Theor Appl Genet (2006) 113: 137–146
DOI 10.1007/s00122-006-0280-8

ORIGINAL PAPER

**James H. Orf · Kevin Chase · James Specht**
**Ik-Young Choi · Perry B. Cregan · Karl G. Lark**

# Abnormal leaf formation in soybean: genetic and environmental effects

Received: 8 August 2005 / Accepted: 24 March 2006 / Published online: 22 April 2006
© Springer-Verlag 2006

**Abstract** Soybean is trifoliolate, but 4-, 5-, or 6-foliolate leaves have been reported and expression of such multifoliolate (MF) leaf forms has been shown to be heritable. Here we analyze the genetic complexity of the MF phenotype and the dependence of its expression on the environment. Recombinant inbred (RI) segregants of soybean were grown in different environments. The frequency of plants expressing the MF phenotype as well as the frequency of nodes exhibiting MF leaves varied with both the environment and the RI segregant genotype. Growth chamber experiments supported field observations suggesting that environment (day length, temperature, etc.) at emergence influenced expression of MF during subsequent growth. Marker facilitated analyses of three RI segregant populations identified quantitative trait loci (QTLs) in 17 regions of the soybean genome. These either directly regulated MF phenotype expression, or were involved in interactions with such loci. Loci, identified in one RI population also were identifiable in another, different, RI population. Most of the loci affected both the frequency of plants expressing MF, and the number of nodes on MF plants that expressed the phenotype. However, a few loci differentiated between these two effects. Many loci affected plants in both field experiments, however, a few differentiated between the two environments. Similar patterns were observed for interactions between loci. QTLs regulating the MF phenotype were located in genome regions that also contained QTLs regulating major agronomic traits—e.g. yield, lodging, etc. This suggests that the loci involved regulate plant growth at some over-arching level, controlling multiple phenotypes or traits.

Communicated by F. J. Muehlbauer

J. H. Orf (✉)
Department of Agronomy and Plant Genetics,
University of Minnesota, 411 Borlaug Hall,
1991 Buford Circle, St. Paul, MN 55108, USA
E-mail: orfxx001@umn.edu
Tel.: +1-612-6258275
Fax: +1-612-6251268

K. Chase · K. G. Lark
Department of Biology, University of Utah,
257 South 1400 East, Room 201,
Salt Lake City, UT 84112, USA

J. Specht
Department of Agronomy and Horticulture,
University of Nebraska, 322 Keim Hall,
Lincoln, NE 68583, USA

I.-Y. Choi · P. B. Cregan
Soybean Genomics Improvement Lab., USDA-ARS,
Beltsville, MD 20705, USA

## Introduction

More than half of all annual plants are inbreeding including most important agronomic species such as soybean [ *Glycine max* (L.) Merrill] (Stebbins 1950; Allard et al. 1968; Allard 1975). As a consequence, they are homozygous and maintain their functional genomes intact during reproduction. Genetic variants that arise, rapidly become homozygous and such variation can be maintained in successive generations. Thus the entire genome can evolve as a unit and interactions between genes distributed throughout the genome can be maintained and selected. Adaptation to changing environments and subsequent selection could utilize the vast reservoir of genetic information with which different loci can interact. In such plants, an interactive network of genes could regulate various traits. Phenotypes associated with a single genetic locus can be masked, enhanced or decreased by interactions with the rest of the genome. These effects of genetic background may alter the appearance of the phenotype and/or change the frequency (penetrance) with which a phenotype is observed.

In previous communications, we have searched for, and found, interactions between soybean quantitative trait loci (QTLs) located on different linkage groups (Lark et al. 1994, 1995; Orf et al. 1999b). In this paper, we take advantage of a qualitative phenotype, presence

or absence of 5- or 7- foliolate leaf forms (Takahashi and Fukuyama 1919; Fehr 1972; Wang et al. 2000), to examine the effect of genetic background on phenotypic expression. Like all of the Glycinae, the normal form of soybean leaves is trifoliolate. However, 4-, 5- or 6-foliolate leaves occasionally are observed at some nodes, on a few plants. We shall refer to the phenotype of five or seven foliolate leaves as multifoliolate (MF) leaf form. The five foliolate phenotype has been assumed to be a qualitative trait with poor penetrance, governed by a dominant allele *Lf1*, on linkage group A2 (U3). A recessive allele at another locus, *Lf2*, generates a seven foliolate phenotype (Fehr 1972). Recently Wang et al. (2000) reported multi-foliolate (MF) phenotypes from a *Glycine soja* (Sieb. and Zucc.) and detected several different loci using crosses between this wild soybean accession and various soybean cultivars. In all, five different loci *Lf1*–*Lf5* have been identified that can control the multi-leaflet trait (MF) in soybean. Although these loci have not been mapped (other than *Lf1*) they have been shown to be genetically distinct.

We have treated the frequency of the MF phenotype as a quantitative trait and searched for loci, or interactions between loci, that regulate phenotypic expression in different environments. For this purpose we have utilized three large populations of recombinant inbred (RI) segregants of soybean each of which has been characterized by a large number of molecular genetic markers.

## Materials and methods

The genetic materials used in the studies were described in detail in Orf et al. 1999a. Briefly, three RI populations from crosses between Minsoy (PI 27890) and Noir I (PI 290136) (MN), Minsoy and Archer (MA), and Noir I and Archer (NA) were evaluated for the multifoliolate leaf form character (MF—i.e. 4-, 5-, or 7-foliolate leaves). The MN population was grown in 1995 in Nebraska, in 2001 at Waseca, Minnesota and in 2002 at Rosemount, Minnesota. In 2001 the MA and NA populations also were grown at Waseca, MN. The Nebraska trial was planted 19 May 1995, the Waseca trial was planted 29 May 2001 and the Rosemount trial was planted 15 June 2002.

The MN RI population was scored for several classical markers, including the MF trait, in Nebraska in 1995 (see, Specht et al. 2001 for details). The populations grown in 2001 and 2002 were scored for the number of plants/row and number of nodes/plant that exhibited the MF trait at growth stage R5 (Fehr and Caviness 1977). Seed from the MN population in 2001 was harvested and used to plant the 2002 trial. The plots were planted as one row plots 3.5 m long with 110 seeds per plot. The plots were spaced 75 cm apart. Weather data were collected at each site and day length recorded based on NOAA published information.

