

STUDY ON FACTORS AFFECTING THE EFFICIENCY OF MARKER-ASSISTED INTROGRESSION

J. Y. Bai

College of Animal Science and Technology, Henan University of Science and Technology,
70, Tianjing Road, Luoyang, China. Postcode, 471003,
Corresponding Author e-mail: byezi@163.com

ABSTRACT

In this paper we simulate the breeding process of marker-assisted introgressing a favorable QTL allele from “donor” to “recipient”. During this process, the foreground selection and background selection were made for introgression population simultaneously: foreground selection is making indirect selection of target gene by its closely linked two flanking markers, four selection methods including random selection, genomic similarity selection, index selection and MBLUP selection are used in background selection, and the MBLUP was first used in background selection. The effect of different factors on the efficiency of marker-assisted introgression a QTL was simulated in this study, the factors include different background selection methods (random selection, genomic similarity selection, index selection and MBLUP selection), population sizes (500 and 1000), ratios of QTL variance (10% and 30%). The following results were obtained: (1) Genomic similarity selection and index selection can make recipient genome recover entirely through 3~4 backcross generations, though MBLUP selection could not make recipient genome recover quickly, it can get the most genetic responses for background traits. (2) Enlarging population size can increase frequency of favourable allele for introgressing QTL and genetic responses for background traits. (3) Enlarging ratio of QTL variance can increase both frequency of favourable allele for background QTL and genetic responses for background traits.

Key words: Marker-assisted introgression, Foreground selection, Background selection, MBLUP.

INTRODUCTION

In animal breeding, we often meet such problem: A breed or line with many superior traits while at the same time it exist some objection, so influence its economic value. For example, Landrace or Large White were feeded worldwide for their fast growing, high feed use efficiency, high muscle ratio and so on, while their meat quality is relatively poor, contrary, many native pig breeds of China possess good meat quality, but other traits are inferior. So if we discover some genes related to meat quality of native pig breeds of China, we could introgress them to Landrace or Large White, improving their meat quality. Marker-assisted introgression is the introgression of some valuable genes from “donor” line to “recipient” line with the information of molecular marker linked to these genes. The marker could be used: i) to control the target gene (foreground selection), ii) to control the genetic background (background selection) (Hospital 2003), restoring the recipient’s genetic background quickly, and eliminating unfavorable gene linked to target gene effectively.

In recent years, DNA molecular markers were used to mapping many major genes or quantitative traits locis related to important economic traits of livestock by candidate gene approach and genome scan approach (Andersson 2001; Chaiwong *et al.*, 2002), these genes or markers closely linked with them then can be used to

resolve some challenge that traditional breeding method can not resolve. With the more profound research of animal functional genome, more and more economics important genes are being found, which give breeders many opportunities, one of the prospective opportunities is to foster ideal breed (Dekkers 2004; Dekkers and Hospital, 2002). There are some successful examples, marker-assisted introgression halothane gene in pigs (Hanset *et al.*, 1995), marker-assisted introgression naked neck gene in chickens (Yancovich *et al.*, 1996), marker-assisted introgression booroola FecB gene in sheep (Gootwine *et al.*, 1998).

From the last period of 1980s, many simulate studies have been done on marker-assisted introgression, mainly concentrate on background selection (Hillel *et al.*, 1990; Hospital *et al.*, 1992, 1997, 2003; Groen and Smith 1995; Visscher *et al.*, 1996; Chaiwong *et al.*, 2002). In these studies, recipient’s genetic background needed restore completely, assuming that two alleles of marker fixed in donor and recipient respectively. These studies indicate with molecular marker, recipient’s genetic background can restore quickly as well as guarantee introgressing donor gene effectively. But to restore recipient’s genetic background completely, all the markers dispersing through genome are needed, this increasing the cost, in practical breeding, the genetic background of recipient need not restore entirely, but some certain background traits need restore, the markers

closely linked with QTLs related to these traits and the phenotypic information can be used to make MBLUP background selection so that the genetic background could restore.

The aim of this paper is : (1) the first used MBLUP for genetic background restoration, analyze the effect of different background selection methods on marker-assisted introgression efficiency, (2) analyze the effect of different population sizes on marker-assisted introgression efficiency, (3) analyze the effect of different ratios of QTL variance to total genetic variance on marker-assisted introgression efficiency, (4) discuss the suitable marker-assisted introgression scheme.

