree is used two times in $F_2$ and $F_5$ generations. Strategy II-M is similar to selected bulk, where pedigree is used only once in $F_5$ generation (Wang et al. 2003).

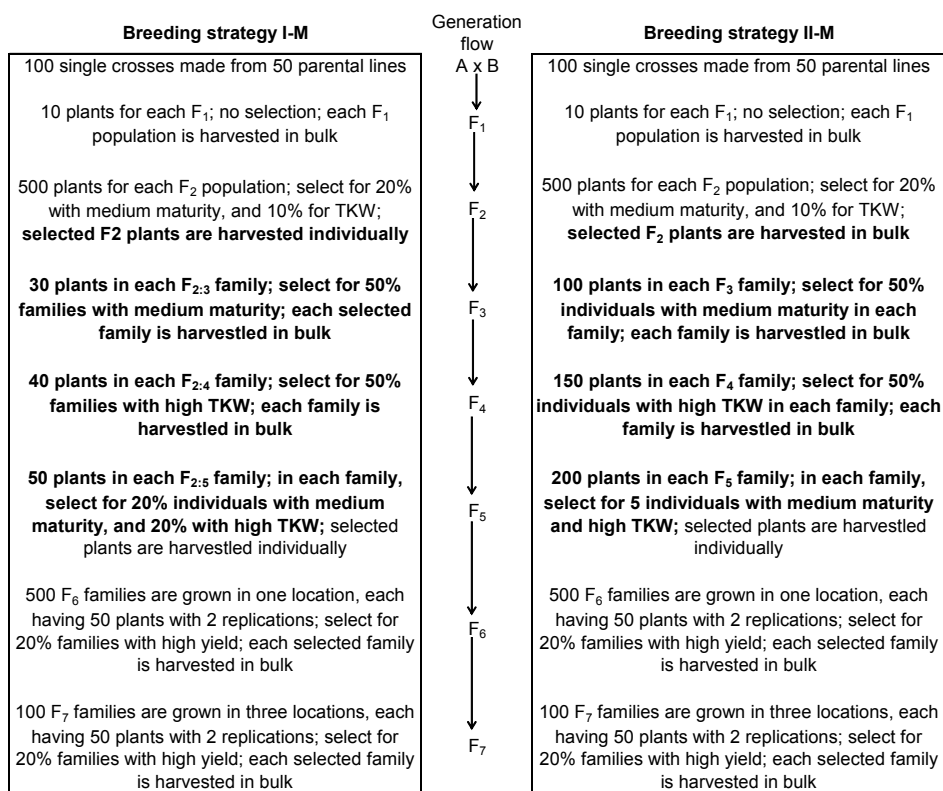| Breeding strategy I-M | Generation flow | Breeding strategy II-M |
|---|---|---|
| 100 single crosses made from 50 parental lines | A x B | 100 single crosses made from 50 parental lines |
| 10 plants for each $F_1$; no selection; each $F_1$ population is harvested in bulk | $F_1$ | 10 plants for each $F_1$; no selection; each $F_1$ population is harvested in bulk |
| 500 plants for each $F_2$ population; select for 20% with medium maturity, and 10% for TKW; **selected F2 plants are harvested individually** | $F_2$ | 500 plants for each $F_2$ population; select for 20% with medium maturity, and 10% for TKW; **selected $F_2$ plants are harvested in bulk** |
| **30 plants in each $F_{2:3}$ family; select for 50% families with medium maturity; each selected family is harvestled in bulk** | $F_3$ | **100 plants in each $F_3$ family; select for 50% individuals with medium maturity in each family; each family is harvestled in bulk** |
| **40 plants in each $F_{2:4}$ family; select for 50% families with high TKW; each family is harvestled in bulk** | $F_4$ | **150 plants in each $F_4$ family; select for 50% individuals with high TKW in each family; each family is harvestled in bulk** |
| **50 plants in each $F_{2:5}$ family; in each family, select for 20% individuals with medium maturity, and 20% with high TKW;** selected plants are harvestled individually | $F_5$ | **200 plants in each $F_5$ family; in each family, select for 5 individuals with medium maturity and high TKW;** selected plants are harvestled individually |
| 500 $F_6$ families are grown in one location, each having 50 plants with 2 replications; select for 20% families with high yield; each selected family is harvested in bulk | $F_6$ | 500 $F_6$ families are grown in one location, each having 50 plants with 2 replications; select for 20% families with high yield; each selected family is harvested in bulk |
| 100 $F_7$ families are grown in three locations, each having 50 plants with 2 replications; select for 20% families with high yield; each selected family is harvested in bulk | $F_7$ | 100 $F_7$ families are grown in three locations, each having 50 plants with 2 replications; select for 20% families with high yield; each selected family is harvested in bulk |

Fig. 6. Planting and selection details in two plant breeding strategies I-M and II-M. Major difference between the two strategies was highlighted in bold.

