Genomic Regions Governing Soybean Seed Nitrogen Accumulation

D.R. Panthee[a], V.R. Pantalone[a,*], C.E. Sams[a], A.M. Saxton[b], D.R. West[a], and W.E. Rayford[c]

Departments of [a]Plant Sciences and [b]Animal Science, University of Tennessee, Knoxville, Tennessee 37996, and [c]ARS, USDA, NCAUR, Peoria, Illinois 61604

ABSTRACT: Nitrogen accumulation in the form of seed protein takes place in soybean [*Glycine max* (L.) Merr.] during the reproductive stages of development. The purpose of this study was to relate genotypic differences in seed nitrogen accumulation with genomic regions controlling nitrogen accumulation in soybean during R_5, R_6, and R_7 growth stages. A population of 101 $F_{6:8}$ recombinant inbred lines (RIL) developed from a cross of N87-984-16 × TN93-99 was utilized. The RIL were grown at the University of Tennessee, Knoxville Experiment Station, in a randomized complete block design with three replications in 2002. Seed nitrogen was determined from pod samples harvested at the R_5, R_6, and R_7 growth stages. A significant ($P < 0.05$) difference among genotypes was found for nitrogen accumulation at all three growth stages. Single-factor ANOVA revealed that quantitative trait loci (QTL) governing nitrogen accumulation in soybean seed were distributed in the linkage groups A2, B2, D1a, D1b, E, G, and M. Phenotypic variation explained by an individual QTL ranged from 5 to 11.6%. These QTL may provide useful marker-assisted selection opportunities for soybean protein improvement.

Paper no. J10696 in *JAOCS 81*, 77–81 (January 2004).

KEY WORDS: Gene-mapping, genomics, *Glycine max* (L.) Merr., QTL, protein.

Soybean [*Glycine max* (L.) Merr.] seeds typically contain approximately 40% protein on a dry weight basis, making soybeans one of the highest protein-containing crop species. Nitrogen (N) is an important constituent of storage protein that is accumulated during reproductive stages of seed development. The rate of storage protein accumulation can be estimated from the rate of N accumulation.

Understanding the physiology of N accumulation and protein synthesis should allow more efficient breeding to develop high-protein soybean lines. According to Shibles *et al*. (1), sources of N include uptake of soil nitrate during seed growth, fixation of atmospheric N_2 during seed growth, and redistribution of N accumulated in vegetative tissues before seed growth begins. Staswick *et al*. (2) determined that large quantities of N are also mobilized from mature organs to developing seeds. Imsande (3) found that leaf proteins are hydrolyzed and that the salvaged N is translocated to the developing pods and seeds during soybean pod filling. An increase in the protein content of the cotyledon commences as early as 15 d after flowering and continues through to maturity. The most rapid rate of accumulation of total protein in cotyledons is between 25 to 40 d after flowering (4). Accumulation of N into growing seeds and formation of seed storage proteins, which contribute to soybean seed yield, have important economic implications.

Kumudini *et al*. (5) demonstrated a genotypic effect in seed N accumulation in soybeans, where an ancestral set of cultivars ('Pagoda' and 'Mandarin') was superior in this process compared with later cultivars ('Maple Glen' and 'OAC Bayfield'). Recently, Grandgirard (6) showed that the seed N accumulation rate in soybeans is determined by the genotype. However, no reports are available about the genomic regions or quantitative trait loci (QTL) controlling this process. Soybean storage protein is inherited as a quantitative trait (7), which is influenced by the growing environment. At the same time, the potential exists for interaction of genes that mediate metabolic pathways. Therefore, it is important to consider genotypic effects on seed composition to provide stable production. For this reason, knowledge of genotypic differences in the timing and rates of constituent deposition (such as N accumulation) during seed development is important (8). Once the rate of N accumulation in the seed is understood in a given genotype, management practices can be adjusted to obtain the maximum N accumulation in soybean seed. QTL analyses can detect important genomic regions governing N accumulation. Furthermore, knowledge of specific QTL provides opportunities for marker-assisted selection in cultivar development programs. Therefore, to gain more insight, this study was conducted to determine whether there are genotypic differences in soybeans for seed N accumulation and to identify genomic regions controlling this process.

