遗传学报 *Acta Genetica Sinica*, July 2006, 33 (7): 617–624 ISSN 0379-4172

Improving QTL Mapping Resolution Based on Genotypic Sampling—a Case Using a RIL Population

YAN Jian-Bing, TANG Ji-Hua, MENG Yi-Jiang, MA Xi-Qing, TENG Wen-Tao, Subhash Chander, LI Lin, LI Jian-Sheng①

National Maize Improvement Center of China, China Agricultural University, Beijing 100094, *China*

Abstract: The QTL mapping results were compared with the genotypically selected and random samples of the same size on the base of a RIL population. The results demonstrated that there were no obvious differences in the trait distribution and marker segregation distortion between the genotypically selected and random samples with the same population size. However, a significant increase in QTL detection power, sensitivity, specificity, and QTL resolution in the genotypically selected samples were observed. Moreover, the highly significant effect was detected in small size of genotypically selected samples. In QTL mapping, phenotyping is a more sensitive limiting factor than genotyping so that the selection of samples could be an attractive strategy for increasing genome-wide QTL mapping resolution. The efficient selection of samples should be more helpful for QTL maker assistant selection, fine mapping, and QTL cloning.

Key words: quantitative trait loci (QTL); selective mapping; recombinant inbred lines (RIL); QTL mapping resolution

Quantitative trait loci (QTL) mapping in the whole genome is one of the most important ways for gene finding until today. One of the key factors contributing to the success of QTL marker assisted selection(MAS) and/or positional cloning is the precision with which QTL position can be estimated. The precision of QTL can be affected by several factors such as, size and type of populations, density and resolution of linkage maps, and methods used for statistical analysis [1]. In general, QTL precision should be higher with larger population size, and higher density of linkage maps. Usually, the population used in QTL study can be divided into two groups: primary population (e.g. F_2, RIL, DH, and BC) and secondary population (e.g. near isogenic lines (NIL) constructed by continuous backcrossing and hybrid). On the basis of the primary population, it is difficult to improve the QTL precision through improving the statistical method [2], with a confidence interval being 10–20 cM [3]. Usually, larger population size is required for fine mapping the target traits on the basis of the secondary population. Both kinds of samples were randomly produced. Melchinger *et al.* [4] used two independent samples (n = 344 and 107), which derived from the testcross (TC) progenies of the corresponding F_3 lines with two testers to perform the QTL experiments in four environments. The results demonstrated that the numbers of QTL and their effects might be underestimated with the smaller size samples (e.g. n<200). However, it is difficult to handle the larger population size in field experiment. Vision *et al.* [5] proposed a method for choosing small mapping samples from a large sample to replace furthest the information of original population by genotypic selection. Selective mapping is an experimental design strategy for genome-wide, high-density linkage mapping of molecular markers in experimental crosses. In the first step, a limited number of markers, which distribute average around the genome, are genotyped in a large base population. From the genotypic data, individuals

Received: 2005-06-27; Accepted: 2005-08-05

This work was supported by the National Basic Research Project (No. 2001CB1088).

① Corresponding author. E-mail: lijs@163bj.com; Tel: +86-10-6273 2422

are selected to collectively provide good coverage of the crossover sites in the larger population. In the second step, a large number of markers can be genotyped on the selected sample. Recently, the results through simulations showed that the QTL detection power, sensitivity, specificity, and the precision of estimated QTL positions were increased significantly in the genotypically selected samples compared with the same size of random samples. Here, we compared the differences of these parameters during QTL mapping between the same size of selected and random population on the basis of phenotypic and molecular markers data of a RIL population that should provide some useful information for further construction of high-density linkage map, QTL fine mapping, and map-based cloning.