## 2.4.1 General simulation information

Generation information specifies the number of strategies to be simulated or compared, number of simulation runs, number of breeding cycles, number of crosses to be made at the beginning of each breeding cycle, indicator for crossing block update, and indicators for outputting simulation results. Indicator 0 for crossing block update means that only the final selected lines will be used as the parents for next breeding cycle. The parents in the current crossing block will not be considered for crossing in the following cycles. Indicator 1 means that for the next cycle, some parents come from the current crossing block, and some from the final selected lines. 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.

## 2.4.2 Number of generations and number of selection rounds in each generation

In the breeding program in Fig. 6, the best advanced lines developed from the $F_7$ generation will be returned to the crossing block to be used for new crosses. Therefore, the number of generations in one breeding cycle is seven for both strategies (Figs. 6, 7 and 8). The crossing block (viewed as $F_0$) and the seven generations need to be defined in QuLine. 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. Most generations in plant breeding programs have just one selection round, but some generations may have more than one selection round (Wang et al. 2003). More rounds of selection also allow the selection on traits measured by seeds instead of plants grown in the field. All generations in Strategies I-M and II-M have one round (Figs. 7 and 8).

## 2.4.3 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 nine options for seed propagation, presented here in the order of increasing genetic diversity ($F_1$ excluded): (i) clone (asexual reproduction), (ii) DH (doubled haploid), (iii) self (self-pollination), (iv) single cross (single cross between two parents), (v) backcross (back crossed to one of the two parents), (vi) topcross (crossed to a third parent, also known as three-way cross), (vii) doublecross (crossed between two F1s), (viii) random (random mating among the selected plants in a family), and (ix) no selfing (random mating but self-pollination is eliminated). The seed for $F_1$ 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 (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 in QuLine. Three seed propagation types are used in defining Strategies I-M, and II-M, which were clone, singlecross (only used for $F_1$ generation) and selfing (Figs. 7 and 8).

### 2.4.4 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, 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. The bulk generation advance method will not change the number of families in the following generation if among-family selection is not applied in the current generation, whereas the pedigree method increases the number of families rapidly if among-family selection intensity is weak, 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 advance 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. 7 that pedigree is used in $F_2$ and $F_5$, and bulk is used in the other generations in Strategy I-M. In comparison, pedigree is used only in $F_5$ in Strategy II-M.

### 2.4.5 Field experimental design for each selection round

The parameters used to define the virtual field experimental 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 (Figs. 7 and 8). 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.

### 2.4.6 Among-family selection and within-family selection for each selection round

Three traits have been defined before and now can be used in selection. There are two levels of selection in plant breeding, among-family and within-in family. The definition of these two types of selections is essentially the same: the number of traits to be selected is followed by the definition of each trait (Wang et al., 2004). Apart from the trait code there are two parameters that define a trait used in the selection, selection mode and selected amount. Selected amount can be a proportion of the number of families, individuals in selection, a threshold value or a specified number. The four options for defining selected proportions are (i) T (top), where the individuals or families with highest phenotypic values for the trait of interest will be selected; (ii) B (bottom), where the individuals or families with the lowest phenotypic values will be selected; (iii) M (middle), where individuals or families with medium trait phenotypic values will be selected; and (iv) R (random), where individuals or families will be randomly selected. The two options for defining threshold selection are (i) TV (top value), where the individuals or families whose phenotypic values are higher than the threshold will be selected; and (ii) BV (bottom value), where the individuals or families