MATERIALS AND METHODS

Plant materials. A total of 101 $F_{6:8}$ recombinant inbred lines (RIL) were developed from a cross of N87-984-16 × TN93-99. The N87-984-16 parent is one of two sister lines whose blend constitutes the high-protein commercial variety 'Prolina' (9). The TN93-99 parent is a high-yielding and well-adapted University of Tennessee breeding line and is currently registered

*To whom correspondence should be addressed at Department of Plant Sciences, University of Tennessee, 2431 Joe Johnson Dr., 252 Ellington Plant Science Bldg., Knoxville, TN 37996-4561. E-mail: vpantalo@utk.edu

as a germplasm (10). All RIL, the two parents, and the check cultivars 'Hutcheson,' '5002T,' and '5601T' were planted in a randomized, complete block design with three replications at the Knoxville Experiment Station of the University of Tennessee (Knoxville, TN) in May 2002. Each line was planted in a four-row plot of 6-m length with a spacing of 75 cm between rows.

Sample collection. Pods were collected at R_5, R_6, and R_7 reproductive growth stages. Fehr and Caviness (11) regarded the R_2 stage as "days to flowering." Once days to flowering for all the lines were recorded, days to R_5, R_6, and R_7 growth stages were estimated for a tentative visit for pod collection. Pod collection was estimated to occur at 5, 7, and 9 wk after R_2 for R_5, R_6, and R_7 growth stages, respectively. Actual sample collection was done when the individual genotype reached the reproductive growth stages as specified in Fehr and Caviness (11). At R_5, R_6 and R_7, pods were sampled from the 7th to 9th node of the plant to maintain uniformity. Approximately 25 pods were collected at R_5, 15 pods at R_6, and 10 pods at R_7 so that at least 1 g of dried ground sample would be available. The seeds were separated from the pods and dried at 60°C for 24 h. Dried samples were ground with a coffee grinder for about 20 s and transferred to 2-mL Eppendorf tubes.

N determination. N content of the sample was determined using a LECO CHN 2000 Analyzer (LECO Corporation, St. Joseph, MI) following AOAC Official Method 990.03 (12). The principle involved in this analysis is that nitrogen freed by combustion at high temperature in pure oxygen is measured by thermal conductivity detection and can be converted to equivalent protein by an appropriate numerical factor (6.25 in the case of soybean protein).

Amino acid analysis. Soybean seeds were ground in a water-cooled (20°C) Knifetec 1095 Sample Mill (FOSS Tecator, Hogana, Sweden) for 20 s. This process produced soybean flour with a uniform particle size. The ground sample was analyzed *via* NIR reflectance, using an FOSS 6500 instrument and amino acid prediction equation (FOSS North America, Eden Prairie, MN).

DNA extraction and polymerase chain reaction (PCR). DNA was extracted from all RIL and the two parental lines by following the Qiagen Plant Easy DNA Extraction Kit (Qiagen, Hilden, Germany). PCR consisted of 7.4 μL of ddH_2O, 1 μL of 10× PCR Buffer, 1 μL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 μL of 20 μM forward and reverse primer, 0.1 μL of 5 units/μL of Klentaq (Ab Peptides Inc., St. Louis, MO), and 2 μL of 20 ng/μL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were (i) 94°C for 5 min followed by 35 cycles at 94°C for denaturation for 25 s, (ii) 47°C for annealing for 30 s, (iii) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for 5 min. Parents were screened with a total of 568 $(ATT)_n$ type of simple sequence repeat (SSR) genetic markers, which spanned all 20 molecular linkage groups (LG) of soybean (13). The DNA sequence information for soybean SSR markers is publicly available from the USDA Internet site http://www.soybase.org (verified December 23, 2003). A total of 86 SSR markers, distributed among all 20 LG were found to be polymorphic between the parents and among the RIL and were used in QTL analyses.

A 6% nondenaturing PAGE was used to separate the PCR product. The PAGE consisted of 28.5 mL (40% bis-acrylamide solution), 160.2 mL [0.60× Tris borate EDTA buffer (TBE)], 1.33 mL (10% ammonium persulfate), and 66.5 μL tetramethylethylenediamine. Two microliters loading buffer was added to PCR product, and 10 μL sample was loaded in the gel. Running buffer was 0.5× TBE. The gel was run at a constant 300 V for 3 h. A fan was used to keep the glass plates cool during gel running. The bands were visualized by staining the gel with 50 μL of ethidium bromide (10 mg/mL) and photographing it under exposure to UV light, and scored using 1 for parent 1 (N87-984-16), 2 for heterozygote, and 3 for parent 2 (TN93-99).

