Theor Appl Genet (2000) 101:789–795

A.J. Hayes · Y.G. Yue · M.A. Saghai Maroof

Expression of two soybean resistance gene candidates shows divergence of paralogous single-copy genes

Received: 29 November 1999 / Accepted: 22 December 1999

Abstract The cloning of several plant genes directly involved in triggering a disease resistance response has shown that numerous resistance genes in the nucleotide binding site (NBS)/leucine-rich repeat (LRR) class have similar conserved amino acid sequences. In this study, we used a short soybean DNA sequence, previously cloned based on its conserved NBS, as a probe to identify full-length resistance gene candidates. Two homologous, but genetically independent genes were identified. One gene maps to the soybean molecular linkage group (MLG) F and a second is coded on MLG E. The first gene contains a 3,279 nucleotide open reading frame (ORF) sequence and possesses all the functional motifs characteristic of previously cloned NBS/LRR resistance genes. The N-terminal sequence of the deduced gene product is highly characteristic of other resistance genes in the subgroup of NBS/LRR genes which show homology to the Toll/Interleukin-1 receptor genes. The C-terminal region is somewhat more divergent as seen in other cloned disease resistance genes. This region of the F-linked gene contains an LRR region that is characterized by two alternatively spliced products which produce gene products with either a four-repeat or a ten-repeat LRR. The second cloned gene that maps to soybean MLG E contains 1,565 nucleotides of ORF in the N-terminal domain. Despite strong homology, however, the 3´ region of this gene contains several in-frame stop codons and apparent frame shifts compared to the F-linked gene, suggesting that its functionality as a disease resistance gene is questionable. These two disease resistance gene candidates are shown to be closely related to one another and to the members of the NBS/LRR class of disease resistance genes.

Key words Nucleotide binding site (NBS) · Leucine-rich repeat (LRR) · Disease resistance

Communicated by G. Wenzel

A.J. Hayes · M.A. Saghai Maroof (✉)
Department of Crop and Soil Environmental Sciences,
Virginia Tech, Blacksburg, VA 24061-0404, USA
e-mail: smaroof@vt.edu, Fax: +1-540-231-3431

Y.G. Yue
Dow AgroSciences LLC, 9330 Zionsville Rd., Indianapolis,
IN 46268, USA

Introduction

Ground-breaking work employing both map-based and transposon tagging strategies has resulted in the cloning of a number of disease resistance genes (for a review see Baker et al. 1997). Several of these genes, including *N* from tobacco (*Nicotiana tabacum*) (Whitham et al. 1994), *L*6 from flax (*Linum usitatissimum*) (Lawrence et al. 1995), and *Rps*2 from *Arabidopsis* (Bent et al. 1994) contain short amino acid sequences that are highly conserved.

Yu et al. (1996) designed degenerate primers from two of these conserved sequences corresponding to the p-loop and kinase-3a coding regions. These conserved motifs are part of the nucleotide binding site (NBS) portion of disease resistance genes, *N* and *Rps*2. At least 11 different NBS classes were identified by Yu et al. (1996) by amplifying corresponding sequences of about 340 bp from soybean containing similar conserved sequences. In a similar study conducted by Kanazin et al. (1996), additional soybean sequences were characterized by amplifying soybean products using degenerate primers corresponding to the p-loop and undefined GLPLAL regions. Additional homology-based studies have been published by groups working with such diverse species as potato (*Solanum tuberosum*) (Leister et al. 1996), tomato (*Lycopersicon esculentum*) (Ohmori et al. 1998), common bean (*Phaseolus vulgaris*) (Rivkin 1999), *Arabidopsis* (Aarts et al. 1998; Speulman et al. 1998), sunflower (*Helianthus annuus*) (Gentzbittel et al. 1998), lettuce (*Lactuca sativa*) (Shen et al. 1998), and the grain crops, rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and maize (*Zea mays*) (Collins et al. 1998; Leister et al. 1998; Seah et al. 1998).

In soybean, several NBS sequence classes have been mapped to regions of the genome where resistance genes have previously been identified. At least five classes have been shown to be associated with one region of molecular linkage group (MLG) J, near resistance genes *Rmd*, and *Rps*2 (Kanazin et al. 1996; Yu et al. 1996). Similarly, two NBS classes map to a resistance gene cluster on MLG F that includes the *Rsv*1, *Rps*3, *Rpv*, and *Rpg*1 loci (Yu et al. 1996).

