Theor Appl Genet (1994) 88:486–489

K. G. Lark · J. Orf · L. M. Mansur

Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles

Received: 3 May 1993 / Accepted: 9 September 1993

Abstract Quantitative trait values for seed oil and protein content or for maturity were measured in recombinant inbred lines (RIL) of soybean derived from a cross between two soybean cultivars: 'Minsoy' PI 27890 and 'Noir 1' PI 290136. Seed oil was found to be inversely correlated to protein content. By analyzing DNA from plants with extreme phenotypes, we were able to identify quantitative trait loci (QTL) for these traits as being linked to several restriction fragment length polymorphism (RFLP) loci, including R183 for oil and protein content and R79 for maturity. Cumulative distributions of trait values were graphed for those RIL with 'Minsoy' alleles and for those with 'Noir 1' alleles. As already suggested by the alleles found associated with extreme phenotypes, the distributions were consistent with an independent and additive expression of the maturity QTL linked to R79. That is, the cumulative distributions for plants with 'Minsoy' alleles and for plants with 'Noir 1' alleles were similar in shape, but the entire 'Noir 1' curve had been shifted to later maturity dates. In contrast, the trait distributions for a locus affecting oil and protein content linked to R183 were not compatible with an additive model. These results suggest that this approach can be used for rapid identification of QTLs with epistatic expression.

Key words RFLP · QTL · Epistasis · Soybean · Recombinant inbreds

Communicated by A. L. Kahler

K. G. Lark (✉) · L. M. Mansur
Department of Biology, University of Utah, Salt Lake City, UT 84112, USA

J. Orf
Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

Introduction

The simplest models of genetic control of quantitative traits assume the loci act in an additive and independent manner (Falconer 1989). However, as the majority of complex biochemical and cellular processes involve interactions between several components, one may expect genetic expression to be epistatic. We describe here a simple experimental analysis that can provide insight into whether or not a locus is expressed additively or epistatically.

When a quantitative locus is acting independently and additively, we can expect that opposite phenotypes will be correlated with alternate parental alleles of a linked qualitative marker. In particular, we expect this to be true for the two extreme phenotypes (e.g, the greatest will correlate with allele A, the least with allele a). When epistasis occurs, this need not be true. For example, a regulatory gene may condition the expression of only one of the alleles of a structural gene, affecting only one end of the range of possible phenotypic variation.

The use of molecular markers has greatly simplified the genetic analysis of quantitative traits, providing a reliable and extensive framework of qualitative markers to which quantitative trait loci (QTL) can be linked (Stuber 1989). Because quantitative traits are controlled by many loci, the expression of which may be influenced by environmental factors, it is not possible to focus on individual loci with the degree of accuracy that can be brought to bear on qualitative markers. Molecular markers linked to quantiative trait loci can help to define QTL and identify different alleles controlling the quantitative trait (Robertson 1985). We have used these tools to examine a population of recombinant inbred soybeans that show extreme variations resulting from transgressive segregation (Mansur et al. 1993a, b)

In a previous study, we obtained data (i.e., Fig. 2 Mansur et al. 1993a) that suggested that extreme phenotypes were not always associated symmetrically with both parental restriction fragment length polymorphism

(RFLP) alleles (least with one allele, most with the other). In the work described herein, we describe how such observations can be used to identify QTL that are candidates for epistatic expression. For our examples we have chosen maturity and seed oil and protein content, which segregate in this recombinant inbred population soybeans derived by crossing a 'Minsoy' (PI 27890) with 'Noir 1' (PI 290136). The extreme variation encountered in this population makes it ideal for such studies.

Materials and methods

The origin of the 284 recombinant inbred lines (RIL) and the two parents from which they were derived, PI 27890, ('Minsoy'), and PI 290136, ('Noir 1'), have been described previously (Mansur et al. 1993a). A F_9 bulk sample of seeds was used for field experiments in Minnesota during the summer of 1991, and F_{10} seed were used in Chile during the winter of 1991–1992. Details of the design and execution of the field experiments have also been described previously (Mansur et al. 1993a). Maturity was measured as the number of days after 31 July in Minnesota (seeds planted 25 May 1991) or after 29 February in Chile (seeds planted 13 December 1992) when 95% of pods on the main stem had reached mature pod color. Seed oil and seed protein, on a 13% moisture basis, were measured on a 7-g sample of whole seed taken from each plot and ground in a Stine mill before analysis. (Analysis was made at the University of Minnesota with a model 101S Pacific Scientific NIR grain analyzer).

