Power of tests for QTL detection using replicated progenies derived from a diallel cross

A. Rebai, B. Goffinet
Laboratory of Biometry, National Institute of Agronomy Research, Centre de Toulouse, Chemin Borde Rouge, Auzeville BP 27, 31326 Castanet-Tolosan, France

Received: 7 October 1992 / Accepted: 28 January 1993

Abstract. In crop species, most QTL (quantitative trait loci) mapping strategies use segregating populations derived from an initial cross between two lines. However, schemes including more than two parents could also be used. We propose an approach using a high-density restriction fragment length polymorphism (RFLP) map established on six F₂ populations derived from diallel crosses among four inbred lines and the phenotypic performances of two types of replicated progenies (F₃ and topcross). The QTL is supposed to be on the marker locus considered. Three linear model tests for the detection of QTL effects (T₁, T₂ and T₃) are described and their power studied for the two types of progeny. T₁ tests the global genetic effects of the QTL (additivity and dominance) and T₂ tests only additive effects assuming dominance is absent when it could exist. The models of these two tests assume that the main effects of QTL alleles are constant in different genetic backgrounds. The additive model of test T₃ considers the six F₂ populations independently, and T₃ is the equivalent of the classical mean comparison test if we neglect dominance; it uses only contrasts between the homozygote marker classes. The results show that T₂ is much more powerful than T₃. The power of T₁ and T₂ depends on the relative sizes of the additive and dominance effects, and their comparison is not easy to establish. Nevertheless, T₁ seems to be the more powerful in most situations, indicating that it is often more interesting to ignore dominance when testing for a QTL effect. For a given size of genetic effects, the power is affected by the total number of individuals genotyped in F₂ and the recombination rate between the marker locus and the putative QTL. The approach presented in this paper has some drawbacks but could be easily generalized to other sizes of dialles and different progeny types.

Key words: Diallel – Restriction fragment length polymorphism markers – Replicated progenies – Linear models – Power of test

Introduction

Most traits of economic importance in plants show a continuous variation in phenotype that is the result of the collective action of multiple genetic factors and environmental effects. These factors are located at different quantitative trait loci (QTLs) and, in general, have not been resolved individually. Quantitative genetic studies have dealt largely with a global characterization of genetic factors by developing powerful biometrical approaches.

Sax (1923) was the first to propose the detection of specific genes affecting quantitative traits by studying their associations with marker genes. Numerous contributions to the theory and applications of the marker-based detection of QTLs have since been made (reviewed by Thompson and Thoday 1979), but the restricted number of effective markers has since been made (reviewed by Thompson and Thoday 1979), but the restricted number of effective markers rapidly limited this kind of investigation until the development, in the 1980s, of a new biomolecular tool: restriction fragment length polymorphism (RFLP). This technique provides a potentially unlimited number of codominant markers dispersed along the length of the genome (see Beckmann and Soller 1986 for review).

In crop species, most QTL mapping strategies start with the recognition of two parental genotypes, gen-
eraly inbred lines homozygous for alternative alleles at the loci of interest, followed by the generation of hybrid $F_1$ progeny. The classical schemes continue by either selfing the $F_1$ or backcrossing it to one of the parental lines to generate a segregating $F_2$ or backcross (BC) population. The latter is then genotyped for the markers and scored for phenotypic performance in order to examine marker-phenotype relationships. Several methods have been advanced in the literature for the detection of linkage between a marker locus and a QTL by linear model or maximum likelihood analysis, have been described for various progeny types (Jayakar 1970; Soller et al. 1976; Soller and Genizi 1978; Weller 1986; Luo and Kearsey 1989, 1991). The traditional method uses contrasts among marker genotype means (mean comparison method) or a one-way analysis of variance (ANOVA) to test for the presence of a QTL in the vicinity of the marker locus (Soller et al. 1976; Tanksley et al. 1982; Edwards et al. 1987). It has been widely applied in crop species (Edwards et al. 1987, 1992; Stuber et al. 1987; Weller et al. 1988) and extended to other types of progeny (Bechmann and Soller 1988; Cowen 1988; Ellis 1986; Simpson 1989; Soller and Beckmann 1990). Even if the ANOVA approach (contrary to maximum likelihood methods) provides no information about the recombination between markers and putative QTLs, it does provide simple tests the powers of which can be easily compared for various experimental designs (Soller et al. 1976; Knapp and Bridges 1990).

Lander and Botstein (1989) proposed a promising method (interval mapping) based on maximum likelihood analysis that examines intervals between neighbouring markers to provide a confidence interval within which the QTL may be found. Statistical details of their approach and its advantages relative to the ANOVA method in $F_2$ populations are given by Carbonell et al. (1992) and van Ooijen (1992). Although no complete studies of the power of interval mapping have been done, many authors state that it has a greater power than traditional methods. Whatever the method used for QTL detection [ANOVA, interval mapping, method of Knapp et al. (1990), etc.], the power of a test can be calculated by considering, first, that there is an infinite number of markers along the genome. Consequently, one will calculate the power of a test by assuming that the QTL is on the marker and using an appropriate threshold which takes into account the fact that we will perform this same test all along the genome. In the present case we are concerned with an ANOVA test. In the discussion we give some indications about the exact power of different tests. We can say that the loss in power, compared to the case where we have an infinite number of markers, will be approximately 10–20% and that the proportion between the power of different tests will be conserved.

The aim of this paper is to compare, by analytical calculations, the powers of different linear model tests for QTL mapping using two progeny types derived from a diallel cross between four inbred lines.

Modelling

Assumptions

We consider a diallel cross between four homozygous lines $L_1$, $L_2$, $L_3$, $L_4$ (without selfings nor reciprocals). The six $F_2$ populations obtained by self-fertilizing $F_1$ hybrids are genotyped with RFLPs in order to construct a high-density marker linkage map. $F_2$ progenies are then selfed and crossed to two testers, which are the complementary parental lines (the $F_2$ coming from the cross $L_1 \times L_2$ is crossed with $L_3$ and $L_4$ as testers). The $F_3$ and topcross progenies are then grown in replicate and scored for a quantitative trait. We suppose that the number of individuals is the same for each $F_2$ population, and we denote by $N$ the total $F_2$ population size. The polymorphic marker locus studied ($M$) is supposed to have four alleles, $M_1$, $M_2$, $M_3$, $M_4$. This assumption is quite restrictive because only a limited number of marker loci will have four alleles (about 50% of the RFLP markers in maize [Zea mays (L.) if the parents are well sampled] but does facilitate the statistical treatments. The segregation of the markers in each $F_2$ is supposed to be Mendelian (an $F_1$ which if $M_1 M_1$ gives a $F_2$ descendance that is $\frac{1}{4} M_1 M_1, \frac{1}{4} M_1 M_2, \frac{1}{4} M_2 M_2, \frac{1}{4} M_2 M_1$). We made no assumption about the number of QTL alleles. This point will be discussed later.

Genetic and statistical models

Two models that consider marker-linked effects as being independent of the genetic background are described. A more complete model will be discussed later.

$F_3$ progenies

We used a linear model neglecting genotype × environment interactions and assuming no epistasis. So, we wrote (1):

$Y_{ijk}(M_i M_j) = \mu_{ij} + 2a_i + g_{ijk} + e_{ijk}$.

where $Y_{ijk}(M_i M_j)$ is the phenotypic mean (over all replications) value of the $F_3$ offsprings of the $k^{th}$ indi-