whose phenotypic values are lower than the threshold will be selected. The three options for defining number selection are (i) TN (top number), where a specified number of the individuals or families with highest phenotypic values for the trait of interest will be selected; (ii) BN (bottom number), where a specified number of the individuals or families with lowest phenotypic values for the trait of interest will be selected; and (iii) RN (random number), where a specified number of the individuals or families will be selected randomly. Independent culling is used if multiple traits are considered for among-family or within-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. 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 with the selected amounts for traits.

```
!********************General information for the simulation experiment********************************
!NumStr NumRun NumCyc NumCro CBUpdate OutGES OutPOP OutHIS OutROG OutCOE OutVar Cross RMtimes PopSize
 2       5       10     100    0        0      0      0      0      0      0     random 0       0


!********************Information for selection strategies to be simulated*****************************
!StrategyNumber StrategyName  NumGenerations
 1              StrategyI-M    7

!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      clone       bulk    1   1   1   1                                  
                                           0
                                           0
 1  0  F1      singlecross bulk    1   10  1   1
                                           0
                                           0
 1  0  F2      self        pedigree 1  500 1   1
                                           0
                                           2   1  M  0.20    2  T  0.10
 1  0  F3      self        bulk    1   30  1   1
                                           1   1  M  0.50
                                           0
 1  0  F4      self        bulk    1   30  1   1
                                           1   2  T  0.50
                                           0
 1  0  F5      self        pedigree 1  50  1   1
                                           0
                                           2   1  M  0.20    2  T  0.20
 1  0  F6      self        bulk    2   50  1   1
                                           1   3  T  0.20
                                           0
 1  0  F7      self        bulk    2   50  3   1  1  1
                                           1   3  T  0.20
                                           0
```

Fig. 7. General simulation information and definition of strategy I-M in QuLine

Taking $F_2$ generation of Strategy I-M as an example, no among-family selection is conducted, but two traits are used for within-family selection, i.e. maturity (ID=1), and TKW (ID=2). Selection mode is M for maturity, and selected amount is 0.2, indicating 20% of the 500 F2 individuals (i.e. 100) with medium maturity will be first selected. Selection mode is T for TKW, and selected amount is 0.1, indicating 10% of the 100 retained $F_2$ individuals (i.e. 10) with highest TKW will be selected. The ten final selected $F_2$ individuals will be harvested individually, as "pedigree" is defined as the generation advance method (Fig. 7). For comparison, two other strategies were defined, where the selection mode is B for maturity, denoted by I-B and II-B. Other selection details are the same as those in I-M and II-B, respectively.

```
!*******************Information for selection strategies to be simulated*****************************
!StrategyNumber StrategyName  NumGenerations
 2              StrategyII-M      7

!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     clone       bulk      1  1  1  1
                                          0
                                          0
 1  0  F1     singlecross bulk      1  10 1  1
                                          0
                                          0
 1  0  F2     self        bulk      1  500 1  1
                                          0
                                          2  1  M  0.20    2  T  0.10
 1  0  F3     self        bulk      1  50 1  1
                                          0
                                          1  1  M  0.50
 1  0  F4     self        bulk      1  50 1  1
                                          0
                                          1  2  T  0.50
 1  0  F5     self        pedigree  1  200 1  1
                                          0
                                          2  1  M  0.20    2  T 0.125
 1  0  F6     self        bulk      2  50 1  1
                                          1  3  T  0.20
                                          0
 1  0  F7     self        bulk      2  50 3  1  1  1
                                          1  3  T  0.20
                                          0
```

Fig. 8. Definition of strategy II-M in QuLine

## 2.5 Simulation experimental design

A G×E system called "WheatModel" (Figs. 2, 3, and 4), three starting populations called Pop02, Pop05, and Pop08 (Fig. 5), and four breeding strategies called I-M, II-M, I-B, and II-B (Figs. 7 and 8) have been defined previously. A total of 12 simulation experiments are designed (Table 1). Each experiment was repeated for 1000 times, and mean across the 100 times will be used to compare the efficiency of different strategies.

| Experiment | G×E system | Population | Breeding strategy |
|---|---|---|---|
| 1 | WheatModel | Pop02 | Strategy I-M |
| 2 | WheatModel | Pop02 | Strategy II-M |
| 3 | WheatModel | Pop05 | Strategy I-M |
| 4 | WheatModel | Pop05 | Strategy II-M |
| 5 | WheatModel | Pop08 | Strategy I-M |
| 6 | WheatModel | Pop08 | Strategy II-M |
| 7 | WheatModel | Pop02 | Strategy I-B |
| 8 | WheatModel | Pop02 | Strategy II-B |
| 9 | WheatModel | Pop05 | Strategy I-B |
| 10 | WheatModel | Pop05 | Strategy II-B |
| 11 | WheatModel | Pop08 | Strategy I-B |
| 12 | WheatModel | Pop08 | Strategy II-B |

