Optimal Design for Marker-assisted Gene Pyramiding in Cross Population

L. Y. Xu[1,a], F. P. Zhao[1,2,a], X. H. Sheng[1], H. X. Ren[1], L. Zhang[1], C. H. Wei[1] and L. X. Du[1,*]

[1] Institute of Animal Science, Chinese Academy of Agricultural Sciences,
National Center for Molecular Genetics and Breeding of Animal, Beijing 100193, China

ABSTRACT: Marker-assisted gene pyramiding aims to produce individuals with superior economic traits according to the optimal breeding scheme which involves selecting a series of favorite target alleles after cross of base populations and pyramiding them into a single genotype. Inspired by the science of evolutionary computation, we used the metaphor of hill-climbing to model the dynamic behavior of gene pyramiding. In consideration of the traditional cross program of animals along with the features of animal segregating populations, four types of cross programs and two types of selection strategies for gene pyramiding are performed from a practical perspective. Two population cross for pyramiding two genes (denoted II), three population cascading cross for pyramiding three genes(denoted III), four population symmetry (denoted IIII-S) and cascading cross for pyramiding four genes (denoted IIII-C), and various schemes (denoted cross program-A-E) are designed for each cross program given different levels of initial favorite allele frequencies, base population sizes and trait heritabilities. The process of gene pyramiding breeding for various schemes are simulated and compared based on the population hamming distance, average superior genotype frequencies and average phenotypic values. By simulation, the results show that the larger base population size and the higher the initial favorite allele frequency the higher the efficiency of gene pyramiding. Parents cross order is shown to be the most important factor in a cascading cross, but has no significant influence on the symmetric cross. The results also show that genotypic selection strategy is superior to phenotypic selection in accelerating gene pyramiding. Moreover, the method and corresponding software was used to compare different cross schemes and selection strategies. (**Key Words:** Gene Pyramiding, Evolutionary Computation, Cross Population, Population Hamming Distance)

INTRODUCTION

Gene pyramiding aims to design a superior trait through combining favorite alleles into an ideal genotype. Currently, molecular dissection of complex traits has striven to explain the genetic architecture of agronomic traits in plants or economic traits in animals (Doerge, 2002; Ljungberg et al., 2002; Chen and Kendziorski, 2007). Many quantitative trait loci and linked markers have been identified. The rapidly growing molecular information will provide great opportunities for practical applications of crops and farm animals using marker assisted selection as well as marker assisted gene pyramiding (Fadiel et al., 2005).

Marker assisted gene pyramiding is an important branch

of marker assisted selection. It has been successfully applied in many plant breeding schemes most of which involved pyramiding of disease resistance genes with the main effects (Huang, 1997; Singh et al., 2001; Saghai Marrof, 2008; Kameswara Rao et al., 2010). Although, in recent years, some theoretical studies of marker assisted selection have been done (Lande and Thompson, 1990; Ruane and Colleau, 1995; Moreau et al., 1998; Lange and Whittaker, 2001; Hu, 2007), the theoretical study of marker assisted gene pyramiding has just begun.

Servin et al. (2004) investigated the theoretical issues of gene pyramiding and proposed general principles for designing gene pyramiding schemes in plants. They proposed that if the location and a series of genes of interested were known, the selection problem may be reduced to a "building block" problem. The estimate of pyramiding efficiency is based on gene transmitted probability and the minimum population size needed for obtaining the individual with an ideal genotype. In consideration of the features of an animal population, such

as the long generation interval and limitation of fertility, Zhao et al. (2009) extended these theories to design some representative gene pyramiding schemes for pyramiding three and four target genes, and proposed two criteria to select the optimal scheme in certain conditions. However, these theoretical studies did not take into account the initial target gene frequencies in the base population and selection strategies. In practice, animal breeding populations are segregating populations. So the likelihood that a favorable allele is completely absent in a breed is small. Hence, gene pyramiding breeding theory for animals needs further study.

Within the field of evolutionary computation, there have been some studies using animal breeding strategies to design algorithms to search for optimal solutions to problems (Muhlenbein and Voosen, 1993; Podlich and Cooper, 1998). Inspired by the science of evolutionary computation (David, 1989), the algorithms of gene pyramiding breeding are developed based on the same theoretical foundation of the building block hypothesis from the evolutionary algorithms perspective. Selection over several generations promotes the superior allele pyramided at all target loci. Considering the segregating population in animal breeding practices, we designed four types of cross programs for pyramiding two, three and four target genes. In these programs, we used the population hamming distance and superior genotype frequencies to measure the pyramiding efficiencies in the process of gene pyramiding. There are also some other factors considered, which include

the initial favorite allele frequencies in base populations, base population sizes and the selection strategies.

MATERIALS AND METHODS

General concept of gene pyramiding breeding

Marker-assisted gene pyramiding aims to produce individuals with superior economic traits in optimal breeding schemes through selecting and pyramiding favorite target alleles or linked markers into a single genotype. Servin et al. (2004) proposed that gene pyramiding breeding consisted of two basic steps, the pyramiding step and the gene fixation step. In our studies, we designed four types of cross programs for pyramiding two, three and four target genes in the pyramiding step (Figure 1), and in this step the target genes existing in different populations with different favorite allele frequencies are cumulated into one cross population. The fixation step begins with the cross population, then the selected parents are intercrossed to fix all the target genes into an ideal genotype individual (Servin et al., 2004; Zhao et al., 2009).

Population and individual genotype simulation

Our studies assume gene pyramiding design is a process of searching the optimal genotype combination, the target trait was mainly controlled by several major genes, and the individual's genotype was coded by a string of 0 or 1. We
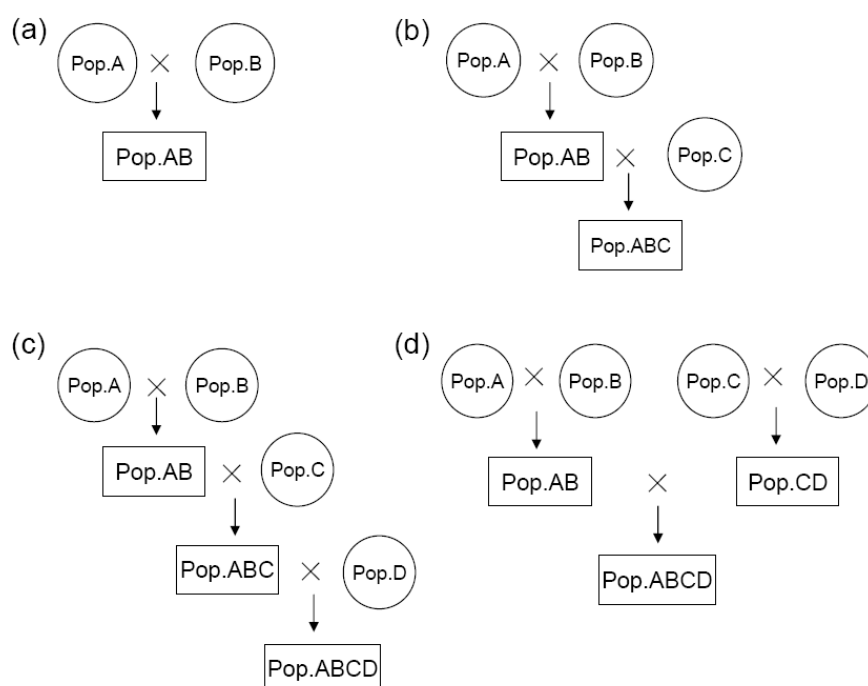

Figure 1. Four types of cross programs in the pyramiding step. a) Two populations cross aiming to two genes pyramiding. b) Three populations cascading cross aiming to three genes pyramiding. c) Four populations cascading cross aiming to four genes pyramiding. d) Four populations symmetric cross aiming to four genes pyramiding.

coded the genotype of one locus using two characters (0 or 1), and the string population represents the genotype of all individuals in the population. Initial base population is represented by N×M Matrix (N denotes the number of individuals in the base population, M/2 denotes the number of loci). The favorite allele frequencies of the initial base population are set at various levels. In each generation, individuals are evaluated by genotypic scores and phenotypic values using different selection strategies.