The greenhouse/growth chamber experiments were conducted in two growth chambers and a greenhouse bench. The growth chambers were set to simulate the day length at the planting dates for Waseca 2001 and Rosemount 2002. The greenhouse was set at 15.5 h of day length. Twelve genotypes from the MN population (four that had few or no multifoliolate leaves in both years—Group 1; four that had a large number of multifoliolate leaves in 2002 but few in 2001—Group 2; and four that had a large number of multifoliolate leaves in both 2001 and 2002—Group 3) were planted on 15 October 2003 in the growth chambers and greenhouse in peat pots. Seeds of each genotype were individually planted in peat pots with ten replications. The growth chamber parameters were: Growth Chamber 1–beginning day length 15.35 h, temperature 23°C day/15°C night, light intensity 7000 microEinsteins, relative humidity 80%; Growth Chamber 2—beginning day length 15.19 h, temperature 23°C day/17°C night, light intensity 7000 microEinsteins, relative humidity 80%. The greenhouse was set at 23°C day/17°C night but had more variability than growth chambers, light intensity varied from full sunlight ~ 8500 microEinsteins to supplemental lights ~2100 microEinsteins. The plants in peat pots were transplanted to 15 cm pots with a soil/sand mix with *Bradyrhizobium* inoculum and placed in the greenhouse on 10 November 2003 when the plants from the growth chambers were at growth stage V2. All the plants were grown in the greenhouse until they reached growth stage V7 at which time the MF phenotype data were recorded.

Quantitative trait loci (QTLs) were identified using the simple interval mapping feature of the computer package PLABQTL (Utz and Melchinger 1996). This program uses a multiple regression approach to interval mapping with marker order and distances determined by Mapmaker (Orf et al. 1999a). We established empirical LOD thresholds for QTL detection using permutation tests (Churchill and Doerge 1994). Five thousand simulations were used to estimate a LOD threshold of 3.2 for a genomewide significance of 0.05. Interactions between QTLs were analyzed using the Epistat computer program (Lark et al. 1995; Chase et al. 1997). This computer software uses maximum likelihood methods for both the identification and evaluation of significance of interactions between pairs of QTLs (Chase et al. 1997). *P* values were Bonferroni adjusted for the number of trials. Analyses of variance were used to partition the total variance into genetic and environmental components. Broad sense heritability estimates (Hanson et al. 1956) were computed as:

$$H^2 = \frac{s_G}{s_G + s_e/r}$$

Where $H^2$ = heritability, $s_G$ = genotypic variance, $s_e$ = error and $r$ = the number of reps for the trait.

## Results and discussion

Three field experiments demonstrated that the expression of the MF character was complex and that this complexity might be attributable to the interaction between the genome and the environment. The MN RI population was scored for classical markers in Nebraska in 1995. QTL analysis of the MF phenotypic data indicated that the MF trait might be polygenic (see below). In 2001, a detailed experiment was carried out in which all three RI populations (MN, MA and NA) were grown in the same Minnesota environment and the phenotype scored as the plants/row and the nodes/plant that exhibited the MF trait. This experiment was repeated at a different Minnesota location in 2002 using only the MN population.

## Phenotypic data

The data from all of these field experiments are summarized in Table 1, which lists the means, standard deviations, and maximum percentage of MF values observed for different populations in different environments. In Nebraska, the segregant lines were scored qualitatively, i.e. 1 if any penta-foliolate plant was observed in the segregant line, 0 if the RI line did not exhibit the MF character. The Minnesota scoring involved measuring the quantitative value of the character as the percentage of the plants (Pl/row) that exhibited the MF phenotype as well as the percentage of nodes/plant that were MF (Nd/Pl). In 2001 the phenotype was scored on two replicates of RI lines being grown for yield tests. In that experiment the repeatability of the scores could be evaluated, comparing the replicates (MF-plants/row: MN, 0.67; NA, 0.75; MA, 0.78). This level of replicability indicates that the large standard deviations in Table 1 were the result of differences between the RI lines— i.e. differences in genotype. The coefficients of variation (implicit in the data in Table 1) indicate that the variation between RI lines in the three RI populations in 2001 was similar, although more MF plants (with more affected nodes) were observed in the MN population. In 2002 only one replicate was evaluated. A heritability value, $H^2 = 0.7$, was calculated for both plants/row and nodes/plant using the 2001 and 2002 MN experiments. This value is about the same as the estimates of repeatability obtained in 2001 and supports the conclusion that genotypic effects regulate the expression of the phenotype.

Details of the Minnesota 2001 experiments are presented in Fig. 1. Expression of the MF trait varied between different RI lines ranging from no MF plants to as many as 85% of the plants in some RI lines of the MN population. Those RI lines in which MF plants exceeded 10% plants/row had varying numbers of MF nodes ranging from 10 to 35% per plant. RI lines with fewer than 10% MF plants/row had a smaller number of MF

**Table 1** Mean values (and standard deviations) of the MF phenotype for the indicated RI populations grown in the given environments. For each RI line the range of values extended from zero to the indicated MF maximum

| RI population: Parents[a] | Minsoy × Noir I | | | | | | | Noir I × Archer | | Minsoy × Archer | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Size: no. segregants | 236 | 237 | | 236 | | 236 | | 237 | | 230 | |
| Environment | Nebraska[b] | Minnesota '01 | | Minnesota '02 | | Average Minnesota '01 + '02 | | Minnesota '01 | | Minnesota '01 | |
| Trait | MF PL/row | MF PL/row | MF node/PL | MF PL/row | MF node/PL | MF PL/row | MF node/PL | MF PL/row | MF node/PL | MF PL/row | MF node/PL |
| Mean (SD) | 0.5 (0.5) | 13 (18) | 10 (8) | 25 (38) | 20 (16) | 19 (26) | 15 (11) | 8 (9) | 8 (12) | 6 (10) | 6 (9) |
| LF max (%) | 31 | 85 | 35 | 100 | 90 | 92.5 | 50 | 40 | 80 | 70 | 30 |