Several recent studies have shown that these sequences obtained by degenerate primer amplification are useful for identifying full-length disease resistance genes. In *Arabidopsis* the *Rpp*8 gene which confers downy mildew (*Pernospora parasitica*) resistance was cloned using a smaller amplified sequence that contains a portion of the leucine-rich repeat (LRR) domain similar to that of the *Arabidopsis* resistance gene, *Rpm*1 (McDowell et al. 1998). This sequence was shown to cosegregate with the *Rpp*8 locus. Physical identification of a full-length gene sequence was facilitated by BAC and YAC analyses, and the gene was shown to be *Rpp*8 by genetic complementation. A structural homolog of *Rpp*8 which does not confer resistance to downy mildew was identified 3.7 kb downstream of *Rpp*8.

In lettuce the chromosomal region containing the *Dm*3 locus conferring resistance to downy mildew (*Bremia lactucae*) was investigated using a degenerate primer-amplified sequence from lettuce which corresponds to the NBS region from previously cloned NBS/LRR resistance genes (Meyers et al. 1998). Genomic BAC analysis revealed no less than 22 unique but highly homologous genes or pseudogenes in a 3.5-Mb region near *Dm*3. One of these open reading frame (ORF) sequences has been identified as a strong candidate to be the *Dm*3 gene.

Following analysis of the various NBS sequences cloned from soybean (Yu et al. 1996), it was determined that one of these clones, called NBS5, could represent a sequence from an important gene involved in disease resistance. NBS5 is 36.4% similar to the NBS region of the *N* gene based on deduced amino acid structural comparison. Its close proximity to a cluster of resistance genes on the soybean MLG F, including *Rsv*1, *Rpv*, *Rpg*1, and *Rps*3, makes it a potentially important candidate sequence that might represent a functional resistance gene. In this genomic region of MLG F, there are two known classes of NBS sequences. One class, represented by NBS61, is a high-copy repeated gene family which maps to multiple loci in this resistance gene cluster. The other class, represented by NBS5, is essentially single copy upon Southern hybridization. This suggests that this particular gene might prove to be more easily cloned than some other high-copy, multi-gene family, NBS clones.

The objectives of our research are to investigate the expression of genes which correspond to the 336-bp NBS5 sequence and to then characterize these genes as candidate members of the NBS/LRR class of disease resistance genes.

Materials and methods

Library screening

Two lambda cDNA libraries and one lambda genomic library were screened in this study. The first cDNA library was constructed by Clontech (Palo Alto, Calif.) using the lambda TriplEx™ vector. The cDNA were produced from mRNA of 10-day-old leaf tissue of greenhouse-grown plants of soybean line L81-4420. The cDNA for construction of this library were obtained by reverse transcription (RT) using both random and poly-T primers. A second library, obtained from Stratagene (La Jolla, Calif.), was constructed in the lambda Uni-zap XR™ vector. This library was produced from mRNA obtained from 12-day-old epicotyl tissue of greenhouse-grown plants of the cultivar Williams 82. In this library, only poly-T primer was used for cDNA construction. The genomic library was constructed from cv. Williams 82 by Stratagene using the lambda FixII™ vector.

Libraries were planted out onto 150-mm plates according to the manufacturers' protocols. Positive plaques were identified by transfer to Nylon membranes, followed by hybridization with [^{32}P]-labeled probe, NBS5 (and later the partial cDNA clone L2–5i). Briefly, Magnalift (MSI, Westboro, Mass.) nylon membranes were positioned onto freshly grown plates for 3–4 min. They were then washed for 5 min in a denaturing buffer (1.5 *M* NaCl, 0.5 *N* NaOH) followed by a 5-min wash in a neutralization buffer (1.5 *M* NaCl, 0.5 *M* Tris-HCl, pH 7.5). Finally, the lifts were rinsed in 2×SSC before being dried and baked at 90°C for 2 h. Lifts were then prehybridized for 20 h (1 *M* NaH_2PO_4, 1 *M* Na_2HPO_4, 1 *M* PB buffer, 5×Denhardt's solution, and 5 mg/ml salmon sperm DNA). Random-primer-labeled probe was added to the buffer and the lifts were incubated for 20 h at 65°C. Lifts were then washed twice for 5 min with cold low-stringency buffer (2×SSC, 0.5% SDS), followed by a 15-min wash with a higher stringency buffer (0.5×SSC, 0.1% SDS) at 65°C, and a second 5-min wash under the same conditions. Lifts were covered with plastic wrap and exposed to Kodak (New Haven, Conn.) Xomat X-ray film for 8–24 h. Positive plaques were excised from primary plates and were screened a second time using 90-mm plates of a low plaque density to isolate single positive plaques.