1 Materials and Methods

1.1 Plant materials and data of field experiment

The population used in this study consisted of 294 F_8 recombinant inbred lines (RIL) derived by single seed descendent (SSD) from an elite single cross (Yuyu22) between Zong3 and 87-1. RIL population along with its parents and the F_1 were planted at three locations, Changping (Beijing), Jinan (Shandong Province), and Xunxian (Henan Province) during summer 2003 and 2004, respectively. A randomized complete block design was employed with three replications at each location. In all locations, each field plot consisted of 20 plants grown in a 5 m single row with a planting density of 45 000 plants/ha. At maturity, the plant height was measured from ten random plants in each row. The pooled mean of plant height was used for QTL analysis.

1.2 Construction of genetic linkage map

Total genomic DNA was extracted from the plant leaves according to Saghai-Maroof *et al.*[7], and the SSR analyses were based on the methods reported by Senior *et al*[8]. Polymorphic markers between two parents were selected to assay the mapping population. A genetic linkage map was constructed using Mapmaker 3.0[9]. The critical *LOD* score for the test of independence of marker pairs was set at 3.0, and the order with the highest *LOD* score was then selected. Finally, 263 polymorphic markers were used to construct the linkage map.

1.3 Data analysis

1.3.1 Production of subpopulation

On the basis of the original RIL population, two methods were used to produce the subpopulation: one was produced randomly by the use of Microsoft EXCEL; whereas another was generated by selection by the software of Mappop (http://www.bio.unc.edu/faculty/visio/lab/mappop). The idea of selection is as follows: a Bin was defined to be an interval along a linkage group within which no breakpoints occur among any members of a given set of individuals but which is bounded by such breakpoints in at least one individual (or by the end or a linkage group)[5]. By minimizing the squares of the bin lengths (SSBL), we could get a subpopulation that had more crossovers frequency than the same size random population. Different samples size from the RIL population such as 30, 60, 90, 120, 150, 180, 210, 240, and 270 were produced by random and selection for next step data analysis.

1.3.2 Chi-square test for segregation distortion of markers

Chi^2 (χ^2) test used to test the marker ratio for normal Mendelian segregation (1:1) at the 0.05 significance level. The formula is as follows:

$$\chi^2 = \sum \frac{(|O-E|-1/2)^2}{E} .$$

1.3.3 Analysis of crossover enrichment

The use of the SSBL objective function is expected to lead to an enrichment of crossovers in a selected sample. The total number of crossovers in the selected sample relative to that of expected in the same size of random sample is referred to as the crossover enrichment, or CE [6]. Xu *et al.*[6] pointed out that when the marker spacing was less than or equal to about 10 cM, crossover enrichment (CE) could very closely be predicted by the following empirical formula: $CE = 1 + 0.5(1-f)\sqrt{A/L}$. Here, f is the sample frac-

tion; A is a constant that is determined by the type of base population. For a RIL population, $A \approx 500$, L is the length of molecular linkage map. The particular parameter range should be (L = [100, 2500], f = [0.1, 0.9] and marker spacing from 1 to 10 cM).

1. 3. 4 QTL analysis

The QTL analysis was performed by single-marker analysis as implemented in QTL Cartographer V 2.0[10]. The method was only used to estimate the effect of QTL detection power in the subpopulations. The QTL detection power was defined to be the probability that the maximum threshold for logarithm of odd (*LOD*) at any marker or position exceeded the significance threshold for a Type Ⅰ error of p=0.05[6]. On the basis of the mean of the three locations, single-locus QTL were mapped with composite interval mapping (CIM) [1] by QTL cartographer 2.0[10]. Model 6 in CIM was employed to map QTL in this study. A window size of 10 cM was used, and cofactors were chosen by stepwise regression (SRmapqtl). We assumed that there should be one QTL if *LOD* $\geqslant$2.5, the position of the highest *LOD* peak within the range was taken to be the QTL position. 1-*LOD* drop support interval was defined as the distance between the two points on either side of the peak where the *LOD* declined by one unit. The QTL mapped in the original RIL population with 294 individuals was counted as true QTL. If the peak of a QTL mapped in the subpopulation located in the 1-*LOD* drop support interval of a true QTL, it was counted as a true positive QTL (TP); if not, it was considered as a false positive QTL (FP). If one QTL can be detected in the original population but not in the subpopulation, it was counted as a false negative QTL (FN); in contrast, if one QTL could not be detected in the original population but in the subpopulations, it was counted as a false positive QTL (FP). Sensitivity (Sn) and specificity (Sp) were calculated as follows:

Sn=TP/ (TP+FN); Sp=TP/ (TP+FP).

