Simple sequence repeat (SSR) markers linked to *E1*, *E3*, *E4*, and *E7* maturity genes in soybean

Stephen J. Molnar, Satish Rai, Martin Charette, and Elroy R. Cober

Abstract: Soybean near isogenic lines (NILs), contrasting for maturity and photoperiod sensitivity loci, were genotyped with approximately 430 mapped simple sequence repeats (SSRs), also known as microsatellite markers. By analysis of allele distributions across the NILs, it was possible to confirm the map location of the *Dt1* indeterminate growth locus, to refine the SSR mapping of the *T* tawny pubescence locus, to map *E1* and *E3* maturity loci with molecular markers, and to map the *E4* and *E7* maturity loci for the first time. Molecular markers flanking these loci are now available for marker-assisted breeding for these traits. Analysis of map locations identified a putative homologous relationship among four chromosomal regions; one in the middle of linkage group (LG) C2 carrying *E1* and *E7*, one on LG I carrying *E4*, one at the top of LG C2, at which there is a reproductive period quantitative trait locus (QTL), and the fourth on LG B1. Other evidence suggests that homology also exists between the *E1* + *E7* region on LG C2 and a region on LG L linked to a pod maturity QTL. Homology relationships predict possible locations in the soybean genome of additional maturity loci, as well as which maturity loci may share a common evolutionary origin and similar mechanism(s) of action.

Key words: soybean, maturity, photoperiod, microsatellite marker, near isogenic line.

Résumé : Des lignées quasi-isogéniques (NIL) du soja, porteuses d'allèles différents pour des locus de maturité et de réponse à la photopériode, ont été génotypées avec 430 microsatellites (SSR) de position chromosomique connue. En comparant la distribution allélique chez ces lignées, il a été possible de confirmer la position du locus *Dt1*, lequel confère une croissance déterminée ou indéterminée. Il a également été possible de préciser l'emplacement du locus *T* (« tawny pubescence »), d'assigner une position aux locus de maturité *E1* et *E3* avec marquers moléculaires et de déterminer, pour la première fois, la position des locus de maturité *E4* et *E7*. Des marqueurs moléculaires bordant ces locus sont maintenant disponibles pour des fins de sélection assistée de marqueurs. L'examen de ces emplacements chromosomiques suggère une possible relation d'homologie pour quatre régions chromosomiques. La première serait située au milieu du groupe de liaison (LG) C2 (portant les locus *E1* et *E7*), la seconde sur le LG1 (portant *E4*), une troisième au sommet du LG C2 (où se trouve un QTL pour la période de reproduction) et une quatrième sur le LG B1. D'autres évidences suggèrent que de l'homologie existerait également entre la région portant *E1* + *E7* sur le LG C2 et une région portant un QTL pour la maturité des gousses, lequel est situé sur le LG L. Ces possibles homologies prédisent l'emplacement possible de locus additionnels contrôlant la maturité de même qu'à prédire quels locus de maturité partagent une même origine évolutive et un mécanisme d'action similaire.

Mots clés : soja, maturité, photopériode, marqueur microsatellite, lignée quasi-isogénique.

[Traduit par la Rédaction]

Introduction

The development of short season soybeans for Canadian and northern regions requires effective use of early maturity genes. Seven loci affecting time to flowering and maturity have been identified, but others may exist. Alleles at these loci act through their effects on photoperiod sensitivity. Deploying these loci can be difficult and linked molecular markers could assist the soybean breeder in selecting for early maturity and in combining specific alleles at maturity loci. The identification of linked markers for single-gene traits can be realized either through targeted approaches such as using near-isogenic lines (NILs) (Muehlbauer et al. 1991) and bulked segregant analysis (BSA) (Michelmore et al. 1991) or through genome-wide searching for QTLs.

Well-characterized maturity genes have been identified by classical methods. Pairs of NILs contrasting for six photoperiod-sensitivity (time-to-maturity) genes (*E1, E2, E3, E4, E5, E7*) are publically available (Cober et al. 1996*a*). Using *Gycine max* (L.) Merr 'Harosoy'-based NILs, it was shown that the *E1*, *E3*, and *E4* alleles respond to long days with delayed flowering and maturity (Cober et al. 1996*a*). The *E1*, *E3*, *E4*, and *E7* loci also show a response to light quality, suggesting that some *E* alleles may be related to phytochrome (Cober et al. 1996*b*; Cober and Voldeng 2001*b*). Maturity and pubescence colour are associated in short season soybeans, partly owing to the tight genetic linkage between *E1* and the *T* locus, which governs tawny vs.

Received 19 September 2002. Accepted 21 July 2003. Published on the NRC Research Press Web site at http://genome.nrc.ca on 13 November 2003.

Corresponding Editor: F. Belzile.

S.J. Molnar,[1] S. Rai, M. Charette, and E.R. Cober. Eastern Cereal and Oilseed Research Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada.

[1]Corresponding author (e-mail: molnarsj@agr.gc.ca).

 doi: 10.1139/G03-079

grey pubescence colour (Cober et al. 1997). A second maturity locus, *E7*, also linked to *T*, has been identified (Cober and Voldeng 2001*a*). In addition, *E6*, a dominant gene conditioning early flowering and maturity, has been reported in unrelated germplasm (Bonato and Vello 1999).

NILs have been used to identify molecular markers linked to classical loci in soybean (Muehlbauer et al. 1991). This study examined the allelic status of 15 restriction fragment length polymorphism (RFLP) probes across 66 'Clark'-based NILs and 50 'Harosoy'-based NILs contrasting for a variety of traits. Whereas NILs contrasting for *E1*, *E3*, *Dt1*, and *T* were included, no linked markers for these specific loci were identified among the limited number of RFLP probes tested.

Using QTL analysis in an interspecific cross of *Glycine max* and *Glycine soja*, Keim et al. (1990) identified maturity QTL on LG M (equivalent to LG C2 on the current integrated map of Cregan et al. 1999) and LG H (equivalent to LG D1) and an unlinked RFLP marker, K472 (LG C1). Two QTL for maturity were reported by Mansur et al. (1993*a*) in the cross *G. max* 'Minsoy' × *G. max* 'Noir1'. One QTL was mapped to LG 14 (LG C2) and the other QTL was associated with an unlinked marker A676 (LG C2). Li and Niwa (1996) analysed an F_2 population derived from the cross between *G. max* 'Heihe 5' (insensitive to photoperiod) and *G. max* 'Akisengoku' (highly sensitive to photoperiod) under short-day conditions, and reported that "days to flowering" was controlled by a major gene and early maturity was partially dominant to late maturity. Further, they reported that a days-to-flowering locus was located between two novel synthetic oligonucleotide microsatellite-type markers, which were not mapped to a standard soybean linkage group. Lee et al. (1996) reported five putative QTLs for maturity, one of which was unlinked and four others that mapped to LGs B1 (two QTL), C1, and L. A major QTL for maturity and photoperiod insensitivity was observed on LG C2 in two populations with minor QTL on LGs A2, G, J, and L (Tasma et al. 2001). These studies offered some guidance as to which genomic regions may contain the specific known maturity genes and could be targeted for identifying molecular markers linked to maturity loci.

This paper reports on the use of NILs to identify SSR molecular markers that are linked to the maturity loci *E1*, *E3*, *E4*, and *E7*, the pubescence colour locus *T*, and the determinate growth habit locus *Dt1*. These molecular markers could be used for marker-assisted breeding for these traits.

Materials and methods

Germplasm

Twenty-five NILs, including five recurrent parental varieties, contrasting for photoperiod sensitivity (time to maturity) and stem type genes were used in the present study (Table 1). Table 1 also includes information on three additional donor parental 'Harosoy' isolines to facilitate tracing the pedigrees of the NILs and to define the source of specific alleles.