RFLP markers were determined by hybridizing radioactive DNA from plasmid clones of R79 or R183 to fragments of plant DNA separated on agarose gels after digestion with *Taq*I or *Bc*I restriction enzymes, respectively. RFLP markers were scored on between 240 and 250 of the RIL plants.

Extreme phenotypes were selected after computing means across locations for each of the 284 RIL. Those within the extreme 8–9% of the upper and lower end of the range of values for which RFLP alleles had been determined (20 plants in each extreme) were chosen as extreme phenotypes. When heterozygotes were present they were excluded from the genotypes listed in Table 1.

A cumulative distribution was constructed for each allele of R79 or R183 as a function of increasing maturity or seed oil/protein content, respectively. Distributions for 'Noir 1' and 'Minsoy' alleles were compared using a Kolmogorov-Smirnov 2 sample test (Conover 1980) as implemented by the software package Systat (Wilkinson 1988). This test compares cumulative distributions by looking at a statistic based on the maximum difference (d) between the heights of two cumulative distribution curves. Regression analyses of maturity and oil or protein also were carried out using the PC SAS statistical package (SAS 1988).

Results

Previously, we reported the extreme transgressive variation found in this recombinant inbred population for maturity, height, lodging, and yield: the parental plants, 'Noir 1' and 'Minsoy', different in maturity by 2 days, whereas RIL plants with extreme phenotypes differed by 22 days (Mansur et al. 1993a). A similar variation was encountered for seed oil and seed protein content.Values of seed oil for 'Minsoy' and 'Noir 1' were 17.4% and 17.8%, respectively. For the extreme 7% of the RIL plants, they were 15.5% and 19.6%. Comparable differences in variation were obtained for seed protein.

A QTL for maturity linked to RFLP locus R79 represents an example of a trait that conforms to the expectations for a QTL that acts independently. A QTL for seed protein and oil content linked to RFLP locus R183 is expressed differently.

Table 1 presents the frequency of parental alleles for these loci found in plants with extreme phenotypes for these traits. By means of the binomial distribution it is relatively easy to calculate the expectation that the observed distribution of alleles occurred by chance. Low oil high protein appears to be linked to R183 (probability of this occurring by chance <0.001) but not to R79. High oil/low protein does not show significant linkage to either R79 or R183. Both late and early maturity appear to be linked to R79 (probability occurring by chance 0.001 and <0.0001, respectively) but not to R183. Regression analysis over the entire poplation confirmed the linkage of maturity to R79 ($R^2 = 0.449$, $P = 0.0001$) and of seed oil/protein in R183 ($R^2 = 0.144$, $P = 0.0018$).

Figure 1 presents the cumulative distributions of 'Noir 1' or 'Minsoy' alleles of the RFLP locus R79 in the recombinant inbred population as a function of increasing maturity. As predicted from Table 1, the 'Minsoy' alleles are associated with early maturity, the 'Noir' alleles with late maturity. The two curves are different as tested by the method of Kolmogorov-Smirnov ($d = 0.39$ $P = <0.0001$). Each of the two halves of the distributions are also different [early maturity ($d = 0.705$, $P < 0.0001$); late maturity ($d = 0.747$, $P = <0.001$)], which is consistent with an additive model in which 'Minsoy' alleles condition early maturity and 'Noir 1' alleles condition late maturity.