Data analyses. Nitrogen content data were analyzed using SAS software (14). Associations between molecular markers and nitrogen accumulation at various growth stages (R_5, R_6, and R_7), and amino acids were analyzed with SAS using single-factor ANOVA. QTL position was determined by analyzing the data with MAPMAKER/EXP 3.0 (15) and QTL Cartographer 2.0 (16).

RESULTS AND DISCUSSION

RIL are fixed in the homozygous state at nearly all loci. Thus, differences in a measured trait (such as seed N content), which are associated with allelic differences at specific DNA marker loci, reveal the underlying genomic regions governing that trait. Seed N content varied from 4.63 to 9.91% (Table 1) among the soybean RIL determined at different growth stages. The frequency distribution of all the RIL is presented in Figure 1. The RIL differed significantly ($P < 0.05$) for N accumulation, indicating that some lines had better genetic potential than others for this character, suggesting an opportunity to select specific RIL for high N accumulation. This finding is consistent with the results of Grandgirard (6). In our study, there was a gradual increase in seed N content at different growth stages. The seed N content differed significantly ($P < 0.05$) between R_5, R_6, and R_7 growth stages. Mean N contents were 6.28% at R_5, 6.35% at R_6, and 6.68% at R_7. Most of the N accumulation took place within 5 wk after flowering in soybean. This has management implications for efforts targeting increases in protein or specific amino acid content *via* nutrient amendments (7,17,18). Depending on the time required for N uptake by the growing soybean plant, time of N application can be adjusted. Our observations verify the findings of Gayler and Sykes (4) that the majority of seed N accumulates by R_5 stage. They found that the protein fraction starts accumulating 15 to 17 d after flowering (DAF) and reaches a maximum by 25 to 30 DAF. They also found that the majority of protein accumulates within 5 wk. Grandgirard (6) mentioned that seed N accumulation rate is highly correlated with the instantaneous available N concentration. Lhuillier-Soundele *et al.* (19) also found that seed N accumulation was positively related to available N in vegetative parts of the pea (*Pisum sativum* L.).

TABLE 1
Descriptive Statistics for Nitrogen and Protein Content on a Dry Weight Basis in a Soybean Recombinant Inbred Line Population (*n* = 101) Measured at Three Reproductive Growth Stages

	R_5 stage		R_6 stage		R_7 stage	
Descriptors	Nitrogen	Protein	Nitrogen	Protein	Nitrogen	Protein
Minimum	5.02	31.38	4.63	28.91	5.11	31.92
Mean	6.29	39.29	6.34	39.60	6.67	41.71
Maximum	8.25	51.53	9.91	61.94	7.79	48.66
SD	0.38	2.39	0.52	3.25	0.36	2.24

Cregan *et al.* (13) developed a large set of SSR molecular markers for soybean that are currently well-mapped to 20 molecular LG, which correspond to the 20 pairs of homologous chromosomes in soybean. After screening a total of 568 SSR markers, from which 86 markers were identified as exhibiting polymorphic DNA banding patterns among the RIL, we identified seven markers linked to QTL governing N accumulation at the R_5 stage in soybean. Those QTL were distributed among five linkage groups (A2, B2, D1b, E, and G). The variation among RIL for N accumulation explained by the QTL ranged from 5.9 to 10.7% (Table 2). Marker Satt274 (LG D1b) was identified as a major QTL for seed N content at the R_5 growth stage.

Three markers were linked to QTL governing N accumulation in seeds at the R_6 growth stage. At R_6, in addition to LG A2 and D1b, a marker from LG M was also detected. The strongest QTL (Satt459 on LG D1b) explained 9.3% of the variation among RIL for N accumulation. At R_7, seven markers were associated with QTL from four LG (B2, D1a, E, and G) controlling N accumulation. The Satt263 marker (LG E) was the strongest QTL and explained 11.6% of the variation among RIL for N accumulation. Few reports are available on genomic regions controlling seed nitrogen content in crop species. Bezant *et al*. (20) identified eight QTL controlling seed N content in barley using RFLP (restriction fragment-length polymorphism) markers. Our results identifying QTL for seed N accumulation at various growth stages in soybean have important implications in soybean crop management and cultivar development.