The QTL mapping resolution was measured by 1-*LOD* drop support interval. The mean of 1-*LOD* drop support interval of QTL in the same size population were used to estimate the QTL resolution in selected and random populations. At this point, only the TP QTL had been considered.

2 Results

2. 1 Frequency for genotypic percentage and segregation distortion of markers

263 SSR markers were employed to construct a linkage map. Using Mapmaker 3.0, polymorphic markers were classified into 10 linkage groups covering 10 maize chromosomes with a total length of 2 360.8 cM with an average interval of 8.98 cM. The chromosome fragments from the parents in the original and subpopulation were all fitted for the theoretical ratio 1:1 by χ^2 tests. When the population size was smaller ($\leqslant$90), the allelic frequencies did not completely fit in normal distribution, although the 1:1 segregation ratio fit both in selected and random populations, respectively. With larger population size ($\geqslant$120), allelic frequencies in both selected and random populations followed normal distribution. Moreover, it looks better in selected population than in random population (Fig. 1.) There were no significant differences for markers segregation distortion both in selected and random populations at $P \leqslant 0.05$ level. As a whole, the markers segregation distortion ratio was increased with the rise in population size, otherwise, the markers segregation distortion ratio was little lower in random population than in selected population (Fig. 2.) For the original RIL populations, the markers segregation distortion ratio was almost same with the F_2 population, which was derived from the same hybrid "Yuyu22" with RIL population [11].

2. 2 Performance of plant heights

Plant height is an appropriate model trait for studying QTL because it had higher broad heritability, heterosis and could be easily measured. In this study, the pooled mean of plant height in 18 replications across three locations over two years was used for final data analysis. The plant height followed normal distribution in both selected and random populations of same size (Fig. 3.) It implied that the genotypic selection had no obvious effects for trait performance.

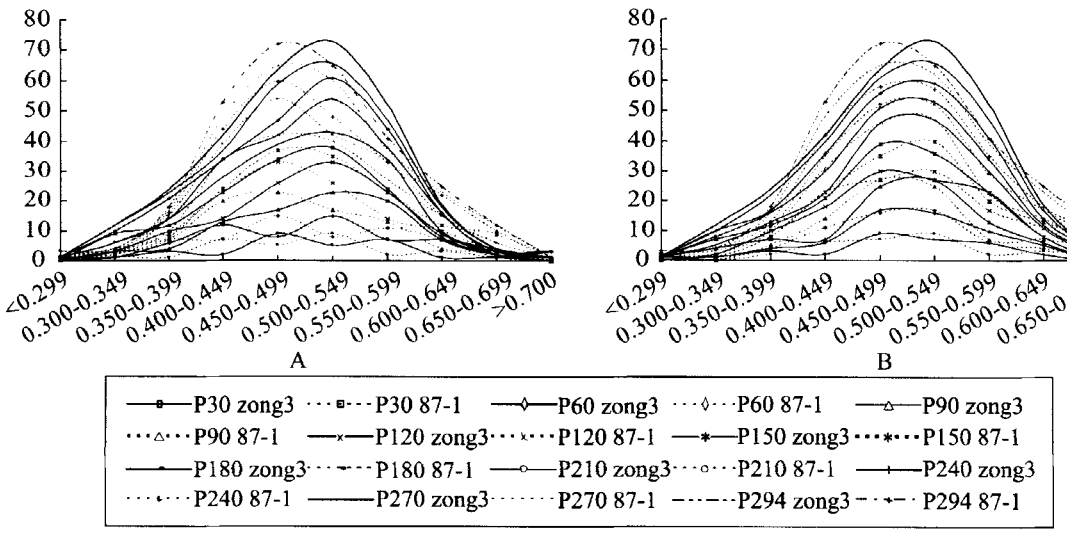

Fig. 1 Frequency for genotypic percentage in selected and random samples in different population size

A: random population; B: selected population.