[a]Parents: (2001 data) Minsoy MF PL/row 5%, MF Node/PL 20%; Noir I MF PL/row 10%, MF Node/PL 10%, Archer none

[b]Nebraska data rated on a 0 or 1 scale

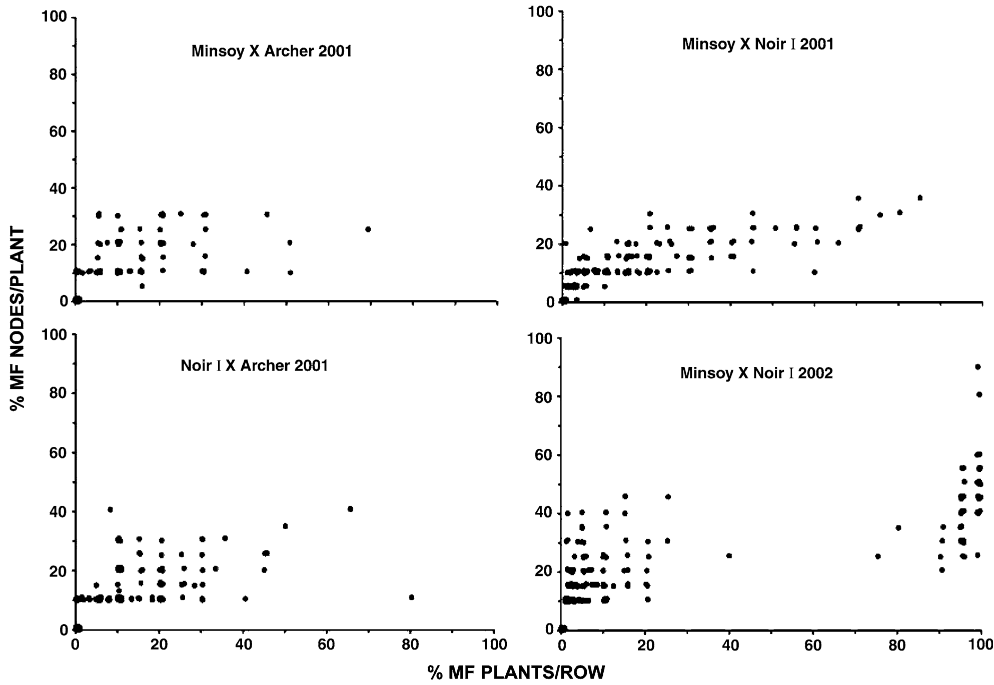


**Fig. 1** Scatter graphs showing the frequency of MF nodes per RI segregant as a function of the frequency of MF plants/row. Data are shown for the MN, MA and NA populations grown in the 2001 environment as well as for the MN population in 2002. The numbers of segregants without MF plants in 2001 were: MN-55; NA-107; MA-147. In 2002 there were 34 MN segregants without MF plants

nodes/plant (ranging from 1 to 10%). It seemed clear that for the majority of the plants in all of the RI populations there was only a small correlation between the MF-nodes/plant and the MF-plants/row. This suggests that MF is not simply the number of nodes per plant row that express the trait, but that expression is determined by at least two loci; the ability of a plant to express the trait and the frequency of nodes subsequently expressed.

When the experiment was repeated in 2002 using the MN population, a different pattern of expression was observed. RI lines were distributed bimodally with respect to the MF plants/row phenotype. Many of the lines expressed the MF character in a large percentage of plants (> 90%), and these plants also had a relatively high frequency of nodes with MF leaves. Another group exhibited the character in only a few plants and fewer nodes in these plants were affected.

Many of the RI lines that expressed a high frequency of MF plants/row in 2002 had exhibited a relatively low expression frequency in 2001 (Fig. 2a). Although there was some correlation between lower MF frequencies in both years, there was little correlation between the two years in the frequency of MF nodes/plant found in individual RI lines (Fig. 2b). A striking observation was the contrast between the distribution of MF plants/row in 2002 (bimodal) and the distribution of MF nodes/plant (unimodal). One explanation, suggested by the data, was that the environment at the time of emergence might influence the ability to form multifoliates during subsequent growth. Differences in the environments at

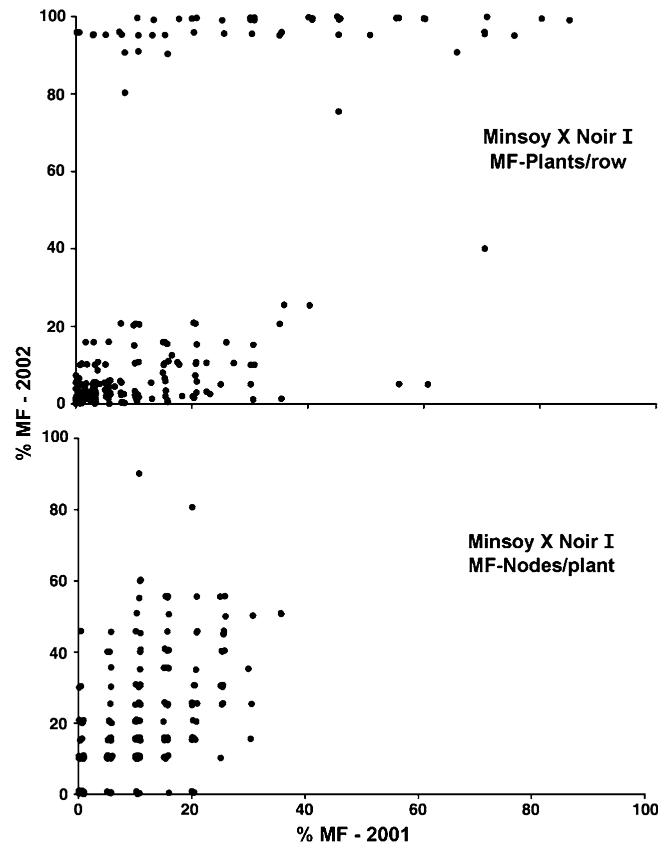


**Fig. 2** Comparison of the 2001 with the 2002 Minnesota environments. MN population: **a** MF plants/row and **b** MF nodes/plant

emergence would include the date of planting, and location, leading to altered photoperiods, as well as the temperature and humidity of the two field environments during the short period in which plants emerged.

