Theor Appl Genet (1997) 95:618–621

S.M. Kaeppler

Power analysis for quantitative trait locus mapping in populations derived by multiple backcrosses

Received: 4 August 1996 / Accepted: 4 April 1997

Abstract Populations derived by multiple backcrosses are potentially useful for quantitative trait locus (QTL) mapping studies. Comparisons of relative power to detect QTL using populations derived by multiple backcrosses are needed to make decisions when mapping projects are initiated. The objective of this study was to theoretically compare the power to detect QTL in populations derived by multiple backcrosses relative to mapping in a recombinant inbred population of equal size. Backcrossing results in a reduction in genetic variance with each generation and also results in an increasing frequency of the recurrent parent marker genotype. The relevant outcome for QTL mapping is a reduction in genetic variance to partition between marker genotype classes and increasing unbalance of the number of individuals contributing to the mean of the marker genotypes. Both of these factors lead to a decrease in the power to detect a QTL as the number of backcross generations increases. Experimental error was held constant with the populations compared. From a theoretical standpoint, backcross-derived populations offer few advantages for QTL detection. If, however, a backcrossing approach is the most efficient method to achieve a desired breeding objective and if QTL detection is an objective of equal or less importance, backcross-derived populations are a reasonable approach to QTL detection.

Keywords Genetic mapping · Molecular marker · Quantitative trait · Backcross

Communicated by A.L. Kahler

S.M. Kaeppler (✉)
Department of Agronomy, University of Wisconsin,
1575 Linden Drive, 455 Moore Hall, Madison WI 53706, USA
Fax: + +1 608-262-5217; e-mail: smkaeppl@facstaff.wisc.edu

Introduction

Quantitative trait locus (QTL) mapping has been accomplished using a number of population structures, including segregating heterogeneous families (e.g., F_2-derived F_3 families), recombinant inbred families, and backcross-derived families. While numerous statistical procedures have been used to analyze data from these mapping experiments, the basic approach is to test for equality of marker genotype means, with a significant deviation from equality indicating linkage between a marker and a QTL. Discussions of statistical issues relating to QTL detection are available for many types of populations (Knapp and Bridges 1990).

Backcrossing is a procedure used in a number of breeding programs to improve a good line or parent especially when an unadapted or poor line is used as the source of favorable alleles. Backcross-derived lines have been used in both trait-based (Osborn et al. 1987; Phillips et al. 1992; Koester et al. 1993) and marker-based (Eshed and Zamir 1995; Tanksley et al. 1996) QTL mapping studies. Advanced backcross populations (two or more backcrosses) have been proposed as population structures that would facilitate QTL detection and integration into elite breeding lines (Tanksley and Nelson 1996).

Statistical issues relating to QTL detection in populations derived by one backcross have been presented by Knapp and Bridges (1991). QTL detection in advanced backcross populations has been previously assessed by computer simulation (Tanksley and Nelson 1996). This manuscript describes a theoretical approach for determining the relative power to detect QTL in advanced backcross populations. This approach allows rapid calculations of power without the effort and imprecision associated with computer simulation. Comparisons were made relative to a recombinant inbred population of equal size and are described for populations derived by one to four backcrosses.

Theory and methods

Analyses were done for both backcross and inbred-backcross populations. Inbred-backcross populations have been selfed to homozygosity following the specified number of backcrosses. Throughout the manuscript, backcross populations will be designated BC followed by the number of backcrosses, and inbred-backcross populations will be designated I-BC followed by the number of backcrosses. Comparisons were made to recombinant inbred populations which are designated RI.

Comparisons were made for unreplicated populations based on the following linear model:

$$y_{ij} = \mu + \tau_i + e_{ij}$$

where:

y_{ij} = the phenotypic value of the jth line ($j = 1, 2, \ldots, n$) with the ith marker genotype ($i = 1$ to 2),

μ = mean value of the population calculated across all QTL except the one under consideration,

τ_i = the effect of the ith marker genotype,

and

e_{ij} = residual error calculated as variation among lines within genotypes.

