Journal of Animal and Veterinary Advances 10 (18): 2421-2433, 2011
ISSN: 1680-5593

Selection for Gene Pyramiding Design in Admixed Population

Lingyang Xu, Hangxing Ren, Xihui Sheng, Li Zhang, Caihong Wei and Lixin Du
National Center for Molecular Genetics and Breeding of Animal,
Institute of Animal Sciences, Chinese Academy of Agricultural Sciences,
No. 2 Yuanmingyuan West Road, Haidian District, 100193 Beijing, P.R. China

Abstract: Gene pyramiding aims to design superior trait through selecting and combining favorite target alleles into a single genotype, thus it was advocated for designing breeding programs via selecting and pyramiding optimal combinations of alleles. In this study, we investigated selection for gene pyramiding design given the animal segregating population and the target trait was controlled by major genes. The admixed population was used as the base population. The mating parents were selected via detecting the favorite genes or linked markers and offspring were produced by the discrete recombination of parents. The phenotypic values were predicted by the genotype-phenotype model. Three selection strategies-genotypic selection, simple phenotypic selection and phenotypic selection integrating molecular information were developed. In genotypic selection, we only considered the favorite allele frequencies and base population sizes in admixed population and in phenotypic selection, we used genotype-phenotype model including trait heritability, gene effect and gene interaction effects to predicate phenotypic values. In each generation, we calculated population hamming distance, average superior genotype frequency and average phenotypic value to comprehensively measure the progress of gene pyramiding. The strategy requires minimum generations to gain gene pyramiding were defined as the optimization strategy. Examples were given for four target genes in order to compare the progress of gene pyramiding. The results indicate that gene pyramiding breeding process was greatly affected by the selection strategy. The gene effect and gene interaction effects information affect the selection of optimal genotype combinations and more precise molecular information was needed to guide the design of effective gene pyramiding breeding programs.

Key words: Gene pyramiding, evolutionary computation, selection strategies, population hamming distance, breeding strategy, China

INTRODUCTION

Recently, rapid technological development in molecular genetics and functional genomics has resulted in new opportunities for the exploitation of quantitative trait loci and linked markers from generally superior populations (Hospital *et al*., 2000; Visscher *et al*., 1996; Wall *et al*., 2005). The opportunities have promoted the development of marker-assisted selection in plant and animals breeding. Some theoretical studies on marker-assisted selection have been conducted (Hu, 2007; Lande and Thompson, 1990; Lange and Whittaker, 2001). In practice, single-gene introgression schemes have been developed for plants such as barley (Jefferies *et al*., 2003) and experimental animals such as mice (Koudande *et al*., 2005) and marker-assisted selection has enormous potential to improve the efficiency and precision of conventional breeding. Gene pyramiding is an important application of marker assisted selection. Gene pyramiding should be divided into two steps-pyramiding step and fixation step (Ishii *et al*., 2008; Ishii and Yonezawa, 2007; Servin *et al*., 2004). Experimental and field data have shown that gene pyramiding is a very effective strategy in crop breeding. The most widespread application for pyramiding has been for combining multiple disease resistance genes (Castro *et al*., 2003; Huang *et al*., 1997; Saghai Marrof *et al*., 2008; Singh *et al*., 2001). Gene pyramiding has also been used in grain production (Ashikari *et al*., 2005). There also exist study demonstrates that gene pyramiding is feasible in animal breeding using Drosophila melanogaster (Jiang *et al*., 2008).

Servin *et al*. (2004) investigated the theoretical issues of gene pyramiding and proposed certain general principles for designing gene pyramiding schemes and theoretical research on this topic has extended to the design and comparison of gene pyramiding schemes in animals (Zhao *et al*., 2009).

Corresponding Author: Lixin Du, National Center for Molecular Genetics and Breeding of Animal, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, No. 2 Yuanmingyuan West Road, Haidian District, 100193 Beijing, P.R. China

Servin *et al*. (2004) had not considered the influence of initial favorite allele frequency and selection effect on gene pyramiding. The studies mainly focused on the gene pyramiding at fixation step. Suppose the gene pyramiding step have been accomplished and multiple target genes have been dispersed in the admixed population with various levels of favorite allele frequencies.

In this study, we firstly regarded gene pyramiding breeding as a metaphor for climbing mountains, inspired by the science of evolutionary computation(Goldberg, 1989; Holland, 1992), individual carrying various genotype combination responds to the various genotypic scores and phenotypic values then these genotypic scores and phenotypic values were used for individual selection. Generation selection and genetic operator promoted more and more individuals carry the optimal genotype combination. Second, we investigated the gene pyramiding progress aiming at pyramiding different number of target genes using genotypic selection over nine generations. Third, examples are given for four target genes and three selection strategies-genotypic selection, simple phenotypic selection and phenotypic selection integrating molecular information were designed and compared. We developed the method and model integrating molecular information from multi-level and gene pyramiding breeding simulation platform incorporating flexible selection strategies would be used to guide gene pyramiding breeding.

MATERIALS AND METHODS

Individual genotype and population genotype simulation: The studies use two characters 0 or 1 to code one genotype at one locus, assumed the target trait was mainly controlled by several major genes. The initial base population was coded by an N×2M matrix (N represents the number of individuals in the population, M represents the number of loci, each loci has two alleles), population with special preset initial favorite allele frequencies were initialized as the base population, the favorite allele denotes the allele 1 in the simulation. We defined an ideal population is the population in which all individuals carrying favorite allele at target loci. Take the case of four genes, the ideal genotype combination of individual is 11-11-11-11.

In the simulations, discrete recombination was used to combine (mates) two individuals (parents) to produce new offspring which inspired by evolution computer (Goldberg, 1989; Holland, 1992), new individuals were produced by the crossover of two selected parents. Discrete recombination uses crossover mask to indicate which parent will supply bits (allele) to the offspring, a crossover mask is as the same length as the individual structure which was randomly generated by 0 or 1 with equal probability. Crossover mask 1 indicates the allele of offspring at this locus inherited from parent 1, crossover mask 0 indicates the allele of offspring at this locus inherited from parent 2. Discrete recombination at each locus produced offspring with new genotype combination. Offspring1 was produced by mast 1 and offspring 2 was produced by mast 2, the allele inherited from parent 1 was marked with underline as fellow:

$$\begin{array}{ll} \text{Parent 1} & \underline{0}\,1\,1\,\underline{1}\,0\,\underline{0}\,\underline{1}\,1 \\ \text{Parent 2} & 1\,0\,\underline{1}\,0\,1\,1\,0\,0 \\ \text{Mast 1} & 0\,1\,1\,0\,0\,0\,1\,1 \\ \downarrow & \\ \text{Offspring 1} & 1\,\underline{1}\,\underline{1}\,0\,1\,1\,\underline{1}\,\underline{1} \\ \text{Mast 2} & 1\,0\,0\,1\,1\,1\,0\,0 \\ \downarrow & \\ \text{Offspring 2} & \underline{0}\,0\,\underline{1}\,\underline{1}\,\underline{0}\,\underline{0}\,0\,0 \end{array} \quad (1)$$

Superior genotype 11 frequency and population hamming distance: The superior genotype 11 frequency and population hamming distance were calculated in each generation to measure the process of pyramiding breeding program. The initial favorite allele frequencies for target genes were set at different levels to represent the difference of base population. In information theory, the Hamming distance, named after Richard Hamming is the number of positions in two strings of equal length for which the corresponding elements are different. Hamming distance have been used to measure the number of nucleotide differences between two genetic sequences (Pilcher *et al*., 2008). In this research, we borrow this idea to measure the distance between two populations which is called the Population Hamming Distance (PHD). PHD is the total number of different alleles at target loci in the population at each generation compared to ideal population. For the following example, pop(t) and pop(ideal) both populations with four target loci (two alleles at each locus) and population size is 6. Matrix column represents target loci, row represents individuals of population. Population hamming distance between pop(t) and pop(ideal) is 19.

$$\text{Pop(t)} = \begin{pmatrix} 10 & 10 & 11 & 00 \\ 01 & 11 & 00 & 01 \\ 00 & 01 & 11 & 10 \\ 01 & 11 & 10 & 00 \\ 11 & 10 & 10 & 01 \end{pmatrix}$$

$$\text{Pop(ideal)} = \begin{pmatrix} 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 \end{pmatrix} \quad (2)$$

Genotypic selection and phenotypic selection: Using genotypic selection, genotype 11 was given a genotypic value 2, genotypic value of 10 or 01 is 1 and genotypic value of 00 is 0. The sum of the genotypic value at all loci was taken as the genotypic selection score for an individual.