To further test the dependence of the MF character on the environment, we compared different MN RI lines during emergence in different growth chamber environments followed by subsequent growth in a common environment. For this purpose, 12 RI lines from the MN population were selected (Table 2) based on their phenotypic behavior in 2001 and 2002. Four of these exhibited infrequent MF plants/row in both years (group 1); another four had exhibited infrequent MF plants/row in 2001 but frequent MF plants/ row in 2002 (group 2); four others exhibited high frequencies of MF plants in both years (group 3). Plantlets emerged in three different environments (two growth chambers and the greenhouse) and were transferred at the second trifoliolate stage to the greenhouse. All were grown in the common greenhouse environment until the seventh to ninth trifoliolate stage and analyzed for occurrence of MF character at different nodes. (See, Materials and methods for numbers of plants, details of potting and growth chamber conditions—Table 2).

As expected, the three groups differed significantly in the expression of the MF trait. Group 1 was extremely low (6/200 MF plants, data not shown) and no conclusions could be drawn about environmental effects. Expression in group 2 (MN 9, 49, 164 and 328) was moderately low but again there were too few MF plants to draw conclusions about the effect of emergence environment on MF expression. However, the difference between lines in group 3 is highly significant (Anova $P < 0.0001$) demonstrating the dependence of the expression of the MF phenotype on genotype.

Little can be said about the effect of emergence environment on groups 1 and 2 because of the low frequency of the trait. However, group 3 (MN 30, 181, 189 and 198) showed significant effects of the three emergence environments on subsequent MF expression during growth in the common greenhouse environment. For both number of MF plants and the number of nodes per MF plant grown in the common greenhouse environment, emergence in the greenhouse was less predictive of MF plants than the first growth chamber environment ($P < 0.0001$), which in turn produced significantly less MF expression than emergence in the second growth chamber ($P < 0.0001$). The effects observed were the same for the number of MF plants and the number of MF nodes per plant. Moreover, the same differences remained significant ($P < 0.0001$) when the MF expression on only the sixth and seventh trifoliolate nodes were compared. Although this experiment did not reproduce the results from the 2002 field experiment, it demonstrated unequivocally that the environment at the time of emergence influences subsequent MF expression during growth of the plant.

## Genotypic data

Only 19 RI segregant lines failed to express the MF phenotype in at least one of the three environments in which the MN population was grown. Thus, less than 10% of the MN population appeared to be genetically restricted to trifoliolate leaves, supporting the assertion of Wang et al. (2000) that several loci may regulate the MF phenotype.

### Identification of QTLs

In each environment expression of the MF character segregated, allowing us to identify a total of ten QTLs (Table 3, bold type). Presence or absence of $> 3$ foliolateness was scored on a per RI line basis in Nebraska, whereas MF nodes/plant and MF plants/row were scored in Minnesota. The results in Table 3 and Fig. 3 present the QTLs as genomic regions, each containing segregating haplotypes with one or more genetic loci.

**Table 2** MF expression in plants from different MN RI lines grown in different environments

| Green- house groups | RIL | Plants/row | | | | | Nodes/plant | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Field | | Greenhouse | | | Field | | Greenhouse | | |
| | | WA 2001 | RO 2002 | GH | WA | RO | WA 2001 | RO 2002 | GH | WA | RO |
| 2 | MN9 | 2.5 | 95 | 0.02 | 0.06 | 0.04 | 5 | 35 | 0.04 | 0.12 | 0.12 |
| | MN49 | 0 | 95 | 0.00 | 0.00 | 0.00 | 0 | 30 | 0.00 | 0.00 | 0.00 |
| | MN164 | 5 | 95 | 0.20 | 0.20 | 0.30 | 10 | 45 | 0.38 | 0.58 | 0.50 |
| | MN328 | 2.5 | 95 | 0.02 | 0.00 | 0.06 | 5 | 45 | 0.02 | 0.00 | 0.10 |
| 3 | MN30 | 80 | 99 | 0.10 | 0.70 | 0.54 | 30 | 50 | 0.24 | 1.68 | 1.10 |
| | MN181 | 60 | 100 | 0.10 | 0.76 | 0.62 | 25 | 40 | 0.56 | 2.38 | 1.72 |
| | MN189 | 60 | 100 | 0.42 | 0.86 | 0.68 | 10 | 50 | 0.88 | 3.40 | 2.42 |
| | MN198 | 85 | 100 | 0.29 | 0.88 | 0.73 | 35 | 50 | 0.90 | 4.18 | 2.43 |

RI lines were selected (see text) based on their MF phenotypes in the field experiments of 2001 and 2002. (The field phenotypes are presented in the first two columns of each group.) After emergence in the three different environments (see Materials and methods) 50 plants from each RI line were transferred to the greenhouse environment where they continued to grow to the 7–9 trifoliolate stage. These were scored for the presence of the MF trait (percentage of plants exhibiting MF) and for the frequency of MF nodes per MF plant

**Table 3** Normalized mean values of traits associated with parental haplotypes. (*M* Minsoy; *N* Noir I; *A* Archer). Haplotypes are described in the first three lines of the table: Linkage group: Utah; (USDA/Iowa), region of significant association and nearest marker. Trait values were normalized to a population mean of 0. Means of parental haplotypes (M, N, or A) are presented. Higher values represent more multifoliolate plants (or leaf nodes)

