Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another

(epistasis/maximum likelihood/recombinant inbreds)

KARL G. LARK[†‡], KEVIN CHASE[†], FRED ADLER[†§], LEVI M. MANSUR[†¶], AND JAMES H. ORF[||]

Departments of [†]Biology and [§]Mathematics, University of Utah, Salt Lake City, UT 84112; [¶]Facultad de Agronomia, Universidad Catolica de Valparaiso, Quillota, Chile; and [||]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108

ABSTRACT   A large recombinant inbred population of soybean has been characterized for 220 restriction fragment-length polymorphism (RFLP) markers. Values for agronomic traits also have been measured. Quantitative trait loci (QTL) for height, yield, and maturity were located by their linkage to RFLP markers. QTL controlling large amounts of trait variation were analyzed for the dependence of trait variation on particular alleles at a second locus by comparing cumulative distributions of the trait for each genotype (four genotypes per pair of loci). Interesting pairs of loci were analyzed statistically with maximum likelihood and Monte Carlo comparison of additive and epistatic models. For each locus affecting height, variation was conditional upon the presence of a particular allele at a second unlinked locus that itself explained little or no trait variation. The results show that interactions between QTL are frequent and control large effects. Interactions distinguished between different QTL in a single linkage group and between QTL that affect different traits closely linked to one RFLP marker—i.e., distinguished between pleiotropy and closely linked genes. The implications for the evolution of inbreeding plants and for the construction of agronomic breeding strategies are discussed.

Polygenic quantitative traits result from interactions between multiple genes and the environment (1). Detailed genetic analysis of these traits has become possible with the availability of a large number of molecular markers to which quantitative trait loci (QTL) can be linked (2). In addition, estimation of trait values can be greatly improved by the use of recombinant inbred (RI) populations in which individual lines represent stable inbred (therefore homozygous) segregant progeny from a single genetic cross (3, 4). With use of such a population, trait values can be measured repeatedly in different environments, and genetic variation can be separated from environmental effects.

Because it is easy to store and maintain seed, large RI populations of plants can be prepared and maintained. This makes possible the isolation and identification of rare genotypes involving combinations of parental alleles that produce phenotypes with values of quantitative traits more extreme than those of the parents (5). In this paper, we analyze a RI population of soybean derived from a cross between two cultivars, Minsoy and Noir 1, that shows transgressive variation among the segregant progeny (6, 7)—i.e., variation that extends beyond the range of the parents.

We have searched for epistatic effects between QTL that produce large amounts of trait variation. As examples, we have considered QTL that explain large amounts of trait variation, but for which the effect is conditional upon an allele at another
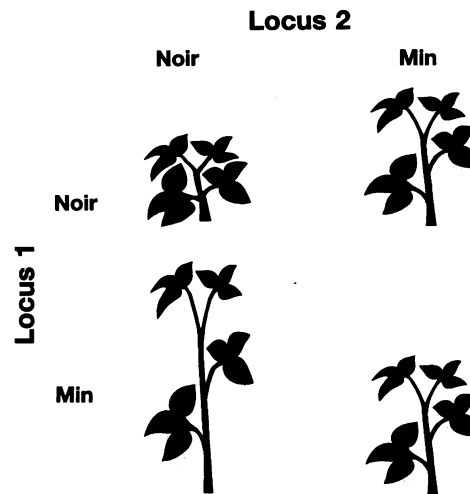

FIG. 1.   Epistasis between QTL regulates trait variation. In this example (see Fig. 2) the trait is height. Variation controlled by alleles at locus 1 is conditional upon the presence of a Noir 1 allele at locus 2.

unlinked locus (which, by itself, is not associated with trait variation). This type of interaction is shown schematically in Fig. 1. In this example, a difference in height produced by different alleles at locus 1 is conditional upon the presence of a Noir 1 allele at locus 2. Only the RI lines with the Noir 1 allele at locus 2 will show differences in height produced by the first locus. To identify such pairs of loci, we chose as the first locus QTL that had been found to explain significant variation in a quantitative trait. We then scanned through unlinked second loci, dividing the population of RI lines into four genotypic classes by pairing each of the alleles at the first locus with each of the alleles at the second locus. These pairwise comparisons have identified second loci with substantial epistatic effects on plant height and on other traits such as seed yield.