MATERIALS AND METHODS

Population structure: Supposing that the original donor population is composed of 10 males, the original recipient population is composed of 10 males and 200 females. Half of the recipient population females were mated to donor population males to produce F1 individuals, the other half were mated to males of the same population to maintain recipient population. During backcross phase, 10 males were selected from introgression population by foreground selection and background selection, all of which were mated to 100 females random selected from recipient population to produce next generation. At the same time males and females of recipient population were random selected to mate each other to produce next generation. During intercross phase, 10 males and 100 females were selected from introgression population by foreground selection and background selection to mate each other for next generation. During the whole breeding process, half-sibs and full-sibs mating would be avoided, the sex of the offspring is assigned at random with a probability of either sex being 0.5, all of generations don't overlap each other.

QTLs and markers genotype : Three traits were researched, each of which were affected by a QTL and polygene respectively, each of the three corresponding QTLs has two alleles (Q and q), Q was favourable allele, q was unfavourable allele, there were no dominant effect between them. Assuming that in original donor line, the genotype of first QTL was QQ, the remaining two QTLs were qq, and adverse as in original recipient line. We need introgress Q allele of first QTL from donor line to recipient line, the trait affected by first QTL was named foreground trait, the two traits affected by the other two QTLs respectively were named background traits. The three QTLs located on 1 chromosome (55cM), 2 chromosome (45cM), 3 chromosome (65cM), assuming the whole genome is composed of 10 chromosomes, 6 markers dispersed in each of the chromosomes evenly, map distances between markers was 20cM, each marker

locus has two alleles, which were fixed in original donor line and recipient line respectively.

The phenotypic values of trait: Phenotypic value y come from model: $y = v + u + e$, v is effect of QTL, u is the polygenic effect, e is random residual effect. For individuals of original population, u sampled from a standard normal distribution $u \sim N(0, \sigma_u^2)$, for progeny individuals, $u = 0.5u_s + 0.5u_d + m$, u_s and u_d stand for paternal and maternal polygenic effect respectively, m stand for deviation due to medelian sampling, which comply with standard normal distribution $m \sim N(0, (\sigma_u^2/2)(1 + (F_s + F_d)/2))$, F_s and F_d stand for inbreeding coefficients of the father and mother. In all the generations, e generated from $N(0, \sigma_e^2)$.

Foreground selection method: Foreground selection is to select individuals carrying introgressed QTL allele from introgression population, assuming the QTL unknown, making indirection selection by its double-flanked markers, the map distance between the two markers and introgressed QTL is 5cM and 15cM. During backcross phase, selected individuals of which two markers are both heterozygosity, during intercrossing phase, selected individuals of which two markers are both homozygosity or heterozygosity.

Background selection methods:

- (1) *Random selection:* Random selecting individuals without making background selection.
- (2) *Genomic similarity selection:* Making selection due to ratio of the number of markers homozygous for the recipient allele over the total number of markers (Groen and Smith, 1995).
- (3) *Index selection:* Making selection due to index derived from combining phenotypes and marker information (Visscher *et al.*, 1996). $I_m = \sum_{i=1}^m a_i v_i$, $I_m < b_p P$,

where I_m is marker score, a_i is weight of i th marker, the weight of marker on both end of chromosome is 0.5, while is 1 on the other locus of chromosome. v_i is value of i th marker, when two alleles of the marker come from recipient line, value is 1, otherwise value is 0. P is phenotypic value of trait, b_m and b_p the weights given to the marker score and phenotypic value, defined as: $b_m = (1 - h^2)$, $b_p = h^2(1 - p)$, h^2 is trait heritability, P is the proportion of QTL contribution in genetic variance.