DNA extraction

Plants were grown in the greenhouse or field. DNA was extracted from frozen leaves using a urea extraction buffer method developed by S. Rai (unpublished). Extraction buffer included 100 mmol Tris–HCl/L (pH 8), 20 mmol EDTA/L (pH 8.0), 700 mmol NaCl/L, 8 mol urea/L, 3% w/v sarcosyl, and 1% w/v SDS. Powdered leaves (250 mg) were incubated at room temperature for 5 min in 500 µL of extraction buffer. After the lysis of leaf samples, 500 µL of phenol – chloroform – isoamyl alcohol (25:24:1) were mixed with the extraction buffer for 5–10 min and then centrifuged at 10 000*g* for 10 min. The upper phase was transferred to a new tube, and 500 µL of chloroform – isoamyl alcohol (24:1) was added with mixing. After centrifugation at 10 000*g* for 10 min, the upper aqueous phase was transferred to a new tube and 250 µL of 5 mol NaCl/L was added to the solution to remove polysaccharides (Fang et al. 1992). DNA was precipitated using one volume of isopropanol and incubated at –20 °C for 10 min. The mixture was centrifuged at 10 000*g* for 10 min to form a pellet, which was twice washed with a 0.75 volume fraction of ethanol, followed by centrifugation. DNA pellets were dried in a Speedvac (Savant, Farmingdale, N.Y.) without heat and subsequently dissolved in 100 µL of sterile TE buffer (10 mM Tris, 1 mM EDTA (pH 8.0)). DNA was treated with 5 µL of a 10 µg/µL solution of RNAse A (Sigma Aldrich, St. Louis, Mo.) for 45 min at 37 °C. DNA concentration was determined by optical density and 5 µg of DNA was tested by electrophoresis on a 1% w/v agarose gel to confirm quality.

SSR markers

Soybean SSR primers designed by Cregan et al. (1999) were used to amplify genomic DNA. SSR primers were purchased from Research Genetics Inc. (Invitrogen, Calsbad, Calif.), or were custom synthesized by the University of Guelph (Guelph, Ont.) or Sigma Genosys (Sigma Aldrich, St. Louis, Mo.). SSR markers were chosen to give uniform high-density coverage of the soybean genome, based on the integrated map of Cregan et al. (1999).

PCR conditions

DNA amplification was carried out using the method described by Cregan et al. (1999) with some modifications in reaction conditions. A custom buffer (S. Rai, unpublished) was used for PCR. A 5× concentration stock solution of this buffer contained 400 mmol Tris–HCl/L (pH 8.3), 100 mmol ammonium sulfate/L, 1.0 mmol of each dNTPs/L, and a 0.05 volume fraction of Triton X-100. To this basic buffer was added either 7.5 mmol $MgCl_2$/L (for a final concentration in the PCR of 1.5 mmol/L) for custom primers from Sigma Genosys or the University of Guelph, or 10 mmol $MgCl_2$/L (for a final concentration of 2.0 mmol/L) for SSR primers from Research Genetics. To amplify DNA, 1.5 or 2.0 mmol $MgCl_2$/L, 1× buffer, 50 ng genomic DNA, 0.15 µL of a 20-µmol/L stock of each primer, 0.75 U *Taq* polymerase, and 200 µmol/L each of dNTPs in a final volume of 15 µL were used. *Taq* polymerase and dNTPs, were purchased from MBI Fermantas (Hanover, Md.). The thermocycler was programmed with 1 denaturing cycle of 95 °C for 3 min, followed by 35 amplification cycles of denaturation at 94 °C for 45 s, followed by the optimal annealing temperature for 50 s, and a 72 °C extension step for 45 s. After the final extension cycle at 72 °C for 5 min, the samples were held at 4 °C until removed from the machine. PCR-cycling was per-

Table 1. Genotypes and pedigrees of soybean near-isogenic lines.

Line	Genotype*	Pedigree	Reference†
Harosoy isolines			
Preliminary screen			
Harosoy	t Dt1 e1 E3 E4 E7	'Mandarin(Ottawa)'*2/AK('Harrow')	
OT89-5	t Dt1 e1 e3 e4 E7	PI 438477/2*'Evans'//7*L62-667	3
OT93-28	t Dt1 E1 e3 e4 E7	OT89-5/L71-802	4
OT94-41	t Dt1 e1 E3 e4 E7	OT89-5/L67-153	4
L62-667	t Dt1 e1 e3 E4 E7	'Harosoy'*6/T204	1
OT94-47	t Dt1 e1 e3 e4 e7	OT89-5//PI 196529/6*L62-667	
Additional six			
OT93-26	T Dt1 E1 e3 e4 E7	OT89-5/L71-802	4
OT94-49	T dt1 E1 e3 e4 E7	L71-802/OT89-5//OT89-6	4
OT94-51	t dt1 E1 e3 e4 E7	L71-802/OT89-5//OT89-6	4
OT94-37	t dt1 e1 e3 E4 E7	OT89-5/L67-153	4
OT94-39	t dt1 e1 E3 e4 E7	OT89-5/L67-153	4
OT89-6	t dt1 e1 e3 e4 E7	OT89-5/L67-153	3
Parental Harosoy isolines			
L71-802	T Dt1 E1 e3 E4 E7	L62-667/L67-2324	1
L67-2324	T Dt1 E1 E3 E4 E7	'Harosoy'*6/PI 196166	1
L67-153	t dt1 e1 E3 E4 E7	'Harosoy'*6/'Higan'	1
'Maple Presto' isolines			
'Maple Presto'	T Dt1 e1 e3 e4 e7	'Amsoy'/'Portage'//PI 438477	
OT89-9	T dt1 e1 e3 e4 e7	L67-153/7*'Maple Presto'	2
OT96-39	T dt1 E1 e3 e4 ?	PI 317334A/7*OT89-9	
OT96-41	T dt1 E1 e3 e4 ?	PI 317333/7*OT89-9	
'Maple Arrow' isolines			
'Maple Arrow'	T Dt1 e1 E3 e4 e7	'Harosoy'/PI 438477	
OT89-1	T Dt1 e1 e3 e4 e7	PI 196529/7*'Maple Arrow'	3
OT89-2	T dt1 e1 e3 e4 e7	PI 196259/7*'Maple Arrow'	3
'Maple Amber' isolines			
'Maple Amber'	T Dt1 e1 e3 E4 e7	'Harosoy 63'/'Altona'//PI 438477	
OT89-4	T Dt1 e1 e3 e4 e7	'McCall'/7*'Maple Amber'	3
OT89-3	T Dt1 e1 e3 e4 e7	'McCall'/7*'Maple Amber'	3
OT97-40	Maple Amber-*Pd1*	'Harosoy'-Pd1/7*'Maple Amber'	
'Evans' isolines			
'Evans'‡	t Dt1 e1 E3/e3 E4 E7?	'Merit'/'Harosoy'	
OT89-7	t Dt1 e1 e3 e4 E7?	PI 438477/2*'Evans'//'Evans'-e3	3

*Locus codes: *T/t*, Tawny/grey pubescence; *Dt1/dt1*, determinate/indeterminate growth habit; *E1/e1, E3/e3*, *E4/e4*, and *E7/e7*, maturity loci. Substituted alleles at a locus did not come from a single source in the development of all sets of NILs. In 'Harosoy' isolines and in parental 'Harosoy' isolines, the source of *E1* and *T* is PI196166, *e4* is PI438477, *e7* is PI196529, and *dt1* is 'Higan'. T204 is known to be the source of *e3* in L62-667 and L71-802 and presumed to be the source in other isolines since it is present in the recurrent parent. In 'Maple Presto' isolines, the source of *E1* is PI317334A and *dt1* is 'Higan'. In 'Maple Arrow' isolines, the source of *e3* and *dt1* is PI196529. In 'Maple Amber' isolines the source of *e4* is 'McCall'.