Discrete recombination is used to combine (mate) two individuals (parents) to produce new offspring by the crossover of the selected parents. Discrete recombination uses a crossover mask to indicate which parents will supply bits (alleles) to the offspring, and a crossover mask is the same length as the individual structure, which is randomly generated by 0 or 1 with equal probability. Crossover mask 1 indicates the allele of offspring at this locus is inherited from parent 1, crossover mask 0 indicates the allele of offspring at this locus is inherited from parent 2. Discrete recombination at each locus is used to produce offspring with a new genotype combination. Offspring1 is produced by mast1, and offspring 2 is produced by mast 2, the allele inherited from parent 1 is marked with underline (see as follow).

```
parent 1        0 1 1 1 0 0 1 1
parent 2        1 0 1 0 1 1 0 0

mast 1          0 1 1 0 0 0 1 1
  |
offspring 1     1 1 1 0 1 1 1 1

mast 2          1 0 0 1 1 1 0 0
  |
offspring 2     0 0 1 1 0 0 0 0
```

In our simulations, the supposed ideal population is the population with fixation of favorite alleles at all target loci. For example, as to four loci, the ideal genotype is 11-11-11-11, and the ideal population is coded as 1s matrix, in which all individuals carry ideal genotypes. 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 has been used to measure the number of nucleotide differences between two genetic sequences (Pilcher, 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 the 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 the population. Population hamming distance between pop (t) and pop (ideal) is 19.

$$
\text{pop(t)} = \begin{pmatrix} 10\ 10\ 11\ 00 \\ 01\ 11\ 00\ 01 \\ 00\ 01\ 11\ 10 \\ 01\ 11\ 10\ 00 \\ 11\ 10\ 10\ 01 \end{pmatrix} \quad \text{pop(ideal)} = \begin{pmatrix} 11\ 11\ 11\ 11 \\ 11\ 11\ 11\ 11 \\ 11\ 11\ 11\ 11 \\ 11\ 11\ 11\ 11 \\ 11\ 11\ 11\ 11 \end{pmatrix}
$$

Genotypic selection and phenotypic selection strategy

In the genotypic selection strategy, genotype 11 is scored 2, genotype 10 is scored 1, and genotype 00 is scored 0. The genotypic selection score is the sum of the score of genotype at all loci, and the score is used as the selection criterion in subsequent generations, and the additive genetic effects are assumed here.

In the phenotypic selection strategy, the phenotypic observation of each individual is modeled as:

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

Where p_i is the phenotypic observation of individual i, μ_0 is the overall mean, g_j is the gene effect at jth locus (j = 1,2,…,m, where m is the number of target genes), x_{ij} is an indicator variable of genotype j with value 0, 1, 2, and is the residual error following the distribution N(0, σ_ε^2). The values of genotypes are defined in terms of the midpoint (m), additive (a) and dominance (d) genetic parameters. The numerical coding of three genotypes 11, 10, 00 are 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 equation (4). From the defined heritability an estimate of σ_ε^2 is obtained by calculating σ_ε^2 and re-arranging equation (4) to (5).

$$
h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2} \tag{4}
$$

$$
\sigma_\varepsilon^2 = \frac{\sigma_g^2}{h^2} - \sigma_g^2 \tag{5}
$$

Cross programs and gene pyramiding design breeding

In this study, we designed four types of cross programs for gene pyramiding breeding, which are represented by II, III, III.C, IIII.S. For each cross program, various schemes are also designed given various levels of initial favorite allele frequencies and trait heritabilities, the schemes are denoted by cross program-X-h/G (X is an indicator variable with letter A, B, C, etc, h denotes trait heritability 0.2, 0.4 or

0.6 and G denotes genotypic selection). II represents pyramiding two target genes from popA and popB (Figure 1a), A1/A2 denotes favorite allele frequencies in the first/second loci in the popA, B1/B2 denotes favorite allele frequencies in the first/second loci in the popB, N denotes the base population size. The base population sizes of popA and popB vary from 500, 1,000 to 2,000. The initial favorite allele frequencies A1/A2 and B1/B2 at first/second loci are set as 0, 0.25, or 0.50, respectively. The popAB is produced by crossing popA with popB. The top 500 individuals based on genotypic score are selected for the next generation and each pair of parents is assumed to produce four offspring with the sex ratio 1:1. Then, the selected parents are randomly intercrossed to produce the subsequent generations until two target genes are pyramided into an ideal genotype.

III represents pyramiding three target genes from popA, popB and popC (Figure 1b), which we called a three population cascading cross, A1/A2/A3 denotes favorite allele frequencies in the first/second/third loci in the popA, B1/B2/B3 denotes favorite allele frequencies in the first/second/third loci in the popB, C1/C2/C3 denotes favorite allele frequencies in the first/second/third loci in the popC. The initial favorite allele frequencies A1/A2/A3, B1/B2/B3 and C1/C2/C3 at first/second/third loci are set as 0, 0.25, or 0.50, respectively. The base population size of popA, popB and popC vary from 500, 1,000 to 2,000. The popA and popB are crossed to produce the popAB, and each pair of parents is assumed to have four offspring with the sex ratio 1:1. The top 500 individuals are selected based on genotype scores for the next generation. The initial population size of popC is set as 2×N, the top 500 of popAB and popC are crossed to produce popABC. Then each pair of parents are randomly intercrossed to produce the subsequent generations until three target genes are pyramided into an ideal individual.

IIII represents pyramiding four target genes from popA, popB, popC and popD, A1/A2/A3/A4 denotes favorite alleles frequencies in the first/second/third/fourth loci in the popA, B1/B2/B3/B4 denotes favorite allele frequencies in the first/second/third/fourth loci in the popB, C1/C2/C3/C4 denotes favorite allele frequencies in the first/second/third/fourth loci in the popC, D1/D2/D3/D4 denotes favorite allele frequencies in the first/second/third/fourth loci in the popD. The base population sizes (N) are set as 500 and 1,000, respectively. Other breeding parameters are as the same as schemes II and III. For four population cascading cross, denoted IIII.C (Figure 1c), the base population size of popA, popB, popC and popD are N, N, 2×N and 4×N, PopA and popB are crossed to produce popAB, the top 500 of popAB cross with popC to produce population popABC, than the top 500 of popABC cross with popD to produce popABCD. For four population

symmetric cross, denoted IIII.S (Figure 1d), the base population size of popA, popB, popC and popD are N, N, N and N respectively, PopA and popB are crossed to produce popAB, and popC and popD are crossed to produce popCD, then the top 500 of popAB cross with the top 500 of popCD to produce popABCD in the next generation. Each pair of parent is assumed to produce four offspring with the sex ratio 1:1. In the population PopAB, PopCD, popABCD, individuals are selected based on genotypic scores or phenotypic values, the top 500 individuals are selected as the parents, the selected parents are randomly intercrossed in the subsequent generations until the four target gene are pyramided into an ideal individual.

In this study, we designed four types of cross programs, the base population size and initial favorite allele frequency are set at different levels in each cross program, and trait heritability is also considered in phenotypic selection. The gene pyramiding generation, population hamming distance and the superior genotype frequency are used to measure the process of gene pyramiding breeding. We performed Monte Carlo simulation for each cross scheme, and simulations are repeated 1,000 times. Our computer programs are implemented via Matlab and run on the Inter(R) Core(TM) 2 Duo CPU. Microsoft Windows XP.