MATERIALS AND METHODS

The soybean RI population used here and the measurement of the traits have been described in detail (7). Briefly, 224 different segregants were used that had been derived from crossing the soybean cultivars Minsoy (PI 27890) and Noir 1 (PI 290136). Traits were measured during two seasons in two locations, three replications each (12 replications total), during the F9 to F12 generation of inbreeding. Heritabilities (H^2) were calculated to be 0.93 for height (cm), 0.84 for maturity (days from planting), and 0.77 for yield [kg/hectare (ha)].

Abbreviations: RFLP, restriction fragment-length polymorphism; QTL, quantitative trait locus or loci; RI, recombinant inbred; lod, logarithm of odds.
[‡]To whom reprint requests should be addressed.

Genetics: Lark *et al.*

Proc. Natl. Acad. Sci. USA 92 (1995)   4657

Measurements of segregating restriction fragment-length polymorphism (RFLP) markers have been described (8). These were linked by using MAPMAKER (9). A preliminary genetic map based on F2 segregants from this cross has been published (8), and a more complete map, derived from the RI population, will be published elsewhere. Because a common genetic map reconciling markers from different crosses with chromosome cytology (in soybean, $n = 20$) is not yet available in soybean, the linkage groups in Fig. 4 have not been assigned to particular chromosomes.

QTL data in Table 1 were determined by using analysis of variance (1, 10). Pairwise comparisons of QTL (see in Table 2; see also *Results*) were made by using a computer program devised by K.C., to be published elsewhere in detail. The program displays the cumulative distributions of trait values for the two alleles at each locus (as in Fig. 2 *Top* or *Bottom*), as well as the four distributions corresponding to the four possible genotypes involving the pairs of loci (e.g., Fig. 2 *Middle*). Comparison of the distributions reveals candidates for interactive effects. With this program, it was possible to make visual and quantitative comparisons of any given QTL to *ca.* 200 other loci within a couple of hours.

We used likelihood ratios (11) to statistically compare an epistatic and an additive model. Let AB indicate a genotype with Noir 1 at locus 1 and Minsoy at Locus 2, etc., and denote the estimated mean by μ_{AB}, the sample mean by $\hat{\mu}_{AB}$, and the sample size by n_{AB}. The epistatic model assumes no constraints on the means and computes the logarithm of odds (lod) score associated with the maximum likelihood estimator of each mean. The additive model assumes $\mu_{AA} + \mu_{BB} = \mu_{AB} + \mu_{BA}$ and computes the lod score associated with the maximum likelihood estimator of the means subject to this constraint. The log likelihood function [$\log(L)$], up to an additive constant, in each case is (under the assumption that the distributions are normal and that the variance is given by the uncorrected sample variances $= \hat{\sigma}^2_{AB}$, etc.) as follows:

$$\log(L) = -n_{AA}(\mu_{AA} - \hat{\mu}_{AA})^2/2\hat{\sigma}^2_{AA},$$

$$-n_{AB}(\mu_{AB} - \hat{\mu}_{AB})^2/2\hat{\sigma}^2_{AB},$$

$$-n_{BA}(\mu_{BA} - \hat{\mu}_{BA})^2/2\hat{\sigma}^2_{BA},$$

$$-n_{BB}(\mu_{BB} - \hat{\mu}_{BB})^2/2\hat{\sigma}^2_{BB}.$$

Under the epistatic model, the maximum likelihood estimators of the means are the sample means. Under the additive constraint, the maximum likelihood estimators are as follows:

$$\mu_{AA} = \hat{\mu}_{AA} - \delta c_{AA}/c, \quad \mu_{AB} = \hat{\mu}_{AB} + \delta c_{AB}/c,$$

$$\mu_{BA} = \hat{\mu}_{BA} + \delta c_{BA}/c, \quad \mu_{BB} = \hat{\mu}_{BB} - \delta c_{BB}/c,$$