†Reference codes: 1, Bernard et al. 1991; 2, Voldeng and Saindon 1991*a*; 3, Voldeng and Saindon 1991*b*; 4, Voldeng et al. 1996.

‡'Evans' is heterogeneous at the *E3* locus. An 'Evans'-*e3* isoline was selected using fluorescent extended 20 h photoperiod.

formed in an Eppendorf Mastercycler gradient machine. Each SSR had its annealing temperature optimized by testing a gradient of temperatures, running the resultant PCR products on a 2% w/v agarose gel, and selecting the highest temperature that gave a strong band.

Gel electrophoresis and staining

The amplified DNA was diluted with a 0.75 volume fraction of formamide loading dye (29.5 mL formamide, 110 µL of 2.75 mol NaOH/L, and 0.05 g of xylene cyanol FF). Samples of 4 µL were separated on a 6% w/v denaturing polyacrylamide gel (prepared from a 19:1 acrylamide–bisacrylamide 40% w/v solution purchased from Bioshop (Burlington, Ont.)) with 8 mol urea/L and 1 × TBE buffer. The molecular weight standard used was a 30- to 330-bp ladder (Gibco BRL, Carlsbad, Calif.). Gels were stained with silver nitrate (Fisher, Hampton, N.H.) using Promega's protocol (Technical manual No. 23 available at http://www.promega.com) and scored manually on a light box.

Data analysis

To score the gel, the band observed in 'Harosoy' was defined as the A band, and alternative bands were recorded using additional letters. Multiple banding patterns were recorded using a combination of the letters that were assigned to the corresponding single bands. Faint bands were recorded by putting the corresponding letter in quotation marks. Tables of molecular genotype data were arranged

with markers in columns, sorted by map order, and NILs were arranged in rows, sorted first by recurrent parent and second by genotype at the classical locus, to facilitate testing for correlations between molecular and classical loci.

Results

Six NILs were selected from available NILs contrasting for photoperiod sensitivity (time to flowering and maturity) and determinate vs. indeterminate growth type (Table 1). This set included the recurrent parent 'Harosoy' (*e1E3E4E7*), a derived NIL OT89-5 (*e1e3e4E7*), and four additional NILs that contrasted with OT89-5 for *E1* (OT93-28: *E1e3e4E7*), *E3* (OT94-41: *e1E3e4E7*), *E4* (L62-667: *e1e3E4E7*), and *E7* (OT94-47: *e1e3e4e7*) maturity genes. In those cases where an NIL pair share an allele that differs from the 'Harosoy' allele at that locus, pedigrees confirm or suggest that both members of the NIL pair received the non-'Harosoy' allele from the same donor. When these NILs were genotyped with 430 mapped SSR markers (Cregan et al. 1999), 46 SSR loci possessed more than one allele. At 15 SSR loci, the molecular genotype correlated with one of the four maturity genotypes across all six NILs. Thus, candidate markers and LGs were identified for each of the four maturity loci under study.

The 15 SSR markers were then used to genotype an additional six 'Harosoy'-derived NILs, as well as four 'Maple Presto', three 'Maple Arrow', four 'Maple Amber', and two 'Evans'-derived NILs (Table 1) to test the potential candidate regions, their applicability across germplasm, and refine the putative assignment of map location. Additional flanking SSR markers, selected from the map, were also tested across the 25 NILs to assist in the analysis.

E1 locus

At the *E1* locus, the initial screening of 430 SSR markers across the six 'Harosoy' NILs identified Satt357 and Satt365 on LG C2 and Satt462 on LG L as being potentially linked. When the genotype of six additional 'Harosoy'-derived NILs was determined, the apparent association between the LG L marker and *E1* was lost. In contrast, there was a perfect correlation for Satt365 and an ambiguous result for Satt357, owing to a multiple banding pattern in OT94-51. These two loci are 53.8 cM apart on LG C2 of the USDA – Iowa State University soybean map and unlinked (Cregan et al. 1999). The genotype of additional 'Maple Arrow' derived NILs also gave a multiple banding pattern with Satt357 that was inconsistent with it being linked to *E1* (data not shown). To fine map around the Satt365 locus, the genotype of the NILs was determined at Satt363, Sat076, Satt286, and Satt277, which are proximal to Satt365, and at Satt557, Satt289, and Satt134, which are distal to Satt365. Sat076 was monomorphic. Data for the closest informative flanking markers and for informative NILs is shown in Table 2. The letter coding of banding patterns present a graphical genotype for each NIL and identify the recombination points between 'Harosoy'-derived chromosomal segments (labeled A) and non-'Harosoy'-derived segments. In two NILs, OT93-28 and OT94-51, the correlation to *E1* was broken for each of the proximal markers and in one NIL, OT94-47, the correlation was broken for each of the distal markers. Thus using an approach analogous to deletion mapping, *E1* was mapped to a 2.1-cM interval between Satt277 and Satt557 containing the diagnostic marker Satt365. The banding pattern for Satt365 is shown in Fig. 1. Testing of all the SSR markers on 'Maple Presto' derived NILs was not informative, since Satt286, Satt277, and Satt365 were monomorphic for a third "C" allele. This was consistent with the derivation of the NILs, since PI 196166 was the donor of *E1* to all 'Harosoy'-derived NILs, but a different donor, PI 317334A, was the source of the *E1* allele for all 'Maple Presto' derived NILs (Table 1). 'Maple Arrow', 'Maple Amber', and 'Evans'-derived NILs all carry the *e1* allele, and Satt365 was monomorphic with C, C, and A alleles, respectively (data not shown).

E3 locus

Potential locations for the *E3* locus on LGs H (Satt293), I (Satt367), K (Sat119), and L (Satt229) were identified in the initial screen of 430 SSRs across six 'Harosoy'-derived NILs. When all 12 'Harosoy'-derived NILs were examined, none of the markers except Satt229 on LG L correlated perfectly with *E3* (Table 3). Distal flanking markers Satt373 and Satt513 were monomorphic. Satt513 is only 5.9 cM from the end of the LG, which is therefore presumed to be the distal boundary of the *E3* region. The closest proximal marker (Satt006) was also monomorphic but the next marker (Sct099) was polymorphic and did not correlate with *E3*. Thus, *E3* was mapped to a 32.4-cM interval, between Satt099 and the end of the linkage group, containing the diagnostic marker Satt229. The Satt229 banding pattern in 'Maple Arrow' derived NILs was consistent with linkage to *E3* (data not shown) even though the donor of *e3* was T204 for 'Harosoy'-derived NILs, but PI 196529 for 'Maple Arrow' derived NILs (Table 1). Satt229 was monomorphic and therefore uninformative in the two 'Evans'-derived NILs (data not shown). Satt229 was not diagnostic in 'Maple Presto' and 'Maple Amber' derived NILs, since Satt229 was polymorphic despite all lines carrying *e3* (data not shown).

E4 locus

The preliminary screening of SSR markers across six 'Harosoy' NILs identified three markers Satt367, Satt587, and Satt354, all on linkage group I, as potentially linked to *E4*. When the additional 6 'Harosoy' NILs were tested, the correlation for Satt367 and Satt587 with *E4* broke down. In contrast, the correlation between Satt354 and *E4* was perfect across all 12 'Harosoy'-derived NILs (Table 4). The closest proximal flanking markers to Satt354, which were the coincident Satt239 (data not shown) and Satt496, did not correlate with *E4*. With respect to the distal flanking markers, Satt105, Satt270, and Satt049 were monomorphic, such that the distal boundary of the genomic region carrying *E4* could not be determined. These results suggest that *E4* maps close to the diagnostic marker Satt354. However, Satt354 was not diagnostic across NILs based on other recurrent parents, despite PI438477 being the donor of *e4* to all but the 'Maple Amber' derived NILs (data not shown).