In the simple phenotypic selection strategy, the phenotypic observation of each individual is modeled as Eq. 3:

$$p_i = \mu_0 + \sum_{j=1}^{m} g_j x_{ij} + \varepsilon_i \quad (3)$$

Where:

p_i = The phenotypic observation of individual i
μ_0 = The population mean
g_j = The gene effect at jth locus (j = 1, 2, … , m)
x_{ij} = An indicator variable of genotype k (11, 10, 00)
ε_i = The residual error following the distribution N (0, σ_ε^2).

The genotypic values were defined in terms of the midpoint (m), additive (a) and dominance (d) genetic parameters. The numerical coding of three genotypes 11, 10, 00 were 5, 4, 1, respectively in the model 3. For an analysis of genotypes in a single environment, heritability on an individual basis will be estimated as Eq. 4. From the defined heritability an estimate of σ_ε^2 is obtained by calculating σ_ε^2 and re-arranging Eq. 4-5:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2} \quad (4)$$

$$\sigma_\varepsilon^2 = \frac{\sigma_g^2}{h^2} - \sigma_g^2 \quad (5)$$

In the phenotypic selection strategy integrating molecular information, we use two genotype-phenotype models to predict phenotype value, model 1 and 2 (Eq. 6):

$$p_i = \mu_0 + \sum_{j=1}^{m} g_j x_{ij} + gim_{C(m,x_{ij})} + \varepsilon_i \quad (6)$$

The variable p_i, μ_0, x_{ij}, ε_i were denoted as the same as the model 3 in simple phenotypic selection strategy, the difference among them depends on level of molecular information integration. In genotype-phenotype model 1, only gene effects were integrated, the values of genotypes could be set to any quantitative values according to specific genetic background. In genotype-phenotype model 2 both gene effects and gene interaction effects were integrated and the values of them could also be set specifically to represent the given molecular information. We supposed that target genes effects and interaction effects have been identified as available information with quantitative values and these are prerequisites for implementing of model 6.

In the model II, $gim_{(l\text{-}1,\ l\text{-}2,\ l\text{-}3,\ l\text{-}4)}$ denotes the genotype interaction effect, similar to polygenic effect but represents the actual information integration in model 2. For each specific genotype combination, the gim(l-1, l-2, l-3, l-4) corresponds to a scalar. Considered the genotype interaction effects were unknown, so in the simulation they were taken as random variables and $gim_{(l\text{-}1,\ l\text{-}2,\ l\text{-}3,\ l\text{-}4)} \sim N(0, 1)$ as in the results section rand 1 and 2 denote four genotype interaction effects using 4D matrices as data format in simulation, the values of rand 1 and 2 were sampled from $gim_{(l\text{-}1,\ l\text{-}2,\ l\text{-}3,\ l\text{-}4)}$ (Appendix).

Gene pyramiding at fixation steps: In this study, we investigated gene pyramiding at fixation steps, admixed population was taken as the base population, the target genes and linked markers were detected for selection in each generation. We performed simulation from admixed population with various levels of favorite allele frequencies and different selection strategies over several generations. Moreover, the progress of gene pyramiding were investigated and compared by the population hamming distance, superior genotype 11 frequency and average phenotypic values. In the simulations, the gene pyramiding process at fixation step is as follows (Fig. 1):

Step 1: Create an initial string population as the basic admixed population.

Step 2: Selection based on genotype score or phenotypic value according to different selection strategies.

Step 3: Select the top 500 as the parents.

Step 4: Random mating to produce the next generation.

Step 5: Return to step 2.

These studies firstly investigated the gene pyramiding progress aiming at pyramiding different number of target genes using genotypic selection, set the number of target genes at 4, 8, 16 or 20, the initial favorite allele frequency (denotes allele 1) at each locus was set 0.5, base population and parent population in each generation have a constant size 500. Random mating produces four offspring in a sex ratio of 1:1, each male

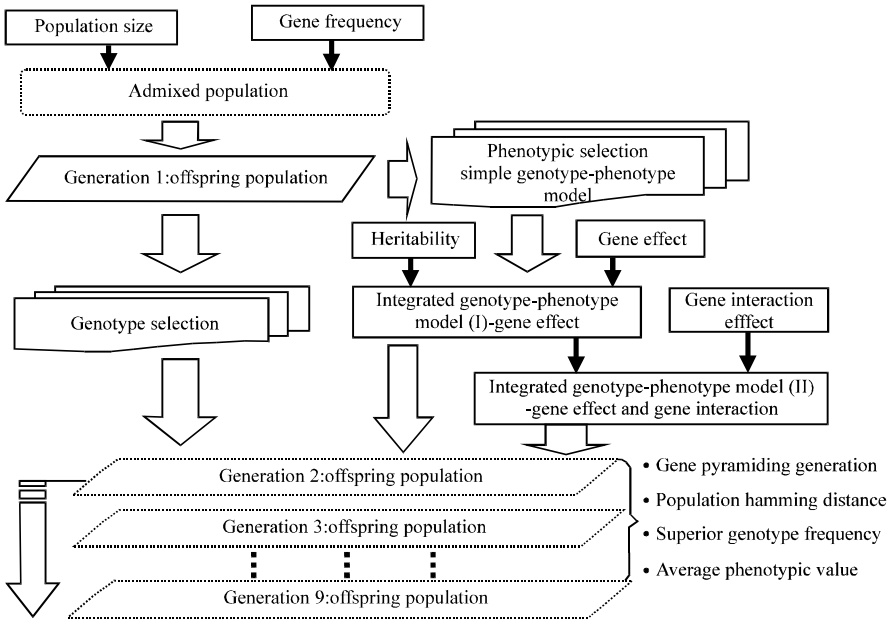


Fig. 1: Framework of gene pyramiding at fixation step

only mate with one female, the top 500 individuals were selected for the next generations. Second, examples were given for four target genes to compare the three types of selection strategies. In the genotypic selection strategy, we performed simulation with initial superior gene frequencies at 0.2, 0.4 or 0.6 for all four loci in base population. We assumed that the base population and parent population have a constant size 500. Random mating produces four offspring in a sex ratio of 1:1 each male only mate with one female. The top 500 individuals were selected for the next generations. In phenotypic selection strategies, we used two phenotypic selections and the breeding process was similar to genotypic selection. Once, the base population had been established, the phenotypic values were calculated using genotype-phenotype model 1 and 2; 500 individuals with the highest phenotypic values were selected for mating in subsequent generations. For each generation, the population hamming distance, average phenotypic value and average genotype 11 frequencies were calculated to measure the process of four target genes pyramiding. In this study, Monte Carlo method was used to simulate the gene pyramiding design in order to obtain an average assessment of the breeding process. Each simulation was repeated 1000 times. The computer programs were written in Matlab and run on an Inter (R) Core (TM) 2 Duo CPU, Microsoft Windows XP.