**Bold: Significant discovery (Identified a QTL)**
QTL effective in one environment, not in another

Underlined: difference > 0.25 (Confirming an identified QTL)
QTL more effective in one environment than in another

| **Linkage Group** | | | U2 (E) | U3 (A2) | U6 -1 (N) | U6-2 (N) | U6 -3 (N) | U12-1 (D2) | U12-2 (D2) | U14-2 (L) | U14-3 (L) | U22 (C1) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Map Region** | | | 94-104 | 104-118 | 40-62 | 75-90 | 102-110 | 70-78 | 102-110 | 100-112 | 130-138 | 116-130 |
| **Nearest Marker** | | | Satt045 | Satt508 | Satt080 | Sat_091 | Satt022 | Satt397 | L204_2 | G173_1 | A802_2 | Satt180 |
| **Environment** | **Trait** | **Parent** | | | | | | | | | | |
| **NE** | | **N** | 0.13 | **0.30** | **-0.29** | -0.23 | **-0.34** | 0.16 | 0.09 | 0.13 | 0.04 | **-0.25** |
| | | **M** | -0.15 | **-0.35** | **0.28** | 0.30 | **0.36** | -0.15 | -0.08 | -0.17 | -0.05 | **0.19** |
| **MN 2001** | **Plants/row (P/R)** | **A** | -0.06 | -0.18 | -0.01 | 0.18 | 0.2 | 0.01 | -0.01 | -0.09 | -0.05 | 0.01 |
| | | **N** | 0.06 | 0.19 | 0.01 | -0.13 | -0.14 | -0.05 | 0.00 | 0.12 | 0.07 | -0.01 |
| | | **A** | 0.18 | 0.02 | 0.07 | -0.10 | -0.09 | 0.00 | 0.02 | 0.02 | 0.02 | **-0.30** |
| | | **M** | -0.19 | -0.08 | -0.06 | 0.05 | 0.13 | 0.00 | -0.06 | -0.02 | 0.05 | **0.27** |
| | | **N** | 0.17 | **0.28** | **-0.38** | **-0.33** | **-0.42** | 0.17 | 0.05 | **0.23** | 0.15 | -0.16 |
| | | **M** | -0.19 | **-0.32** | **0.33** | **0.35** | **0.35** | -0.15 | -0.05 | **-0.29** | -0.21 | 0.12 |
| **MN 2002** | **Pl/R** | **N** | 0.18 | **0.24** | -0.24 | -0.21 | **-0.31** | **0.29** | **0.26** | 0.03 | 0.01 | -0.21 |
| | | **M** | -0.21 | **-0.29** | 0.23 | 0.25 | **0.32** | **-0.27** | **-0.23** | -0.04 | -0.01 | 0.15 |
| | **N/P** | **N** | **0.24** | **0.24** | **-0.3** | **-0.29** | **-0.31** | **0.23** | 0.13 | 0.07 | 0.03 | **-0.31** |
| | | **M** | **-0.28** | **-0.29** | **0.27** | **0.29** | **0.35** | **-0.21** | -0.12 | -0.09 | -0.04 | **0.23** |
| **MN 2001** | **Nodes/plant (N/P)** | **N** | 0.16 | **0.25** | **-0.31** | **-0.37** | **-0.38** | 0.10 | 0.07 | 0.16 | 0.09 | **-0.21** |
| | | **M** | -0.19 | **-0.29** | **0.35** | **0.37** | **0.38** | -0.09 | -0.07 | -0.21 | -0.12 | **0.16** |
| | | **A** | 0.14 | -0.07 | -0.01 | -0.03 | -0.08 | -0.01 | -0.03 | 0.04 | 0.08 | -0.19 |
| | | **M** | -0.15 | -0.06 | 0.01 | 0.03 | 0.11 | 0.01 | -0.01 | -0.05 | -0.06 | 0.17 |
| | | **A** | -0.08 | **-0.30** | 0.08 | **0.17** | 0.22 | -0.06 | -0.04 | -0.21 | **-0.20** | 0.00 |
| | | **N** | 0.08 | **0.3** | -0.11 | **-0.20** | -0.15 | -0.14 | 0.02 | 0.29 | **0.26** | 0.00 |

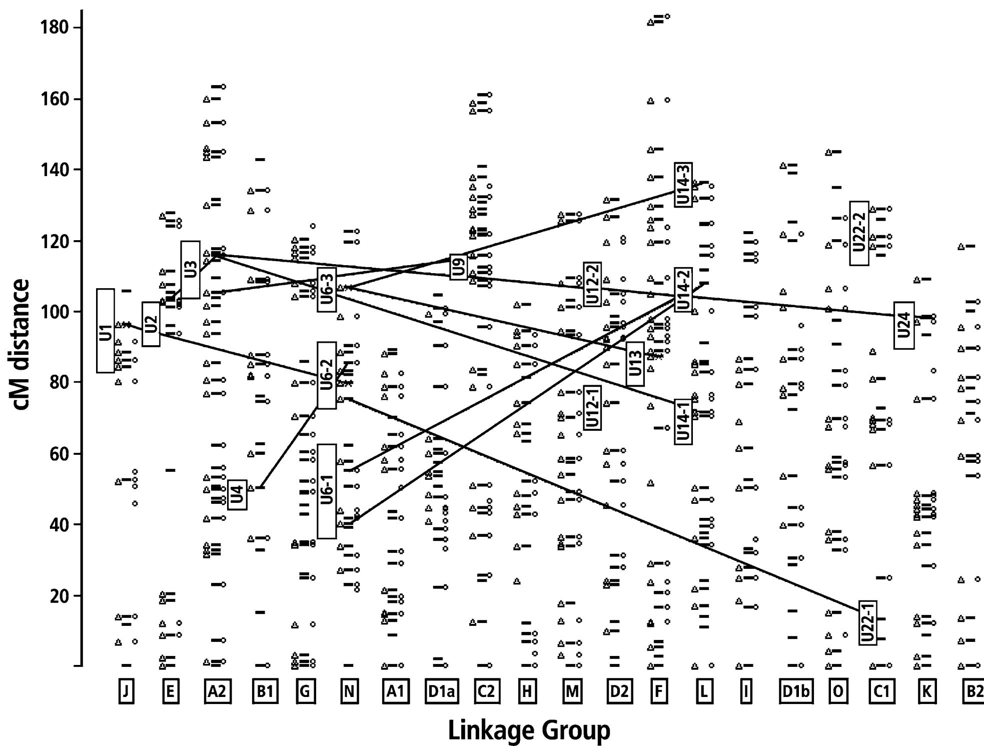


**Fig. 3** A map of the 20 soybean linkage groups showing the distribution of genotypic regions regulating the MF phenotype [MA (*triangle*); MN (*dash*); NA (*circle*)]. Multiple regions were identified on linkage groups U6 (N), U12 (D2) and U14 (L). Interactions between loci are indicated by lines connecting regions of the genome. For further details, see text and Tables 3 and 4