RESULTS

Gene pyramiding through genotypic selection

In the genotypic selection strategy, we firstly designed three schemes for two target genes pyramiding program (II). Table 1 shows changes of population hamming distance over generations (1-6). For scheme II-B, initial base population size is 500, and the population hamming distance at G4 and G5 are 490 and 196, but for population size 2,000, it goes up to 1,921 and 739. Another factor affecting gene pyramiding progress is the initial favorite allele frequency. For the base population with 500 individuals, see scheme II-C (A1/A2(0.5/0.25), B1/B2(0.25/0.5)), all the target genes are fixed at G4, with the initial favorite allele frequency decrease, such as scheme II-A (A1/A2(0.5/0), B1/B2(0/0.5)) and II-B (A1/A2(0.25/0), B1/B2(0/0.25)), two target genes are pyramided until the G5 and G6 (Table 1).

As to the cross program for three genes pyramiding (III), base population size is set to 500, and the initial favorite allele frequency is 0.5 (III-A), three target genes are pyramided at G7, when the population size increases to 2,000 and the allele frequency decreases to 0.25 (III-B), three genes are pyramided at G8 (Table 2). Under the same population size, a cross scheme with initial gene frequency 0.25 needs more generations than that of 0.5. For the scheme III-C, three populations with two loci carrying favorite genes, the population hamming distance for

Table 1. Changes of population hamming distance over generations (1-6)* for II

Cross scheme	Population size	A1/A2[1]	B1/B2[2]	G1	G2	G3	G4	G5	G6
II-A	500	0.50/0.00	0.00/0.50	6,000	786	445	170	0	0
II-B	500	0.25/0.00	0.00/0.25	7,000	1,161	816	490	196	0
II-C	500	0.50/0.25	0.25/0.50	5,000	422	195	0	0	0
II-A	1,000	0.50/0.00	0.00/0.50	12,002	1,567	876	325	0	0
II-B	1,000	0.25/0.00	0.00/0.25	14,002	2,321	1,631	972	383	0
II-C	1,000	0.50/0.25	0.25/0.50	10,000	843	384	0	0	0
II-A	2,000	0.50/0.00	0.00/0.50	23,999	3,131	1,735	633	0	0
II-B	2,000	0.25/0.00	0.00/0.25	28,001	4,632	3,253	1,921	739	0
II-C	2,000	0.50/0.25	0.25/0.50	19,997	1,686	759	0	0	0

* Population hamming distance of zero indicating the fixation of favorite alleles at both loci.
[1] Allele frequencies in first/second loci in population A. [2] Allele frequencies in first/second loci in population B.

population size 500, 1,000 and 2,000 are 107, 194 and 367 at G5, respectively, and three target genes are pyramided at G6. We also compared scheme III-D with III-E, for III-D, the population C with the target locus carrying higher frequency favorite allele (0.5) is taken as the third cross population, for III-E, the population C with the target locus carrying lower frequency favorite allele (0.25) is taken as the third cross population, the results show that population hamming distances in III-D are lower than that of III-E at the first four generations, but for the subsequence generations they show the opposite trend. So for population size 500, 1,000 and 2,000, the population hamming distance does not change significantly for schemes III-D and III-E (Table 2).

We designed two cross programs (symmetric and cascading) for four genes pyramiding from four donor populations. Table 3 shows the changes of population hamming distance over generation (1-10) for symmetric cross program (IIII.S). Population size is set as 500, four target genes pyramided at G8 (IIII.S-A) when the initial favorite allele frequency at each locus in each population is 0.5, compared to one in which the frequency is 0.25, the genes are pyramided at G10 (IIII.S-B). When population size is 1,000 and initial favorite allele frequency is 0.5, the results show that the population hamming distances are 443 and 501 at the G7 and G8, respectively. When the population size is 500, the population hamming distances are 232 and 264 (Table 3). We investigated the schemes with different favorite allele frequencies at each locus in four population, such as schemes IIII.S-C and IIII.S-D. Both schemes show the similar results, the population hamming distances are 133, 132 and 213, 226 for base

Table 2. Changes of population hamming distance over generations (1-8)* for III

Cross scheme	Population size	A1/A2/A3[1]	B1/B2/B3[2]	C1/C2/C3[3]	G1	G2	G3	G4	G5	G6	G7	G8
III-A	500	0.50/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.50	19,995	1,466	1,134	735	350	59	0	0
III-B	500	0.00/0.00/0.50	0.00/0.25/0.00	0.00/0.00/0.25	21,999	1,925	1,621	1,214	791	408	104	0
III-C	500	0.50/0.25/0.00	0.00/0.50/0.25	0.25/0.00/0.50	17,998	1,177	779	402	107	0	0	0
III-D	500	0.25/0.00/0.00	0.00/0.25/0.00	0.00/0.00/0.50	21,001	1,806	1,462	1,095	705	329	44	0
III-E	500	0.25/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.25	21,496	1,853	1,507	1,109	698	315	29	0
III-A	1,000	0.50/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.50	39,996	2,929	2,262	1,463	680	90	0	0
III-B	1,000	0.00/0.00/0.50	0.00/0.25/0.00	0.00/0.00/0.25	43,999	3,849	3,235	2,421	1,561	788	180	0
III-C	1,000	0.50/0.25/0.00	0.00/0.50/0.25	0.25/0.00/0.50	36,000	2,352	1,552	787	194	0	0	0
III-D	1,000	0.25/0.00/0.00	0.00/0.25/0.00	0.00/0.00/0.50	42,001	3,612	2,920	2,181	1,396	633	49	0
III-E	1,000	0.25/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.25	43,001	3,706	3,011	2,210	1,385	612	30	0
III-A	2,000	0.50/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.50	79,999	5,846	4,507	2,906	1,325	135	0	0
III-B	2,000	0.00/0.00/0.50	0.00/0.25/0.00	0.00/0.00/0.25	88,000	7,695	6,458	4,826	3,091	1,528	309	0
III-C	2,000	0.50/0.25/0.00	0.00/0.50/0.25	0.25/0.00/0.50	71,998	4,702	3,102	1,558	367	0	0	0
III-D	2,000	0.25/0.00/0.00	0.00/0.25/0.00	0.00/0.00/0.50	83,999	7,217	5,822	4,339	2,765	1,234	47	0
III-E	2,000	0.25/0.00/0.00	0.00/0.50/0.00	0.00/0.00/0.25	86,000	7,410	6,010	4,405	2,750	1,203	23	0

* Population hamming distance of zero indicating the fixation of favorite alleles at three loci.
[1] Allele frequencies in first/second/third loci in population A. [2] Allele frequencies in first/second/third loci in population B.
[3] Allele frequencies in first/second/third loci in population C.

Table 3. Changes of population hamming distance over generations (1-10)* for IIII.S

Cross scheme	Population size	A1/A2/A3/A4[1]	B1/B2/B3/B4[2]	C1/C2/C3/C4[3]	D1/D2/D3/D4[4]	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
IIII.S-A	500	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	28,000	2,404	2,018	1,550	1,070	617	232	0	0	0
IIII.S-B	500	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	29,999	2,907	2,563	2,112	1,621	1,127	664	264	3	0
IIII.S-C	500	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	28,998	2,722	2,289	1,838	1,361	894	470	133	0	0
IIII.S-D	500	0.50/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.25	28,999	2,722	2,290	1,837	1,362	894	469	132	0	0
IIII.S-A	1,000	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	55,999	4,804	4,026	3,089	2,124	1,215	443	0	0	0
IIII.S-B	1,000	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	59,997	5,813	5,116	4,215	3,228	2,234	1,301	501	0	0
IIII.S-C	1,000	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	57,998	5,440	4,563	3,651	2,690	1,743	886	213	0	0
IIII.S-D	1,000	0.50/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.25	57,999	5,444	4,567	3,660	2,702	1,757	901	226	0	0

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

[1] Allele frequencies in first/second/third/fourth loci in population A. [2] Allele frequencies in first/second/third/fourth loci in population B.

[3] Allele frequencies in first/second/third/fourth loci in population C. [4] Allele frequencies in first/second/third/fourth loci in population D.

population size 500 and 1,000 at the G8, respectively. And four target genes are pyramided at the G9.