where $c_{AA} = \hat{\sigma}^2_{AA}/n_{AA}$, etc., $c = c_{AA} + c_{AB} + c_{BA} + c_{BB}$, and $\delta = (\hat{\mu}_{AA} + \hat{\mu}_{BB}) - (\hat{\mu}_{AB} + \hat{\mu}_{BA})$. δ represents the "deviation from additivity."
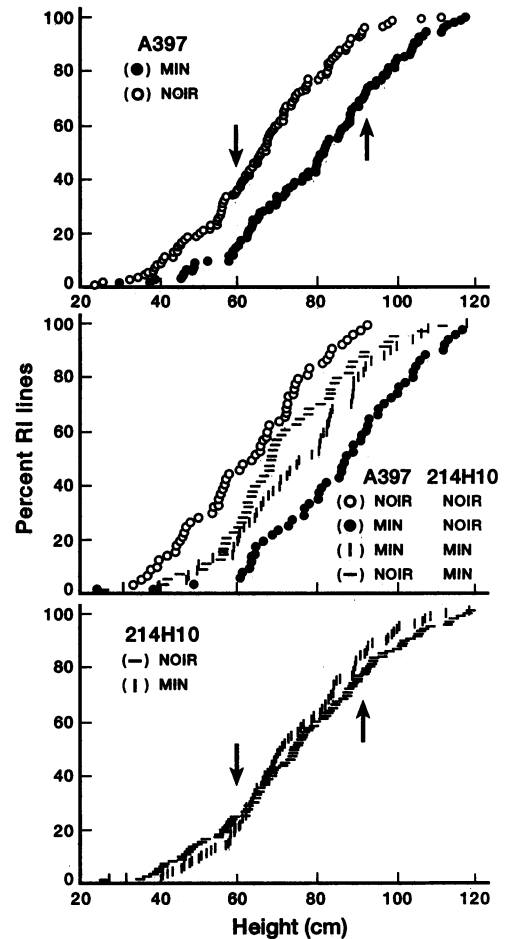

FIG. 2. Cumulative distributions of height for plants with particular genotypes at the pair of RFLP loci, 214H10 and A397. The genotypes are noted in the figure. Mean values of height for Minsoy (MIN) ($\downarrow$) and Noir 1 ($\uparrow$) also are indicated. Similar distributions were obtained when the locus BLT29 or A109a was substituted for A397. These loci are closely linked to A397.

The numbers reported as "lod scores against the additive model" subtract the lod score of the additive model from that of the epistatic model. A small difference indicates that the data can be effectively explained by the additive model, while a large difference indicates that the data are inconsistent with the additive constraint. Because some genotypes and traits do not satisfy the normality and variance assumptions, we conducted Monte Carlo simulations by splitting the distributions representing the effect of the first locus into four "genotypes" to find the probability $P_{(lod)}$ of exceeding the measured difference in lod scores by chance. As a check, we also calculated the probability $P_{(\delta)}$ of exceeding the measured deviation from

Table 1. QTL affecting height

QTL	$P > F$	R^2	Δcm	QTL	$P > F$	R^2	Δcm
QG173tb*	$<10^{-4}$	0.28	21	QA295	0.13	0.02	4.3
QA363b*	$<10^{-4}$	0.13	16	QBLT531	0.79	<0.01	0.7
QA397	$<10^{-4}$	0.13	14	Q214H10	0.72	<0.01	1.9
QL50t3*	$<10^{-4}$	0.11	13	QBLT4	0.44	0.01	3.1
QR79	$<10^{-4}$	0.08	11	QA343	0.03	0.03	5.0
QA60	2×10^{-4}	0.08	11	QL50dr13	0.26	0.01	0.6
QBLT7	0.0016	0.06	09	QG173ta	0.39	<0.01	1.0