E7 locus

The preliminary screening of SSR markers across the six 'Harosoy'-derived NILs identified LGs C2 (Satt100, Satt319,

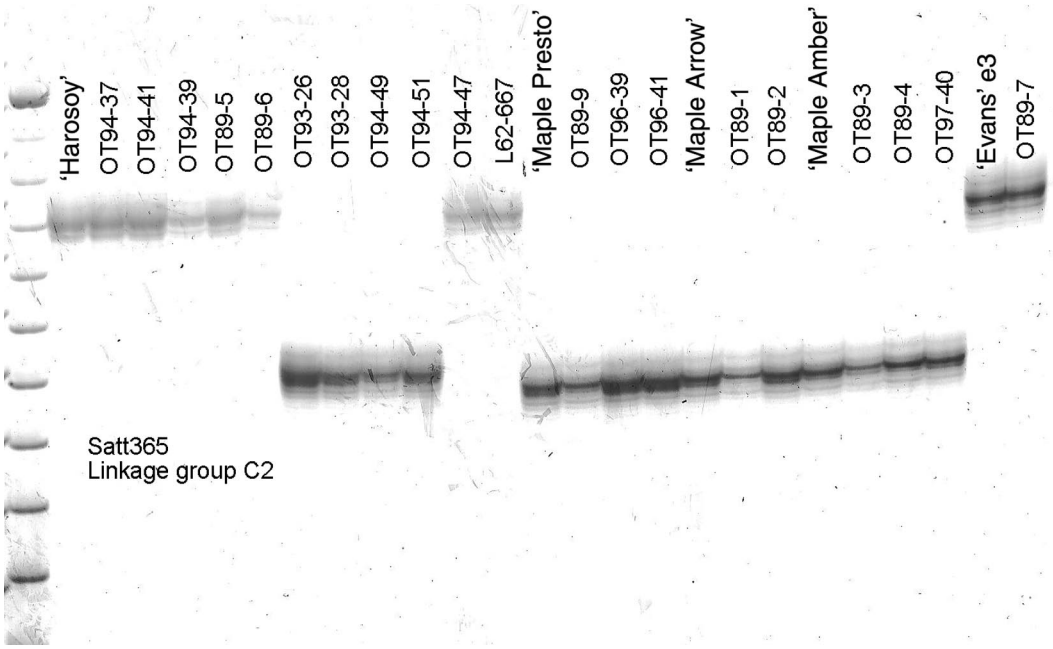
Table 2. SSR genotypes of NILs relative to the *E1* maturity locus.

			LG, SSRs in map order, and distance between SSRs (cM)*				
			C2	C2	C2	C2	C2
Lane in			Satt286	Satt277	Satt365	Satt557	Satt289
Fig. 1	Line	*E1* or *e1*	10.3	0.9	1.2	0.8	
1	'Harosoy'	*e1*	A	A	A	A	A
3	OT94-41	*e1*	A	A	A	A	A
5	OT89-5	*e1*	A	A	A	A	A
11	OT94-47	*e1*	A	A	A	B	B
12	L62-667	*e1*	A	A	A	A	A
8	OT93-28	*E1*	A	A	B	B	B
2	OT94-37	*e1*	A	A	A	A	A
4	OT94-39	*e1*	A	A	A	A	A
6	OT89-6	*e1*	A	A	A	A	A
7	OT93-26	*E1*	B	B	B	B	B
9	OT94-49	*E1*	B	B	B	B	B
10	OT94-51	*E1*	A	A	B	B	B

Notes: The results are presented in Table 2 with NILs in rows and markers in columns sorted by map order, such that each row is analogous to a graphical genotype of an NIL. Further, the NILs in rows were sorted first by recurrent parent and second by genotype at the *E1* locus, to facilitate comparison of the molecular genotypes of *e1* versus *E1* NILs within and between recurrent parent families. Dotted rectangle indicates correspondence between SSR and *E1/e1* genotypes.

*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999)

Fig. 1. The banding pattern for Satt365 (diagnostic for *E1* locus) across NILs. The primer name indicates that it targets a locus having ATT trinucleotide repeat units. Alleles at this locus differ by one or more repeat units. Minor "stutter bands" having one additional or one fewer repeat units flank the main band in the gel photograph and are a PCR amplification artifact common to most SSR markers. The genotype scoring of this gel is presented in Table 2.



Satt460), H (Satt442), and L (Satt497) as potential locations for the *E7* locus (Table 5). Each marker also correlated with *E7* over the remaining six 'Harosoy' NILs. The size of the putative introgressed segment was examined at each candidate location. Satt192, tightly linked to Satt442 on LG H, did not correlate with *E7*, and the other flanking marker Sctt009 was monomorphic; however, this candidate region may be quite small. Likewise, for additional SSR markers flanking Satt497 on LG L, one did not correlate with *E7* and the other was monomorphic so the candidate region on LG L may be of moderate size. However, the three SSR markers on LG C2 that do correlate with *E7* are linked, spanning a 22.2-cM region. Therefore, based on 'Harosoy'-derived NIL data, a region of at least 22.2 cM on LG C2 correlates with *E7* and suggests that LG C2 is the probable location for *E7*. The proximal boundary of the candidate region was defined by flanking markers Satt557, Satt289, and Satt134 that did not correlate with *E7* in four NILs (OT93-28, OT93-26,

Table 3. SSR genotypes of NILs relative to the *E3* maturity locus.

		LG, SSRs in map order, and distance between SSRs (cM)*				
		L	L	L	L	L
		Sat099	Satt006	Satt229	Satt373	Satt513
Line	*E3* or *e3*	15.3	3	10.3	0.6	
'Harosoy'	*E3*	A	A	A	A	A
OT94-41	*E3*	B	A	A	A	A
OT89-5	*e3*	A	A	B	A	A
OT94-47	*e3*	A	A	B	A	A
L62-667	*e3*	A	A	B	A	A
OT93-28	*e3*	A	A	B	A	A
OT94-39	*E3*	A	A	A		
OT94-37	*e3*	A	A	B		
OT89-6	*e3*	A	A	B		
OT93-26	*e3*	A	A	B		
OT94-49	*e3*	AC	A	B		
OT94-51	*e3*	AC	A	B		

Note: Dotted rectangle indicates correspondence between SSR and *E3/e3* genotypes.
*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999).

Table 4. SSR genotypes of NILs relative to the *E4* maturity locus.

		LG, SSRs in map order, and distance between SSRs (cM)*				
		I	I	I	I	I
		Satt496	Satt354	Sat105	Satt270	Satt049
Line	*E4* or *e4*	11.9	20.7	0	5.5	
'Harosoy'	*E4*	A	A	A	A	A
L62-667	*E4*	A	A	A	A	A
OT94-41	*e4*	A	B	A	A	A
OT89-5	*e4*	C	B	A	A	A
OT94-47	*e4*	C	B	A	A	A
OT93-28	*e4*	A	B	A	A	A
OT94-37	*E4*	B	A	A		
OT94-39	*e4*	A	B	A		
OT89-6	*e4*	C	B	A		
OT93-26	*e4*	A	B	A		
OT94-49	*e4*	C	B	A		
OT94-51	*e4*	C	B	A		

Note: Dotted rectangle indicates correspondence between SSR and *E4/e4* genotype.
*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999).

OT94-49, and OT94-51). The distal boundary of this genomic region could not be defined because flanking markers Satt307, Satt316, and Satt202 were monomorphic. Upon genotyping the additional non-'Harosoy' NILs, the LG H marker Satt442 did not correlate for 'Maple Amber' derived NILs, perhaps eliminating LG H as a candidate map location of *E7*. The candidate marker Satt497 on LG L was monomorphic and non-informative on the non-'Harosoy' NILs, thus a candidate location for *E7* on LG L has not been excluded. The LG C2 marker Satt100 did not correlate for 'Maple Presto' derived NILs; Satt319 was monomorphic across the 'Maple Presto', 'Maple Arrow', and 'Maple Amber' derived NILs; and Satt460 did not correlate for 'Maple Arrow' derived NIL OT89-1. Taken together, these results suggested that *E7* maps to a 22.2-cM interval on LG C2 between Satt100 and Satt460 containing the diagnostic marker Satt319.