RESULTS

Genotypic selection for different number of target genes: Table 1 shows the changes of population hamming distance using genotypic selection over generations given 2, 4, 8 or 10 target loci in base population. The initial favorite allele frequency at each locus is set to 0.5 approximately. The genotypic selection promotes the target genes pyramided in subsequent generations through selecting the top 500 individuals as parents which were sorted by genotypic scores, population hamming distance decreases to zero when all the target genes pyramiding into an ideal genotype. As for 10 target genes, some individuals retained certain loci not pyramided until the ninth generation. However, as for 2, 4 and 8 target genes, gene pyramiding was successfully realized at the 5, 6 and 9th generation.

Simple genotype-phenotype model: When initial favorite allele frequency of each locus is set at the same level, results indicate that simulation series for trait heritability 0.2 need more generations than trait heritability 0.4 and 0.6. The population hamming distance decreases with generations (Fig. 2).

As to allele frequency 0.6 and trait heritability 0.6, four target genes pyramided at the seventh generation. Supposed trait heritability 1.0, the selection strategy is equivalent to genotypic selection. The least efficient gene pyramiding scheme corresponded to the one with initial gene frequency of 0.2 and trait heritability of 0.2. When the initial favorite allele frequency is set to 0.6 and heritability is set to 0.6, the increasing trend is more significant than others. Using simple genotype-phenotype model, the changes of the average phenotypic value over generations is contrary to that of population hamming distance (Fig. 3). Thus, optimization

Table 1: Changes of population hamming distance using genotypic selection over generations (1-9)*

Generations	G1	G2	G3	G4	G5	G6	G7	G8	G9
PHD-10	5000.57	4150.24	3334.57	2560.28	1826.49	1165.56	606.16	190.28	0.05
PHD-8	3998.29	3241.04	2509.66	1832.76	1193.08	641.11	218.96	0.27	0.00
PHD-4	2000.48	1472.00	977.38	528.49	163.73	0.00	0.00	0.00	0.00
PHD-2	999.44	638.51	289.62	8.75	0.00	0.00	0.00	0.00	0.00

*Population hamming distance of zero indicating the fixation of favorite alleles at target loci

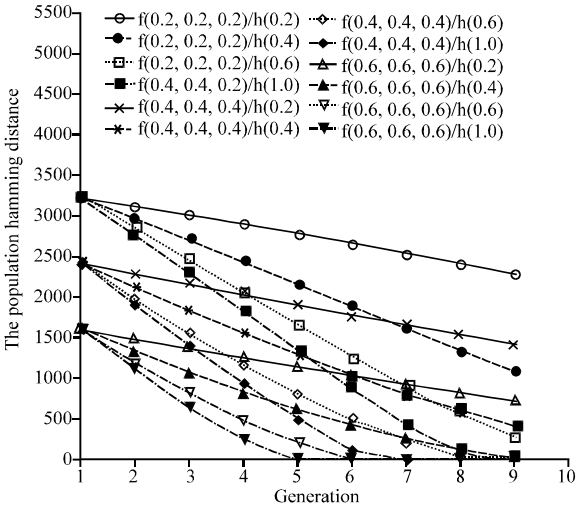


Fig. 2: Changes of population hamming distance over generations (1-9); f (0.2, 0.2, 0.2, 0.2) represents the favorite allele frequency at first/second/third/fourth locus and h (0.2) represents trait heritability 0.2

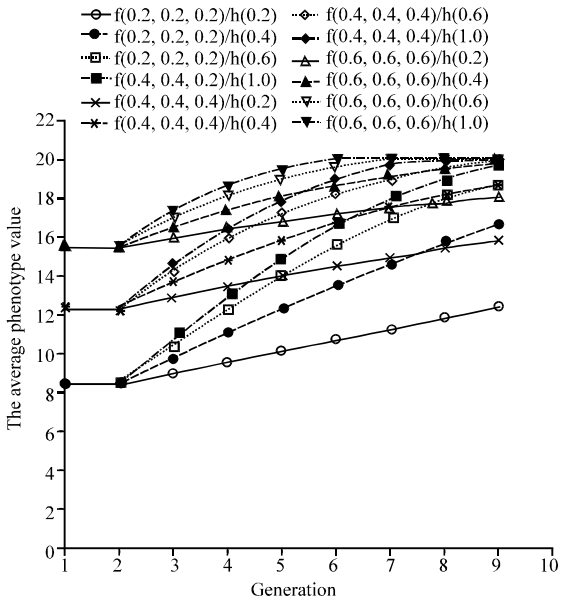


Fig. 3: Changes of population average phenotypic value over generations (1-9); f (0.2, 0.2, 0.2, 0.2) represents the favorite allele frequency at first/second/third/fourth locus and h (0.2) represents trait heritability 0.2

of the trait via gene pyramiding largely depends on the trait heritability which was controlled by the target genes and initial favorite allele frequency in base population.

Figure 4 shows the changes of superior genotype 11 frequency over nine generations. As to base population with the same favorite allele frequency at each locus, the target genes controlled higher trait heritability could easily reach to fixation. Supposed the same trait heritability, we compared the gene pyramiding process give different levels of initial favorite allele frequencies (Fig. 4e, b, f). The genotype 11 frequency for the f (0.6, 0.6, 0.6, 0.6)/h(0.4) easily reach a value of one at the ninth generation (Fig. 4f) which indicate that all the loci had pyramided to the ideal genotype.

We also compared the result of phenotypic selection with genotypic selection, the initial favorite allele frequency of four targets genes are set to 0.2, 0.4 or 0.6, respectively. Simulation results show that when initial favorite allele frequencies is 0.2, the target gene were pyramided at the ninth generation using genotypic selection but as to phenotypic selection with trait heritability 0.2, the average genotype 11 frequency is 0.176. When the initial favorite allele frequencies were set to 0.6, four target genes pyramided at the 5th generation using genotypic selection, compared to the traditional phenotypic selection, the average genotype 11 frequency are 0.4950, 0.7130 and 0.9180, respectively when the trait heritability are 0.2, 0.4, 0.6.

Phenotypic selection using genotype and phenotype model 1 and 2: As to phenotypic selection using model 1 only considered the gene effect, selection prompted the optimal genotype combination pyramided into the ideal individual. In this simulation, the initial favorite allele frequency for each locus is set at 0.3, the genotype 11 has a larger predefined gene effect than genotypes 10 or 01 and 00 and trait heritability was set at 0.2, 0.4 or 0.6. The most optimal genotype combination is the one with maximum summary of the genotypic values, so optimal genotype combination for four target locus is 11 11 11 11. We investigated four target genes with the same trait heritability of 0.4 (Fig. 5a-d), the genotype 11 frequency increases more significant than others when the gene