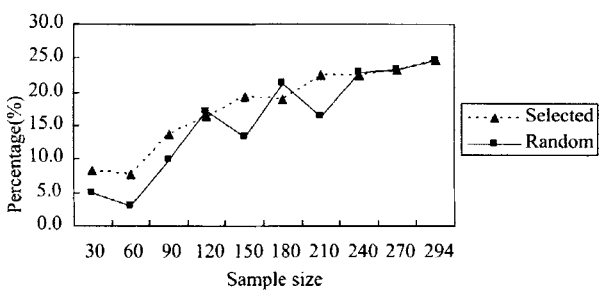

Fig. 2 Chi-square test for segregation distortion of markers in selected and random samples in different population size ($P<0.05$)

2.3 Crossover enrichment in both selective and random population

In general, the crossover enrichment (*CE*) was negative linear with the population size (Fig. 4). When the population size was 30 (about 1/10 of original population), *CE* ratio in selected population was 1.45 times higher than in the same size random population. However, when the population size was 270 (about 9/10 of original population), *CE* ratio as almost same in both selected and random populations (1.04). It showed that the effect of selection decreased with the increase in the size of the selected population.

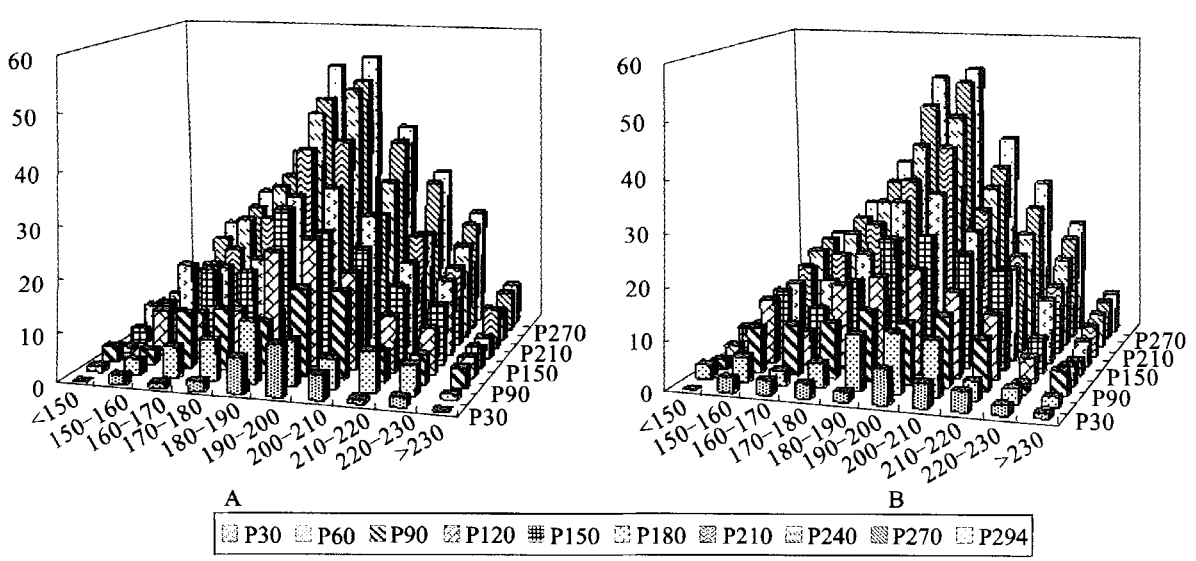

Fig. 3 Frequency distribution for plant height in selected and random samples in different population size

A: selected population; B: random population.