T locus

The set of six 'Harosoy'-derived NILs that were initially extensively genotyped do not contrast at the *T* locus; however, the full set of 12 'Harosoy'-derived NILs do. Analysis of the available data identified a region on LG C2 as a possible map location for the *T* locus (Table 6). The adjacent markers Satt363, Satt286, and Satt277 correlated perfectly with *T*; however, the distal flanking markers Satt365 and Satt557 did not. There are no proximal flanking markers close to Satt363 that could be tested. On 'Maple Arrow' derived NILs, Satt363 did not correlate for the OT89-2 NIL and on the 'Evans'-derived NIL, a faint "A" band in addition

Table 5. SSR genotypes of NILs relative to the *E7* maturity locus.

		LG, SSRs in map order, and distance between SSRs (cM)*												
		C2	C2	C2	C2	C2	C2	C2	H	H	H	L	L	L
		Satt289	Satt134	Satt100	Satt319	Satt460	Satt307	Satt316	Satt192	Satt442	Sctt009	Satt398	Satt497	Satt462
Line	*E7* or *e7*	0	1.9	2.2	20	2.5	3.6	2	1.2	2.9		7.7	7.0	
'Harosoy'	*E7*	A	A	A	A	A	A	A	A	A	A	A	A	A
L62-667	*E7*	A	A	A	A	A	A	A	A	A	A	A	A	A
OT94-41	*E7*	A	A	A	A	A	A	A	A	A	A	A	A	A
OT89-5	*E7*	A	A	A	A	A	A	A	B	A	A	A	A	A
OT93-28	*E7*	B	B	A	A	A	A	A	A	A	A	A	A	B
OT94-47	*e7*	B	B	B	B	B	A	A	A	B	A	A	B	A
OT94-37	*E7*	A	A	A	A	A	A	A		A			A	
OT94-39	*E7*	A	A	A	A	A	A	A		A			A	
OT89-6	*E7*	A	A	A	A	A	A	A		A			A	
OT93-26	*E7*	B	B	A	A	A	A	A		A			A	
OT94-49	*E7*	B	B	A	A	A	A	A		A			A	
OT94-51	*E7*	B	B	A	A	A	A	A		A			A	
'Maple Presto'	*e7*	B	B	B	B	A	B	B		C			B	
OT89-9	*e7*	B	B	B	B	A	B	B		C			B	
OT96-39	*e7*	B	B	B	B	A	BC	B		C			B	
OT96-41	*e7*	B	B	C, "B"	B	A	B	B		C			B	
'Maple Arrow'	*e7*	B	B	B	B	A	B	B		C			A	
OT89-1	*e7*	B	B	B	B	C	BA	B		C			A	
OT89-2	*e7*	B	B	B	B	A	B, "A"	B		C			A	
'Maple Amber'	*e7*	B	B	B	B	A	B	A		D			A	
OT89-3	*e7*	B	B	B	B	A	B	A		A			A	
OT89-4	*e7*	B	B	B	B	A	B	A		D			A	
OT97-40	*e7*	B	B	B	B	A	B	A		D			A	

Note: Dotted rectangles indicate correspondence between SSR and *E7/e7* genotypes. Letters in quotation marks indicate a faint band.
*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999).

Table 6. SSR genotypes of NILs relative to the *T* maturity locus.

		LG, SSRs in map order, and distance between SSRs (cM)*				
		C2	C2	C2	C2	C2
		Satt363	Satt286	Satt277	Satt365	Satt557
Line	*T* or *t*	0.9	10.3	0.9	1.2	0.8
'Harosoy'	*t*	A	A	A	A	A
OT94-41	*t*	A	A	A	A	A
OT89-5	*t*	A	A	A	A	A
OT94-47	*t*	A	A	A	A	B
L62-667	*t*	A	A	A	A	A
OT93-28	*t*	A	A	A	B	B
OT94-37	*t*	A	A	A	A	A
OT94-39	*t*	A	A	A	A	A
OT89-6	*t*	A	A	A	A	A
OT94-51	*t*	A	A	A	B	B
OT93-26	*T*	B	B	B	B	B
OT94-49	*T*	B	B	B	B	B
'Maple Arrow'	*T*	C	C	C	C	B
OT89-1	*T*	C	C	C	C	B
OT89-2	*T*	D	C	C	C	B
Evans	*t*	D	D	A	A	B
OT89-1	*t*	D	D, "A"	A	A	A

Note: Letters in quotation marks indicate a faint band.
*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999).

to the expected "D" band suggests that Satt286 also did not correlate with *T*. Satt277 was monomorphic on all additional NILs. The results indicated that the *T* locus mapped into an interval of 11.2 cM between Satt286 and Satt365, which included the diagnostic marker Satt277. Sat076, which co-maps with Satt286, was monomorphic on all isoline families.

Dt locus

In the present study, *E3* was located on LG L near Satt229. In the same genomic region, others have reported that the *Dt1* locus is located between Sat099 and Satt006 (Cregan et al. 1999). Because our NILs also contrast for *Dt1*, SSR genotypes were examined to independently determine a map location of *Dt1* (Table 7). Across the 12 'Harosoy'-derived NILs, none of the six SSRs tested showed a perfect correlation with *Dt1*. However, the correlation peaked with Sat099, for which 8 of 12 NILs were consistent with the expected genotype. Sat099 does not have close flanking markers that could be tested; however, the correlations were 7/12 for proximal markers (Satt527, Satt166, Satt448), 6/12 for the distal marker (Satt229), and Satt006 was monomorphic. Results with 'Maple Presto' and 'Maple Arrow' derived NILs were consistent for all SSRs except Satt229. Both Satt229 and Satt006 were excluded by results with 'Maple Amber' derived NILs, whereas 'Evans'-derived NILs excluded Satt527 and Satt448. Taken together, there was weak evidence for locating *Dt1* in the 26.7-cM interval between Satt448 and Satt006, which contains Sat099.

Discussion

The present study located the *E3* locus to LG L, near Satt229, between Satt099 and the end of the LG (Table 8). This refines the position of *E3* relative to the previous reports that linked *E3* to *Dt1* (Cober and Voldeng 1996) and mapped *E3* at 28.0 cM from the *Dt1* locus on the classical map (Cregan et al. 1999). The genomic region around Satt229 does contain multiple related QTL (SoyBase 2002), including days to first flower (Mansur et al. 1993*a*, 1996; Orf et al. 1999), reproductive period (Mansur et al. 1996; Orf et al. 1999), R5 (equivalent to beginning seed) (Mansur et al. 1993*a*), and seed pod maturity (Mansur et al. 1996; Orf et al. 1999; Specht et al. 2001).

As stated above, the *Dt1* locus is also known to be located on LG L, linked to *E3* (Table 8). The *Dt1* locus has been mapped at 0.4 cM from Satt006 on the University of Nebraska map and 12.5 and 2.6 cM, respectively, from the flanking SSR markers Sat099 and Satt006 on the University of Utah map (Cregan et al. 1999). In the present study, the set of NILs did not clearly identify a map location for *Dt1*, but suggested a location between Sat448 and Satt006 that is consistent with both the Utah and Nebraska maps. This imprecision may be due to the relatively large distance between adjacent markers. Alternatively, the NILs may differ in this region owing to chromosomal rearrangements such that a single map location cannot be deduced by comparing the genotypes of the NILs. One approach to investigating the first hypothesis is to consider if between the actual marker loci there could be a hypothetical location that did have a genotype that correlates with *Dt1/dt1*. Consider a hypothetical marker locus X between Satt448 and Sat099 (Table 7). For an NIL such as OT89-5, which had the 'Harosoy' A allele at both Satt448 and Sat099, the 'Harosoy' A allele could be assumed at locus X, unless a rare double recombination had occurred. For an NIL such as OT94-41, where both flanking markers were non-'Harosoy', the allele at X could also be

Table 7. SSR genotypes of NILs relative to the *Dt1* maturity locus.