Oil and protein content are inversely related in this RIL population with a correlation of -0.934 ($P = 0.0001$). Figure 2 presents the cumulative distributions of 'Noir 1' and 'Minsoy' alleles as a functions of increasing seed oil or protein. It can be seen that the 'Minsoy' alleles are distributed between the two extremes in a symmetrical fashion and that the curves for oil and

Table 1 Alleles of RFLP loci found in RIL plants with extreme phenotypes for oil/protein or maturity (see Materials and methods for selection of plants with extreme phenotypes)

Extreme phenotype	Low oil high protein		High oil low protein		Early maturity		Late maturity	
Allele Locus	Noir	Minsoy	Noir	Minsoy	Noir	Minsoy	Noir	Minsoy
R79	9	11	7	13	1	19	16	4
R183	3	17	13	7	7	13	12	8

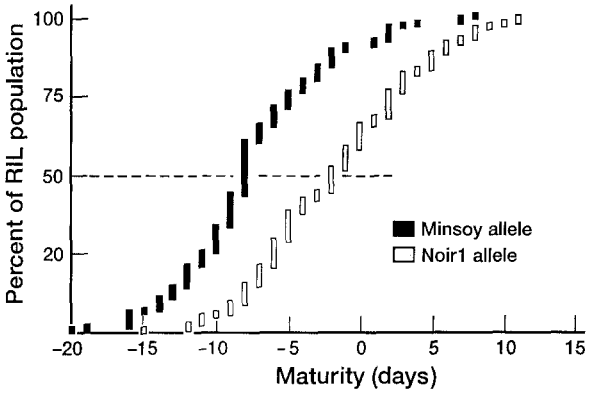

Fig. 1 Cumulative distributions of maturity in RIL plants with 'Minsoy' or 'Noir 1' alleles of the RFLP locus R79. (For measurement of maturity see Materials and methods) On the *ordinate* we have graphed the percentage of plants with a maturity less than or equal to the maturity value shown on the *abscissa*. Plants with 'Minsoy' alleles were graphed separately from plants with 'Noir 1' alleles. The *dashed line* divides the two halves of the population, which were tested separately for differences between the distributions (see text). Note that only a few plants with 'Minsoy' alleles of R79 are earlier than those with 'Noir 1' alleles and that only a few plants with 'Noir 1' alleles are later than those with 'Minsoy' alleles

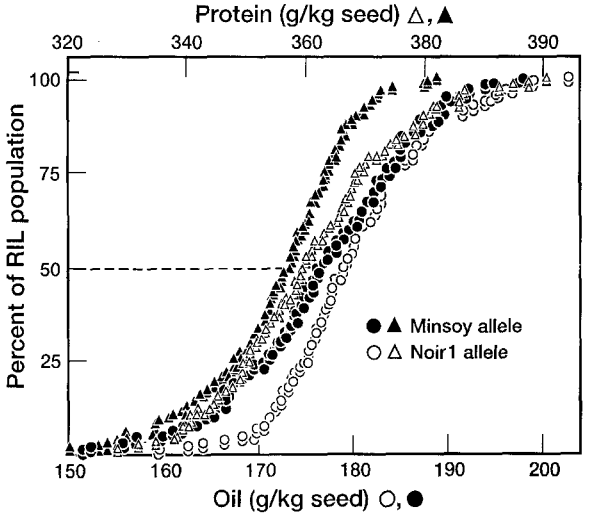

Fig. 2 Cumulative distributions of seed oil and protein content in RIL plants with 'Minsoy' or 'Noir 1' alleles of the RFLP locus R183. (For measurement of seed oil or protein see Materials and methods) On the *ordinate* we have graphed the percentage of plants with an oil or protein content less than or equal to the value shown on the *abscissa*. Plants with 'Minsoy' alleles were graphed separately from plants with 'Noir 1' alleles. The *dashed line* divides the two halves of the population, which were tested separately for differences between the distributions (see text). Note that two types of extreme plants are found with the 'Minsoy' alleles of R183 (high protein-lowoil, or low protein-high oil). However, no plants with the 'Noir 1' allele were found with extremely high protein-low oil

protein content are identical. This is in contrast to the 'Noir 1' alleles for which the curves are asymmetric: at low protein concentrations, the distribution of 'Noir 1' and 'Minsoy' alleles is the same ($d = 0.217$, $P = 0.109$), but at high protein levels the distributions differ ($d = 0.415$, $P = 0.0001$). Conversely, 'Noir 1' and 'Minsoy' alleles share similar distributions at high oil values, but the distributions differ at low values of oil.