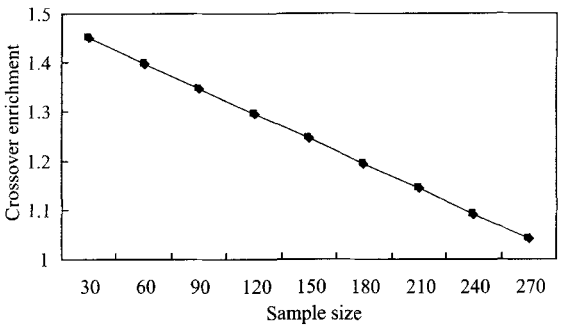

Fig. 4 Crossover enrichment after selection in different population size based on the RIL population

2.4 QTL analysis

QTL detection power: The probability of the maximum threshold for logarithm of odd (*LOD*) at any marker or position exceeded the significance threshold reflected partly the QTL detection power [6]. In general, the QTL detection power improved with the increase in size of both the selected and random populations (Fig. 5). Taking selected populations as the example, the QTL detection power was 0.36 when the population size was 30; the QTL detection power reached to 0.71 which was close to the value (0.73) of original population when the population size increased to 270. However, as a whole, the QTL detection power was better in selected population than the same size random population (Fig. 5). The results implied that the QTL detection power could be improved by genotypic selection in the same size population.

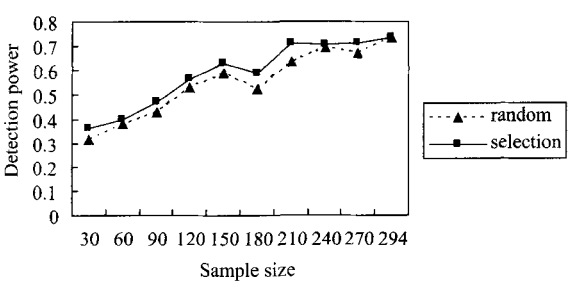

Fig. 5 QTL detection power in selected and random samples in different population size based on the RIL population

Sensitivity: The proportion that the true QTL number detected in the subpopulation took among QTL number detected in the original population reflected the sensitivity of QTL detection. As showed in Fig. 6 (left), as a whole, the sensitivity should increase with the increase of population size both in selected and random populations. For example, the sensitivity were 0.13 and 0.50 when the population size was 90 for random and selected populations, respectively, that means only 1.3 QTL could be detected in random population but 5 QTL could be detected in selected population if 10 QTL could be detected in the original RIL population. However, there were no differences of sensitivity of QTL detection between sub and original populations when the selected population size increased to 270. The sensitivity of QTL detection was higher in selected population than in random population when the population size was smaller than 210 (except population size=30). There were no obvious differences for sensitivity between selected population and random population when the population size was bigger than 210 (except population size=270).

Specificity: The proportion of true QTL number among all QTL numbers detected in the subpopulation reflected the specificity of QTL detection. As a whole, the specificity was increased with the increase in size of both selected and random populations as shown in Fig. 6 (right). Regard selected population as the example: the specificity was 0.33 when the population size was 30 that means only 3.3 QTL could be true if 10 QTL could be detected in the subpopulation. However, there were no differences of specificity between sub and original populations when the population size increased to 270. The specificity of QTL detection was higher in selected population than in random population when the population size was smaller than 210. There were no obvious differences for specificity between selected and random populations when the population size was bigger than 210. Another character for selected population is that there were no obvious differences for specificity when the population size ranged from 90 to 240.

2.5 QTL mapping resolution

We only compared the 1-*LOD* drop support

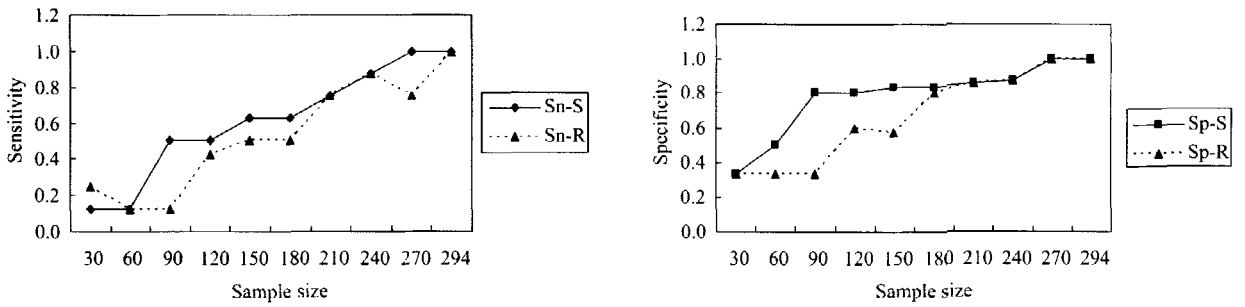

Fig. 6 Sensitivity and specificity of QTL detection in selected and random samples in different population size based on the RIL population