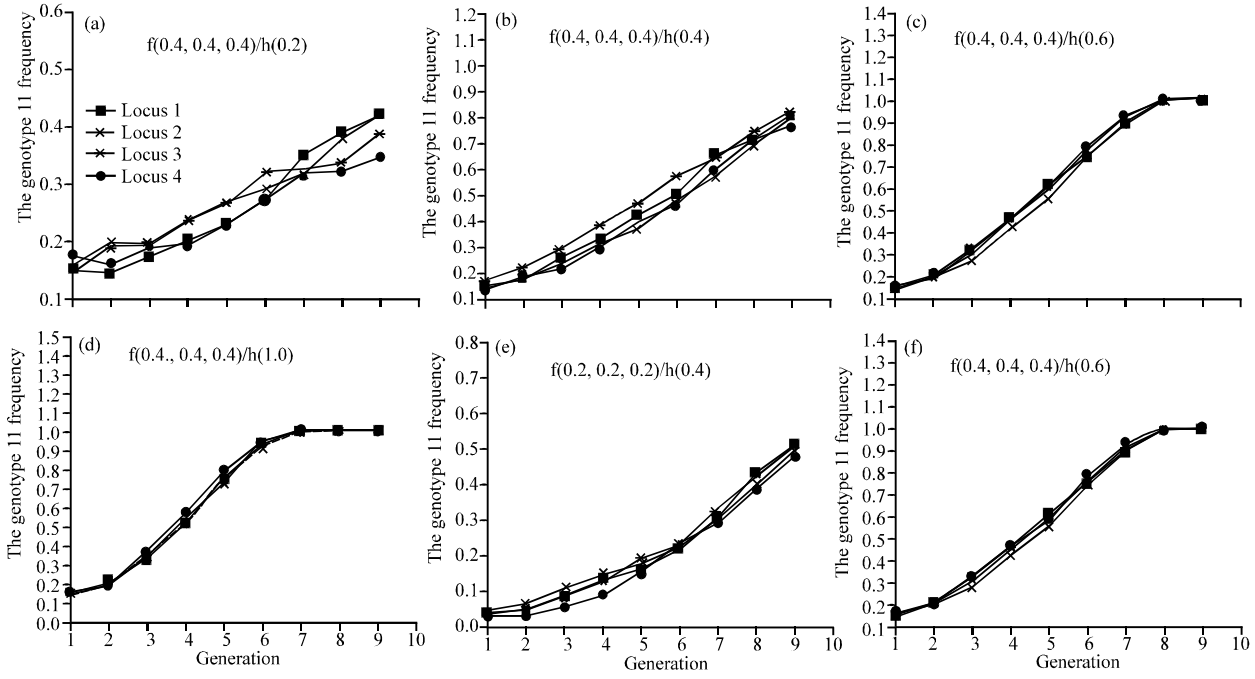


Fig. 4: Changes of genotype 11 frequency using simple model over generations (1-9). a-d, each locus was initialed with the same favorite allele frequency, f (0.4, 0.4, 0.4, 0.4), e, b and f set the same trait heritability, h(0.4); locus 1-4 represent the genotype 11 frequency at 1-4th locus, respectively; a) f (0.4, 0.4, 0.4, 0.4)/h(0.2); b) f (0.4,0.4,0.4,0.4)/h(0.4); c) f (0.4, 0.4, 0.4, 0.4)/h(0.6); d) f (0.4,0.4,0.4,0.4)/h(1.0); e) f(0.2, 0.2, 0.2, 0.2)/h(0.4), F) f (0.6,0.6,0.6,0.6)/h(0.4)

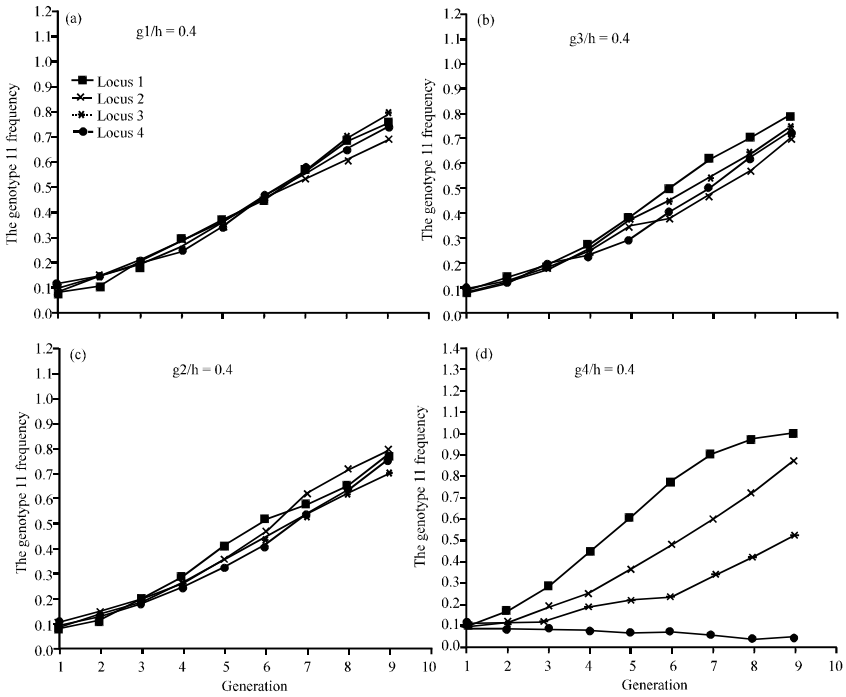


Fig. 5: Continued

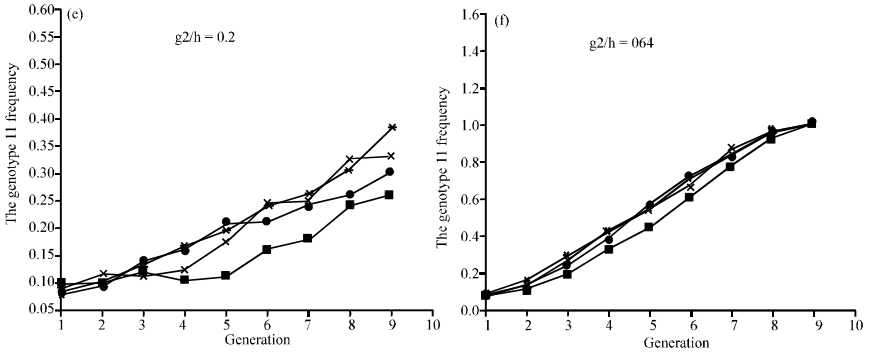


Fig. 5: P Changes of genotype 11 frequency using model 1 that integrates only gene effects over generations (1-9), the initial favorite allele frequency for each locus was set to 0.3. a-d show the same trait heritability h(0.4) but different gene effects g 1-4. b, e and f show the same gene effects but different trait heritability h = 0.2, h = 0.4 or h = 0.6; a) g 1/h = 0.4; b) g 2/h = 0.4; c) g 3/h = 0.4; d) g 4/h = 0.4, E) g 2/h = 0.2; f) g 2/h = 0.6; g 1 = [0.3 0.2 0.1; 0.3 0.2 0.1; 0.3 0.2 0.1; 0.3 0.2 0.1]; g 2 = [0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2]; g 3 = [0.9 0.6 0.3; 0.9 0.6 0.3; 0.9 0.6 0.3; 0.9 0.6 0.3]; g 4 = [0.9 0.6 0.3; 0.6 0.4 0.2; 0.3 0.2 0.1; 0.5 0.5 0.5]