Notably, the present study was conducted in only one environment, but with sampling over three growth stages. To confirm QTL and to identify environmentally stable QTL (21), additional molecular mapping can be performed in multiple locations and in additional populations by other researchers. Although the results presented in this study are limited to one environment, soybean protein is considered to be a highly heritable trait; i.e., genetic factors (rather than environmental factors) primarily govern measurable variation for protein. Thus, the present study has provided potentially important information for marker-assisted selection to increase protein concentration in soybeans by identifying QTL associated with N accumulation at reproductive growth stages.

Interestingly, few common LG were associated with N accumulation across different growth stages (Table 2). This suggests that the same gene may not express throughout reproductive development. Rather, total seed N accumulated may be the result of cumulative effects of several QTL. Significant differences existed among RIL for amino acid composition (data not shown), which enabled detection of QTL for specific amino acids. This was likely because the N87-984-16 parent ('Prolina') has greater levels of the sulfur-containing amino acids (cysteine and methionine) than normal cultivars (8). The QTL for seed N accumulation found at different growth stages were also found to govern specific amino acids in this population (Table 3). For example, QTL for methionine were detected at Satt133 (LG A2) and Satt 235 (LG G), which were also R_5 seed N QTL; at Satt459 (LG D1b) and Satt201 (LG M), which were also R_6 seed N QTL; and at Satt45 (LG E) and Satt268 (LG E), which were also R_7 seed N QTL (Tables 2 and 3). This suggests that different amino acids are synthesized throughout different growth stages, or that assembly of specific amino acids requires different genes that are regulated during reproductive growth stages.

All the QTL detected governing seed N accumulation were located on seven LG (A2, B2, D1a, D1b, E, G, and M). Brummer *et al*. (21) mapped QTL controlling protein content in eight different soybean populations using RFLP markers. They identified QTL for protein linked to markers located on nine different LG including A2, B2, D1, E, and G, which are consistent with our observations. Different QTL may be identified in different populations because of genetic background effects. Qui *et al*. (22) used a population from 'Peking' × 'Essex' for mapping various characters including protein. They used RFLP markers and found QTL associated with protein on LG F and

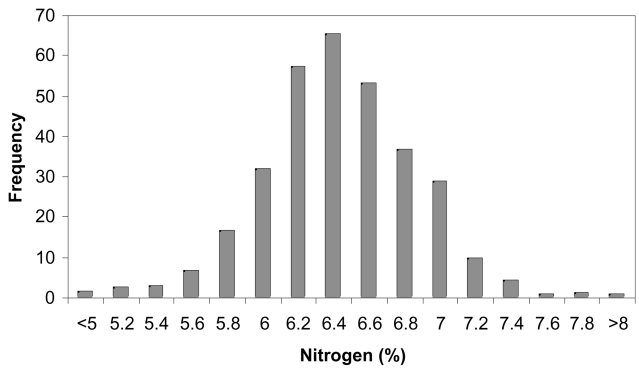

FIG. 1. Frequency distribution for a population of soybean recombinant inbred lines for seed nitrogen content class means (constructed at intervals of 0.2% nitrogen), averaged over three reproductive growth stages.

TABLE 2
Molecular Markers and Molecular Linkage Groups for Quantitative Trait Loci (QTL) Governing Nitrogen Accumulation in Developing Seeds of Soybean at Three Reproductive Growth Stages