		LG, SSRs in map order, and distance between SSRs (cM)*							
		L	L	L		L		L	L
		Satt527	Satt166	Satt448		Sat099		Satt006	Satt229
Line	*Dt1* or *dt1*	1.4	0.9	11.4	X	15.3	Y	3	
'Harosoy'	*Dt1*	A	A	A	A	A	A	A	A
OT94-41	*Dt1*	B	B	B	B	B	A (or B)	A	A
OT89-5	*Dt1*	A	A	A	A	A	A (or B)	A	B
OT94-47	*Dt1*	A	A	A	A	A	A (or B)	A	B
L62-667	*Dt1*	A	A	A	A	A	A (or B)	A	B
OT93-28	*Dt1*	AB	AB	A	A	A	A (or B)	A	B
OT93-26	*Dt1*	B	B	B	A (or B)	A	A (or B)	A	B
OT94-39	*dt1*	B	B	B	B (or A)	A	A	A	A
OT94-37	*dt1*	B	B	B	B (or A)	A	B (or A)	A	B
OT89-6	*dt1*	A	A	A	A	A	B (or A)	A	B
OT94-49	*dt1*	A	A	A	C (or A)	C	B or C	A	B
OT94-51	*dt1*	A	A	A	AC (or A)	AC	B (or AC)	A	B

Note: At hypothetical loci X and Y, the SSR genotype is given as either the single probable allele, or if two alleles are possible, then as the allele that correlates with *Dt1/dt1* with the alternative allele in parentheses. Dotted rectangles indicate correspondence between SSR and *Dt1/dt1* genotypes.

*Recombination distances from USDA – Iowa State University map (Cregan et al. 1999).

Table 8. Map positions of *E3* and *Dt1* loci on soybean LG L.

	Marker name and distance between markers (cM) on four partial maps of LG L (modified from Cregan et al. 1999)							
Results of present study	USDA – Iowa State University		University of Utah		University of Nebraska		Classical	
	Satt398	7.7	Satt398	8.6				
	Satt497	7.0	Satt462	3.9				
	Satt462	26.4	L050_7	36.8				
	Satt527	1.4	Satt527	0.4	Satt398	3.3		
	Satt166	0.9	L050_8	5.6	Satt497	21.6		
	Satt448	11.4	Sat099	12.5	*L1*	40.1	*L1*	37.0
Dt1	Sat099	15.3	*Dt1*	2.6	*Dt1*	0.4	*Dt1*	28.0
	Satt006	3.0	Satt006	15.2	Satt006	16.3	*E3*	
E3	Satt229	10.3	Satt513	0.4	Satt373			
	Satt373	0.6	Satt373					
	Satt513							

presumed to be non-'Harosoy'. For an NIL such as OT93-26, where one flanking marker had a 'Harosoy' allele and the other a non-'Harosoy' allele, a recombination must have occurred in that interval and either allele could be expected at locus X. The allele that correlates best with the *Dt1/dt1* genotype of the NIL and the alternate allele (in brackets) are both shown in Table 7. A hypothetical marker and (or) locus X could have a genotype that correlated with *Dt1/dt1* in 10 of 12 NILs. A hypothetical marker Y between Sat099 and Satt229, ignoring the monomorphic genotype of Satt006, could have correlated with *Dt1/dt1* for 11 of 12 NILs. Both of these are higher correlations than observed with actual markers and favour mapping *Dt1* to the 18.3-cM interval between Sat099 and Satt229.

Analysis of the SSR genotypes of NILs contrasting for *E1/e1* alleles mapped this maturity locus to a 2.1-cM interval on LG C2 defined by Satt277 and Satt557 and containing the tightly linked diagnostic marker Satt365 (Table 9). This refines the location and is consistent with the report that *E1* is on LG C2 approximately 4.0 cM from the *T* locus on the classical map, which compares to a distance of 4.9 cM between the *T* locus and Satt277 on the University of Nebraska map (Cober et al. 1997; Cregan et al. 1999).

The use of NILs in the present study also mapped the *T* locus onto LG C2 within an interval of 11.2 cM between Satt286 and Satt365, which includes the diagnostic marker Satt277 (Table 9). Previously, the *T* locus had been mapped to LG C2 4.0 cM distal to *E1* on the classical map and the *T* locus was mapped at 4.9 cM from Satt277 and approximately 24.1 cM from Satt307 on the University of Nebraska map. The Satt277 to Satt307 interval is unusual in that it covers 29 cM on the University of Nebraska map, only 9.1 cM on the more dense University of Utah map, and 29.5 cM on the even more dense USDA – Iowa State University map (Cregan et al. 1999). The linear order of markers, however, is substantially maintained. The *T* locus has also been mapped close to Satt205 on LG C2 (Tasma et al. 2001), and Satt205 is placed co-incident with Satt557 on the SoyBase

Table 9. Map positions of *E1*, *E7*, and *T* loci on soybean LG C2.

Results of present study	Marker name and distance between markers (cM) on four partial maps of LG C2 (modified from Cregan et al. 1999)							
	USDA – Iowa State University		University of Utah		University of Nebraska		Classical	
	Sat062	98.4	Sat062	66.2	A109_3	16.5		
	Satt363	0.9	Satt363	1.2	Sat062	91.2		
	Sat076	0.0	Sat076	1.2	Sat076	2.7		
	Satt286	10.3	Satt286	5.4	Satt286	5.0		
T	Satt277	0.9	Satt277	4.5	Satt277	2.5		
E1	Satt365	1.2	Satt557	0.0	K011_3	2.4	*E1*	4.0
	Satt557	0.8	Satt365	0.6	*T*	24.1	*T*	
	Satt289	0.0	Satt319	0.0	Satt307	3.4		
	Satt134	1.9	Satt134	0.0	Satt316	3.3		
	Satt100	2.2	Satt289	0.0	Satt202			
E7	Satt319	20.0	A109_2	0.2				
	Satt460	2.5	L050_2	0.0				
	Satt307	3.6	Satt100	1.5				
	Satt316	2.0	Satt460	2.3				
	Satt202	19.6	Satt307	3.3				
	Satt357		Satt316	1.2				
			Satt202	31.3				
			Satt357					

Note: Satt205 is only weakly linked to LG C2 on the USDA – Iowa State University map (Cregan et al. 1999); however, it is mapped to LG C2 in two other populations (Tasma et al. 2001) and placed coincident with Satt557 on the SoyBase composite map.

composite map. There is strong evidence that the *T* gene encodes flavonoid 3′-hydroxylase and that the recessive *t* allele carries a single-nucleotide deletion that results in a truncated and non-functional protein (Toda et al. 2002).

In the present study, the *E7* locus was also mapped to a 22.2-cM interval between Satt100 and Satt460 on LG C2 containing the diagnostic marker Satt319 (Table 9). This confirms that *E7* is closely linked to *E1* (Cober and Voldeng 2001*a*), at an approximate distance of 6.1 cM.