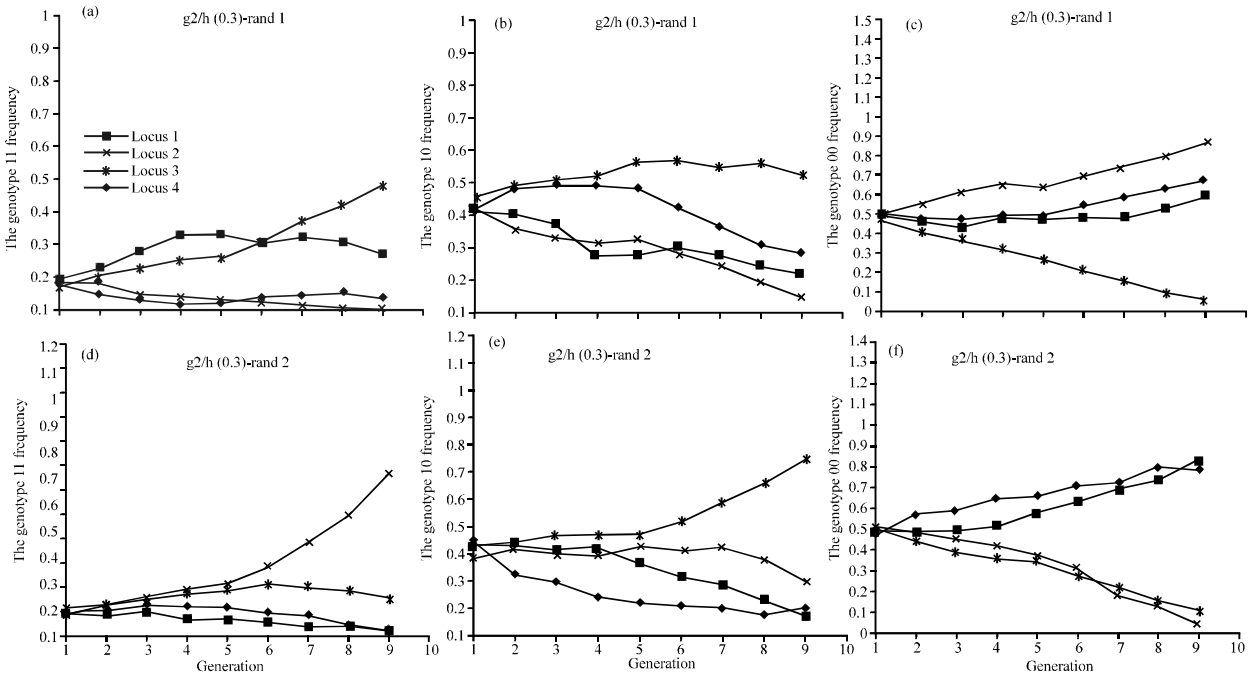


Fig. 6: Changes of Genotype 11 10 00 frequencies using model 2 that integrates both gene effects and gene interaction effects over generations (1-9). The initial favorite allele frequency at each locus was set at 0.3, the trait heritability was 0.3 and the randomly generated gene interaction effects were taken as rand 1 or 2 and rand1 and 2 represented 4D matrices from gim, g 2 is the gene effect as the same as Fig. 5

effects are larger (Fig. 5d). In the case of gene effects g 1-3, the genotype 11 frequency shows the same increasing trend over nine generations (Fig. 5a-c). Supposed the same gene effects at four target loci, the trait heritability is 0.6, the genotype 11 frequency changes more significantly than that of 0.2 and 0.4 (Fig. 5 e, b and f). The genotype 00 and 10 or 01 frequencies with small gene effects decrease with generations. The individual carrying the superior genotype combination were selected as the parents and inferior ones were eliminated. So, the superior genotype at each locus is close to fixation over generations. However as to the model integrating gene effect and gene interaction, the changes of genotype 11 frequency over nine generations appeared to be ambiguous (Fig. 6). As to the model 1, the population hamming distance exhibited the same trend with the

different gene effects g 1-3 with the same trait heritability (Table 2). However, different gene effects greatly affect the average phenotypic value (Table 3). Supposed trait heritability was 0.4, the gene effect at four target locus was set g 1, the average phenotypic value at the 9th generation is 1.6502 as to gene effect g 2, 3 and 4, the average phenotypic value are 4.1952, 6.2911 and 4.3311, respectively.

In the genotype-phenotype model 2 integrating gene effects and gene interaction effects, the initial favorite allele frequency at each locus was set to 0.3 and trait heritability is 0.3 or 0.6. The gene interaction effects were randomly generated as randl or rand 2. The results show that trait heritability and the gene interaction effect had no significant impact on the average phenotypic value (the results were not presented here) more over, the population hamming distance over nine generations did not decrease significantly, the target genes can be pyramided into one optimize genotype under the selection of this integrate model (Table 4). The results show that average phenotypic value increase with generations for both rand 1 and 2 with only slight difference between them (the results were not presented here). Figure 6 shows the changes of three type genotypes frequencies given rand 1 and 2. When the model integrating the rand 1 as the gene interaction effects, the highest frequencies of genotypes are genotype 00/00/10/00 at the first/second/third/fourth loci, responds to 0.596/0.854/0.524/0.674, the preset favorite allele don not pyramided over several generations

Table 2: Compared genotypic selection with phenotypic selection over nine generations

Allele frequency	Generation	G	P-h = 0.2	P-h = 0.4	P-h = 0.6
f (0.2, 0.2, 0.2, 0.2)[&]	G9[iG]	1.0[iU]	0.1760[§1]	0.4970[§2]	0.8440[§3]
f (0.4, 0.4, 0.4, 0.4)	G7	1.0	0.3260	0.6175	0.9075
f (0.6, 0.6, 0.6, 0.6)	G5	1.0	0.4950	0.7130	0.9180

[&]the favorite allele frequencies at first/second/third/fourth locus are 0.2, 0.2, 0.2, 0.2; [iG] four target genes were pyramided at generation 9th using genotypic selection; [iU] the average genotype 11 frequency at four locus using genotypic selection; [§1]the average genotype 11 frequency at four locus using phenotypic selection given trait heritability 0.2; [§2]the average genotype 11 frequency at four locus using phenotypic selection given trait heritability 0.4; [§3] the average genotype 11 frequency at four locus using phenotypic selection given trait heritability 0.6

Table 3: Changes of average phenotypic value using model integrating only gene effect over nine generations (1-9), gene pyramiding scheme considering gene effects g 1-3 or 4 and trait heritability 0.2, 0.4 or 0.6

	g											
	g1[§1]			g2[§2]			g3[§3]			g4[§4]		
	Heritability											
Generations	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
G1	1.2802	1.2798	1.2800	2.5603	2.5598	2.5601	3.8405	3.8399	3.8404	2.9201	2.9201	2.9202
G2	1.2804	1.2801	1.2793	2.5626	2.5609	2.5598	3.8398	3.8395	3.8402	2.9195	2.9190	2.9192
G3	1.3299	1.3952	1.4535	2.6586	2.7915	2.9073	3.9897	4.1849	4.3630	3.0154	3.1360	3.2482
G4	1.3805	1.5145	1.6288	2.7647	3.0294	3.2587	4.1445	4.5408	4.8895	3.1120	3.3642	3.5831
G5	1.4333	1.6359	1.8069	2.8706	3.2721	3.6150	4.2997	4.9074	5.4243	3.2110	3.5952	3.9087
G6	1.4879	1.7579	1.9800	2.9749	3.5178	3.9629	4.4608	5.2733	5.9451	3.3126	3.8173	4.1903
G7	1.5428	1.8775	2.1396	3.0827	3.7559	4.2812	4.6224	5.6318	6.4230	3.4123	4.0181	4.3984
G8	1.5960	1.9914	2.2745	3.1899	3.9838	4.5508	4.7849	5.9729	6.8258	3.5124	4.1914	4.5228
G9	1.6505	2.0971	2.3739	3.2991	4.1952	4.7489	4.9482	6.2911	7.1232	3.6123	4.3311	4.5950

[§1]genotypic value of 11,10,00 at fist/second/third/fouth loci, g1 = [0.3 0.2 0.1; 0.3 0.2 0.1; 0.3 0.2 0.1; 0.3 0.2 0.1]; [§2] genotypic value of 11,10,00 at fist/second/third/fouth loci, g 2 = [0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2]; [§3] genotypic value of 11,10,00 at fist/second/third/fouth loci, g 3 = [0.9 0.6 0.3; 0.9 0.6 0.3; 0.9 0.6 0.3; 0.9 0.6 0.3]; [§4] genotypic value of 11, 10, 00 at fist/second/third/fouth loci, g 4 = [0.9 0.6 0.3; 0.6 0.4 0.2; 0.3 0.2 0.1; 0.5 0.5 0.5]