Table 4 shows the changes of population hamming distance over generations (1-10) for cascading cross program (IIII.C), population size is 500 and initial gene frequency is 0.5, the target genes pyramided at G9 (IIII.C-A). When the initial favorite allele frequency varies to 0.25, the gene pyramided at the G10 (IIII.C-B). For population size 1,000, simulations show the same results. The changes of population hamming distances are also compared base on different population sizes. If the population size are 500 and 1,000, the population hamming distance are 44 and 41 with favorite allele frequency 0.5 (IIII.C-A) at the G8. At the G9, the population hamming distances are 79 and 100 with initial favorite allele frequency 0.25, and four target genes are pyramided at the G10 (IIII.C-B) (Table 4).

For a cascading cross program, we also designed a serials of schemes with various levels of favorite allele frequencies at the four target loci, allele frequencies at two loci are 0.25, and allele frequencies at another two loci are 0.5, such as IIII.C-C, IIII.C-D and IIII.C-E, the population hamming distances are respectively 220, 262, 280 and 407, 494, 523 for population size 500 and 1,000 at G8.

For the four population symmetric cross program, it is not necessary to consider the cross parents order. But for cascading cross, we investigated the parent population given different levels of favorite allele frequencies, corresponding to the schemes IIII.C-C, IIII.C-D and IIII.C-E (Table 4). The results show that when the population size is 500 and 1,000, the population hamming distances in scheme IIII.C-E are lower than those of IIII.C-C and IIII.C-D at the first five generations. But for the subsequent generations, the population hamming distances show no significant differences.

Gene pyramiding through phenotypic selection

Many economic traits in animals are quantitative traits controlled by multiple major genes with low heritability. In addition to a genotypic selection strategy, we employed the phenotypic selection strategy based on different heritabilities of the trait, in order to compare genotypic selection with traditional phenotypic selection in gene pyramiding breeding.

The phenotypic selection strategy also includes four types of hybrid schemes. The population size is set to 500, others breeding simulation parameters are the same as those of the genotypic selection strategy. The frequency of superior genotype 11 is calculated and compared (the results of average phenotypic values and population hamming distances are not presented here).

Table 4. Changes of population hamming distance over generations (1-10)* for IIII.C

Cross scheme	Population size	A1/A2/A3/A4[1]	B1/B2/B3/B4[2]	C1/C2/C3/C4[3]	D1/D2/D3/D4[4]	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
IIII.C-A	500	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	28,001	2,447	2,090	1,649	1,187	742	343	44	0	0
IIII.C-B	500	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	30,001	2,920	2,601	2,173	1,707	1,239	788	386	79	0
IIII.C-C	500	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	29,500	2,833	2,453	2,005	1,524	1,044	597	220	0	0
IIII.C-D	500	0.50/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.25	29,250	2,792	2,397	1,965	1,507	1,052	629	262	11	0
IIII.C-E	500	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	28,499	2,655	2,266	1,866	1,466	1,054	649	280	16	0
IIII.C-A	1,000	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	56,001	4,882	4,165	3,282	2,352	1,450	639	41	0	0
IIII.C-B	1,000	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	60,001	5,839	5,196	4,340	3,399	2,447	1,532	715	100	0
IIII.C-C	1,000	0.50/0.00/0.00/0.00	0.00/0.50/0.00/0.00	0.00/0.00/0.25/0.00	0.00/0.00/0.00/0.25	59,001	5,666	4,903	4,005	3,038	2,066	1,164	407	0	0
IIII.C-D	1,000	0.50/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.25	58,501	5,581	4,783	3,912	2,992	2,078	1,227	494	3	0
IIII.C-E	1,000	0.25/0.00/0.00/0.00	0.00/0.25/0.00/0.00	0.00/0.00/0.50/0.00	0.00/0.00/0.00/0.50	57,001	5,304	4,526	3,711	2,906	2,078	1,264	523	5	0

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

[1] Allele frequencies in first/second/third/fourth loci in population A. [2] Allele frequencies in first/second/third/fourth loci in population B.

[3] Allele frequencies in first/second/third/fourth loci in population C. [4] Allele frequencies in first/second/third/fourth loci in population D.
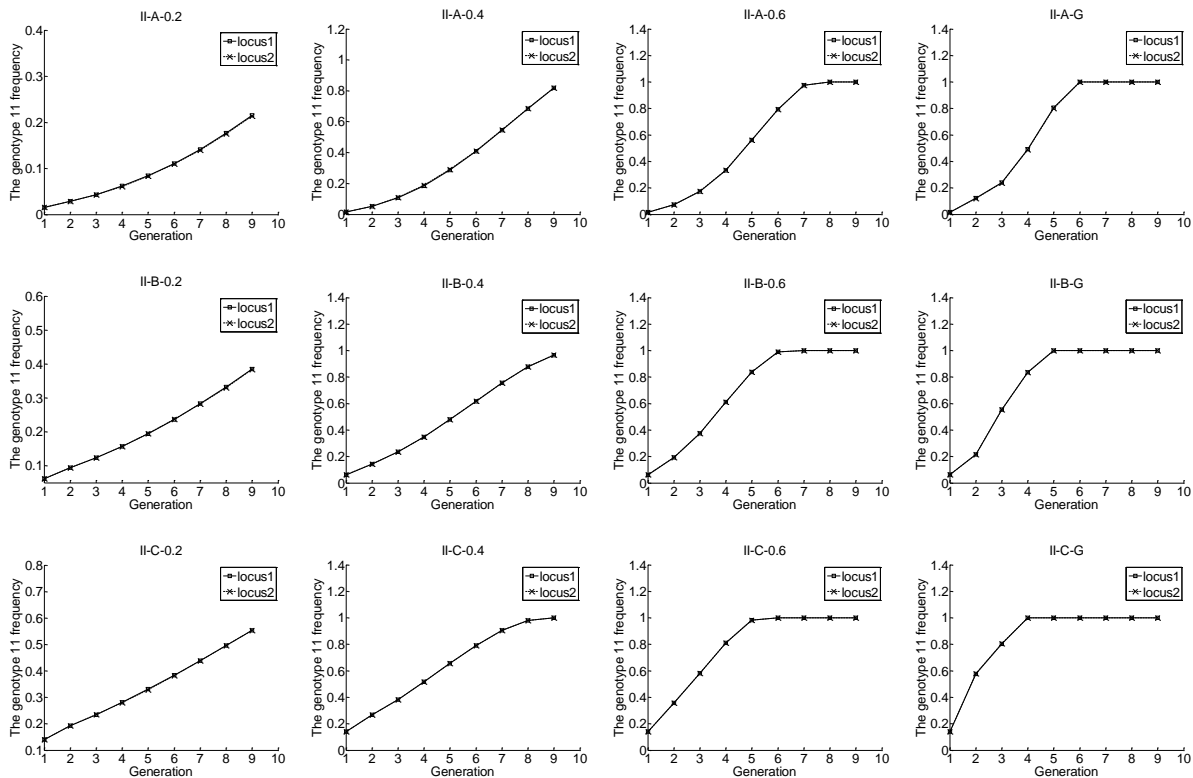
Figure 2. Genotype 11 frequencies for two populations cross. Locus1 denotes the changes of genotype 11 frequency at first target locus from popA. Locus2 denotes the changes of genotype 11 frequency at second target locus from popB. 0.2, 0.4, 0.6 represent heritability in phenotypic selection, and G represent genotypic selection. Locus1 denotes the changes of genotype 11 frequency at first target locus in popA. Locus2 denotes the changes of genotype 11 frequency at second target locus in popB. II-A, II-B, II-C represent three types of cross schemes. II-A, A1/A2[0.25/0], B1/B2[0/0.25]; II-B, A1/A2[0.5/0], B1/B2[0/0.5]; II-C, A1/A2[0.5/0.25], B1/B2[0.25/0.5].