(4) *MBLUP selection :* Integrating phenotype information and QTL information of background traits

(by linked marker) to estimate individual's breeding values of two background traits respectively, making two estimated breeding values standard, through equal weight of weights, become breeding index, select due to the index. The model of MBLUP: $y = N W v + Z u + e$, where y is vector of trait phenotype value, v is random vector of QTL allelic effects with mean 0 and variance-covariance matrix $G \uparrow_v^2$ (G is the gametic relationship matrix at the QTL, \uparrow_v^2 is the QTL allelic variance), u is the random vector of breeding values with mean 0 and variance-covariance matrix $A \uparrow_u^2$ (A is the additive genetic relationship matrix, \uparrow_u^2 is the additive genetic variance), e is vector of residual effects with mean 0 and variance-covariance matrix $I \uparrow_e^2$ (I is a vector of one's), W and Z are the incidence matrices of v and u . The mixed model equations are

$$\begin{array}{ccc} Z \uparrow_e^2 < k_1 A^{-1} & Z W & \hat{u} \\ W \uparrow_e^2 & W W < k_2 G^{-1} & \hat{v} \end{array} \quad \begin{array}{c} N \\ N \end{array} \quad \begin{array}{c} Z y \\ W y \end{array}$$

where k_1 and k_2 are equal to $\uparrow_e^2 / \uparrow_u^2$ and $\uparrow_e^2 / \uparrow_v^2$, respectively.

The method of Wang *et al.*, (1995) and Liu *et al.*, (2001) used to calculate G^{-1} directly, the equations are solved through the algorithm of iteration on data and the convergence criterion 10D-9.

Parameters setting: Assuming that the phenotype variance of foreground trait and two background traits is 20, 50, 10 respectively, the heritability of three traits is 0.50, 0.30, 0.40. Background selection methods: random selection, genomic similarity selection, index selection and MBLUP selection. Population sizes: 500 and 1000. Ratios of QTL variance to total genetic variance: 10% and 30%. For each parameters combination, 30 replications of simulation were carried out

RESULTS

Frequency of favourable allele for introgressing QTL:

Different background selection methods have little effect on frequency of favourable allele for introgressing QTL, so the change of frequency in different generations under different population sizes was given only by MBLUP selection method (figure 1).

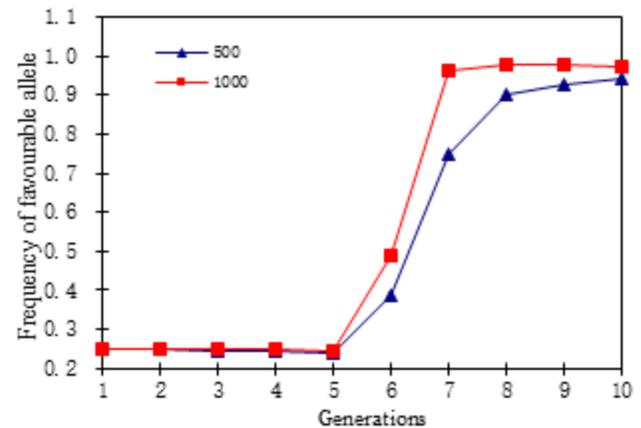


Figure 1. Frequency of favourable allele for introgressing QTL with respect to different population sizes

Note: From 1 to 5 generations is backcross phase, from 6 to 10 generations is intercross phase

It clearly show that during backcross phase, frequency of favourable allele for introgressing QTL keep at 0.25, and there is no difference between two population sizes. During intercross phase, the frequency of favourable allele for introgressing QTL show ascending trend with the intercross generation's increasing, and there is significant difference between two population sizes, intercrossing two generations, frequency of favourable allele for introgressing QTL was 0.7476 and 0.9628 with population size was 500 and 1000 respectively, when intercross 5 generations, frequency of favourable allele for introgressing QTL under two population sizes reached 0.9444 and 0.9748. So during intercross phase, enlarging population size can increase frequency of favourable allele for introgressing QTL.

Restoration of genetic background: Hospital 2003 used ratio of homozygous markers (two alleles both from recipient) to all markers to show restoration of recipient genetic background, unlike Hospital, we used percentage of alleles inherited from the recipient parent to estimate restoration of recipient genetic background, the restoration status of recipient genetic background under different background selection methods show in figure 2: backcrossing two generations, the restoration degree of recipient genetic background reached 93% by genomic similarity selection and index selection, backcrossing four generations, restoration degree of recipient genetic background reached 100%. Random selection and MBLUP selection show similar result, backcrossing two generations make restoration degree of recipient genetic background reached 87%, backcrossing four generations, it reached 96%. So for restoration degree of recipient genetic background, genomic similarity selection and

index selection get better effect than random selection and MBLUP selection.