Table 4: Changes of population hamming distance over generations (1-9)*, gene pyramiding scheme consider the trait heritability was 0.3 or 0.6 gene effect g 2 and 4 and gene interaction effects rand 1 and 2

	h = 0.3[+]		h = 0.3		h = 0.6		h = 0.6	
	g2[§2]	g2	g4[§4]	g4	g2	g2	g4	g4
Genotypes	rand 1	rand 2	rand 1	rand 2	rand 1	rand 2	rand 1	rand 2
G1	2800.00	2800.00	2800.00	2800.00	2800.00	2800.00	2800.00	2800.00
G2	2724.65	2750.65	2750.30	2776.69	2666.22	2711.57	2695.27	2743.36
G3	2642.46	2714.97	2696.61	2755.94	2574.59	2681.67	2636.68	2705.54
G4	2569.43	2703.38	2649.33	2745.19	2532.88	2693.38	2617.25	2712.62
G5	2502.23	2708.21	2605.16	2739.98	2505.54	2727.24	2614.60	2732.14
G6	2439.04	2727.45	2559.43	2735.69	2484.44	2788.05	2616.93	2759.85
G7	2378.86	2760.44	2511.52	2731.02	2466.68	2873.14	2625.29	2786.53
G8	2323.11	2811.96	2455.97	2726.55	2448.34	2970.93	2638.06	2804.79
G9	2267.44	2881.62	2394.44	2724.10	2419.85	3060.92	2651.84	2819.85

*Population hamming distance of zero indicated the fixation of favorite alleles at four loci; [+]The trait heritability is 0.3; [§2]Genotypic value of 11,10,00 at fist/second/third/fouth loci, g2 = [0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2; 0.6 0.4 0.2]; [§4]Genotypic value of 11,10,00 at fist/second/third/fouth loci, g4 = [0.9 0.6 0.3; 0.6 0.4 0.2; 0.3 0.2 0.1; 0.5 0.5 0.5]

selection because of the gene interaction effects, the optimal genotype combination are 00 00 10 00 but not 11 11 11 11 which are preset in the simulations. As to the rand 2, the highest frequencies of genotypes are genotype 00/11/10/00 at the first/second/third/fourth loci, responds to 0.82/0.6624/ 0.744/0.782. Different gene interaction effects result in the different optimal genotype combination by selecting the superior individuals over generations.

DISCUSSION

Results from the studies clearly show that substantial benefits can be expected from use of the selection on gene pyramiding in admixed population, provided the information on the major gene is incorporated properly in selection models. The process of gene pyramiding can be predicted given various selection strategies which would guild the breeder to make properly selection decisions.

Ideal genotypic selection strategy: Genotypic selection is an ideal strategy in practical breeding programs, breeders select the optimize genotype combination individuals to realize the optimization of target economic traits. The results clearly show that gene pyramiding risk and difficulty increase with number of target genes (Table 1). When the base population size was fixed at 500, target genes pyramided at the earlier generation under the same condition. The initial favorite allele frequency has significant influence on genotypic selection because each target gene would provide adequate and effective information for decision-making of breeders. When applying the genotypic selection strategy, we presumed that several target genes effect were additive, this presume may be suitable for the situation that all target genes controlling quality trait. However in practice breeding, this assumption may not be entirely valid because the agricultural traits of crops and economic traits of animals are quantitative traits. Therefore, we investigated the gene pyramiding breeding strategies using the models (1 and 2) linking genotype and phenotype and focused on one trait with specific heritability which was controlled by several major genes.

Phenotypic selection using the simple model and the integrated models 1 and 2: In the simple genotype-phenotype model, trait heritability is the main factor that affecting the effective gene pyramiding breeding. When the genotypic value was preset, trait heritability would have a direct impact on the average phenotypic value predicted by the model. In this simple model, the preset genotypic values of three genotypes are 5, 4, 1, respectively so, the optimal genotype combination is 11 11 11 11. As to the trait given more larger trait heritability, the dominant components in the model is the gene effect so, gene pyramiding breeding would be a process of select individual with the maximum genotypic value combination over generations (Fig. 4). The results also indicate that the target genes with larger heritability are pyramided at the earlier generation using this simple genotype-phenotype model.

Figure 5 shows the changes of genotype 11 frequency using model 1 that integrates only gene effects over generations (1-9), the main difference between this model and simple genotype-phenotype model are the set of genotypic values. The integrated model 1 can set the genotypic value at various levels. The aim of this model design is to investigate what optimal genotype combination is given different genotypic value.

As a special example for Fig. 5d, the genotypic value of 11, 10 and 00 are set to 0.5 at the fourth locus, results indicate that selection has no effect on the gene pyramiding >9 generations.

For the model 2 integrating the gene effect and gene interaction effects which was defined as molecular information integration selection, the results show that phenotypic selection tend to select individuals with superior values responds to the several different genotype combinations. This would make gene pyramiding a difficult issue and these target genes are hard to pyramid into one ideal genotype combination. The studies only give two randomly generated gene interaction effects rand 1 and 2, supposed the gene effects and gene interaction effects, trait heritability have been identified, selection would promote several favorite allele represents the optimal genotype combination fixed in the subsequent generations. We designed the integrated model in this study supposed that gene effect and gene interaction effects have been determined with quantitative values representing the effective molecular information, thus optimal genotype combination can be predicted on the gene pyramiding simulation platform. Also, the effectiveness of gene pyramiding can be shown by the changes of genotypes frequencies.

Simulations based on evolutionary computation: The studies use the metaphor of hill-climbing to model the dynamic behavior of gene pyramiding, inspired by the science of evolutionary computation (Goldberg, 1989; Holland, 1992), this strategy can use inexact solutions to deal with difficult tasks such as the solution of NP-complete problems. This strategy is most appropriate

for studying the combinatorial optimization of genotypes. As for gene pyramiding breeding, we considered a complex trait controlled by a series of major genes as the NP problem and genes were regarded as objects, gene pyramiding is select the optimize genotype combination individuals to realize the optimization of target economic traits. In recently years, theoretical and experimental studies on system biology would provide a new perspective for understanding complex traits (Benfey and Mitchell-Olds, 2008; Sauer *et al*., 2007; Sieberts and Schadt, 2007). As the information from the analysis of complex phenotypes becomes more and more precise, the relationship between gene networks at a micro-level would also become more and more clear. Further development of accurate and practical models is necessary to link the genotype and phenotype in order to increase the accuracy of model prediction. Evolutionary computation technology will help exploit useful information and guide the precise optimal design of breeding by gene pyramiding.

Discrete recombination (Goldberg, 1989) from the evolutionary computation was used to produce the genotype of offspring through the genotype of parents. This strategy disregard the recombination rate based on location of target gene in comparison to the more commonly used mapping functions such as Haldane's and Kosambi's. Discrete recombination may have application certain limitations but suitable for the simulation of multi-gene over multi-generations. Various types of recombination can be extended in the simulations, we used discrete recombination to investigate the transmission of genotype from parent generation to offspring generation for the sake of demonstration and also, it was expected that the Monte Carlo simulations would balance the crossover uncertainty.