Figure 2 shows the changes of genotype 11 frequency for a two population cross program, the initial allele frequency is set as 0.5, 0.25 or 0, respectively, and A1/A2 and B1/B2 are the favorite allele frequencies for a pair of cross parent combination. Under the same preset initial allele frequency, we supposed larger the heritability of the trait, the more quickly will average phenotypic value increase to the maximum value. As to scheme II-A, two target genes are pyramided at G8 using phenotypic selection supposing that the trait heritability is 0.6 (II-A-0.6), while at G6 using the genotypic selection (II-A-G). Considered the scheme II-B, two genes are pyramided at G6 (II-B-0.6) and G5 (II-B-G) respectively. In the genotypic selection strategy, two genes are pyramided at G5, which compared with II-B-0.6, II-B-0.4 and II-B-0.2, the average superior genotype 11 frequencies are 0.82, 0.44 and 0.41. We also designed scheme II-C with different no-zero initial allele frequencies at two loci. We set A1/A2 as 0.5/0.25, B1/B2 as 0.25/0.5, and TGPG (TGPG denotes target genotype pyramided generation) are G7 and G5 respectively for phenotypic selection given trait heritability is 0.6 and genotypic selection. Comparing three types of two population cross schemes, TGPG are G6, G5 and G4 with the using genotypic selection, and the trait heritability is 0.6,

the TGPG reduce to G8, G6, G5, respectively.

We investigated three genes which pyramided from three donor populations in four types of schemes (denotes III-A, B, C, D), and found that when both trait heritability and the initial favorite allele frequency of each locus are at lower level it is very difficult for three target genes to fix at G10, such as in schemes III-A-0.2, III-B-0.2, III-A-0.4, and III-B-0.4 (Figure 3). TGPGs are G9 (III-A-0.6), G8 (III-B-0.6), and G7 (III-C-0.6) using phenotypic selection with trait heritability 0.6, while TGPGs are G8 (III-A-G), G7 (III-B-G), and G6 (III-C-G) using genotypic selection.

We compared the two schemes III-C (A1/A2/A3(0.25/0/0), B1/B2/B3(0/0.25/0), C1/C2/C3(0/0/0.5)) and III-D ((A1/A2/A3(0.25/0/0), B1/B2/B3(0/0.5/0), C1/C2/C3(0/0/0.25)) (Figure 3). The results show that the breeding progress with the higher favorite allele frequency 0.5 in the third cross population as similar to that of allele frequency 0.25 (III-C and III-D), and the results also show that the genotype 11 at the first locus from popA and the second locus from popB share the same increasing trend, and genotype frequency 11 at the third locus is higher than those of the first two loci. Moreover, with the increase of initial favorite allele frequencies at all three loci, the aim of the gene pyramiding is achieved at the earlier generations.
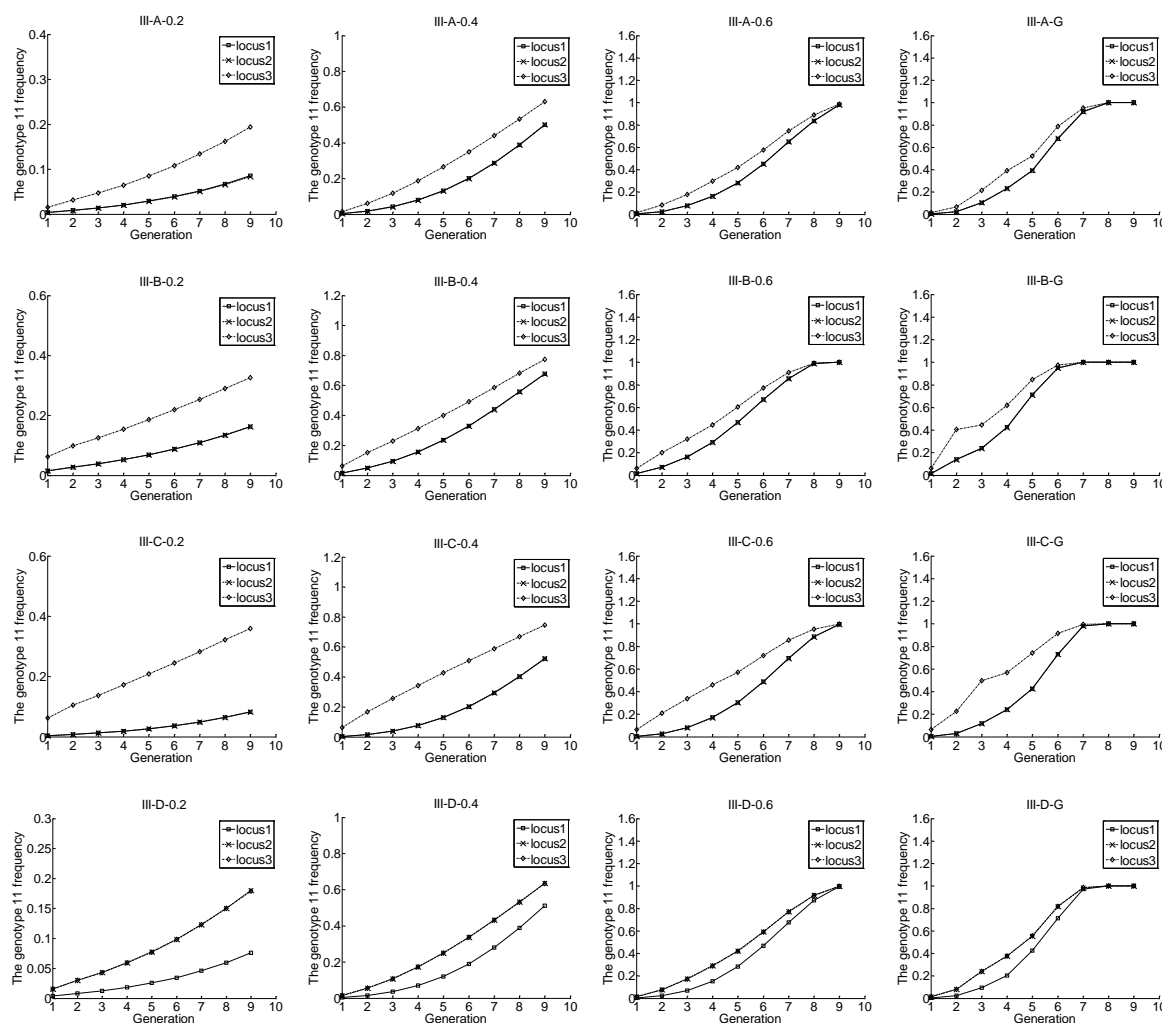
Figure 3. Genotype 11 frequencies for three populations cascading cross. 0.2, 0.4, 0.6 represent heritability in phenotypic selection strategies, and G represents genotypic selection. Locus1 denotes the changes of genotype 11 frequency at first target locus in popA. Locus2 denotes the changes of genotype 11 frequency at second target locus in popB. Locus3 denotes the changes of genotype 11 frequency at third target locus in popC. III-A, III-B, III-C and III-D represents four types of cross schemes. III-A, A1/A2/A3[0.25/0/0], B1/B2/B3[0/0.25/0], C1/C2/C3[0/0/0.25]; III-B, A1/A2/A3 [0.5/0/0], B1/B2/B3[0/0.5/0], C1/C2/C3 [0/0/0.5]; III-C, A1/A2/A3 [0.25/0/0], B1/B2/B3[0/0.25/0], C1/C2/C3 [0/0/0.5]; III-D, A1/A2/A3[0.25/0/0], B1/B2/B3[0/0.5/0], C1/C2/C3 [0/0/0.25].