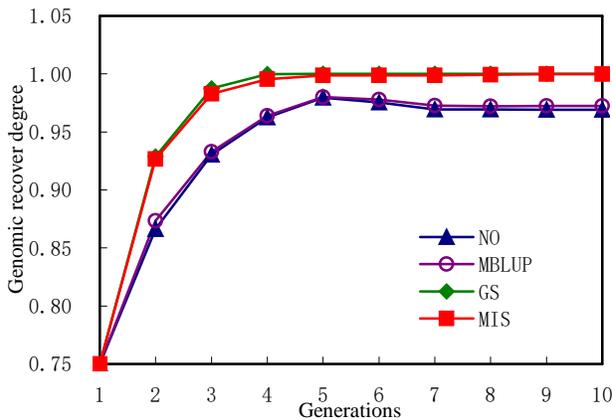


Figure 2. Genomic recover degree with respect to different background selection methods

Note: NO: random selection, MBLUP: MBLUP selection, GS: genomic similarity selection, MIS: index selection

Frequency of favourable allele for background QTLs under different background selection methods:

There exists difference in frequency of favourable allele for background QTLs among different selection methods (figure3): during backcross phase, frequency of favourable allele for two background QTLs show increase by degrees trend with the increase of backcross generations, backcrossing two generations, the frequency of favourable allele for two background QTLs under random selection, MBLUP selection, genomic similarity selection and index selection methods were 0.86, 0.92, 0.93 and 0.95 respectively, backcrossing five generations, under the same four selection methods, frequency of favourable allele for two background QTLs rise to 0.98, 1.00, 1.00 and 1.00 respectively. During intercross phase, frequency of favourable allele for two background QTLs basically keep constant. In all, frequency of favourable allele for two background QTLs show increase by degrees trend with the random selection, MBLUP selection, genomic similarity selection and index selection methods ordinal during backcross phase.

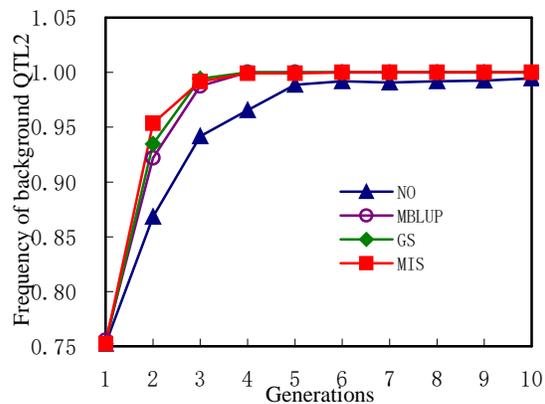
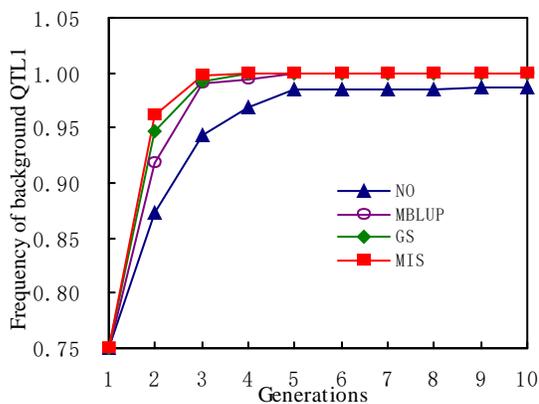


Figure 3. Frequency of favourable allele for background QTLs with respect to different background selection methods.

Frequency of favourable allele for background QTLs under different ratios of QTL variance:

MBLUP was background selection method, the result show in figure 4: during backcross phase, frequency of favourable allele for two background QTLs differ greatly among different ratios of QTL variance, backcrossing two generations, frequency of favourable allele for two background QTLs were 0.88 and 0.92 with the ratios of QTL variance, being 10% and 30% respectively, backcrossing five generations, the frequency rise to 0.98 and 1.00 under the same corresponding ratios of QTL variance. During intercross phase, frequency of favourable allele for two background QTLs basically keep constant. So enlarging ratios of QTL variance can improve frequency of favourable allele for two background QTLs.