QTL are linked to the RFLP loci for which they are named. Δcm was calculated as the difference in mean heights between populations with Minsoy-derived and Noir 1-derived alleles for each locus. Values of $P > F$ and R^2 were obtained from analysis of variance (10).
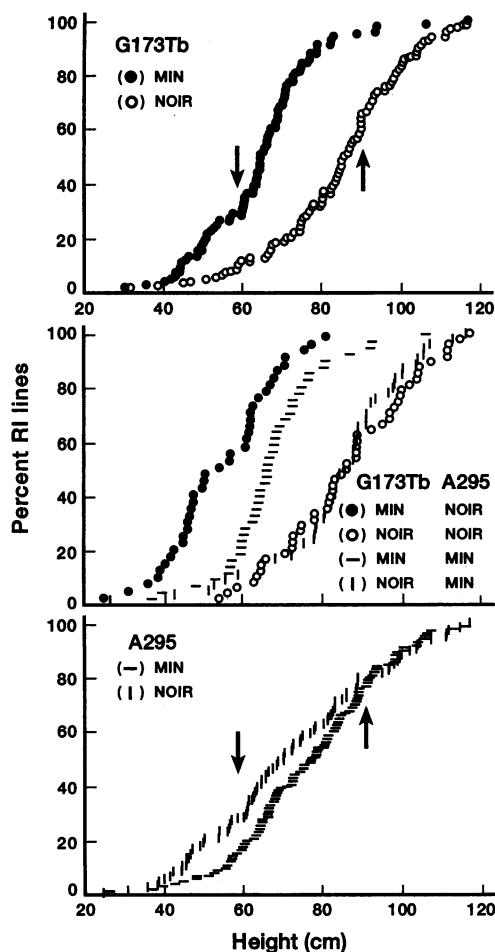*Linked loci (see Fig. 4).

FIG. 3. Cumulative distributions of height for plants with particular genotypes at the RFLP loci G173Tb and A295. Heights of the parents, Minsoy (↓) and Noir 1 (↑), are indicated.

additivity, δ. To correct for the fact that we searched through 32 linkage groups or 220 loci, we transformed the probability P that a single trial exceeded the observed value into $1 - (1 - P)^{32}$ or $1 - (1 - P)^{220}$ to estimate statistical significance.

RESULTS

Interactions Affecting Height. Table 1 lists major QTL that determine height. On the left are listed seven loci linked to QTL that together account for most of the variation in the

segregants obtained from this cross. On the right are seven loci associated with QTL that alone have no significant effect on height. However, as described below, each of these QTL regulates the variation explained by the corresponding locus on the left.

Two examples are presented in detail. Fig. 2 presents cumulative distributions showing the dependence of height on the interaction of specific QTL alleles at loci linked to RFLP markers A397 and 214H10. (Henceforth, we shall refer to QTL by the names of the RFLP loci to which they are linked—i.e., Q214H10 is linked to 214H10.) When analyzed as a single locus, QA397 clearly controls variation in height (Table 1 and Fig. 2*Top*). The Noir 1 allele is associated with short plants, and the Minsoy, with tall plants. In contrast, there is no demonstrable effect of Q214H10, because the distributions of height for populations with the Noir 1 and the Minsoy alleles at 214H10 nearly coincide (Fig. 2*Bottom*). However, Fig. 2 *Middle* shows that the effect of QA397 is conditional upon Q214H10. Plants with the Minsoy allele at 214H10 show no significant variation in height associated with A397, while plants with the Noir 1 allele at 214H10 show a large variation (also see Table 2). Therefore, variation in height due to QA397 is conditional upon the presence of a Noir 1 allele at Q214H10.

Fig. 3 presents an example of a QTL in which only one extreme of variation (short plants) is conditioned by interaction with another locus. Considered in isolation, QA295 has little effect on height and would be discarded in a screen for QTL (Fig. 3*Bottom* and Table 1). In contrast, QG173Tb controls a large fraction of the variation in the population (Table 1 and Fig. 3*Top*). The cumulative distributions in Fig. 3*Middle* show that the Noir 1 allele of QA295 interacts with the Minsoy allele of QG173Tb to produce very short plants, but it has no effect on the Noir 1 allele of QG173Tb (and therefore QA295 has no effect on tall plants). In this interaction, the allele specificity is greater than that shown in Fig. 2 because only the Minsoy allele of QG173Tb is affected.

Table 2 summarizes different interactions that affect height. The mean heights for the genotypes are listed together with the log of the likelihood of explaining the observations with an additive model. In each case the additive model is rejected by a substantial lod score, and the difference between means for alleles at the first locus depends upon the allele at the second locus. For example, in the case of QL50t3 and QBLT4, Δ_M (the difference when BLT4 is Minsoy) is only 1 cm, but Δ_N (the difference when BLT4 is Noir 1) is 23 cm. If the alleles interacted additively, these two values would coincide. However, in all but one case, the difference $\delta = \Delta_M - \Delta_N$ (the deviation from additivity) exceeds the value of Δ in Table 1 (the effect of the first locus alone).