Two cross programs (cascading and symmetric) are investigated for four genes pyramiding in our study. We compared schemes IIII-C-A-0.2, IIII.C-A-0.4, IIII.C-A-0.6, and IIII.C-A-G with IIII.S-A-0.2, IIII.C-S-0.4, IIII.S-A-0.6, and IIII.S-A-G, compared IIII.C-B-0.2, IIII.C-B-0.4, IIII.C-B-0.6, and IIII.C-B-G with IIII.S-B-0.2, IIII.S-B-0.4, IIII.S-B-0.6, and IIII.S-B-G, and also compared IIII-C-C-0.2, IIII.C-C-0.4, IIII.C-C-0.6, and IIII.C-C-G with IIII.S-C-0.2, IIII.S-C-0.4, IIII.S-C-0.6, and IIII.S-C-G (Figure 4, Figure 5). The results show that cascading cross and symmetric cross have no significant difference in the gene pyramiding under certain conditions. The four target genes are pyramided at a similar generation. Comparing schemes IIII.C-A-G, IIII.C-B-G and IIII.C-C-G with IIII.S-A-G, IIII.S-B-G and IIII.S-C-G, it shows that the TGPG are G9, G10, G9 and G8, G9, G9. Under the same condition, the

symmetric cross program was found to be slightly superior to the cascading cross program.

As to the symmetric cross program, the genotype 11 frequencies in popA and popB share the consistent increasing trend under the phenotypic selection strategy, so does the popC and popD (Figure 5). But for the cascading cross program, the popC and popD are taken as the third and the fourth cross population, and the genotype 11 frequencies of the third and the fourth locus are higher than those of the first and the second locus (Figure 4). Our results show that cross order has slight influence on the cascading cross. When the third and the fourth cross population are given the higher favorite allele frequency, as to lower heritability, the superior genotype 11 frequency is higher than the population with lower favorite allele frequency. Scheme IIII.C-E is slightly superior to IIII.C-D
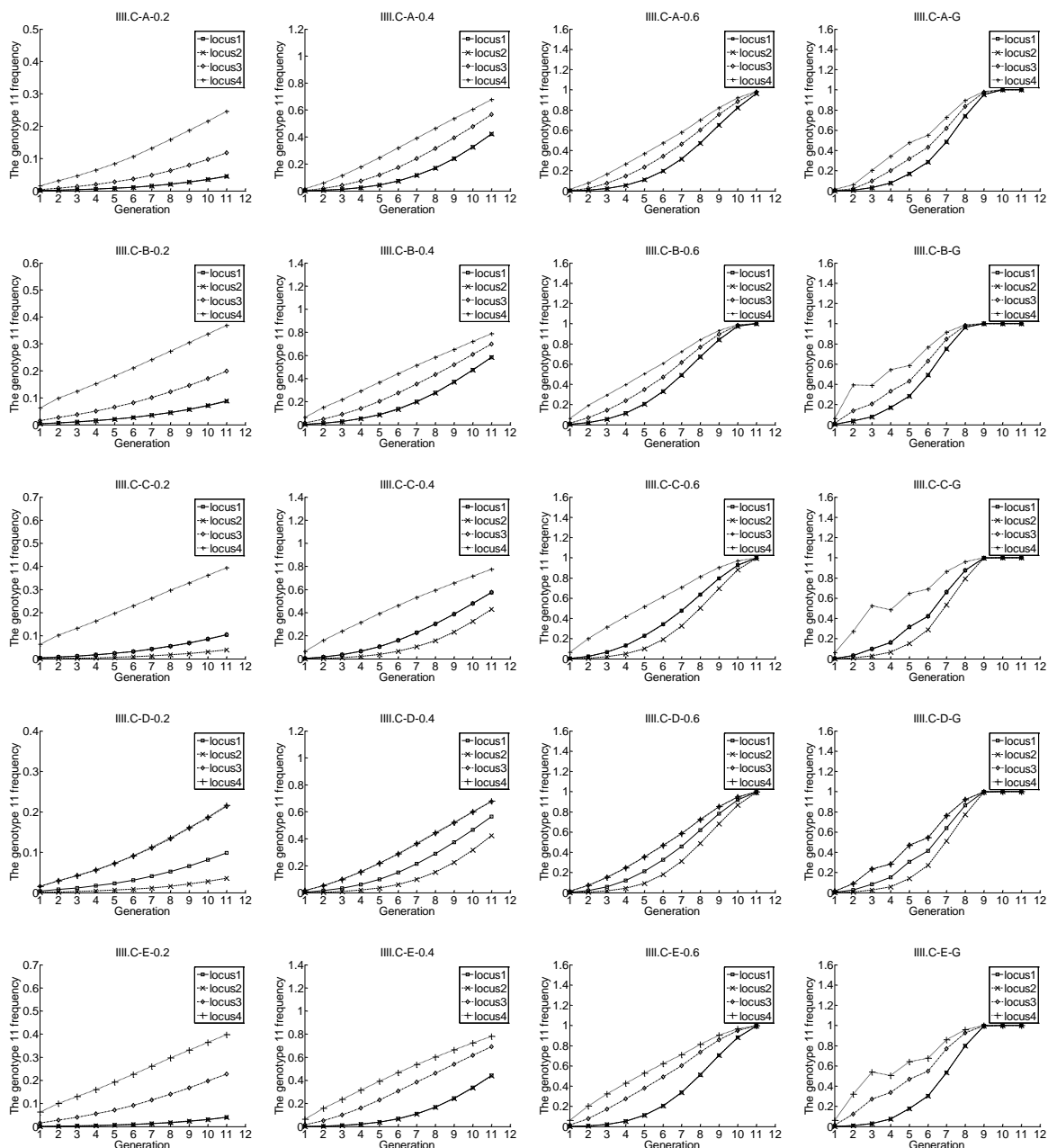
Figure 4. Genotype 11 frequencies for four populations cascading cross. IIII.C-(A-E) represents five types of schemes. Locus1 denotes the changes of genotype 11 frequency at first target locus from popA. Locus2 denotes the changes of genotype 11 frequency at second target locus from popB. Locus3 denotes the changes of genotype 11 frequency at third target locus from popC. Locus4 denotes the changes of genotype 11 frequency at third target locus from popD. IIII.C-A, A1/A2/A3/A4[0.25/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.25/0], D1/D2/D3/D4[0/0/0/0.25]; IIII.C-B, A1/A2/A3/A4[0.5/0/0/0], B1/B2/B3/B4[0/0.5/0/0], C1/C2/C3/C4[0/0/0/0.5], D1/D2/D3/D4 [0/0/0/0.5]; IIII.C-C, A1/A2/A3/A4[0.5/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.25/0], D1/D2/D3/D4 [0/0/0/0.5]; IIII.C-D, A1/A2/A3/A4[0.5/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.5/0], D1/D2/D3/D4 [0/0/0/0.25]; IIII.C-E, A1/A2/A3/A4[0.25/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.5/0], D1/D2/D3/D4 [0/0/0/0.5].

and IIII.C-D, but as to high heritability, the three schemes seem to have no significant differences.