Discussion

As found in previous studies of F_3 families (Mansur et al. 1993b) maturity as well as seed oil and seed protein content vary transgressively in segregating progeny derived from hybrids produced by crossing 'Minsoy' and 'Noir 1'. The extent of this variation is great, thereby facilitating the use of plants with extreme phenotypes for identifying QTL linked to particular RFLP markers. In contrast to the results of this previous study, seed oil and seed protein content were found in the present study to be inversely correlated. This is in agreement with data of others (Brim 1973), and the failure to observe this previously is probably due to limiations in the F_3 family data [In Mansur et al. (1993b), plants were examined at one location during only one season].

We have used plants with extreme phenotypes (Table 1) to rapidly identify loci linked to QTLs controlling maturity as well as seed oil and protein content. The data also suggests that different allele distributions were associated with trait values for maturity on the one hand and seed oil or protein content on the other. Examination of the RIL population confirmed this difference, and the distributions suggest that whereas maturity could be controlled by a locus acting independently and additively, seed oil and protein content could not. The asymmetric distribution in Fig. 2 could mean that extremely high values of protein (low values of oil) may be suppressed by the interaction of the 'Noir 1' allele gene product produced by a QTL linked to R183, with the product of some other locus; whereas the 'Minsoy' allele of the same QTL does not interact. As the location of other QTLs for oil/protein are uncovered, cross correlations between these and R183 should serve to clarify the possible epistatic effects of R183 in regulating seed oil and protein. Our results are presented as an example of a rapid method of distinguishing between possible additive or epistatic roles of quantitative trait loci in controlling trait values. By first ascertaining linkage and possible asymmetries in the allele distribution among plants with extreme phenotypes, attention can be focused on a QTL of interest. A detailed examination of the cumulative distribution of linked RFLP alleles can then suggest whether or not the locus is acting epistatically.

Our analysis was facilitated by the use of extreme phenotypes in the population. For this tool to be useful, two parameters of the population are essential. (1) The genotypic variation must be large relative to environmental effects, thereby assuring that alleles of interest will indeed be associated with the extreme phenotypes. For this reason, traits with high heritability are preferable. (2) The population under study should be large, insuring that many plants will be found that have an extreme

phenotype. Homozygous, near-inbred populations are desirable for such studies since heterozygous plants are not informative.

Acknowledgments We thank Fred Adler for his suggestions for analysis of the data in Figs. 1 and 2 and David Berrigan for the statistical analyses of these data. We thank Kevin Chase for technical assistance in RFLP analyses. This research was supported by a grant from the American Soybean Association, by Grant GM42337 to K.G.L. from the NIH, and founds from the Minnesota Soybean Research and Promotion Council.

References

Brim CA (1973) Quantitative genetics and breeding. In: Caldwell BE (ed) Soybeans: improvement, production and uses. Am Soc Agron, Madison Wis., pp 155–186

Conover WJ (1980) Practical nonparametric statistics J. Wiley and Sons, New York

Falconer DS (1989) Introduction to quantitative genetics 3rd edn. Langman group UK, Harlow, England, pp 349–352

Mansur LM, Lark KG, Orf J (1993a) Determining linkage of quantitative trait loci to RFLP markers using bulked DNA from extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). Theor Appl Genet 86:914–918

Mansur LM, Lark KG, Kross H, Oliveira A (1993b) Interval mapping of quantitative trait loci for reproductive, morphological and seed traits of soybean (*Glycine max* L. Merr.). Theor Appl Genet 86:907–913

Robertson DS (1985) A possible technique for isolating genic DNA for quantitative traits in plants J Theor Biol 117:1–10

SAS Institute (1988) SAS procedures guide release, 6.03 edn. SAS Institute, Gary N.C.

Stuber CW (1989) Marker-based selection for quantitative traits. In: Proc XII Congr EUCARPIA. Parey, Berlin Hamburg, pp 31–49

Wilkinson L (1988) Systate: the system for statistics. Systat, Evanston, Ill.