Frequency of favourable allele for background QTLs under different population sizes:

MBLUP was background selection method, the figure 5 show that during backcross phase, the difference of frequency of favourable allele for background QTLs is a little between two population sizes, backcrossing three generations, under two different population sizes: 500 and 1000, frequencies of favourable allele for the first background QTL were 0.9858 and 0.9912, and they were 0.9868 and 0.9876 for the second background QTL, backcrossing five generations, under two population sizes, frequencies of favourable allele for two background QTLs both reached 1. During intercross phase, frequency of favourable allele for two background QTLs keep constant. So during backcross phase, frequency of favourable allele

for two background QTLs shows a little increase with the

population size's enlarging.

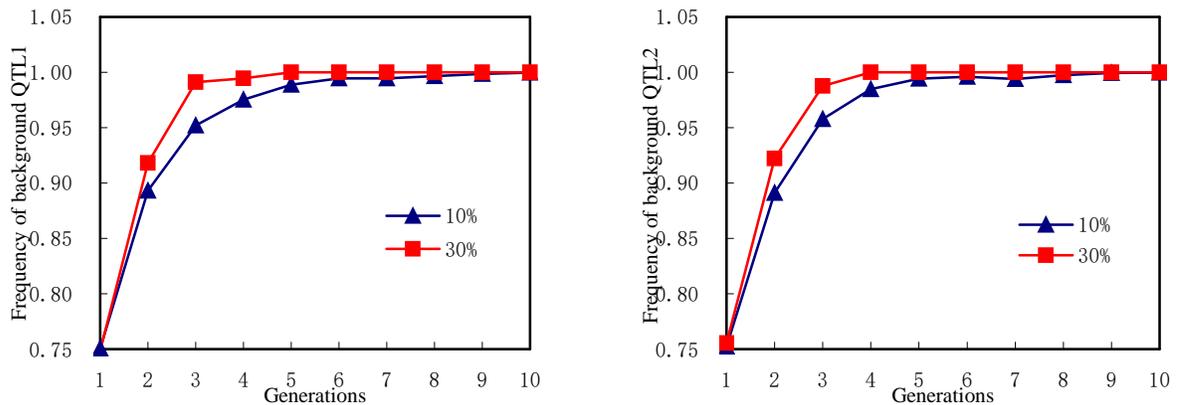


Figure 4. Frequency of favourable allele for background QTLs with respect to different ratios of QTL variance

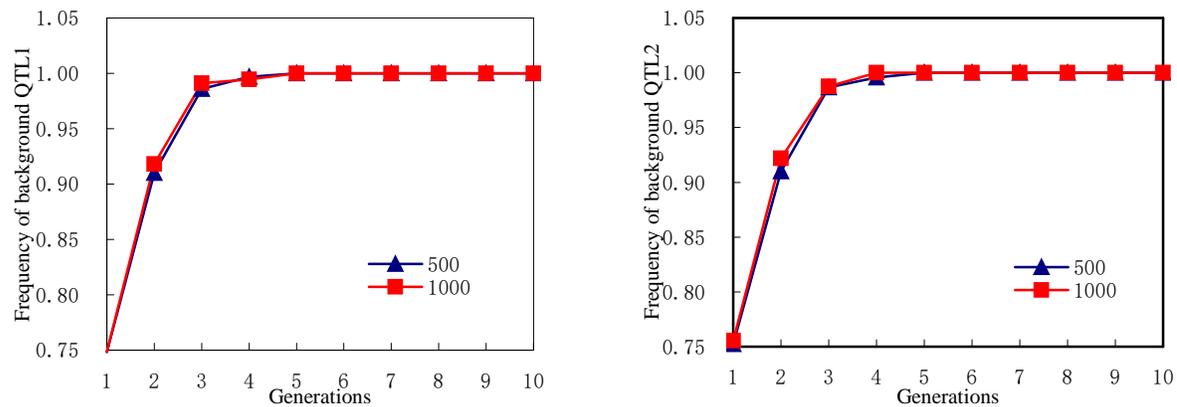


Figure 5. Frequency of favourable allele for background QTLs with respect to different population sizes.

Genetic responses for background traits under different background selection methods: The genetic responses for two background traits under different background selection methods was given in figure 6: during backcross phase, the genetic responses for two background traits show slowly increase by degrees trend with the increase of backcross generations, MBLUP selection obtained the most genetic responses for background traits, followed it were index selection, genomic similarity selection and random selection orderly. During intercross phase, the genetic responses for two background traits show quick increase trend under MBLUP selection and index selection, while under genomic similarity selection and random selection little genetic responses was obtained. In all, the genetic responses for two background traits show increasing trend with random selection, genomic similarity selection, index selection, MBLUP selection orderly, the genetic responses for two background traits under MBLUP selection is far higher than that under other selection methods.