Bounds for the segregating haplotypes are listed in Table 3 and indicated by the boxed regions in Fig. 3. In the MN population, four QTLs were identified from the Nebraska data (bold), associated with three different linkage groups (LGs) (the Utah linkage group is given first and the USDA/Iowa linkage group in parentheses, Cregan et al. 1999). One of these, on U3 (A2), was the locus previously identified as *Lf1*. Two additional loci were found on U6-1 and 3 (N) and one on U22 (C1). MN data from the Minnesota experiments in 2001 and 2002 confirmed these and identified an additional five QTLs (bold) one each on U2 (E), U6-2 (N) and U14-2 (L) as well as two on U12-1& −2 (D2). Noir I alleles at six loci, including the A2 locus *Lf1*, increased the number of multifoliolates, in contrast to four loci located on U6 (N) and U22 (C1) at which Noir I alleles decreased multifoliolates.

In 2001, the MA and NA populations also identified QTLs. However, there were few: three in the NA but only one in the MA population that lacked the *Lf1* allele on U3 (A2). Three of these (on U3 (A2), U6-2 (N) and U22 (C1)), already had been identified in the MN population, either in 2001 or 2002, the fourth, on U14–3 (L) was identified (bold) in the NA population. Archer haplotypes were similar to Minsoy at five loci and to Noir I at the other two.

## Confirmation in other RI populations

Table 3 presents normalized mean values for the multifoliolate trait associated with haplotypes at the different loci. It can be seen that six of the seven loci identified in Minnesota in 2001 were confirmed in one of the other RI populations grown that year (underlined or bold). Loci U3 (A2), U6-2 (N), U6-3 (N), U14-2 (L) and U14-3 (L) could be identified or confirmed either as nodes/plant or plants/row in the MN and NA RI populations (bold and underlined values) and the locus on U22 (C1) was identified in both the MN and MA populations. Moreover, the locus on U2 (E) identified in 2002, was confirmed in the MA population in 2001. Thus seven loci were confirmed in a second, genotypically different, segregant population. The fact that three loci U6-1 (N), U12-1 (D2) and U12-2 (D2) did not segregate in either Archer RI population suggests a dependence of these loci on the rest of the genome for expression.

## Quantitative trait loci regulation of plants/row or nodes/plant in different environments

Of the ten QTLs identified, seven exhibited segregation in all three environments (Table 3, those with bold or underlined values; U2 (E), U3 (A2), U6-1 (N), U6-2 (N), U6-3 (N), U12-1 (D2) and U22 (C1)). Of the remaining three (Table 3, heavy boxes) one, U12-2 (D2), only segregated in Minnesota in 2002; another, U14-3 (L), only in Minnesota in 2001 and a third, 14–2 (L), in both Nebraska and in Minnesota 2001 but not in Minnesota 2002. In addition, three other loci, [(U6-1 (N), U6-2 (N) and U12-1 (D2) (Table 3,light boxes)] exhibited substantial differences between the amount of phenotypic variation that they controlled in Minnesota 2001 and 2002. These data suggest that these six different QTLs contributed to differences in phenotypic expression (hence penetrance) between Minnesota 2001 and 2002 (Fig. 2).

In the MN population nine of the ten QTLs affected both plants/row and nodes/plant, either in 2001 or 2002. The remaining locus, U14-3 (L), affected the number of plants/row in the MN population in 2001 and the number of nodes/plant in the NA population. Thus, all of the loci regulate variation in both aspects of the phenotype, plants/row and nodes/plant. However, the extent to which these aspects are influenced varied between environments.

## Interactions between loci

Using epistat we identified several significant interactions between QTLs. These are presented in Table 4 and shown in Fig. 3. Seven significant interactions ($P < 0.05$) were identified (Table 4 **). In addition three possible interactions ($0.1 > P > 0.05$) are noted (Table 4 *). Two of the significant interactions were confirmed in a second RI population—either MN or NA (Table 4; ** and +). One, between U3 (A2) and U14-1 (L), could only be found in the MN population (Table 4; * and −). Possible interactions between U2 (E) and U3 (A2) and between U3 (A2) and U9 (C2) also only were found in the MN population (Table 4; * and −). The presence or absence of the other five interactions could not be determined in the NA or MA populations (Table 4; 0), because of the absence of markers in critical regions of their genetic maps. The lower marker density in the NA and MA maps can be seen in Fig. 3 by comparing the marker frequencies on the left [MA(*triangle*)] or right [NA(*circle*)] with that in the center [MN(*dash*)].

Two loci in Table 4 are involved in multiple interactions: U3 (A2), the region containing *Lf1*, interacts with four other unlinked loci on U2 (E), U9 (C2), U14-2 (L) and U24 (K); and U6-2 (N) interacts with the loci on U1 (J), U4 (B1) and U22 (C1). Table 5 presents normalized phenotypic means for the four homozygous genotypes involved in each of these two locus interactions. The interaction between loci on U3 (A2) and U2 (E) appears to be haplotype specific, in that Noir I haplotypes at both loci interact to increase the frequency of multifoliolates. In the interactions between loci on U3 (A2) and U9 (C2) or U3 (A2) and U24 (K) variation between haplotypes on U3 is conditional on the specific haplotype of the U9 (C2) or U24 (K) locus (Noir I or Archer respectively). The situation is reversed for the interaction between the U3 (A2) and U14-1 (L) loci. In this case, variation between haplotypes of the U14-1 (L) locus is