The ratio of genetic to non-genetic variation ($\sigma_g^2/\sigma_g^2 + \sigma_e^2$) was determined in a theoretical recombinant inbred population and was defined as heritability for purposes of this study. Genetic variance includes the effect of the QTL under analysis plus the effect of all other QTL. Values for the following variance components were established after defining the portion of variation accounted for by the QTL of interest:

σ_e^2 = experimental error

Φ_q^2 = fixed effect of the QTL in the genomic region under analysis,

and

$\sigma_u^2 = (\sigma_g^2 - \Phi_q^2)$ = variation due to all QTL not linked to the QTL under analysis.

Using an unreplicated model, σ_e^2 cannot be separated from σ_u^2, and a residual error term, σ_E^2, was defined as $\sigma_e^2 + \sigma_u^2$. Experimental error, σ_e^2, was held constant for all population types. Φ_q^2 and σ_u^2 were modified based on the relative coefficient of genetic variance in advanced backcross populations as defined in Table 1. For example, the coefficient 1/4 was used to quantify Φ_q^2 and σ_u^2 in the backcross 1 population relative to the reference recombinant inbred population. Table 1 was produced based on Mather and Jinks (1982) and Hallauer and Miranda (1986) using a completely additive genetic model. Molecular markers and QTL were assumed to be coincident.

The critical test to determine independence of a molecular marker and a QTL in a **backcross** population is Ho: $\mu_{MM} = \mu_{Mm}$, where μ_{MM} and μ_{Mm} are the means of the homozygous recurrent parent and heterozygous marker genotypes, respectively. The F-statistic for testing this hypothesis is $F = MS_{MM\ vs.\ Mm}/MS_E$, and the non-centrality parameter (Knapp and Bridges 1990) is:

$$\lambda = \frac{E(MS_{MM\ vs.\ Mm})}{E(MSE)} = \frac{\sigma_E^2 + \bar{n}\Phi_q^2}{\sigma_E^2}$$

The critical test to determine independence of a molecular marker and a QTL in an **inbred-backcross** population is Ho: $\mu_{MM} = \mu_{mm}$, where μ_{MM} and μ_{Mm} are the means of the homozygous recurrent and donor parent marker genotypes, respectively. The F-statistic for testing this hypothesis is $F = MS_{MM\ vs.\ mm}/MS_E$, and the non-centrality parameter is:

$$\lambda = \frac{E(MS_{MM\ vs.\ mm})}{E(MSE)} = \frac{\sigma_E^2 + \bar{n}\Phi_q^2}{\sigma_E^2}$$

The parameter $\bar{n}$ is used to represent the unbalance of marker genotype means in advanced backcross generations. The variable $\bar{n}$ is the mean number of replications (i.e., individuals or lines) in the two marker classes. This parameter is calculated as:

$$\bar{n} = \frac{N - \frac{\sum_{i=1}^{m} n_i^2}{N}}{m - 1}$$

where:

N = total population size,

n = population members in marker group i (for $i = 1$ to m marker genotypes),

m = number of marker genotypes.

Values for $\bar{n}$ are included in Table 1 for the respective populations based on a population size of 100 individuals.

Table 1 Coefficients for variance among backcross-derived progenies based on an additive genetic model (based on Mather and Jinks 1982 and Hallauer and Miranda 1988). Genotypic frequencies for a single locus are given for each generation. Mean number of replications given for the molecular marker genotype means ($\bar{n}$) is based on a population size of 100

Population structure	Coefficient of genetic variance[a]	Frequency MM[b]	Frequency Mm or mm[b]	$\bar{n}$
RI	1A	1/2	1/2	50.0
BC_1	1/4A	1/2	1/2	50.0
BC_2	3/16A	3/4	1/4	37.5
BC_3	7/64A	7/8	1/8	21.9
BC_4	15/256A	15/16	1/16	11.7
$I\text{-}BC_1$	3/4	3/4	1/4	37.5
$I\text{-}BC_2$	7/16A	7/8	1/8	21.9
$I\text{-}BC_3$	15/64A	15/16	1/16	11.7
$I\text{-}BC_4$	31/256A	31/32	1/32	6.1

[a]Coefficients based on an additive genetic model

[b]Genotypes designated MM, Mm, or mm correspond to molecular marker genotypes, with MM corresponding to the genotype of the recurrent backcross parent. Genotypes MM and Mm are present in backcross (BC) populations, and genotypes MM and mm are present in inbred-backcross (I-BC) populations

Function PROBF and FINV (SAS 1987) were used to calculate critical values and probabilities for F-distributions. Comparisons were made using a factorial of heritabilities from 0.1 to 0.9 incremented by 0.2, and QTL percentages of genetic variation of 5% and 10% in the reference recombinant inbred population. A Type 1 error rate of $\alpha = 0.01$ was used for all comparisons.