Genetic responses for background traits under different ratios of QTL variance: The genetic responses for background traits under different ratios of QTL variance by MBLUP background selection were given in figure 7: whether in backcross phase or in intercross phase, the genetic responses for background traits is higher under 30% ratio of QTL variance than that under 10% ratio of QTL variance, especially in backcross phase, the trend is more significant. So enlarging ratios of QTL variance can increase the genetic responses for background traits.

Genetic responses for background traits under different population sizes: The genetic responses for two background traits under different population sizes by MBLUP background selection were given in figure 8: for the second background trait, whether in backcross phase or in intercross phase, the genetic responses obtained under population size 1000 is higher than that obtained under population size 500. For the first trait, in intercross phase, the genetic responses obtained under population

size 1000 is higher than that obtained under population size 500, in backcross phase, there is hardly any

difference between two population sizes.

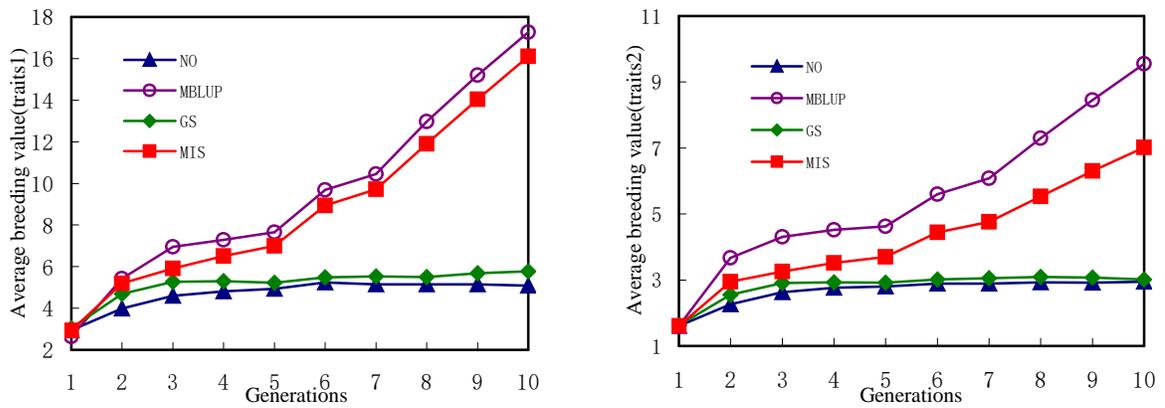


Figure 6. Genetic responses for background traits with respect to different background selection methods

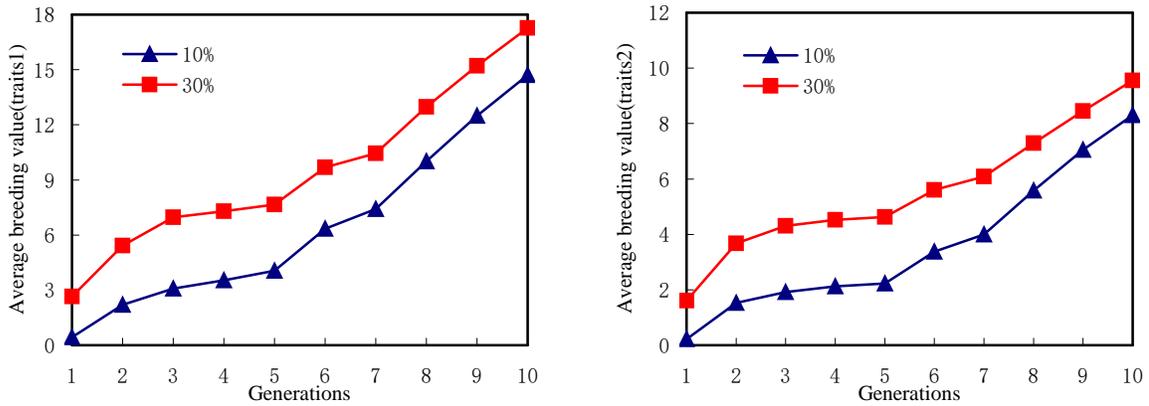


Figure 7. Genetic responses for background traits with respect to different ratios of QTL variance.