Average phenotypic progress for genotypic and phenotypic selection strategies

Table 5 shows the average phenotypic progress using

genotypic selection and phenotypic selection. In the case of the population size of being 500, we first used genotypic selection to get the gene pyramiding generation G(t). At generation t, we investigated the average phenotypic progress using phenotypic selection given the trait with different heritabilities, the average phenotypic progress is
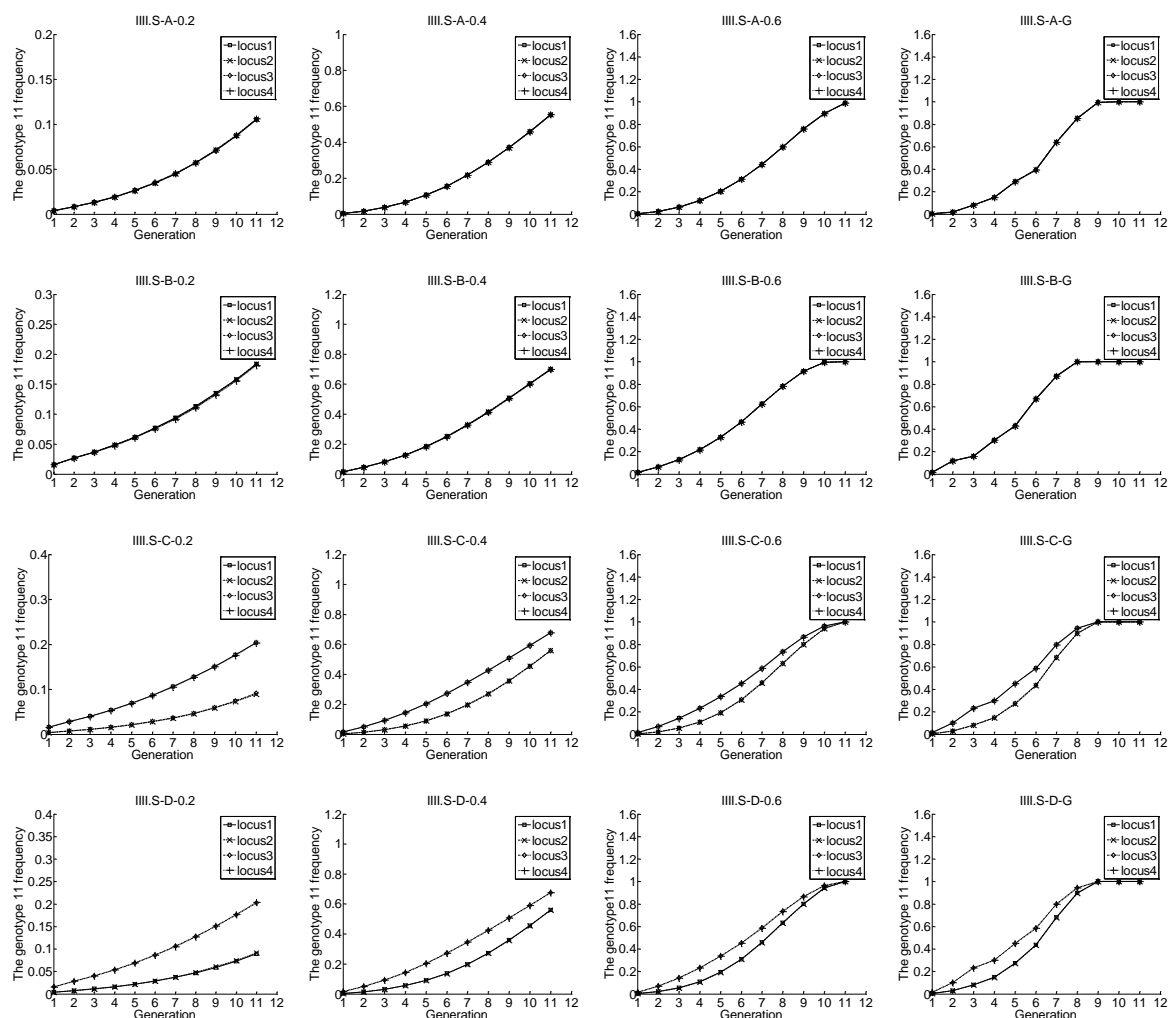
Figure 5. Genotype 11 frequencies for four populations symmetric cross. IIII-S-(A-D) represents four types of schemes. IIII-S-A, A1/A2/A3/A4[0.25/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.25/0], D1/D2/D3/D4[0/0/0/0.25]; IIII-S-B, A1/A2/A3/A4[0.5/0/0/0], B1/B2/B3/B4[0/0.5/0/0], C1/C2/C3/C4[0/0/0/0.5], D1/D2/D3/D4[0/0/0/0.5]; IIII-S-C, A1/A2/A3/A4[0.5/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.25/0], D1/D2/D3/D4[0/0/0/0.5]; IIII-S-D, A1/A2/A3/A4[0.25/0/0/0], B1/B2/B3/B4[0/0.25/0/0], C1/C2/C3/C4[0/0/0.5/0], D1/D2/D3/D4[0/0/0/0.5].

calculated by $(p(t)-p(1))/t$, where $p(t)$ denotes the average phenotype value at the generation t, and $p(1)$ denotes the average phenotype value at the generation 1. In the cross programs II, III and IIII, genotypic selection strategy is superior to phenotypic selection in accelerating gene pyramiding. The trait with lower heritability is more appropriate for using genotypic selection to pyramid target genes (Table 5). The phenotypic selection strategy for heritability 0.6 is the same results with genotypic selection strategy. Comparing the scheme IIII-C with IIII-S, the results of G(t) and average phenotypic progress show that IIII-S is superior to IIII-C. Our simulation also investigates influences of cross order on the schemes in cascading cross via calculating the value of average phenotypic progress, and the scheme IIII.C-C is slightly superior to IIII.C-D and IIII.C-E.

DISCUSSION

Our studies provide a new insight into the pyramiding of multiple genes into a single genotype from evolutionary perspectives. The objective of gene pyramiding breeding is to improve the trait for an entire population by selecting the most optimal genotype combinations. Evolutionary computation (David, 1989; John, 1992) is most appropriate for studying the combinatorial optimization of genotypes. As for gene pyramiding breeding, we assumed a complex trait was controlled by a series of major genes, and gene pyramiding aimed to select individuals with the optimal genotype combination to realize the optimization of a target economic trait. Inspired by the science of evolutionary computation (David, 1989), we used the metaphor of hill-climbing to model the dynamic behavior of gene pyramiding and to build the connection between gene

Table 5. Compare average phenotypic progress using phenotypic selection and genotypic selection

Cross scheme	Generation (t) t	Phenotype selection			Genotype selection
		0.2	0.4	0.6	
II-A	7*	0.34[1]	0.72[2]	0.88[3]	0.87[4]
II-B	6	0.34	0.67	0.82	0.94
II-C	5	0.31	0.58	0.73	0.81
III-A	9	0.43	0.92	1.12	1.17
III-B	8	0.51	0.98	1.14	1.15
III-C	9	0.42	0.88	1.07	1.10
III-D	9	0.44	0.92	1.10	1.13
IIII-C-A	11	0.46	1.04	1.27	1.32
IIII-C-B	10	0.49	1.04	1.27	1.32
IIII-C-C	10	0.46	1.03	1.29	1.37
IIII-C-D	11	0.49	1.03	1.24	1.27
IIII-C-E	11	0.46	0.99	1.20	1.23
IIII-S-A	11	0.49	1.06	1.28	1.32
IIII-S-B	9	0.52	1.10	1.36	1.46
IIII-S-C	10	0.50	1.07	1.07	1.38
IIII-S-D	10	0.50	1.31	1.32	1.38

* The generation gene pyramided at using genotypic selection.

[1] The average phenotypic progress over t generations using phenotypic selection with trait heritability 0.2.

[2] The average phenotypic progress over t generations using phenotypic selection with trait heritability 0.4.

[3] The average phenotypic progress over t generations using phenotypic selection with trait heritability 0.6.

[4] The average phenotypic progress over t generations using genotypic selection.

The average phenotypic progress calculated by $(p(t)-p(1))/t$. $p(t)$ denotes the average phenotype value at the generation t, and $p(1)$ denotes the average phenotype value at the generation 1.

pyramiding and evolutionary computation.

Servin et al. (2004) designed the algorithm for the theory of marker-assisted gene pyramiding based on probability and statistics. They calculated gene transmission probabilities through a pedigree and minimum population sizes necessary to obtain the individual with the ideal genotype. Zhao et al. (2009) extended these theories to design some representative gene pyramiding schemes in animals by taking their reproductive capacity into account. However, their studies made some simplifying assumptions that the genotype of founding parents was homozygous for the favorable allele at each target locus. The assumptions are suitable for laboratory animals rather than farm animals. In practice, animal breeding populations are segregating populations. Therefore, our studies start the base population with various levels of favorite allele frequencies at each target locus. Allele frequencies are set to be 0, 0.25, 0.5 to represent zero, low and medium allele frequency levels in the base population, and it is possible to study gene pyramiding from an arbitrary population given the variable allele frequencies and population sizes.