Table 1. Designing a simulation experiment

## 3. Explanation of simulation results

Various kinds of information can be output by setting appropriate outputting indicators (Fig. 7). These information includes genetic variance, correlation among traits for each environment, correlation among environment for each trait, number of crosses retained after each round of selection, mean genotypic values, percentage of fixed genes for all traits and the percentage of fixed genes for each trait, gene frequency, Hamming distance, selection history, number of families, number of individual plants in each generation for each simulated strategy, etc. Not all outputs are required in any simulations.

### 3.1 Genetic gains from different breeding strategies

Table 2 clearly indicated that the genetic gain on yield from Strategy II was either equal to or higher than the genetic gain from Strategy I. For the starting population Pop02, yield is 4.20 t/ha in the parental population (Table 2). When families and individuals with medium maturity are selected in breeding (i.e. I-M and II-M), Strategy I increased yield to 8.35 t/ha after 10 cycles, and Strategy II to 8.44 t/ha. This is 1.08% higher than the yield from Strategy I. When short maturity is selected (i.e. I-B and II-B), Strategy I increased yield to 7.77 t/ha after 10 cycles, and Strategy II to 7.80 t/ha that is 0.34% higher than the yield from Strategy I. The difference between medium and short maturity selections is caused by the pleiotropic effects of maturity genes on yield (Figs. 1 and 4).

For the starting population Pop05, yield is 6.00 t/ha in the parental population (Table 2). When families and individuals with medium maturity are selected in breeding, Strategy I increased yield to 8.95 t/ha after 10 cycles and Strategy II increase it to 8.96 t/ha. When short maturity is selected, Strategy I increased yield to 8.20 t/ha after 10 cycles and Strategy II increased it to 8.26 t/ha. Difference of genetic gains from the two strategies is minor. For the starting population Pop08, yield is 7.80 t/ha in the parental population (Table 2). When families and individuals with medium maturity are selected in breeding, both strategies increased yield to 9.00 t/ha after 10 cycles. When short maturity is selected, Strategy I increased yield to 8.72 t/ha after 10 cycles and Strategy II increases it to 8.86 t/ha. That is 1.54% higher than the yield from Strategy I.

| Cycle | Pop02 | | | | Pop05 | | | | Pop08 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I-M | II-M | I-B | II-B | I-M | II-M | I-B | II-B | I-M | II-M | I-B | II-B |
| 0 | 4.20 | 4.20 | 4.20 | 4.20 | 6.00 | 6.00 | 6.00 | 6.00 | 7.80 | 7.80 | 7.80 | 7.80 |
| 1 | 5.04 | 5.05 | 4.95 | 4.92 | 7.06 | 7.08 | 6.90 | 6.92 | 8.50 | 8.49 | 8.40 | 8.40 |
| 2 | 5.80 | 5.84 | 5.62 | 5.60 | 7.67 | 7.70 | 7.39 | 7.42 | 8.77 | 8.76 | 8.58 | 8.62 |
| 3 | 6.45 | 6.52 | 6.21 | 6.21 | 8.11 | 8.14 | 7.73 | 7.76 | 8.90 | 8.90 | 8.66 | 8.72 |
| 4 | 6.97 | 7.04 | 6.67 | 6.68 | 8.42 | 8.45 | 7.94 | 7.97 | 8.96 | 8.96 | 8.70 | 8.78 |
| 5 | 7.38 | 7.45 | 7.03 | 7.04 | 8.64 | 8.66 | 8.07 | 8.10 | 8.99 | 8.98 | 8.71 | 8.81 |
| 6 | 7.71 | 7.78 | 7.30 | 7.31 | 8.78 | 8.79 | 8.14 | 8.17 | 9.00 | 8.99 | 8.72 | 8.83 |
| 7 | 7.96 | 8.04 | 7.50 | 7.52 | 8.87 | 8.88 | 8.18 | 8.21 | 9.00 | 9.00 | 8.72 | 8.84 |
| 8 | 8.14 | 8.23 | 7.64 | 7.66 | 8.92 | 8.93 | 8.19 | 8.24 | 9.00 | 9.00 | 8.72 | 8.85 |
| 9 | 8.27 | 8.35 | 7.72 | 7.75 | 8.94 | 8.95 | 8.20 | 8.25 | 9.00 | 9.00 | 8.72 | 8.86 |
| 10 | 8.35 | 8.44 | 7.77 | 7.80 | 8.95 | 8.96 | 8.20 | 8.26 | 9.00 | 9.00 | 8.72 | 8.86 |