Results and discussion

The recombinant inbred population had more power to detect QTL than any of the backcross populations compared (Fig. 1). Power decreased in the backcross populations with each additional generation of backcrossing. The explanation for the decrease in power as backcross generations increase is obvious by inspection of Table 1. Genetic variance decreases each backcross generation, resulting in a decrease in the important variance component Φ_q^2 relative to σ_e^2. In addition, the mean number of replications of molecular marker genotypes, $\bar{n}$, is reduced in each backcross generation, further reducing the power to detect differences between the molecular marker genotype means. Functionally, the recurrent parent marker class mean becomes better estimated as the number of backcrosses increases, and the donor parent marker class mean becomes more poorly estimated due to the increasing unbalance of the number of members in each group.

An observational comparison of the relative importance of $\bar{n}$ and Φ_q^2 can be made from Fig. 1. For example, $\Phi^2_{q:I-BC_2}$ is greater than twice $\Phi^2_{q:BC_2}$: $\bar{n}_{I-BC_2}$ is 21.9 and $\bar{n}_{BC_2}$ is 37.5. Power for the two populations is similar, indicating that a decrease in the variance due to the QTL of interest is proportionally counteracted by an increase in $\bar{n}$.

A recombinant inbred population was used as a basis for comparison in this study because: (1) it is among the most statistically powerful types of populations for detection of QTL, and (2) only two genotypes are present, analogous to the backcross population structures presented. F_2-based populations such as F_2 individuals or F_2-derived F_3 lines have less power to detect the additive effect of a QTL than recombinant inbred populations for a given population size because each marker class is replicated fewer times (Knapp and Bridges 1990). Specific power comparisons need to be made if other types of populations are to be considered. However, for a given population size, few population structures equal the power of recombinant inbred populations for detecting the additive effect of a QTL.

Power estimates are given based on an additive genetic model. With increasing amounts of dominance for the recurrent parent allele, the genotypic values in backcross populations become increasingly similar until equality is expected with complete dominance. Therefore, dominance will decrease power in backcross populations but will not affect power estimates given for recombinant inbred or inbred-backcross populations.

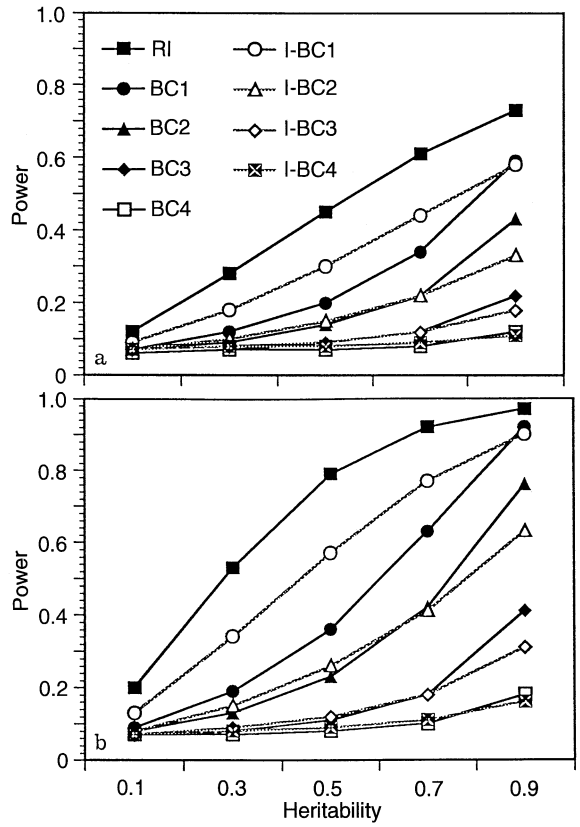

Fig. 1a, b Power $(1 - \beta)$ curves for detection of a QTL in recombinant inbred (RI), backcross (BC), and inbred-backcross (I-BC) populations assuming coincidence of a marker and the QTL. Curves shown for a QTL representing 5% (**a**) and 10% (**b**) of the genetic variation in the recombinant inbred population. Calculations were made for $\alpha = 0.01$ and a population size of 200 individuals