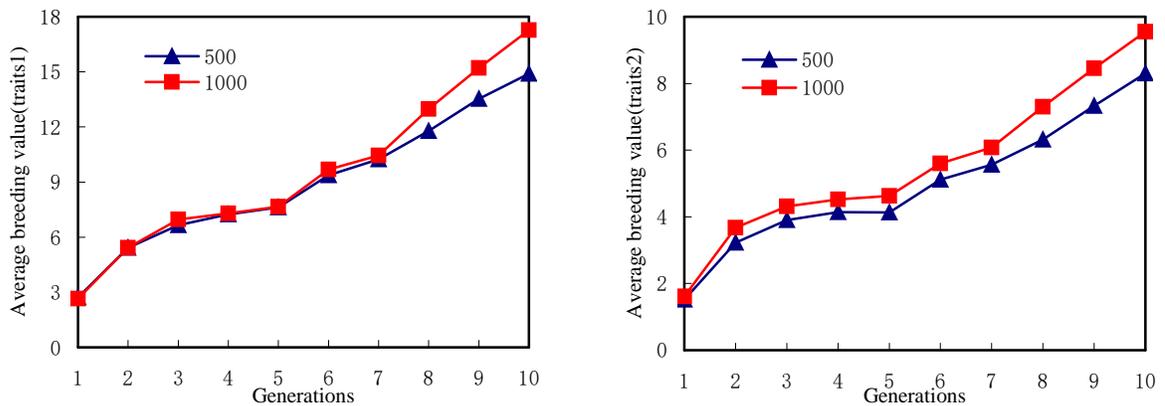


Figure 8. Genetic responses for background traits with respect to different population sizes.

DISCUSSION

Background selection : Background selection is a important factor affecting marker-assisted introgressing

efficiency, for whether a background selection method is appropriate or not can directly affect restoration speed of recipient genetic background. Our study shows that if adopt strategy of full-restoration recipient genetic

background, all the markers covering with recipient genome would be used to make genomic similarity selection or index selection, through backcrossing two generations, the restoration rate of recipient genetic background can reach 93%, through backcrossing 3~4 generations, recipient genome recover entirely, this shorten two generations compared to those without making background selection, which accord with other studies (Groen and Smith,1995; Visscher *et al.*,1996; Hospital *et al.*,1992 ; Hospital 2003). During the marker-assisted introgression process in practical breeding, the genetic background of recipient need not restore entirely, but some certain background traits need restore. So these some traits can be aimed at to make background selection, if only need restore QTL for these background traits, through backcrossing 3generations, frequency of favourable allele for background QTL can reach 1 under index selection. If breeder wouldn't like some important recipient background traits staying at fore-introgressing level, but keeping genetic responses all along, the markers linked to QTLs affecting these traits (not all the markers covering with recipient genome) and phenotypic information of these traits can be used to make MBLUP selection to accomplish his aim, for MBLUP selection not only make background traits get most genetic responses, but also its cost is more lower than the fore two selection methods, its operation is relatively simple, in all, MBLUP is a more practical choice.

Population sizes: Population size is also an important factor affecting marker-assisted introgressing efficiency, our study indicated that enlarge population size can increase frequency of favourable allele for introgressing QTL, when population size is larger, the selection space become broader, too, so could easily select individuals suitable for foreground and background, but too large a population size is difficult to accomplish in practice. Relatively, if the population size is smaller, individuals suitable for foreground selection and background selection were fewer, the selection strategy should be modulated, leading to lower frequency of favourable allele for introgressing QTL. Koudandé *et al.*, (1999) simulated certain inbreeding selfing experiment, deduce formula total number of favourable animals at generation t : $E(N_t) = 2n \{ (p) \{ (rs < 1 > s) \} \}^t$, n =number of males and number of females initially selected from each strain to start the introgression process (i.e. $2n$ donor animals and $2n$ recipient animals), p =probability of inheriting the desired chromosomal region(s), i.e. favourable, r =number of times each favourable male is used for breeding, s =sex ratio in offspring at breeding age, λ =mean number of animals produced in a litter and surviving to breeding age. These formulas aimed at some especial experiment designs and selection methods, lacking certain agility.