Table 2. Genetic gains on yield from four breeding strategies and three starting populations

In simulation, genotypic value of an individual plant (denoted as F for fitness) in each environment type is calculated from the genetic effects defined in the G×E system. The adjusted genotypic value is define as

$$F_{ad} = \frac{F - TG_l}{TG_h - TG_l} \times 100 \ ,$$

where $TG_l$ and $TG_h$ are the genotypic values for the two extreme target genotypes with the lowest and the highest trait values in the G×E system, respectively. This standardization is useful specifically when diverse G×E systems are used to compare the performances of different breeding strategies. The adjusted genetic gains on the three traits were shown in Fig. 9 form medium maturity selection and in Fig. 10 for bottom maturity selection.
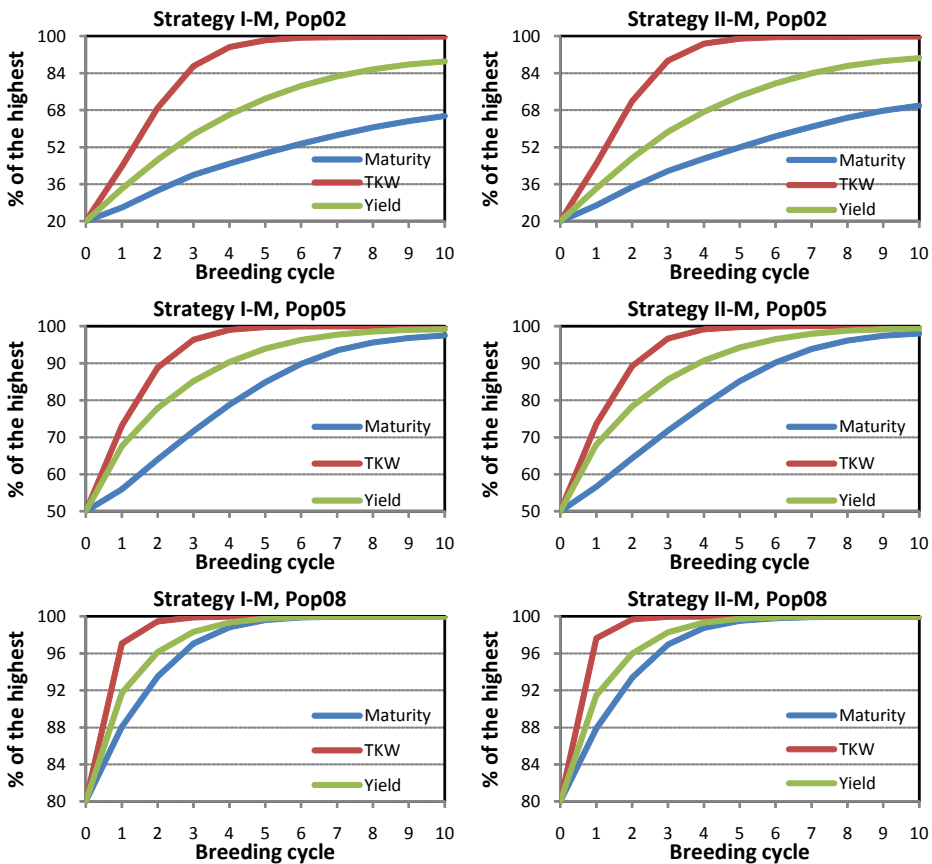


Fig. 9. Adjusted genetic gains from breeding strategies I-M and II-M

When medium maturity is selected (Fig. 9), TKW reaches to the highest value after 5 breeding cycles for Pop02, after 4 cycles for Pop05, and after 2 cycles for Pop08, for both strategies. TKW genes have pleiotropic effects on yield in definition (Figs. 1 and 4), and

TKW and yield were both selected for top performance (Figs. 6, 7 and 8). The selection on TKW and yield both helps increase the frequency of favourable TKW alleles. If there is no correlation between maturity and yield, maturity should keep unchanged. The increase in maturity is due to the selection for high yield. From the genetic model defined in Figs. 1 and 4, the longer the maturity, the higher the yield. Therefore, the selection for high yield retained the alleles of long maturity.