Evaluating positive clones

Lambda cDNA positives were converted to plasmid according to the manufacturer's protocol for further analysis. Lambda DNA from the FixII™ genomic library was extracted using the Promega Wizard™ lambda DNA extraction kit, or conversely by $ZnCl_2$ precipitation according to Ming-Tsan et al. (1998).

Linkage mapping

Positive cDNA clones were mapped in two soybean populations. The population V71-370×PI407162 (hereafter referred to as the VP population) was primarily used for identifying the map location of positive clones. This population of 149 individuals (Maughan et al. 1996) contains over 400 mapped loci covering all 20 molecular linkage groups (Saghai Maroof, unpublished). A second population from the cross PI96983×Lee68 (hereafter referred to as the PL population) was used for fine mapping of MLG F around the disease resistance gene cluster. This population contains 243 F_2 individuals segregating for the *Rsv*1 SMV resistance locus and has at least 40 markers placed in this region of MLG F.

Parental diagnostic and F_2 Southern blots were prepared as previously described (Yu et al. 1996). Briefly, 8 µg of parental and/or F_2 DNA was digested with enzymes *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Xba*I, and *Taq*I according to manufacturer's protocols (Gibco-BRL, Rockville, Md.). Digested DNA was then separated on a 1% agarose gel at 70–90 mAmps for 14–16 h. The DNA was transferred to Hybond™ nylon membrane (Amersham, Piscataway, N.J.) by Southern blotting with 0.4 *M* NaOH buffer. Screen-

ing blots were hybridized with α-[^{32}P]-dCTP random-primer labeled probe (Ambion, Austin, Tex.). Hybridizing bands were visualized by autoradiography on Kodak Xomat™ film. Bands were mapped based on genetic distances calculated by MAPMAKER 4.0 at LOD 3.0 (Lander et al. 1987).

DNA sequencing

Positive cDNA clones were sequenced either at the University of Georgia Molecular Genetics facility using an ABI373, or in-house using an ABI377 DNA sequencer (PE Applied Biosystems, Foster City, Calif.). Plasmid template was prepared by a standard alkaline-lysis procedure followed by purification using QiaexII (Qiagen, Valencia, Calif.). Sequencing was performed based on PE Applied Biosystems' recommendations using dye-labeled terminators in a cycle sequencing reaction. Sequence assembly and analysis was conducted using Lasergene software from DNAstar (Madison, Wis.).

Results

Identification of two types of cDNA clones mapping to E and F homoeologous chromosomal regions

In an initial screening of a cDNA library from the line L81-4420 with the NBS5 probe, a 5′ partial length cDNA was isolated and characterized. This partial cDNA was used for subsequent hybridization to a second cDNA library of cv. Williams 82. Approximately 1.44 million plaques were screened from this library. A total of 19 putative positives were excised for secondary analysis. Several identical duplicated cDNA clones were detected in the screen. Of the 19 original positives, 8 unique large-insert positives were evaluated by hybridization to diagnostic blots and by sequence analysis.

The 8 positives evaluated were L9, L10, L20a, L21, L26, L28, L29, and L33. All of the clones evaluated showed similar banding patterns when hybridized to a set of diagnostic blots, but the clones could be separated into two distinct groups. Clones L20a, L26, and L28 hybridized strongly to the same polymorphic restriction fragment (Southern-blotted DNA from the parents of the VP population digested with *Eco*RI), which had previously been mapped to MLG F using NBS5 as a probe (Fig. 1A). The other 5 clones hybridized weakly to this same restriction fragment. Conversely, these 5 clones hybridized strongly to a second polymorphic fragment to which L20a, L26, and L28 hybridized very weakly (Fig. 1B). This band mapped to a previously identified homoeologous region on MLG E (Shoemaker et al. 1996). Based on these observations we deduced that these 8 cDNA clones represent sequences from two unique loci. One gene sequence corresponding to NBS5 is coded for on MLG F, while a second gene sequence, homoeologous to NBS5, is coded for on MLG E.