Table 2. Effects of interacting QTL on mean plant heights in cm

		Mean plant height, cm									
		Locus 2 is Minsoy			Locus 2 is Noir 1						
Pairs of loci		Locus 1	Locus 1		Locus 1	Locus 1					
Locus 1	Locus 2	MIN	NOIR 1	Δ_M	MIN	NOIR 1	Δ_N	δ	lod	$P_{(\text{lod})} \times 10^3$	$1 - (1 - P)^n$
G173tb	A295	68	84	16	54	86	32	16	6.8	0.8	0.11 → 0.02**
A363b	BLT531	71	77	6	61	85	23	17	5.3	2.1	0.31 → 0.06*
A397	214H10	75	71	4	86	62	24	20	8.1	0.1	0.02 → 0.004***
L50t3	BLT4	72	73	1	63	86	23	22	8.6	0.1	0.01 → 0.002***
R79	A343	62	84	22	76	79	3	19	5.7	1.4	0.20 → 0.03**
A60	L50dr13	65	83	18	72	75	3	15	4.9	1.9	0.34 → 0.058*
BLT7	G173ta	63	81	18	74	74	0	18	5.4	0.9	0.29 → 0.05*

Pairs of QTL are designated by the RFLP Loci (1 and 2) to which they are linked. The lod scores (naperian logarithms) against the additive model are presented. The probability of obtaining that lod value [$P_{(\text{lod})}$] by random chance was estimated by simulation (see *Materials and Methods*). This probability was adjusted in the far-right column for the number of loci or linkage groups searched (loci → linkage groups). Δ_M, Difference between plant height means when locus 2 is Minsoy; Δ_N, difference when locus 2 is Noir 1; δ, deviation from additivity (the absolute value of the difference between Δ_M and Δ_N); ***, very significant; **, significant; *, marginal significance.
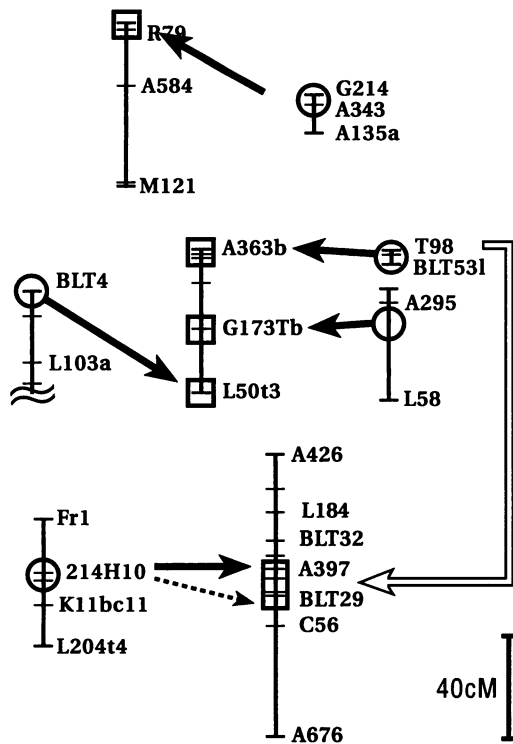
FIG. 4. Interactions between pairs of unlinked loci. Loci that by themselves show trait variation (□) are connected by arrows from loci that regulate them (○). Traits are height (→), yield (⸱⸱⸱→), and maturity (⇒). Linkage groups are not specified (see *Materials and Methods*).

Thus, large deviations from additivity (δ) were observed for all seven pairs of loci in Table 2. As described in *Materials and Methods*, we used Monte Carlo simulation to estimate the probabilities of obtaining the "lod" values shown and to estimate the deviations from additivity (δ) shown. These values $P_{(lod)}$ and $P_{(\delta)}$ were very similar (almost identical) in all but one case [BLT7, in which $P_{(\delta)}$ was 1/10 of $P_{(lod)}$]. The somewhat more conservative values of $P_{(lod)}$ are presented in Table 2. For four of the pairs (*** and **), the deviation from additivity is statistically significant. The deviation of the remaining three is marginal. Tests of these in a new population should confirm or refute their significance. The relationships between five of these seven major height loci and the QTL that regulate their action are shown in Fig. 4. (For simplicity, the interactions involving the less important loci, QA60 and QBLT7, have been omitted.)