In practice, the breeders may want to select for short maturity cultivars. When short maturity is selected (Fig. 10), there is no much difference for TKW. For Pop02 and Pop05, both strategies reduced maturity. For Pop08, strategy I reduced maturity, but strategy II increased maturity slowly, indicating strategy II may result in less selection intensity on maturity.
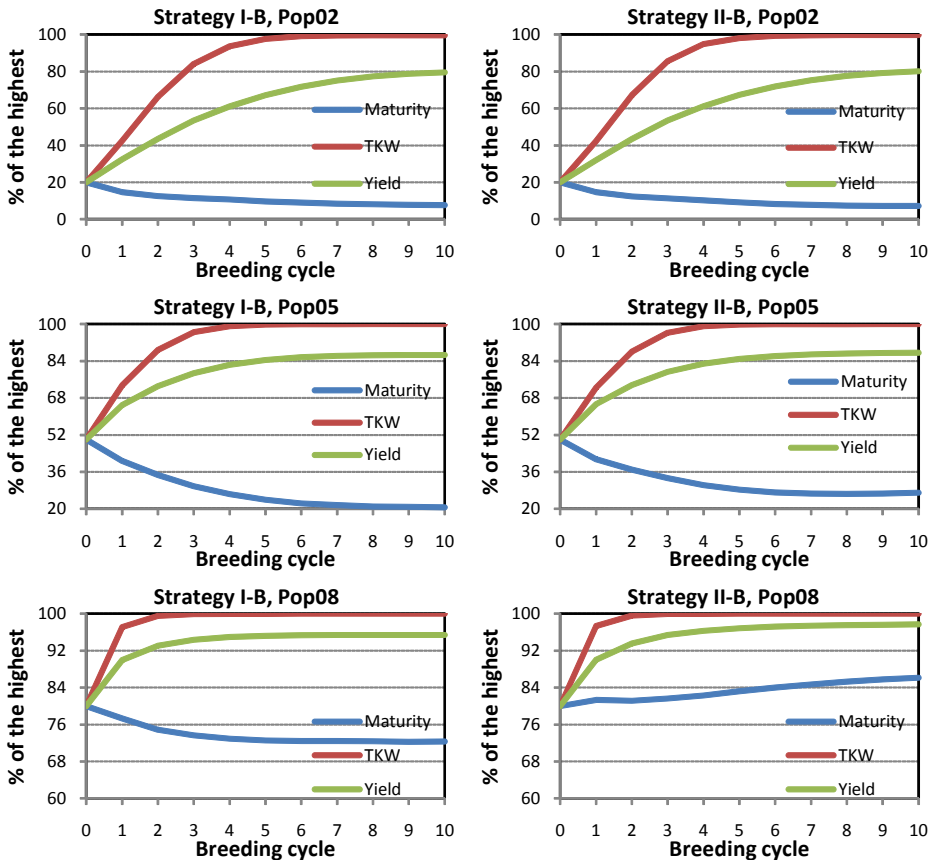


Fig. 10. Adjusted genetic gains from breeding strategies I-B and II-B

## 3.2 Cost and benefit analysis

Previous results showed that the genetic gain on yield from Strategy II was either equal to or higher than the genetic gain from Strategy I (Table 2, Figs. 9 and 10). How much cost will be

needed to run each strategy? For this purpose, we compared the number of families and individual plants to be grown in the two strategies (Table 3). Less families means less seed lots to be prepared by labor for planting, and less individuals means less land to be used. In one breeding cycles, the number of families generated from strategy II is 43.14% of the number of families generated from strategy I. The number of plants to be grown in strategy II is 85.41% of the grown plant number in strategy I. Therefore when strategy II is used, fewer seed lots need to be handled at both harvest and sowing and less land is used, resulting in a significant saving in time, labor and cost.