The three loci *T*, *E1*, and *E7* cluster on LG C2. Based on their respective diagnostic markers, the order of the loci is *T* (Satt277), *E1* (Satt365), and *E7* (Satt319) (Table 9). Note that the recombination distances between these markers differ markedly between maps. In a different population, an alternate order of *E1, T* and *E7* was suggested by F_3 segregation data, whereas consideration of crossover families predicted a gene order of *T*, *E1*, and *E7* (Cober and Voldeng 2001*a*), which is consistent with the current study. Thus, this region may be unstable with a relatively high recombination rate. This region is also rich in QTL, including those for maturity-related traits. Based on those QTLs, it has been suggested that clustering or accumulation of genes for agronomic traits in this region could have arisen through successive chromosomal rearrangements that were selected for during domestication and cultivation (Mansur et al. 1993*a*).

A major maturity QTL on LG C2 was reported in two populations at the *T* or Satt205 locus (Tasma et al. 2001), which is close to Satt557, and the authors speculate that one or a cluster of genes was responsible. A review of SoyBase (2002) shows that the genomic region containing the LG C2 cluster is rich in QTL for days to first flower (Keim et al. 1990; Mansur et al. 1993*a*, 1996; Orf et al. 1999; Yamanaka 2000), seed pod maturity (Keim et al. 1990; Mansur et al. 1993*b*, 1996; Orf et al. 1999; Specht et al. 2001), reproductive period (Orf et al. 1999), R5 (equivalent to beginning seed) (Mansur et al. 1993*a*), and seed filling period (Mansur et al. 1993*a*).

This study is the first to assign *E4* to an LG or map its location. Our results indicate that *E4* is on LG I and is linked to the diagnostic marker Satt354 (Table 10). A review of SoyBase indicated that this region of the soybean genome did not contain a QTL for maturity. However, at approximately 15 and 20 cM from Satt354 there are seed pod maturity QTL (Sebolt et al. 2000).

Soybean is believed to be an ancient tetraploid, which has become a functional diploid. Large duplications have been identified by Shoemaker et al. (1996) using RFLP probes that map to multiple loci. Duplicated regions can be expected to carry genes having similar function, and this was demonstrated to be the case for oil and protein QTL (Shoemaker et al. 1996). Therefore, the regions on LGs L, C2, and I identified in the current study as the location of maturity genes *E3*, *E1* + *E7*, and *E4*, respectively, were examined using the composite map in SoyBase to identify possible homologous relationships. Interesting results were obtained by analysing the RFLP probe A109, which maps to four loci, and, to a lesser extent, K011. K011 has sequence similarity to retrotransposons (Jarvik and Lark 1998) and maps to many locations, therefore results with K011 should be interpreted with caution. The A109_2 and K011_3 loci are tightly linked on LG C2 between *E1* and *E7* (Table 9). The A109_4 and K011_1 loci are on LG I, close to *E4* (Table 10). The A109_3 locus maps near the top of LG C2, tightly linked to a reproductive period QTL (Orf et al. 1999). The A109_1 and K011_2 loci map to LG B1, with no known maturity loci near them. Thus, three regions carrying loci affecting maturity appear to be homologous. This is a

Table 10. Map position of *E4* locus on soybean LG I.

Results of present study	Marker name and distance between markers (cM) on four partial maps of LG I (modified from Cregan et al. 1999)							
	USDA – Iowa State University		University of Utah		University of Nebraska		Classical	
	Satt571	2.4						
	Satt419	7.4	Satt571	2.1				
	Satt367	5.7	Satt419	5.4				
	Satt127	1.0	Satt367	2.0				
	Satt587	7.0	Satt587	2.1	Satt587	6.6		
	K011_1	1.8	Satt127	2.0	*Enp*	21.8	*Enp*	9.0
	Satt239	0.0	Satt239	1.6	*Ln*	6.0	*Ln*	
	Satt496	11.9	A109_4	12.4	Satt049			
E4	Satt354	20.7	Satt354	0.8				
	Satt105	0.0	Satt270	0.0				
	Satt270	5.5	Satt105	10.5				
	Satt049		Satt049					

significant observation for several reasons. Participation in a homologous relationship is an independent criterion that supports the map position of *E4* identified in this study. These results also predict that an as yet unknown maturity locus may reside on LG B1 near the A109_1 locus. Finally, since homologous loci have evolved from a common progenitor gene, these five maturity loci likely share (a) biological mechanism(s) of action.

The initial search strategy of screening SSR markers on a small number of NILs identified a single candidate map location for *E4*, as might be expected, but multiple candidate map locations for *E1*, *E3*, and *E7*. These additional genomic regions were examined in an effort to explain their recovery through SSR screening. *E1* and *E7* were found to be tightly linked on LG C2. The single alternate candidate location for *E1* was on LG L at Satt462 and one of the three alternate candidate locations for *E7* was also on LG L at Satt497. The latter two SSR loci are tightly linked (7.0 cM) on LG L (Table 8), and also tightly linked to a pod maturity QTL (Lee et al. 1996). Furthermore, the RFLP probe L050 maps to a locus L050_7 in this region of LG L as well as to a locus L050_2 in the *E1* plus *E7* region of LG C2 (Table 9). As with the earlier RFLP probe K011, L050 contains inverted terminal repeats suggesting a transposable element and cautious interpretation is advised (Jarvik and Lark 1998). However, taken together, this suggests that a region on LG L has homology with a region on LG C2 that carries the *E1* and *E7* loci, and this region on LG L has one or more additional loci affecting maturity. One explanation of our observations is that the six NILs used may have also contrasted for maturity at this region on LG L. Alternatively, markers for LG L may have been recovered through the screen simply because the homologous relationship between the two regions on LG C2 and LG L has resulted in their coevolution to become complementary and, therefore, their coselection during development of the six NILs. The hypothesis of such "shadow regions" has been proposed to explain the recovery of markers in homologous regions when using NILs and BSA in oat (O'Donoughue et al. 1996) and BSA in barley (Molnar et al. 2000).

For single-gene traits, the use of NILs to map loci has both advantages and disadvantages relative to using recombination mapping. Full recombination mapping is usually successful, but requires phenotyping many lines (50–200) and genotyping all lines at many loci. A NIL strategy is efficient if it exploits previously mapped markers. It is therefore most efficient in a species with an existing high-density consensus map containing relatively polymorphic markers, and in a species where chromosomal rearrangements are rare such that the marker order in the consensus map can be assumed for the NILs. These conditions are met for soybeans and the integrated soybean SSR map (Cregan et al. 1999). Early in the mapping process NILs are much more efficient than recombination mapping. As illustrated by the present study, genotyping 6 to 12 NILs can identify one or a few candidate locations for most loci, whereas the equivalent genotyping of 6 to 12 random lines from a segregating population would give very little useful information. However, NILs can locate a locus only to an interval between mapped polymorphic markers, whereas recombination mapping is more precise because it can locate a locus to a specific map location between polymorphic markers. For molecular marker-assisted breeding purposes, it is sufficient for markers to be within 5 cM of the target locus to be extremely useful, although greater precision can be required for other purposes. An NIL approach is strictly based on correlations and can not differentiate between multiple candidate map positions, when such occur. This was the common situation when 6 'Harosoy'-derived NILs were used in the present study, and persisted for 1 of the 4 maturity loci even when 12 NILs were considered. Three candidate regions where identified for *E7* on LGs C2, H, and L and the ambiguity resolved by comparison of the length of each candidate region and subsequently confirmed by consistency with published classical genetic linkage studies indicating *E7* was on LG C2. Both the issues of mapping precision and identifying multiple candidate map locations diminish with increasing numbers of related NILs and neighbouring SSRs studied. Both issues also diminish if other complementary genetic information is available. The availability of the requisite NILs or mapping population(s) and careful phenotype data is obvious, since developing these genetic resources is costly and time consuming. BSA has been developed as an alternative to NILs in those cases where NILs are not available but the

phenotype of a segregating population has been determined and individual lines with extreme phenotypes can be used to produce two contrasting bulks (Michelmore et al. 1991). The features of BSA closely parallel those discussed for NILs. In other studies, we have employed an efficient hybrid approach of using NILs and (or) BSA to identify a short list of candidate regions and then recombination mapping to analyse those in detail as necessary (O'Donoughue et al. 1996; Molnar et al. 2000).