Three of the QTL—QL50t3, QG173tb, and QA363b—are in one linkage group. Each of these interacts with a different QTL linked to BLT4, A295, or BLT531, respectively. The interactions are quite specific (data not shown). Thus, QL50t3

is conditional on QBLT4 but is not affected by QA295 or QBLT531, QG173Tb is conditional on QA295 but is not affected by QBLT4 or QBLT531, etc. Because these three linked QTL interact with different loci, they represent separate and distinct QTL controlling height.

Differentiating Traits Linked to a Single RFLP. Interactions also can distinguish between effects on different traits of what might have been regarded as a single QTL on the basis of its linkage to a single RFLP. In Fig. 4, the linked RFLP loci A397 and BLT29 are linked to one (or possibly more) QTL for height ($P > F$, <0.0001). These loci also are linked to QTL for maturity ($P > F$, <0.0001) and yield ($P > F$, =0.001). A simple explanation would be to assume that plants that mature late will be tall and bear more seed. That is, that a single genetic change has pleiotropic effects.

However, interactions occur between QBLT29 and QBLT531 that affect yield but not maturity or height (Fig. 4 and Table 3). This effect is large ($\delta = 393$ kg/hectare) and statistically significant ($P_{(lod)} = 5.3 \times 10^{-4}$). The absence of an effect of QBLT531 on maturity or height indicates that a separate gene may exist which only regulates yield. Different genetic interactions can thus serve to distinguish between different, closely linked, QTL.

DISCUSSION

The interactions described in Figs. 2, 3, and 4 and in Tables 1, 2, and 3 were chosen because they involved one locus that controlled significant trait variation (e.g., the data in the left columns of Table 1) and a second locus that by itself had no demonstrable effect on the trait (e.g., the loci in the right hand columns of Table 1). Thus, the interactive effects that were observed could not be due to independent and additive effects, a conclusion born out by the maximum likelihood rejection of an additive model. An unexpected result was that a large number of interactions were found and that each of these controlled a large amount of variation (Table 2). Clearly such epistatic effects are quite significant.

Fig. 4 summarizes most of the interactions that we have described (A60 with L50dr13 and BLT7 with G173Ta are omitted). The data on height show significant, large effects due to interactions involving different pairs of loci, which account for most of the variation in height segregating in this cross.

Interactions between QTL implicate the action of several genes, where single genes otherwise might have been inferred. Interactions involving the three QTL linked to L50t3, G173Tb, and A363b distinguish these as separate loci. Similarly, the interactions of distinct loci with QTL for yield linked to BLT29 suggest the existence of different QTL linked to BLT29 that may affect either yield or height and maturity, producing a pattern that might be interpreted to be the pleiotropic effects of a single locus. (Without data on epistasis, it would have been simplest to assume pleiotropic effects of a single gene—i.e.,

Table 3. The effects of interactions between QBTL29 and QBTL531 on yield, height, and maturity

| Means of measurement | Single locus effect of BLT29 | | | Interacting loci | | | | | | | | | |
| | | | | BLT531 is Minsoy | | | BLT531 is Noir 1 | | | | | | |
	MIN	NOIR	Δ	BLT29 MIN	BLT29 NOIR 1	Δ_M	BLT29 MIN	BLT29 NOIR 1	Δ_N	δ	lod	$P_{(lod)}$	$1 - (1 - P)^n$
Yield, kg/ha	2054	1846	218	2192	1764	428	1961	1926	35	393	7.2	5.3×10^4	$0.09 \to 0.02$**
Height, cm	82	66	16	86	65	21	79	67	12	9	1.5	—	—
Maturity, days	108.1	103.8	4.3	108.9	104.1	4.8	107.5	103.6	3.9	0.9	0.3	—	—

For comparison, data for allelic effects at the single locus, QBLT29, are presented as well as effects of interactions when BLT531 is Minsoy (MIN) and BLT531 is Noir 1. The means for traits are given as well as the difference (Δ) between the means. The lod scores (naperian logarithms) against the additive model are presented, and the probability of obtaining that lod value [$P_{(lod)}$] by random chance was estimated by simulation. This probability was adjusted in the far-right column for the number of loci or linkage groups searched (see Table 2). Δ_M, Difference between measurements when BLT531 is Minsoy; Δ_N, difference when BLT531 is Noir 1; δ, deviation from additivity (the absolute value between Δ_M and Δ_N); **, significant; ha, hectare.

that a single QTL linked to BLT29 exerts multiple effects on maturity, height, and yield. It is of course always possible that a single locus interacts with a number of different QTL to produce these effects. However, it is difficult to imagine a single gene product with these capabilities.)