Markers	Linkage group[a]	Probability	R^2 (%)	Allele mean N (%) N87 (P1)	Allele mean N (%) TN93 (P2)	Difference (P1 – P2)	Contribution (%)
R_5 growth stage							
Satt177	A2	0.027	6.1	6.27	6.38	–0.11	1.75
Satt133	A2	0.027	6.6	6.24	6.35	–0.11	1.76
Satt556	B2	0.033	5.9	6.38	6.26	+0.12	1.92
Satt274	D1b	0.004	10.7	6.24	6.39	–0.15	2.40
Satt185	E	0.011	9.0	6.34	6.22	+0.12	1.93
Satt263	E	0.026	7.0	6.35	6.23	+0.12	1.93
Satt235	G	0.019	7.1	6.37	6.25	+0.12	1.92
R_6 growth stage							
Satt437	A2	0.015	8.2	6.27	6.42	–0.15	2.39
Satt459	D1b	0.100	9.3	6.44	6.28	+0.16	2.55
Satt201	M	0.028	6.5	6.33	6.46	–0.13	2.05
R_7 growth stage							
Satt168	B2	0.049	5.0	6.75	6.64	+0.11	1.66
Satt383	D1a	0.025	5.9	6.75	6.64	+0.11	1.66
Satt268	E	0.019	7.0	6.63	6.75	–0.12	1.81
Satt263	E	0.004	11.6	6.62	6.77	–0.15	2.27
Satt45	E	0.004	9.9	6.62	–0.14	2.11	
Satt570	G	0.028	6.3	6.74	6.62	+0.12	1.81
Satt131	G	0.023	6.3	6.77	6.65	+0.12	1.80

[a]Linkage group reported by Cregan *et al.* (13).

TABLE 3
Soybean Seed Nitrogen Accumulation QTL at Three Reproductive Growth Stages That Were Also Detected as Amino Acid QTL

Growth stage	Linkage group[a]	Marker	Amino acids
R_5	A2	Satt177	Cys
	A2	Satt133	Met, Gly, Pro, Val, Iso, Phe, His, Tyr
	D1b	Satt274	Cys, Thr
	E	Satt185	Asp, Ser, Arg, His
	E	Satt263	Asp, Glu, Ser, Val, Iso, Leu, Phe, His, Tyr
	G	Satt235	Met, Arg
R_6	A2	Satt437	Gly, Val, Iso
	D1b	Satt459	Met, Glu, Gly, Iso
	M	Satt201	Met, Tyr
R_7	B2	Satt168	Thr
	E	Satt268	Met, Arg
	E	Satt263	Asp, Glu, Ser, Val, Arg, Iso, Leu, Phe, His, Tyr
	E	Satt45	Met, Asp, Glu, Arg
	G	Satt570	Thr, Arg
	G	Satt131	Pro

[a]Linkage group reported by Cregan *et al.* (13). See Table 2 for abbreviation.

H. SSR markers, associated with protein QTL, were found to be located on LG D1a and M by Csanadi *et al.* (23). This is consistent with our findings.

Our study has established that N accumulation is affected by the genotype of soybean and that there are genomic regions controlling this process. The QTL identified in this study provide potentially useful marker-assisted selection opportunities for soybean protein improvement.

ACKNOWLEDGMENTS

Funding from the United Soybean Board (Project 1240) and from the Tennessee Soybean Promotion Board is greatly appreciated in support of this project. We appreciate the excellent support provided by the Tennessee Agriculture Experiment Station and by the USDA National Center for Agricultural Utilization Research (Peoria, IL) for seed N analyses.

Technical assistance provided by Beth Meyer, Yamika Stokes, Dr. Madeleine Spencer, and Debbie Landau-Ellis is greatly appreciated.