In conclusion, the present study demonstrated that NILs are a powerful and efficient tool for marker discovery and mapping to intervals of those single-gene traits for which NILs are available. Using 25 NILs, it was possible to remap the *Dt1* locus, to refine the SSR mapping of the *T* locus, to map *E1* and *E3* with molecular markers, and to map the *E4* and *E7* loci for the first time. As additional SSR or other markers become available in the target regions, each of these loci could be mapped to even shorter intervals. Molecular markers flanking these loci are now available for marker-assisted breeding for these traits, which is significant for short-season breeding programs in Canada and elsewhere. Analysis of map location identified a putative homologous relationship between four chromosomal regions: one in the middle of LG C2 carrying maturity loci *E1* and *E7*, one on LG I carrying *E4*, one at the top of LG C2 harbouring a reproductive period QTL, and the fourth on LG B1 where there is as yet no reported QTL or loci affecting maturity. Other evidence suggests that homology also exists between the *E1* + *E7* region on LG C2 and a region on LG L linked to a pod maturity QTL. These results predict the location of additional maturity loci.

Addendum

A recent 2003 recombination mapping study by Abe et al. (Crop Sci. **43**: 1300–1304) associated the *E1* and *E4* loci with SSR markers consistent with those identified in the present study.

Acknowledgements

The authors gratefully acknowledge the detailed pedigree information provided by Harvey D. Voldeng, the critical review of the manuscript by Doris E. Luckert, Nick A. Tinker, and Ken C. Armstrong, and the financial assistance of the Ontario Soybean Growers and of the Agriculture and Agri-Food Canada Matching Investment Initiative.

References

Bernard, R.L., Nelson, R.L., and Cremeens, C.R. 1991. USDA soybean genetic collection: isoline collection. Soybean Genet. Newsl. **18**: 27–57.

Bonato, E.R., and Vello, N.A. 1999. E_6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. **22**: 229–232.

Cober, E.R., and Voldeng, H.D. 2001*a*. A new soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. Crop Sci. **41**: 698–701.

Cober, E.R., and Voldeng, H.D. 2001*b*. Low R:FR light quality delays flowering of *E7E7* soybean lines. Crop Sci. **41**: 1823–1826.

Cober, E.R., and Voldeng, H.D. 1996. *E3* and *Dt1* linkage. Soyb. Genet. Newsl. **23**: 56–57.

Cober, E.R., Tanner, J.W., and Voldeng, H.D. 1996*a*. Genetic control of photoperiod response in early-maturing, near-isogenic soybean lines. Crop Sci. **36**: 601–605.

Cober, E.R., Tanner, J.W., and Voldeng, H.D. 1996*b*. Soybean photoperiod-sensitivity loci respond differentially to light quality. Crop Sci. **36**: 606–610.

Cober, E.R., Voldeng, H.D., and Morrison, M.J. 1997. Maturity and pubescence color are associated in short-season soybean. Crop Sci. **37**: 424–427.

Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., Van Toai, T.T., Lohnes, D.G., Chung, J., and Specht, J.E. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. **39**: 1464–1490.

Fang, G., Hammar, S., and Grumet, R. 1992. A quick and inexpensive method for removing polysaccharides from plant genomic DNA. BioTechniques, **13**: 52–56.

Jarvik, T., and Lark, K.G. 1998. Transposable elements in soybean: *Mariner* takes a voyage to the plant kingdom. *In* Proceedings of the 8th Gatlinburg Symposium and 7th Biennial Conference on Molecular and Cellular Biology of the Soybean, 26–29 July 1998, University of Tennessee, Knoxville, Tenn. *Edited by* P. Gresshoff. University of Tennessee, Knoxville, Tenn. [Abstr.]

Keim, P., Diers, B.W., Olson, T.C., and Shoemaker, R.C. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics, **126**: 735–742.

Lee, S.H., Bailey, M.A., Mian, M.A.R., Carter, T.E., Jr., Ashley, D.A., Hussey, R.S., Parrott, W.A., and Boerma, H.R. 1996. Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci. **36**: 728–735.

Li, J., and Niwa, M. 1996. Microsatellite DNA markers linked to a gene controlling days to flowering in soybean (*Glycine max*) under shortday conditions. Breed. Sci. **46**: 81–84

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

Mansur, L.M., Orf, J., and Lark, K.G. 1993*b*. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). Theor Appl Genet. **86**: 914–918.

Mansur, L.M, Orf, J.H., Chase, K., Jarvik, T., Cregan, P.B., and Lark, K.G. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. **36**: 1327–1336.

Michelmore, R., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genome regions by using segregating populations. Proc. Natl. Acad. Sci. U.S.A. **88**: 9828–9832.

Molnar, S.J., James, L.E., and Kasha, K.J. 2000. Inheritance and RAPD tagging of multiple genes for resistance to net blotch in barley. Genome, **43**: 224–231.

Muehlbauer, G.J., Staswick, P.E., Specht, J.E., Graef, G.L., Shoemaker, R.C., and Keim, P. 1991. *RFLP* mapping using near-isogenic lines in soybean [*Glycine max* (L.) Merr.]. Theor. Appl. Genet. **81**: 189–198.

O'Donoughue, L.S., Chong, J., Wight, C.P., Fedak, G., and Molnar, S.J. 1996. Localization of stem rust resistance genes and associated molecular markers in cultivated oat. Phytopathology, **86**: 719–727.

Orf, J.H., Chase, K., Jarvik, T., Mansur, L.M., Cregan, P.B., Adler, F.R., and Lark, K.G. 1999. Genetics of soybean agronomic

traits: I. Comparison of three related recombinant inbred populations. Crop Sci. **39**: 1642–1651.

Sebolt, A.M., Shoemaker, R.C., and Diers, B.W. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. **40**: 1438–1444.

Shoemaker, R.C., Plolzin, K., Labate, J., Specht, J., Brummer, E.C., Olsen, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Lochert, G., and Boerma, H.R. 1996. Genomic duplication in soybean (*Glycine* subgenus *soja*). Genetics, **144**: 329–338.

Soybase. 2002. A USDA–ARS Plant Genome Program soybean database. Available at http://soybase.org.

Specht, J.E., Chase, K., Macrander, M., Graef, G.L., Chung, J., Markwell, J.P., Germann, M., Orf, J.H., and Lark, K.G. 2001. Soybean response to water: a QTL analysis of drought tolerance. Crop Sci. **41**: 493–509.

Tasma, I.M., Lorenzen, L.L., Green, D.E., and Shoemaker, R.C. 2001. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. Mol. Breed. **8**: 25–35.

Toda, K., Yang, D., Yamanaka, N., Watanabe, S., Harada, K., and Takahashi, R. 2002. A single-base deletion in soybean flavonoid 3′-hydroxylase gene is associated with gray pubescence color. Plant Mol. Biol. **50**: 187–196.

Voldeng, H.D., and Saindon, G. 1991*a*. Registration of four pairs of 'Maple Presto' derived soybean genetic stocks. Crop Sci. **31**: 1398–1399.

Voldeng, H.D., and Saindon, G. 1991*b*. Registration of seven long-daylength insensitive soybean genetic stocks. Crop Sci. **31**: 1399.

Voldeng, H.D., Cober, E.R., Saindon, G., and Morrison, M.J. 1996. Registration of seven early- maturing Harosoy near-isogenic soybean lines. Crop Sci. **36**: 478.

Yamanaka, N. 2000. Quantitative trait locus analysis of flowering time in soybean using a RFLP linkage map. Breed. Sci. **50**: 109–115.