Of the 8 cDNA clones, 5 were completely sequenced (L20a, L21, L28, L29, and L33), and 3 were partially sequenced (L9, L10, and L26). Sequence analysis indicates that in fact the 8 sequenced clones do fall into the same two distinct sequence classes as previously determined by hybridization pattern. L20a, L26, and L28 share an identical 3′ sequence, and L9, L10, L21, L29 and L33 similarly share an identical 3′ sequence. Six of the clones contain only a partial 3′ sequence and not the 5′ start site of the postulated full-length gene. Based on homology comparisons with previously cloned genes, 2 of the clones, L20a and L33, were postulated to be full-length clones, representing the F-linked and E-linked classes, respectively.

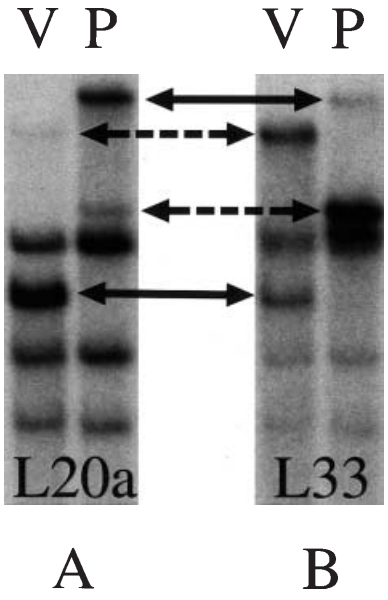

Fig. 1A, B Southern blot of genomic DNA from the mapping parents V71–370 (*V*) and PI407162 (*P*) digested with the restriction enzyme *Eco*RI. **A** Hybridization with the putative F-linked cDNA L20a shows strong hybridization to a single polymorphic band mapping to the F-linkage group (*solid arrows*) and weak hybridization to a polymorphic band mapping to the E-linkage group (*broken arrows*). **B** Hybridization with the putative E-linked cDNA L33 shows weak hybridization to the F-linked band (*solid arrows*) and strong hybridization to the E-linked band (*broken arrows*)

Evaluation of the full-length, F-linked gene

The F-linked resistance gene candidate is best represented by the full-length cDNA clone, L20a, which has an insert size of 3,697 nucleotides, including a 19-nucleotide poly-A tail. The putative start site is located at position 73, and the ORF extends to a stop codon at position 3,354 for a total length of 3,279 nucleotides. This translates to a 1,093-amino acid protein. The clone contains 344 nucleotides of 3′ untranslated sequence. Hybridization of L20a with diagnostic Southern blots shows the expected F-linked restriction fragment length polymorphism (RFLP) pattern.

Sequence analysis indicates that L20a shares numerous conserved defined and undefined motifs with NBS-LRR-type resistance genes (Table 1). It contains the Toll/Interleukin-1 receptor (TIR) homologous region that

Table 1 Regions of conservation shared by soybean cDNA L20a and other NBS/LRR-type disease resistance genes

Consensus sequence[a]	L20a AA position	Homology motif	Homology shared with
YDVFLSFRGxDTR	14–22	Toll homology	*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
KGIxTFxDD	40–48		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
LxxAIxxS (RK) αxααααSx (NG) YAxSxWCLxEL (AV) x1	61–92	Toll homology	*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
VαPαFYDVDPSxQTGx (YF) xKAFxK	103–128		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
WRxALxxαAxαxG	144–156		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
DαVGα (ED) xHα (ED) xααxxL	188–204		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
ααGIxGMGGαGKTTαA (RK) Aα (FY)	214–233	p-loop	*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
LQxxααSEααR	261–271		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
LxxK (RK) αLαVLDDV (DW)	291–304	Kinase-2	*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
GS (RK) αααTTR	322–330	Kinase-3a (putative)	*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
αYEVxxαSxxx (AS) αxLFxx (HK) AF	343–366		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
GLPLALK (VT) αGSxL	385–397		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
L (KR) αSYDxL	424–431		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
KxαFL (DY) (IC) A (ICL) F (FP)	439–446		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4, *I*2, *Rpm*1, *Rps*2, *Cre*3, *Rps*1D
Lxx (RK) SLαxα	473–481		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
αxMHDLαDαGREIVR	487–502		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
PGKRSRαxxxEDxxαLxxxT	509–529		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4
αxxαxxLRxα	555–564		*N*, *M*, *L*6, *Rpp*1, *Rpp*5, *Rps*4