**Table 4** Interactions between MF QTLs. The first two columns identify pairs of interacting QTLs (see Fig. 3). These were identified in the RI populations shown (MN, MA, or NA)

| | | MN | MA | NA |
|---|---|---|---|---|
| U1 (J) (Satt431) | U6-2 (N) (Satt549) | ** | – | + |
| U2 (E) (B124_3) [Sat_045][a] | U3 (A2) (Sat_131b) [Sat_040][b] | * | – | – |
| U3 (A2) (Sat_131b) [Sat_040] | U14-1 (L) (Satt156) | ** | – | – |
| U3 (A2) (Sat040) | U24 (K) (A661_1) | + | – | ** |
| U3 (A2) (Satt329) | U9 (C2) (Satt277) | * | – | – |
| U4 (B1) (Satt197) | U6-2 (N) (BLT015_1) | * | – | 0 |
| U6-1 (N) (Satt530) | U14-2 (L) (G173_1) | ** | 0 | 0 |
| U6-2 (N) (Satt521) | U22 (C1) (A463_1) | ** | 0 | 0 |
| U6-3 (N) (Satt022) | U13 (F) (A186_1) | ** | – | 0 |
| U6-3 (N) (Satt022) | U14-3 (L) (A802_2) | ** | 0 | 0 |

(+) Confirmation of pre-identified interaction, (–) no such confirmation, (0) insufficient data for confirmation
**Significant discovery < 0.05; *suggestive discovery *P* value 0.05–0.1
[a]Sat_045 closely linked to B124_3 and could be used as an alternative marker
[b]Sat_040 closely linked to Sat_131b and could be used as an alternative marker

conditional on the presence of a Minsoy haplotype at the U3 (A2) locus. Similar conditional relationships are observed in the interactions involving the U6-2 (N) locus. Variation at the U6-2 (N) is conditional upon the Minsoy haplotype of the locus on U1 (J) as well as the Noir I haplotype of the locus on U22 (C1).

In all the MN examples involving either U3 (A2) or U6-2 (N) loci, the same genotype to phenotype relationship was observed in both of the Minnesota environments (2001 and 2002). This was not the case for the interaction between the loci at U6-3 (N) and U14-3 (L). In this case, a haplotype specific interaction between the Minsoy haplotype at the U6-3 (N) locus and the Noir I haplotype at the U14-3 (L) locus increased multifoliolate frequency in the 2001 environment, whereas variation between haplotypes at the U6-3 (N) locus was conditional upon the Noir I haplotype of the locus on 14-3 (L). An identical difference between environments was observed for the interaction between the U6-1 (N) and the U14-2 (L) loci (data not shown). Of these four loci, three (U6-1 (N), U14-2 (L) and U14-3 (L)) had been observed to control different amounts of variation in 2001 as opposed to 2002 (boxed values in Table 3). Thus, interactions between loci as well as the independent actions of individual loci are responsible for the multifoliolate phenotypic variation observed in different environments.

The QTLs that we have identified, account for about 30% of the total MF variation in the MN population but about 50% of the heritable variation. QTLs in the NA or MA populations accounted for a much smaller fraction of the variation (5–10%) despite the fact that

**Table 5** Normalized mean values of traits associated with interacting parental haplotypes

| | | Minn '01 | | Minn '02 | |
|---|---|---|---|---|---|
| Pairwise Genotypes | | $N_1N_2$ $M_1N_2$ | $N_1M_2$ $M_1M_2$ | $N_1N_2$ $M_1N_2$ | $N_1M_2$ $M_1M_2$ |
| Locus 1 | Locus 2 | | | | |
| U2 (E) | U3 (A2) | 0.62 | −0.29 | 0.59 | −0.27 |
| | | −0.18 | −0.21 | −0.18 | −0.19 |
| U3 (A2) | U9 (C2) | 0.41 | 0.02 | 0.47 | 0 |
| | | −0.43 | −0.11 | −0.45 | −0.14 |
| U3 (A2) | U1 (J) | 0.26 | 0.21 | 0.19 | 0.25 |
| | | −0.09 | −0.53 | −0.05 | −0.55 |
| U3 (A2) | U24 (K) | 0.55[a] | 0.05[c] | | |
| | | −0.65[b] | −0.01[d] | | |
| U3 (A2) | U14-1 | 0.26 | 0.21 | 0.19 | 0.25 |
| | | −0.09 | −0.53 | −0.05 | −0.55 |
| U1 (J) | U6-2 (N) | −0.18 | −0.18 | −0.14 | −0.06 |
| | | −0.42 | 0.61 | −0.31 | 0.4 |
| U6-2 (N) | U22 (C1) | −0.42 | −0.11 | −0.28 | −0.08 |
| | | 0.48 | −0.09 | 0.44 | −0.21 |
| U6-3 (N) | U14-3 (L) | −0.29 | −0.2 | −0.46 | 0.05 |
| | | 0.53 | −0.22 | 0.42 | −0.09 |

Notation for Noir–Archer genotypes (U3(A2)-U24(K)) are presented below. QTL interactions are identified (see, Table 4) as Locus 1 and 2 and mean values for pairwise genotypes are presented. Trait values were normalized as in Table 3
*M* Minsoy, *N* Noir I, *A* Archer
[a]$N_1N_2$, [b]$A_1N_2$, [c]$N_1A_2$, [d]$A_1A_2$

the coefficients of variation of these RI populations (see Table 1) were similar to or larger than that of the MN population. It would appear that in the Archer populations many genetic loci remain to be discovered that either occur in regions of the genome that lack genetic markers (gaps in the maps for these populations) or are loci that regulate amounts of variation too small to be identified in populations with these numbers of plants. Whatever the reason, the results with these populations indicate that the MF phenotype probably involves many more loci than the ten that we have identified.

## Hox genes

Hox genes have been implicated in leaf morphology in several studies (Doebley and Lukens 1998; Gourlay et al. 2000; Bharathan et al. 2002). SNPs were developed for five such markers using sequence information in genbank. One of these, segregating in the MA population, mapped to U21 at 86 cM midway between Satt_478 and Sat_592 (85 cM on linkage group O of the USDA/Iowa map (Cregan et al. 1999)). Another, at 44 cM on U 17 (42 cM on linkage group I of the USDA/Iowa map), segregated in both the MA and MN populations. Segregation of the MF phenotype could not be linked to either of these markers. The remaining three SNPs did not segregate in any of the RI populations.