Evolutionary Implications. All of the interactions that we have reported occur between loci on different linkage groups. Soybean is an obligate inbreeding plant. In such an organism, pairs of alleles selected at interacting loci belonging to different linkage groups can be maintained. In outcrossing plants, we expect interacting loci either to exert less of an effect or to be linked, the linkage distance reflecting the degree to which selection favors interactions between different alleles.

We have seen that combinations of pairs of loci can exert large effects on quantitative phenotypes. It is tempting to speculate about the role of such interactions in inbreeding plants. If interactions also were conditioned by environmental signals, such loci would provide the plant with a large amount of flexibility in response to environmental change. This type of flexibility could substitute for the flexibility conferred on outcrossing plants by heterosis. Future studies of the response of specific genotypes to environmental stress should establish whether interacting loci play such a role.

Implications for Plant Breeding. Although the results described here were obtained with soybean, it seems likely that other inbred crops, such as rice or wheat, may show similar epistatic effects. Brim (12) noted that different breeding strategies are required when dealing with epistatic QTL. Our results emphasize this, especially for strategies using marker-directed selection. Clearly the phenotypes of offspring will be strongly influenced by the genetic background into which any trait is being moved. For example, an attempt to transfer the QTL for height, linked to A397 (Fig. 2), could result in no change in phenotype without simultaneous transfer of the appropriate allele of the unlinked QTL associated with 214H10.

A detailed analysis of this RI population for epistatic effects on several agronomic traits (i.e., seed protein or oil, seed weight, yield, lodging, etc.) will be published later. However, preliminary analyses indicate that the number and extent of these interactions are similar to those seen with plant height. These substantial contributions to trait values suggest that with selection of appropriate RI genotypes, further genetic crosses can increase or decrease trait values to extremes much greater than those found in this population. Breeders interested in dissecting complex phenotypes such as yield, seed protein, etc., may need to establish and analyze a number of RI populations both to determine epistasis and to provide useful genotypes to serve as breeding stock. Our results for yield (Table 3) suggest that the establishment of such populations could be very valuable.

1. Tanksley, S. D. (1993) *Annu. Rev. Genet.* **27,** 205–233.
2. Dudley, J. W. (1993) *Crop Sci.* **33,** 660–668.
3. Burr, B., Burr, F. A., Thompson, K. H., Albertson, M. C. & Stuber, C. W. (1988) *Genetics* **118,** 519–526.
4. Carrillo, J. M., Rousset, M., Qualset, C. O. & Kasarda, D. D. (1990) *Theor. Appl. Genet.* **79,** 321–330.
5. Mansur, L. M., Lark, K. G. & Orf, J. (1993) *Theor. Appl. Genet.* **86,** 914–918.
6. Mansur, L. M., Lark, K. G., Kross, H. & Oliveira, A. (1993) *Theor. Appl. Genet.* **86,** 907–913.
7. Mansur, M. & Orf, J. (1994) *Crop Sci.*, in press.
8. Lark, K., Weissemann, J. M., Mathews, B. F., Palmer, R., Chase, K. & Macalma, T. (1993) *Theor. Appl. Genet.* **86,** 901–906.
9. Lincoln, S. E. & Lander, S. L. (1993) MAPMAKER/EXP 3.0 and MAPMAKER/QTL 1.1 (Whitehead Inst. of Med. Res. Tech. Rep., Cambridge, MA).
10. SAS Institute (1988) *SAS Procedures Guide Release 6.03* (SAS Inst. Inc., Gary, NC).
11. Bickel, P. J. & Dokson, K. A. (1977) *Mathematical Statistics* (Holden Day, Oakland, CA).
12. Brim, C. A. (1973) in *Soybeans: Improvement, Production and Uses*, ed. Caldwell, B. E. (Am. Soc. Agron., Madison, WI).