interval of the same QTL between the selected and random populations with the same size because the QTL numbers and locations in different size populations were not always the same and the 1-*LOD* drop support intervals for different QTL were also not same. As showed in Fig. 7, there were obvious differences for QTL mapping resolution between the selected and random populations when the population size was smaller than 150. For example, when the population size was 90, the means of QTL 1-*LOD* drop support interval was 7.53 cM in selected population and 13.82 cM in random population. There was no obvious difference for QTL mapping resolution between selected and random populations when the population sizes were 150, 180, and 240, respectively. But in general, the QTL mapping resolution could be improved significantly by genotypic selection.

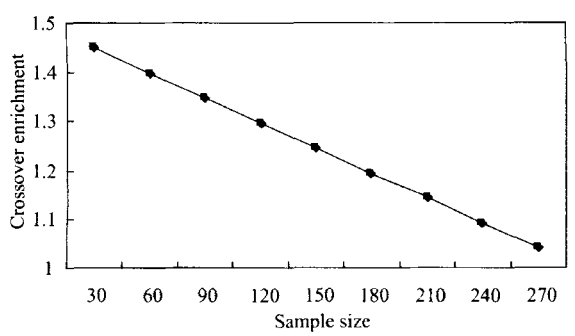

Fig. 7 Confidence intervals based on average of the same QTL obtained using the 1-*LOD* drop method in random and selected samples in different population size

3 Discussion

The results in this study demonstrated that there were no obvious differences in the trait performance and markers segregation distortion between the selected and random populations with same size. However, the QTL detection power, sensitivity, specificity, and QTL mapping resolutions were improved significantly in selected population than in random population. Moreover, the effects of selection should be more significant as the intensity of selection being greater. In the past two decades, gene/QTL mapping had become a hot research area with the rapid development of molecular markers. However, because of the limitation of funding, experimental design, and so on, the population of QTL mapping was often smaller than 500 individuals. Since the 1990's, with the development of PCR-based molecular markers techniques, the expenses were reduced constantly that made it possible to handle a large population such as bigger than 3 000. On the other hand, the acquisition of the phenotypic data had already become a limiting factor at this moment. The strategy of genotypic selection should be considered in future studies. Firstly, we can genotype in a large population using limited markers, then select the appropriate sample size for the next step; secondly, the high-density linkage map can be constructed in the selective population with large numbers of markers; thirdly, the accurate QTL locations and resolutions should be evaluated combined with the phenotypic data. Without doubt, QTL mapping on the basis of genotypic-selected population should provide useful information for QTL marker assisted selection, fine mapping, and map-based cloning.

Recently, Hua *et al.*[12] proposed "immortalized

F_2" population for studying the genetic basis of heterosis in rice. Large numbers of F_1 individuals produced by random cross of a RIL/DH population resembled the F_2 population. The genotypes of F_2 population can be calculated by the genotypes of RIL/DH populations. For example, 43071 (43071=(294 × 293)/2) unrepeated F_1 individuals can be gained on the basis of the present RIL population containing 294 individuals. The genotypes of 43071 individuals can be easily gained through the genotypes of the RIL population containing 294 individuals, but it should be difficult to perform the field experiments with such large size population. Theoretically, we can get an appropriate subpopulation by genotypic selection for the next study.

References:

[1] Zeng Z B. Precision mapping of quantitative trait loci. *Genetics*, 1994, 136 : 1457-1468.