| Gene-ration | Families before selection | | Plants before selection | | Families after selection | | Plants after selection | |
|---|---|---|---|---|---|---|---|---|
| | Strategy I | Strategy II | Strategy I | Strategy II | Strategy I | Strategy II | Strategy I | Strategy II |
| F1 | 100 | 100 | 1000 | 1000 | 100 | 100 | 1000 | 1000 |
| F2 | 100 | 100 | 50000 | 50000 | 1000 | 100 | 1000 | 1000 |
| F3 | 1000 | 100 | 30000 | 5000 | 500 | 100 | 15000 | 2500 |
| F4 | 500 | 100 | 15000 | 5000 | 250 | 100 | 7500 | 2500 |
| F5 | 250 | 100 | 12500 | 20000 | 500 | 500 | 500 | 500 |
| F6 | 500 | 500 | 50000 | 50000 | 100 | 100 | 10000 | 10000 |
| F7 | 100 | 100 | 30000 | 30000 | 20 | 20 | 6000 | 6000 |
| Total | 2550 | 1100 | 188500 | 161000 | | | | |

Table 3. Families and individual plants to be grown in strategies I and II

The simulation results (Tables 2 and 3; Figs. 9 and 10) clearly indicated that strategy II resulted in similar genetic gain on yield, but was more cost-effective compared with strategy I. Strategy I is called MODPED and II is called SELBLK in CIMMYT's wheat breeding. By applying bulk, we may not know which $F_2$, $F_3$ or $F_4$ individual derives which final fixed line, but parental lines deriving each fixed line are still known, which provides the most important information for the next cycle of breeding.

## 4. Conclusion

Conventional plant breeding largely depends on phenotypic selection and breeder's experience; therefore, the breeding efficiency is low and the predictions are inaccurate. Along with the fast development in molecular biology and biotechnology, a large amount of biological data are available for genetic studies of important breeding traits in plants, which in turn allows the conduction of genotypic selection in the breeding process. However, gene information has not been effectively used in crop improvement because of 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. Thus, the best performing crosses and effective breeding strategies can be identified. On the basis of the results from simulation experiments, breeders can optimize their breeding methodology and greatly improve the breeding efficiency.

On the other hand, a great amount of studies on QTL mapping have been conducted for various traits in plants and animals in recent years (Dekkers and Hospital, 2002; Peleman and Voort, 2003; Wang et al., 2005, 2007b, 2009). 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 for the improvement of crop performance. Breeding simulation allows the definition of complicated genetic models consisting of multiple alleles, pleiotropy, epistasis, and genes by environment interaction that provides a useful tool to breeders, who can efficiently use the wide spectrum of genetic data and information available. This approach will be very helpful when the breeders want to compare breeding efficiencies from different selection strategies, to predict the cross performance with known gene information, and to investigate the efficient use of identified QTLs in conventional breeding.

## 5. Acknowledgments

## 6. References

Allard, R.W. (1999). *Principles of plant breeding*, 2nd Edition, John Wiley & Sons, ISBN 0-471-02309-4, New York, USA

Bernardo, R. & Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: A simulation appraisal. *Crop Science*, Vol.46, No.2, (March 2006), pp. 614–621

Bonnett, D.G.; Rebetzke, G.J. & Spielmeyer, W. (2005). Strategies for efficient implementation of molecular markers in wheat breeding. *Molecular Breeding*, Vol.15, No.1, (January 2005), pp. 75-85

Dekkers, J.C.M. & Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Review Genetics*, Vol.3, No.1, (January 2002), pp. 22-32

Dubcovsky, J. (2004). Marker-assisted selection in public breeding programs: the wheat experience. *Crop Science*, Vol.44, No.6, (December 2004), pp. 1895-1898

Falconer, D.S. & Mackay, T.F.C. (1996). Introduction to quantitative genetics, 4th edition, Longman, ISBN 0582-24302-5, Essenx, England

Hallauer, A.R.; Carena, M.J. & Miranda Filho, J.B. (1988). *Quantitative Genetics in Maize Breeding*, 2nd Edition, Springer, ISBN 978-1-4419-0765-3, New York, USA

Kempthorne, O. (1988). An overview of the field of quantitative genetics, In: *Proceedings of the 2nd International Conference on Quantitative Genetics*, B.S. Weir, E.J. Eisen, M.M. Goodman, G. Namkoong (eds), 47-56, Sinauer Associates, Sunderland, Massachusetts, USA

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative genetics*, Sinauer Associates, ISBN 0-87893-481-2, Sunderland, Massachusetts, USA

Peleman, J.D. & Voort, J.R. (2003). Breeding by design. *Trends in Plant Science*, Vol.8, No.7, (July 2003), pp. 330-334