The comparisons presented assume a constant σ_e^2 over all populations considered. In some cases, homogenization of the population by backcrossing may reduce this parameter. For instance, plant height and maturity differences among entries can result in plot-to-plot variation. Tall entries may shade and inhibit growth of shorter neighboring plots. Early-maturing entries may senesce and allow more light to reach late-maturing neighboring plots. Considerations such as these would enhance QTL detection in the more homogeneous advanced backcross populations relative to recombinant inbred populations from the same parents.

The results presented are based on the assumption of the absence of epistasis. The presence and magnitude of epistatic effects have been difficult parameters to detect and quantify. However, epistasis is likely to be substantial for certain traits and populations. If the genetic potential of a donor parent line to improve an elite line is being assessed, and if genetic background (epistatic) effects contribute appreciably to trait expression, then it is prudent to assess donor parent allele effects in the background of the elite parent. In this instance,

homogenization of the population by backcrossing is a useful approach to QTL characterization in a specific genetic background.

From a theoretical standpoint, backcross-derived populations offer few advantages for QTL detection. In selfing species, backcrossing is a particularly labor-intensive procedure whereas selfing is much easier. Similar numbers of selfs and backcrosses are required to reach homozygosity for recombinant inbred or inbred-backcross populations. If QTL detection is a primary objective of a research project, choosing the population structure with the most statistical power is preferred especially if the resources required to produce the population are reasonable. Advanced backcrossing would be preferred if: (1) a backcrossing approach is the most efficient method to achieve a desired breeding objective, QTL effects are relatively large, and QTL detection is an objective of equal or less importance, or (2) if epistasis is important in controlling a trait, and QTLs can be analyzed in a "reference" genetic background.

QTL detection is often attempted to enable marker-assisted selection procedures. With increasing numbers of backcrosses, only QTL with relatively large effects will be efficiently detected. In general, phenotypic selection will likely to be as efficient as marker-assisted selection when selecting for QTL with relatively large effects. Therefore, advanced backcrossing may only allow the detection of QTL which can be efficiently selected without the use of molecular markers.

Population structures for genetic mapping projects are chosen for numerous reasons. This research considered intrinsic factors associated with population development by backcrossing. Reduction in genetic variance and genotypic unbalance are inextricably a part of advanced backcrossing procedures. Trade-offs between breeding objectives and statistical issues are present in all areas of plant improvement. Whatever population structure is chosen, QTL detection will always be enhanced by minimizing experimental error and maximizing marker genotype replication.

References

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141: 1147–1162

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Knapp SJ (1991) Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred, and doubled haploid progenies. Theor Appl Genet 81: 333–338

Knapp SJ, Bridges WC (1990) Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progenies. Genetics 126: 769–777

Koester RP, Sisco PH, Stuber CW (1993) Identification of quantitative trait loci controlling days to flowering and plant height in two isogenic lines of maize. Crop Sci. 33: 1209–1216

Mather K, Jinks JL (1982) Biometrical genetics: the study of continuous variation, 3rd edn. Chapman and Hall, New York

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73: 350–356

Phillips RL, Kim TS, Kaeppler SM, Parentoni SN, Shaver DL, Stucker RE, Openshaw SJ (1992) Genetic dissection of maturity using RFLP's Proc. In: 47th Annu Corn Sorghum Indust Res Conf., D. Wilkinson (ed.), American Seed Trade Association, Washington D.C., pp 135–150

SAS (1987) SAS/STAT guide for personal computers, version 6. SAS Institute. Cary, N.C.

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92: 191–203

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92: 213–224