Design of breeding programs using gene pyramiding: The object of gene pyramiding breeding is to improve the trait for entire population by selecting the most optimal genotype combination. Servin *et al*. (2004) had shown the gene pyramiding risk by statistical modeling. Their designed several possible succession of pair crosses leading to the target genotype (Servin *et al*., 2004) but did not consider the selection effects and initial gene frequency in base population so, the studies mainly focus on these problems. In this study, the method and model were provided would be used to integrate molecular information from multi-level such as gene effect and gene interaction effects. We built simulation platform for the gene pyramiding breed and different level of population size, initial gene frequency and flexible selection strategies can be designed on this platform, the gene pyramiding breeding progress would be predicted for us as valuable references for decision-making of breeders.

CONCLUSION

The studies provide relatively simple models linking the genotype information to phenotype and it offers a new perspective to use information available from marker and Quantitative Trait Loci (QTL) to guide the gene pyramiding design breeding. We hope more optimized models will be developed in future studies as the development of system biology and high-throughout array technology. Moreover, different cross schemes and selection strategies can be designed and compared based on this gene pyramiding simulation platform.

ACKNOWLEDGEMENTS

This study was supported by the National High Technology Development Project of China (863 Program) (Grant No. 2006AA10Z199) and Key Projects in the National Science and Technology Pillar Program during the 11th 5 Years Plan Period (Grant No. 2006BAD01A11).

APPENDIX

In four *loci* gene effect matrix 1, three genotype effects at four loci were represented by g (i, k) where, i denotes the locus and k = 1-3 denotes three genotypes 11, 10 or 01 and 00, respectively. For example, loc1-1 refers to the genotype 11 effect at the 1st locus, loc2-2 refers to the genotype 10 effect in the 2nd locus and loc4-3 refers to the genotype 00 effect in the fourth locus:

$$g(i,k)=\begin{bmatrix} loc1-1 & loc1-2 & loc1-3 \\ loc2-1 & loc2-2 & loc2-3 \\ loc3-1 & loc3-2 & loc3-3 \\ loc4-1 & loc4-2 & loc4-3 \end{bmatrix} \quad (1)$$

In four *loci* gene interaction effects matrix 2, $gim_{(l\text{-}1,\, l\text{-}2,\, l\text{-}3,\, l\text{-}4)}$ denotes the combinational interaction effects of loci l-1, l-2, l-3, l-4 and l-1, l-2, l-3, l-4 were coded by 123 to represent the three genotypes 11, 10 or 01 and 00. The value of $gim_{(l\text{-}1,\ l\text{-}2,\ l\text{-}3,\ l\text{-}4)}$ represents the gene interaction effects of four genotype combination in genotype and phenotype model 2 which were sampled from standard normal distribution N (0, 1). For example, $gim_{(1111)}$ represents the gene interaction effects for the

genotype combination of 11-11-11-11, the 1st locus genotype is 11, 2nd genotype is 11, 3rd genotype is 11 and 4th genotype is 11. $gim_{(1213)}$ represented the gene interaction effects for the combination of 11-10 or 01-11-00 in which the 1st locus genotype is 11, 2nd genotype is 10 or 01, 3rd genotype is 11 and 4th genotype is 00. $gim_{(3333)}$ indicates the gene interaction effects for the combination of 00-00-00-00 in which the 1st locus genotype is 00, 2nd genotype is 00, 3rd genotype is 00 and 4th genotype is 00:

$$gim_{(1-1,1-2,1-3,1-4)} = \left[\begin{array}{ccc|ccc|ccc} 1111 & 1211 & 1311 & 2111 & 2211 & 2311 & 3111 & 3211 & 3311 \\ 1112 & 1212 & 1312 & 2112 & 2212 & 2312 & 3112 & 3212 & 3312 \\ 1113 & 1213 & 1313 & 2113 & 2213 & 2313 & 3113 & 3213 & 3313 \\ \hline 1121 & 1221 & 1321 & 2121 & 2221 & 2321 & 3121 & 3221 & 3321 \\ 1122 & 1222 & 1322 & 2122 & 2222 & 2322 & 3122 & 3222 & 3322 \\ 1123 & 1223 & 1323 & 2123 & 2223 & 2323 & 3123 & 3223 & 3323 \\ \hline 1131 & 1231 & 1331 & 2131 & 2231 & 2331 & 3131 & 3231 & 3331 \\ 1132 & 1232 & 1332 & 2132 & 2232 & 2332 & 3132 & 3232 & 3332 \\ 1133 & 1233 & 1333 & 2133 & 2233 & 2333 & 3133 & 3233 & 3333 \end{array}\right]$$

In the model 2, $gim_{(1-1,1-2,1-3,1-4)}$ denotes the genotype interaction effect, similar to polygenic effect but represents the actual information integration in model 2. Considered the genotype interaction effects were unknown, so in the simulation they were taken as random variables and $gim_{(1-1,\ 1-2,\ 1-3,\ 1-4)} \sim N\ (0,\ 1)$ as in the results section rand 1 and 2 denote four genotype interaction effects using (4D) matrices, the values of rand 1 and 2 were sampled from $gim_{(1-1,1-2,1-3,1-4)}$. The following codes were detail description of gim4 in matlab:

```
>> gim4(:,:,1,1) = [-0.5839  0.8986  0.5768;  -0.0029 1.6396-1.4264; -0.1237-0.2090 0.6300];

gim4(:,:,1,2)=[-1.5620 -0.2955  -1.0372; 0.1886  0.0579 -0.7288; 1.0144  -0.5551  0.2263];
gim4(:,:,1,3)=[0.3766  1.3061  0.4233; 0.8616  -1.0946 0.0018; -1.8306  -0.4980  -0.6359];
gim4(:,:,2,1)=[0.9025  1.0415  2.4823; 0.1726  -0.9162 0.9498 ; -0.2932  -0.7769  -0.1439];
gim4(:,:,2,2)=[-0.1699 -1.4072  1.5671; 0.3425  -1.1499 -0.0336; 0.9685  -0.7011  -0.5948];
gim4(:,:,2,3)=[1.0978  0.0615  -0.4146; -1.0844 -0.2463 0.3544; 1.1517  -0.7214  -0.3996];
gim4(:,:,3,1)=[0.6040  0.0489  -0.9313; -0.1045 1.6993  -1.5508; 1.2477  -1.0951  -0.9404];
gim4(:,:,3,2)=[0.5224  -0.7156  2.5740; 0.2894  -0.0006 -0.3670; 1.7875  -2.2961  -0.9563];
gim4(:,:,3,3)=[ 0.5010 -1.4791  -0.2858; 0.8811  0.2508  0.4403; -0.1095  -0.4080  0.8605];
>> gim4

gim4(:,:,1,1) =
   -0.5839   0.8986   0.5768
   -0.0029   1.6396  -1.4264
   -0.1237  -0.2090   0.6300

gim4(:,:,2,1) =
    0.9025   1.0415   2.4823
    0.1726  -0.9162   0.9498
   -0.2932  -0.7769  -0.1439

gim4(:,:,3,1) =
    0.6040   0.0489  -0.9313
   -0.1045   1.6993  -1.5508
    1.2477  -1.0951  -0.9404

gim4(:,:,1,2) =
   -1.5620  -0.2955  -1.0372
    0.1886   0.0579  -0.7288
    1.0144  -0.5551   0.2263

gim4(:,:,2,2) =
   -0.1699  -1.4072   1.5671
    0.3425  -1.1499  -0.0336
    0.9685  -0.7011  -0.5948

gim4(:,:,3,2) =
    0.5224  -0.7156   2.5740
    0.2894  -0.0006  -0.3670
    1.7875  -2.2961  -0.9563
```