Ratios of QTL variance: Our study show enlarging ratios of QTL variance hardly has any effect on frequency of favourable allele for introgressing QTL and restoration of recipient genetic background, but in favor of increasing frequency of favourable allele for background QTL and genetic responses for background traits. When ratio of QTL variance is relatively large, polygenes have little effect on traits, so QTL favourable allele is easily fixed in population, limiting trait's improvement. Although size of ratios of QTL variance has little effect on frequency of favourable allele for introgressing QTL, when ratio of QTL variance is relatively small, the veracity of QTL checking and mapping will be effected.

For long generation interval, low reproductivity, high feed-cost, introgressing genes among animals need long time and high cost, so how to utilize marker information fully to improve gene introgressing efficiency become a research hotspot. There are many factors affecting marker-assisted introgression efficiency such as whether the expected introgressing gene is known or not (direct check), available marker information in genome (density and diversity), request on restoration degree for recipient genetic background and so on, under different conditions were the different ntrogressing schemes needed to obtain most introgression efficiency. For change in marker density has no effect on conclusion of this research, it isn't included in our study. Furthermore, some more complicated factors such as without finely maping the expected QTL, introgressing many QTLs at the same time, which make Marker-assisted introgression more complicated, and farther research will be needed.

Acknowledgements: Sincere gratitude goes to the sponsor of National Natural Science Foundation (31201777).

REFERENCES

- Andersson, L (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.*, 2: 130- 138.
- Chaiwong, N., J. C. M. Dekkers, R. L. Fernando and M. F. Rothschild (2002). Introgressing multiple QTL in backcross breeding programs of limited size. *Proc 7th World Cong Genet Appl Livest Prod, Montpellier, France*, 19-23.
- Dekkers, J. C. M (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *J. Anim. Sci.*, 82(1): 313-328.
- Dekkers, J. C. M and F. Hospital (2002). The use of molecular genetics in improvement of agricultural populations. *Nat. Rev. Genet.*, 3: 22-32.

- Gootwine E., S. Yossefi, A. Zenou and A. Bor (1998). Marker assisted selection for FecB carriers in Booroola Awassi crosses. Proc 6th World Cong Genet Appl Livest Prod, Armidale, Australia, 24, PP: 161-164.
- Groen, A. F and C. Smith (1995). A stochastic simulation study on the efficiency of marker-assisted introgression in livestock. J. Anim. Breed. Genet. 112: 161-170.
- Hanset, R., C. Dasnoi, S. Scalais, *et al.* (1995). Effets de l'introgression dans le genome Pie'train de l'allele normal aux locus de sensibilite' a l'halothane. Genet. Sel. Evol., 27: 77-88.
- Hillel, J., T. Schaap, A. Haberfeld, A. J. Jeffreys, Y. Plotzky, A. Cahaner and U. Lavi (1990). DNA fingerprints applied to gene introgression in breeding programs. Genetics., 124:783-789.
- Hospital, F and A. Charcosset (1997). Marker-assisted introgression of quantitative trait loci. Genetics., 147:1469-1485.
- Hospital, F., C. Chevalet and P. Mulsant (1992). Using markers in gene introgression breeding programs. Genetics., 132:1199-1210.
- Hospital, F (2003). Marker-assisted breeding. In: H. J. Newbury (ed.), Plant Molecular Breeding. Blackwell Scientific Publishers, London, UK, pp. 30-56.
- Koudandé, O. D., P. C. Thomson and J. A. M. van Arendonk (1999). A model for population growth of laboratory animals subjected to marker-assisted introgression: how many animals do we need? Heredity., 82: 16-24.
- Liu, H. Y., Q. Zhang and Y. Zhang (2001). Relative efficiency of marker assisted selection when marker and QTL are incompletely linked. Chinese. Science. Bulletin., 46(24):2058-2063.
- Visscher, P. M., C. S. Haley and R. Thompson (1996). Marker-assisted introgression in backcross breeding programs. Genetics., 144:1923-1932.
- Wang, T., R. L. Fernando, S. Vander Beek, M. Grossman and J. A. M. Van Arendonk (1995). Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol., 27:251-274.
- Yancovich, A., I. Levin, A. Cahaner and J. Hillel (1996). Introgression of the avian naked neck gene assisted by DNA fingerprints. Anim. Genet., 27: 149-155.