[2] Wu W R, Tang D Z, Li W M. Genetic dissection and molecular dissection of quantitative traits. *Acta Agronomica Sinica*, 2000, 26: 501-507 (in Chinese with an English abstract).

[3] Alpert K B, Tanksley S D. High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: A major fruit weight quantitative trait locus in tomato. *Proc Natl Acad Sci USA*, 1996, 93 : 15503-15507.

[3] Melchinger A E, Utz H F, Schön C C. Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics*, 1998, 149 : 383-403.

[4] Vision T J, Brown D G, Shmoys D B, Durrett R T, Tanksley S D. Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics*, 2000 155 : 407-420.

[5] Xu Z, Zou F, Vision T J. Improving Quantitative Trait Loci Mapping Resolution in Experimental Crosses by the Use of Genotypically Selected Samples. Genetics, 2005; 170(1) : 401-408.

[6] Saghai Maroof M A, Soliman K M, Jorgensen R A, Allard R W. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population, and population dynamics. *Proc Natl Acad Sci USA*, 1984, 81 : 8014-8018.

[7] Senior L, Lynn M, Heun M. Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome*, 1993, 36 : 884-889.

[7] Lincoln S, Daly M, Lander E. Mapping genetic mapping with MAPMAKER/EXP3.0. Cambridge: MA: Whitehead institute Technical Report, 1992.

[8] Basten C J, Weir B S, Zeng Z B. QTL Cartographer. North Carolina State University, 2001.

[9]Yan J B, Tang H, Huang Y Q, Zheng Y L, Li J S. Genetics Analysis of Segregation Distortion of Molecular Markers in Maize F_2 Population. *Acta Genetica Sinica*.2003, 30 : 913-918 (in Chinese with an English abstract).

[10] Hua J P, Xing Y Z, Wu W R, Xu C G, Sun X L, Yu S B, Zhang Q F. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA*, 2003, 100 : 2574-2579.

基于基因型选择提高QTL作图的精度—以一个RIL群体为例

严建兵，汤继华，孟义江，马西青，滕文涛，Subhash Chander，李 林，李建生

中国农业大学国家玉米改良中心，北京 100094

摘 要：以PCR为基础的分子标记以及其他检测技术的发展，使得大规模的标记分析成为现实。这也为通过大群体标记分析，然后基于基因型选择挑选合适的小群体，从而提高QTL定位准确性和精度提供了可能。以一个包含294个家系的重组自交系（RIL）群体为例，通过基因型选择和随机选择的办法产生了一系列大小不等的亚群体，比较了两类群体QTL定位的结果。分析表明：相同大小的基因型选择群体与随机群体相比性状的表型分布都符合正态分布；标记的偏分离情况也没有明显的差别，都随着群体大小的增大，偏分离的比例也逐渐增大。但同等大小的基因型选择群体比随机群体的交换富集率（*CE*）要大，且随着选择强度的增大不断增大，如群体大小为270时，*CE*=1.04，群体大小为30时，*CE*=1.45。总体上，随着群体大小的增加，不管是随机群体还是选择群体，其QTL检测能力、灵敏性和特异性也随之增加，但选择群体的检测能力、灵敏性和特异性总体上要好于随机群体。当群体大于或等于240时，其QTL检测能力基本没有差别；群体大小大于或等于210时，其QTL检测的灵敏性和特异性也没有什么差别。这也说明：选择强度越大，效果越明显。以QTL1-*LOD*区间作为衡量QTL精度的一个指标，结果显示所有基因型选择群体都比相同大小随机群体的QTL定位精度高。目前QTL定位研究中，基因型数据较表型数据而言更容易准确获得，因此通过基因型选择可以更好的优化群体结构，减少田间实验的工作量，提高全基因组水平QTL作图的精度，为随后的QTL辅助选择和精细定位以及克隆提供帮助。

关键词：数量性状位点（QTL）；基因型选择；重组自交系(RIL)；QTL定位精度

作者简介：严建兵（1976-），男，博士，副教授，研究方向：玉米分子育种和基因组学。E-mail: yjianbing@cau.edu.cn