gim4(:,:,1,3) =

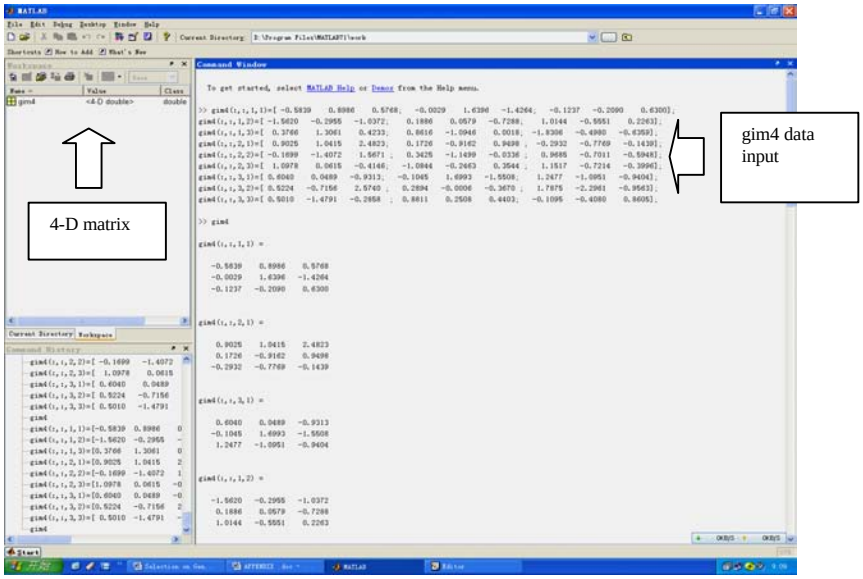
0.3766	1.3061	0.4233
0.8616	-1.0946	0.0018
-1.8306	-0.4980	-0.6359

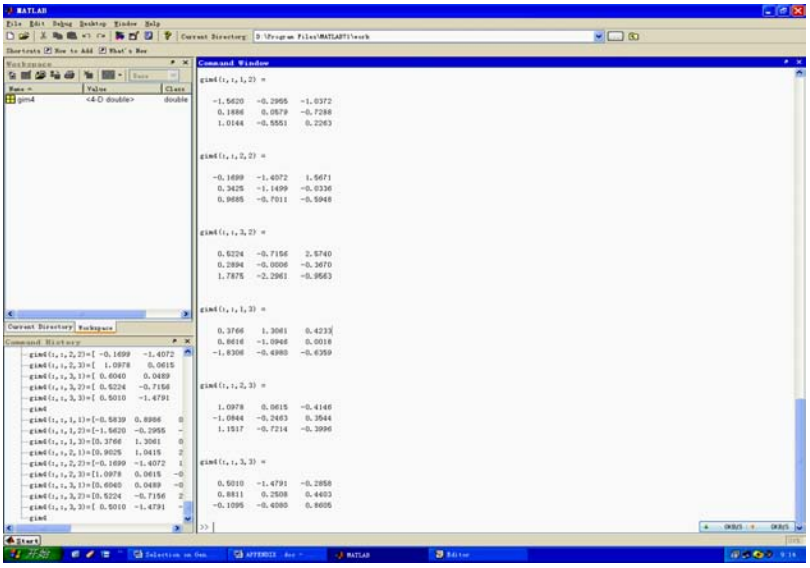
gim4(:,:,2,3) =

1.0978	0.0615	-0.4146
-1.0844	-0.2463	0.3544
1.1517	-0.7214	-0.3996

gim4(:,:,3,3) =

0.5010	-1.4791	-0.2858
0.8811	0.2508	0.4403
-0.1095	-0.4080	0.8605





REFERENCES

Ashikari, M., H. Sakakibara, S. Lin, T. Yamamoto and T. Takashi *et al*., 2005. Cytokinin oxidase regulates rice grain production. Science, 309: 741-745.

Benfey, P.N. and T. Mitchell-Olds, 2008. From genotype to phenotype: Systems biology meets natural variation. Science, 320: 495-497.

Castro, A.J., F. Capettini, A.E. Corey, T. Filichkina and P.M. Hayes *et al*., 2003. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor. Appl. Genet., 107: 922-930.

Goldberg, D.E., 1989. Genetic Algorithms in Search Optimization and Machine Learning. Addison Wesley Publishing Co., Reading, Massachutes.

Holland, J.H., 1992. Adaptation in Natural and Artificial Systems. 2nd Edn., MIT Press, Cambridge, MA, pp: 211.

Hospital, F., I. Goldringer and S. Openshaw, 2000. Efficient marker-based recurrent selection for multiple quantitative trait loci. Genet. Res., 75: 357-368.

Hu, X.S., 2007. A general framework for marker-assisted selection. Theor. Popul. Biol., 71: 524-542.

Huang, N., E.R. Angeles, J. Domingo, G. Magpantay and S. Singh *et al*., 1997. Pyramiding of bacterial blight resistance genes in rice: Marker-assisted selection using RFLP and PCR. Theor. Applied Genet., 95: 313-320.

Ishii, T. and K. Yonezawa, 2007. Optimization of Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: I. Schedule of crossing between the donor lines. Crop. Sci., 47: 537-546.

Ishii, T., T. Hayashi and K. Yonezawa, 2008. Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: III. Multiple-gene assemblage using background marker selection. Crop Sci., 48: 2123-2131.

Jefferies, S.P., B.J. King, A.R. Barr, P. Warner, S.J. Logue and P. Langridge, 2003. Marker-assisted backcross introgression of the Yd2 gene conferring resistance to barley yellow dwarf virus in barley. Plant Breed., 122: 52-56.

Jiang, L., F. Zhao and Q. Zhang, 2008. Simulation of gene pyramiding in *Drosophila melanogaster*. J. Genet. Genomics, 35: 737-742.

Koudande, O.D., J.A. van Arendonk and F. Iraqi, 2005. Marker-assisted introgression of trypanotolerance QTL in mice. Mamm. Genome, 16: 112-119.

Lande, R. and R. Thompson, 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics, 124: 743-756.

Lange, C. and J.C. Whittaker, 2001. On prediction of genetic values in marker-assisted selection. Genetics, 159: 1375-1381.

Pilcher, C.D., J.K. Wong and S.K. Pillai, 2008. Inferring HIV transmission dynamics from phylogenetic sequence relationships. PLoS Med., 5: e69-e69.

Saghai Marrof, M.A., S.C. Jeong, I. Gunduz, D.M. Tucker, G.R. Buss and S.A. Tolin, 2008. Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. Crop Sci., 48: 517-526.

Sauer, U., M. Heinemann and N. Zamboni, 2007. Genetics. Getting closer to the whole picture. Science, 316: 550-551.

Servin, B., O.C. Martin, M. Mezard and F. Hospital, 2004. Toward a theory of marker-assisted gene pyramiding. Genetics, 168: 513-523.

Sieberts, S.K. and E.E. Schadt, 2007. Moving toward a system genetics view of disease. Mamm. Genome, 18: 389-401.

Singh, S., J.S. Sidhu, N. Huang, Y. Vikal and Z. Li *et al*., 2001. Pyramiding three bacterial blight resistance genes (xa5, xa13 and Xa21) using marker-assisted selection into indica rice cultivar PR106. Theor. Applied Genet., 102: 1011-1015.

Visscher, P.M., C.S. Haley and R. Thompson, 1996. Marker-assisted introgression in backcross breeding programs. Genetics, 144: 1923-1932.

Wall, E., P.M. Visscher, F. Hospital and J.A. Woolliams, 2005. Genomic contributions in livestock gene introgression programmes. Genet. Sel. Evol., 37: 291-313.

Zhao, F.P., L. Jiang, H.J. Gao, X.D. Ding and Q. Zhang, 2009. Design and comparison of gene-pyramiding schemes in animals. Animal, 3: 1075-1084.