Servin et al. (2004) and Zhao et al. (2009) described their framework for the design of gene pyramiding by computing the minimum population sizes necessary to obtain the ideal single genotype. The design of these strategies is from an ideal genotype of offspring to minimum population sizes of the base population. From the opposite perspective, our studies predict the offspring genotype by simulating the process of gene pyramiding breeding, given the specialized base populations. Our strategies can be used to integrate various populations (including population size and favorite allele frequency) and different selection strategies.

In comparison with plants, the difficulties in conducting gene pyramiding in animals come from the lower fertility and longer generation intervals. With the development of animal genome projects and new reproduction technologies (artificial insemination and super ovulation), it is possible to produce a large enough number of offspring carrying superior genetic information in each generation to facilitate the selection of subsequent generations. For the sake of demonstration, our studies use discrete recombination to produce offspring with the various recombination types possible for gene pyramiding studies from parental genotypes. Discrete recombination is the basic genetic operator in evolutionary computation; therefore, it is used for the studies of gene pyramiding in order to keep consistency with evolutionary computation.

In order to investigate two-genes, three-genes, four-genes pyramiding, we designed four types of cross programs, II, III, IIII-S and IIII-C, which may represent the general demand in farm animal breeding. There are two

target genes segregating in the population for program II, three target genes segregating for program III. As to program IIII-S and IIII-C, there are four target genes segregating in the population.

Using genotypic selection, the results produced from the simulation of four types of gene pyramiding breeding programs indicate that initial favorite allele frequencies are the most important factor affecting the process of gene pyramiding, rather than the population size, but the larger population size increases the possibility of selecting top individuals as parents at the first generation. As for the two-genes and three-genes pyramiding, initial allele frequency and population size do not have a significant influence on the schemes design of gene pyramiding, but for three gene and four-genes pyramiding, the hybrid parents order must be considered in our schemes design. In four genes pyramiding, our studies show that three generation needed to gain the popABCD (Figure 1c), and only two generations needed using the symmetric cross programs (Figure 1d). For symmetric cross program, it was not necessary to consider the cross order because of the particularly symmetric cross structure. But in a cascading cross program, parent cross order is shown to be not very important factor affecting the gene pyramiding breeding.

In addition to genotypic selection strategy, we also investigated the phenotypic selection strategy as many economic traits of animals are quantitative traits, controlled by several major QTL. The difference between the phenotypic and genotypic selection is selection criterion, genotypic selection based on genotypic score and phenotypic selection based on phenotypic value predicted from a genotype -phenotype model. We use two selection strategies in the consideration of different character of target genes and the trait heritability.

Some geneticists think that traditional mass selection strategies also results in gene pyramiding. Phenotypic selection strategy is used to investigate a target gene controlling a quantitative trait, and moreover, we compare the gene process of gene pyramiding using genotypic selection and phenotypic selection. Initial favorite allele frequencies greatly affect the process of gene pyramiding breeding using phenotypic selection, and another important factor is the trait heritability. From the Figure 2, 3, 4 and 5, we can conclude that for trait with high heritability, gene pyramiding breeding using a phenotypic selection strategy needs fewer generations, while more generations are needed when considering a low heritability trait. In order to achieve gene pyramiding successfully, a breeder should select from a large size base population with high favorite allele frequencies. In phenotypic selection, we set trait heritability to 1, which is equivalent to genotypic selection derived from the formula (3). The results indicate that genotypic selection is superior for gene pyramiding than phenotypic

selection. Design of a cross scheme should concern the initial favorite allele frequency, cross order and the trait heritability. Trait heritability is the main factor affecting the effective gene pyramiding breeding for the quantitative traits. When the genotypic value is preset, trait heritability would have a direct impact on the average phenotypic value predicted by the model and would finally affect the process of gene pyramiding. As to the trait with a larger heritability, the dominant components in the model are the gene effects, so gene pyramiding breeding would be a process of selecting individuals with the optimized genotype combination over generations.

In this paper, genotypic selection and phenotypic selection ignored gene-gene interactions and gene-environment interactions. The current strategies for revealing the genetic basis of complex traits are to carry out a genome wide association studies (Wang et al., 2005; McCarthy et al., 2008; Moore et al., 2010), which would supply us with a amount of genetic information and finally help us to build the precise selection model considering the complex relationship between genotype and phenotype.

The limitation of gene pyramiding in animals is due to the generation intervals and reproductive capability, especially to animals (dairy or beef cattle) with the long generation intervals and low fertility. In our studies, we suppose the potential advantages of gene pyramiding can be applied to any farm animal, but from a practical point of view it may be a challenge.

Our studies made some simplifying assumptions that the animal population is a segregating population and there exist several favorable target genes in different populations. If the multi-tier system (population) meets these assumptions in our studies, we can predict the process of gene pyramiding considering different strategies. Our studies did not take in consideration the positions of most genes, because the location of those genes can be detected through PCR technology. Some examples of gene pyramiding successfully applied can be found in plant breeding. In practice, the position of most genes may be not the key point, how to chose the target gene or linked markers and how to perform selection are of greater importantance.

Our studies provide a flexible simulation platform for exploring gene pyramiding breeding using genotypic selection and phenotypic selection. Base population sizes and the initial favorite allele frequencies can be set at various levels. The results presented by population hamming distance, superior allele frequency and average phenotypic value would provide some theoretical reference for the breeding practice. Further studies can be conducted to build and compare different cross programs and selection strategies.

As to marker-assisted gene pyramiding breeding, how to

design the optimal genotype combinations through different cross schemes and selection strategies would have great practical significance. Animal breeders will be eager to design the optimal cross scheme and selection strategy. We hope that breeding by design would be realized through the collaboration of biologists, bioinformatics and breeding scientists with the aid of powerful computer technology and user-friendly software.

REFERENCES

Chen, M. and C. Kendziorski. 2007. A statistical framework for expression quantitative trait loci mapping. Genetics 177:761-771.

David, E. G. 1989. Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc.

Doerge, R. W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3:43-52.

Fadiel, A., I. Anidi and K. D. Eichenbaum. 2005. Farm animal genomics and informatics: an update. Nucleic Acids Res. 33: 6308-6318.

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

Huang, N., A. E. Domingo, J. Magpantay, G. S. Singh, G. Zhang, N. Kumaravadivel, J. Bennett and G. S. Khush. 1997. Pyramiding of bacterial blight resistance genes in rice, marker-assisted selection using RFLP and PCR. Theor. Appl. Genet. 95:313-320.

John, H. H. 1992. Adaptation in natural and artificial systems. MIT Press.

Kameswara Rao, K., N. Lakshmi, M. Jena and K. Kshirod. 2010. Effective strategy for pyramiding three bacterial blight resistance genes into fine grain rice cultivar, Samba Mahsuri, using sequence tagged site markers. Springer, Heidelberg, ALLEMAGNE.

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.

Ljungberg, K., S. Holmgren and O. Carlborg. 2002. Efficient algorithms for quantitative trait loci mapping problems. J. Comput. Biol. 9:793-804.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis and Joel N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9:356-369.

Moore, J. H., F. W. Asselbergs and S. M. Williams. 2010. Bioinformatics challenges for genome-wide association studies. Bioinformatics 26:445-455.

Moreau, L., A. Charcosset, F. Hospital and A. Gallais. 1998. Marker-assisted selection efficiency in populations of finite size. Genetics 148:1353-1365.

Muhlenbein, H. and D. S. Voosen. 1993. Predictive models for the breeder genetic algorithm. Evol. Comput. 1:25-49.

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

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

Ruane, J. and J. J. Colleau. 1995. Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. Genet. Res. 66:71-83.

Saghai Marrof, M. A., J. S. 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.

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

Singh, S., J. S. Sidhu, N. Huang, Y. Vikal, Z. Li, D. S. Brar, H. S. Dhaliwal and G. S. Khush. 2001. Pyramiding three bacterial blight resistance genes (xa5, xa13 and Xa21) using marker-assisted selection into indica rice cultivar PR106. TAG Theor. Appl. Genet. 102:1011-1015.

Wang, W. Y., B. J. Barratt, D. G. Clayton and J. A. Todd. 2005. Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6:109-118.

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.