REFERENCES

1. Shibles, R.M., D.N. Sundberg, and S.L. Naeve, Physiological Constraints Limiting Seed Protein in Normal and High Seed Protein Soybean Genotypes, *Agron Abstract.*:151 (1994).
2. Staswick, P.E., Z. Zhang, T.E. Clemente, And J.E. Specht, Efficient Down-Regulation of the Major Vegetative Storage Protein Genes in Transgenic Soybean Does Not Compromise Plant Productivity, *Plant Physiol. 127*:1819–1826 (2001).
3. Imsande, J., Selection of Soybean Mutants with Increased Concentrations of Seed Methionine and Cysteine, *Crop Sci. 41*:510–515 (2001).
4. Gayler, K.R., and G.E. Sykes, β-Conglycinins in Developing Soybean Seeds, *Plant Physiol. 67*:958–961 (1981).
5. Kumudini, S., D.J. Hume, and G. Chu, Genetic Improvement in Short-Season Soybeans: II. Nitrogen Accumulation, Remobilization, and Partitioning, *Crop Sci. 42*:141–145 (2002).
6. Grandgirard, D., Analysis and Modeling of Soybean (*Glycine Max* L.) Seed Nitrogen Concentration: Relation Between Nitrogen Remobilization to the Seeds and Crop Yield and Quality, Ph.D. Thesis, Institut National de la Recherche Agrononique (INRA), Dijon, France. 2002. (http://www.inra.fr/Internet/Centres/Dijon/malherbo/theses/grandgi-gb.htm).
7. Nakasathien, S., D.W. Israel, R.F. Wilson, and P. Kwanyuen, Regulation of Seed Protein Concentration in Soybean by Supra-Optimal Nitrogen Supply, *Crop Sci. 40*:1277–1284 (2000).
8. Wilson, R.F., Seed Composition, in *Soybeans: Improvement, Production and Uses*, edited by H.R. Boerma and J.E. Specht, 3rd edn., ASA, CSSA, SSA, Madison, WI, in press.
9. Burton, J.W., T.E. Carter Jr., and R.F. Wilson, Registration of 'Prolina' Soybean, *Crop Sci. 39*:294–295 (1999).
10. Pantalone, V.R., F.L. Allen, and D. Landau-Ellis, Registration of TN93-99 Soybean Germplasm, *Ibid. 43*:1137 (2003).
11. Fehr, W.R., and Caviness, C.E., Stages of Soybean Development, Iowa State University, *Special Report 80*, Cooperative Extension Services, Iowa State University, Ames, 1977, p. 7.
12. Protein (crude) in Animal Feed, *Official Methods of Analysis of ADAC International*, 16th edn., Arlington, VA, 1995, Method 990.03.
13. Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. Vantoai, D.G. Lohnes, J. Chung, and J.E. Specht, `n Integrated Genetic Linkage Map of the Soybean Genome, *Crop Sci. 39*:1464–1490 (1999).
14. SAS Institute, *SAS/STAT User's Guide*, Version 8.0, SAS Institute Inc., Cary, NC, 1999.
15. Lincoln, S.E., M.J. Daly, and E.S. Lander, *Constructing Genetic Linkage Maps with MAPMAKER Version 3: A Tutorial and Reference Manual*, 3rd edn., Whitehead Institute for Biomedical Research, Cambridge, MA, 1993.
16. Wang, S., C.J. Basten, and Z.B. Zeng, Windows QTL Cartographer, Winqtlcart V2.0, North Carolina State University, Raleigh, 2003.
17. Sexton, P.J., S.L. Naeve, N.C. Paek, and R. Shibles, Sulfur Availability, Cotyledon Nitrogen:Sulfur Ratio, and Relative Abundance of Seed Storage Proteins of Soybean, *Crop Sci. 38*: 983–986 (1998).
18. Paek, N.C., P.J. Sexton, S.L. Naeve, and R. Shibles, Differential Accumulation of Soybean Seed Storage Protein Subunits in Response to Sulfur and Nitrogen Nutritional Sources, *Plant Prod. Sci. 3*:268–274 (2000).
19. Lhuillier-Soundele, A., N.G. Munier-Jolain, and B. Ney, Influence of Nitrogen Availability on Seed Nitrogen Accumulation in Pea, *Crop Sci. 39*:1741–1748 (1999).
20. Bezant, J.H., D.A. Laurie, N. Pratchett, J. Chojecki, and M.J. Kearsey, Mapping of QTL Controlling NIR Predicted Hot Water Extract and Grain Nitrogen Content in a Spring Barley Cross Using Marker-Regression, *Plant Breeding 116*:141–145 (1997).
21. Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and, R.C. Shoemaker, Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations, *Crop Sci. 37*:370–378 (1997).
22. Qui, B.X., P.R. Arelli, and D.A. Sleper, RFLP Markers Associated with Soybean Cyst Nematode Resistance and Seed Composition in a 'Peking' × 'Essex' Population, *Theor. Appl. Genet. 98*:356–364 (1999).
23. Csanadi, G., J. Vollmann, G. Stift, and T. Lelley, Seed Quality QTLS Identified in a Molecular Map of Early Maturing Soybean, *Ibid. 103*:912–919 (2001).

[Received July 21, 2003; accepted November 10, 2003]