Podlich, D.W. & Cooper, M. (1998). QU-GENE: a platform for quantitative analysis of genetic models. *Bioinformatics*, Vol.14, No.7, (August 1998), pp. 632-653

Rajaram, S.; van Ginkel, M. & Fischer, R.A. (1994). CIMMYT's wheat breeding mega-environments (ME), *Proceedings of the 8th International Wheat Genetics Symposium*, pp. 1101-1106, China Agricultural Scientech, Beijing, China

Rajaram, S. (1999). Historical aspects and future challenges of an international wheat program. In: *Septoria and Stagonospora Diseases of Cereals: A Compilation of Global Research*, M. van Ginkel, A. McNab, and J. Krupinsky (eds.), 1-17, CIMMYT, D.F., Mexico

Singh, R. P.; Rajaram, S.; Miranda, A.; Huerta-Espino J. & 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*, Vol.100, No. 1, (May 1998), pp35-43

Singh, R.P. & Huerta-Espino, J. (2004). The use of "single-backcross selected-bulk" breeding approach for transferring minor genes based rust resistance into adapted cultivars, *Proceedings of 54th Australian Cereal Chemistry Conference and 11th Wheat Breeders Assembly*, Canberra, Australia, September21-24, 2004

van Ginkel, M.; Trethowan, R.; Ammar, K.; Wang, J. & Lillemo, M. (2002). Guide to bread wheat breeding at CIMMYT (Revised edition). *Wheat Program Special Report*, No.5, (Februrary 2002), CIMMYT, ISSN 0187-7787, D.F., Mexico

Wan, X.; Wan, J.; Su, C.; Wang, C.; Shen, W.; Li, J.; Wang, H.; Jiang, L.; Liu, S.; Chen, L.; Yasui, H. & Yoshimura, A. (2004). QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. *Theoretical and Applied Genetics*, Vol.110, No.1, (December 2004), pp. 71-79

Wan, X.; Wan, J.; Weng, J.; Jiang, L.; Bi, J.; Wang, C. & Zhai, H. (2005). Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theoretical and Applied Genetics*, Vol.110, No.7, (May 2005), pp.1334-1346

Wang, J.; van Ginkel, M.; Podlich, D.; Ye, G.; Trethowan. R.; Pfeiffer, W.; DeLacy, I.H.; Cooper, M. & Rajaram, S. (2003). Comparison of two breeding strategies by computer simulation. *Crop Science* Vol.43, No.5, (September 2003), pp. 1764-1773

Wang, J.; van Ginkel, M.; Trethowan, R.; Ye, G.; DeLacy, I.H.; Podlich, D. & Cooper, M. (2004). Simulating the effects of dominance and epistasis on selecting response in the CIMMYT wheat breeding program using QuLine. *Crop Science* Vol.44, No.6, (November 2004), pp. 2006-2018

Wang, J.; Eagles, H.A.; Trethowan, R. & van Ginkel, M. (2005). Using computer simulation of the selection process and known gene information to assist in parental selection in wheat quality breeding. *Australian Journal of Agricultural Research*, Vol.56, No.5, (May 2005), pp. 465-473

Wang, J. & Pfeiffer, W. H. (2007). Simulation approach and its applications in plant breeding. *Agricultural Sciences in China*, Vol.6, No.8, (August 2007), pp. 908-921

Wang, J.; Chapman, S.C.; Bonnett, D.B.; Rebetzke, G.J. & Crouch, J. (2007a). Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. *Crop Science*, Vol. 47, No.2, (March 2007), pp. 580-588

Wang, J.; Wan, X.; Li, H.; Pfeiffer, W.; Crouch, J. & Wan, J. (2007b). Application of identified QTL-marker associations in rice quality improvement through a design breeding approach. *Theoretical and Applied Genetics*, Vol.115, No.1, (June 2007), pp. 87-100

Wang, J.; Chapman, S. C.; Bonnett, D. G. & Rebetzke, G. J. (2009a). Simultaneous selection of major and minor genes: use of QTL to increase selection efficiency of coleoptile length of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, Vol.119, No.1, (June 2009), pp. 65-74

Wang, J.; Singh, R.P.; Braun, H.-J. & Pfeiffer, W.H. (2009b). Investigating the efficiency of the single backcrossing breeding strategy through computer simulation. *Theoretical and Applied Genetics*, Vol.118, No.4, (february 2009), pp. 683-694