## MF as an indicator of genome response to the environment

Trifoliolate leaf is the normal phenotype of the *Glycinae* including soybean, *G. max*. However, plants with a few nodes giving rise to additional leaflets (MF) are frequently observed in soybean and recently were reported for its wild ancestor *G. soja* (Wang et al. 2000). The fact that only certain of these cultivars exhibited this trait to any pronounced degree already had implied that the complexity of the genetic background of these inbreeding plants played a role in phenotypic expression. Because expression often is limited to a few plants and only some nodes on these plants, it was clear that penetrance would be limited by interaction with the environment during growth. These expectations have been borne out by our results. Striking differences were observed between different genotypes; and phenotypes of individual genotypes were differentially expressed when grown in different environments (Figs. 1 and 2). Additionally, there appears to be an effect of the environment during emergence that affects expression even on much later nodes—e.g. seventh and eighth trifoliates (Table 2).

The MF genotype is complex. In the MN population nine QTLs were identified, including *Lf1*which already had been mapped to A2 (U-3) (Cregan et al. 1999). Not one of these loci is essential for expression of the MF character, yet all contribute to the frequency of its expression. In Minnesota-2001, we obtained evidence for segregation of all of these and found that seven of them also were segregating in either the NA or MA populations. Yet, we could only explain a small percentage of the phenotypic variation in these Archer–derived RIs. Repeatability estimates (Table 1) indicated that little of this variation could be attributed to inaccurate assays or micro-environmental variation. It seems clear, therefore, that in the NA and MA populations there still must be a number of additional loci with small effects on MF variation that we have not detected.

It seems unlikely that the persistence of the MF phenotype is the result of a selective advantage per se that confers some form of fitness. The frequency of such plants is very low in cultivars that have been adapted to any particular environment and the frequency of MF nodes on MF plants also is typically low. We would like to propose that the MF phenotype is incidental to the expression of regulatory genes that adapt soybeans to their environment. Whereas we have failed to associate any *hox* genes with the phenotype in soybean, studies of other plants suggest that regulatory genes with multiple functions control leaf morphology. If genes that regulate the adaptation of soybean to different environments also are involved in regulating leaf morphology, we might expect to find both the variable MF expression and the complex genotype that we have observed.

We note, in passing, that many of the regions of the genome implicated in MF regulation also are involved in major agronomic traits under strong selection (http://www.soybase.ncgr.org). These include major loci that regulate plant growth and flowering such as *Dt1* and *E* as well as loci that regulate shattering, seed weight, seed number, yield, lodging, maturity and several aspects of disease and pest resistance. Minsoy and Noir I have a low level of MF expression (1% or less of all leaves). However, if their adapted genomes are disrupted by recombination and segregation following an intercross, the resulting segregants often have a greatly increased expression of the phenotype and differ from each other when placed in different environments. This would be expected from a genetic network of interacting regulatory elements. The strong effect of emergence on the resulting phenotype (plants/row, Fig. 2) suggests that a two-step process exists that establishes a response to environmental cues during early morphogenesis; and that this response can be subsequently modified during growth (nodes/plant, Fig. 2). If the MF phenotype is responding to a more fundamental regulatory system affecting other agronomic traits, it may be useful as an indicator of plant adaptation. Such studies could lead to an identification of the genetic elements that allow soybean to cope with its environment.

**Acknowledgments** This research was supported in part by grants from the USDA-ARS No. 5812759027, USDA-CSREES No. 2001-35301-10546, United Soybean Board and the Minnesota Soybean Research and Promotion Council.

## References

Allard RW (1975) The mating system and microevolution. Genetics 79(Suppl):115–126

Allard RW, Jain SK, Workman PL (1968) The genetics of inbreeding populations. Adv Genet 14:55–131

Bharathan G, Goliber TE, Moore C, Kessler S, Pham T, Sinha NR (2002) Homologies in leaf form inferred from KNOXI gene expression during development. Science 296:1858–1860

Chase K, Adler FR, Lark KG (1997) Epistat: A computer program for identifying and testing interactions between pairs of quantitative trait loci. Theor Appl Genet 94:724–730

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Van Taoi TT, Lohnes DG, Chung T, Specht JE (1999) An integrated genetic linkage map of the soybean. Crop Sci 39:1464–1490

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082

Fehr WR (1972) Genetic control of leaflet number in soybeans. Crop Sci 12:221–224

Fehr WR and Caviness CE (1977) Stages of soybean development. Special Report 80: Iowa State University, Ames, IA

Gourlay CW, Hofer JM, Ellis TH (2000) Pea compound leaf architecture is regulated by interactions among genes UNIFOLIATA, COCHLEATA, AFILA and TENDRIL-LESS. Plant Cell 12:1279–1294

Hanson CH, Robinson HF, Comstock RE (1956) Biometrical studies of yield in segregating populations of Korean lespedeza. Agron J 48:268–272

Lark KG, Orf J, Mansur LM (1994) Epistatic expression of quantitative trait loci (QTL) in soybean [ *Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. Theor Appl Genet 88:486–489

Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc Nat Acad Sci 92:4656–4660

Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR, Lark KG (1999a) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci 39:1642–1651

Orf JH, Chase K, Adler FR, Mansur LM, Lark KG (1999b) Genetics of soybean agronomic traits: II. Interactions between yield quantative trait loci in soybean. Crop Sci 39:1652–1657

Specht JE, Chase K, Macrander M, Graef GL, Chung J, Markwell JP, Germann M, Orf JH, Lark KG (2001) Soybean response to water: a QTL analysis of drought tolerance. Crop Sci 41:493–509

Stebbins GL Jr (1950) Variation and evolution in plants. Columbia University Press, New York

Takahashi Y, Fukuyama J (1919) Morphological and genetic studies on the soybean. (In Japanese). Hokkaido Agr Exp Sta Rept No 10:1–38

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. J Quant Loci: http://www.probe.nalusda.gov:8000/otherdocs/jqtl/jqtl1996–01/utz.html

Wang K, Li F, Zhou T, Xu Z (2000) Inheritance of a five leaflet character arising from wild soybean (*Glycine soja* Sieb. et. Zucc.) in soybeans (*G. max* (L.) Merr.). Soybean Gen. Newsl 27. http://www.soygenetics.org/articles/sgn2000.005.htm