[a] Amino acids in parentheses denote alternate residues at a postion. α indicates an aliphatic amino acid, and x indicates any amino acid

is characteristic of one subgroup of NBS/LRR genes. The NBS region is highly indicative of this class of resistance genes. In addition the C-terminal region contains two LRRs of six and four repeats, respectively, separated by a small spacer sequence (Fig. 2B). Based on clustal alignment analysis (Higgins and Sharp 1989), L20a is most similar to *N*, a virus resistance gene from tobacco, sharing 26% sequence similarity (Fig. 3). It is least similar to *Xa*1, a bacterial blight resistance gene from rice, sharing only 12% similarity.

Using the cDNA clone L20a as a probe, we isolated Williams 82 genomic clones. Sequencing of the entire L20a genomic region showed that this gene contains three introns occurring between the TIR and NBS regions, the NBS and LRR regions, and the LRR and C-terminus regions. These introns are 107, 262, and 102 nucleotides, respectively. The first two introns occur in locations identical to other NBS/LRR genes, including *N* and *L*6.

There is strong evidence to suggest that a fourth intron is alternately spliced out of the F-linked gene. This sequence of 545 nucleotides which is not spliced out of the L20a cDNA is spliced out of another partial F-linked cDNA clone, L28. Polymerase chain reaction (PCR) amplification of this portion of the F-linked gene from genomic DNA indicated that the missing 545 nucleotides were the result of alternate splicing rather than homologous transcribed genes. Interestingly, this alternatively spliced-out region codes for the first of the two LRR regions which make up the COOH LRR portion of the gene product.

Evaluation of the full-length, E-linked clone

Sequencing of the second positive cDNA clone, L33, which best represents the E-linked gene, indicated that it too represents a full-length sequence which is highly similar to the F-linked gene. L33 is 3,525 nucleotides in length including a 15-nucleotide poly-A tail. The start site for this clone begins at position 82. The open reading frame only extends to position 1,646, however. At this point, there is a single stop codon followed by an intron-like sequence that is highly similar to the second intron of L20a. Following this intron-like sequence are remnant ORFs which are highly similar to the 3′ ORF of the F-linked gene (Fig. 4). With the exception of the first 1,565 nucleotides of the ORF sequence, this clone does not contain a continuous translatable sequence. By comparing the sequence of L20a with L33, one could detect the continuation of the translated sequence that shifts frames multiple times and even includes two point mutations to a stop codon. Notably, the alternative intron which is retained in L20a is not present in the transcribed L33 sequence. These differences are immaterial in the L33 cDNA however, since it is clear that this portion of the gene is not translated. Based on this sequence data the E-linked gene does not code for a full-length resistance gene candidate protein product like the F-linked gene but instead codes for a truncated protein product that contains only the TIR homologous region and the NBS region.

Fig. 2A–C Deduced amino acid sequence of the full-length F-linked resistance gene candidate as represented by the cDNA clone, L20a. **A** N-terminal region showing conservation with those regions that are homologous to the *Drosophila* Toll and human Interleukin-1 genes (*underlined*) and those motifs that make up the nucleotide binding site region including the p-loop, kinase 2, and kinase 3a regions (*bold*). **B** Leucine-rich repeat region of L20a showing ten imperfectly conserved repeats of approximately 24 residues. The L20a consensus sequence is shown *below* the repeats in bold. An *x* denotes any amino acid, and an α denotes any aliphatic amino acid. Consensus sequence of LRRs from the cloned disease resistance genes, *I2*-C, *N*, *Rps2*, and *L6* are shown in *bold* in the *dotted box*. **C** Remaining C-terminal portion of the L20a gene product

A

```
MSKAVSESTDIRVYDVFLSFRGEDTRRSFTGNLYNCLEKRGIHTFIGDYD      50
FESGEEIKASLSEAIEHSRVFVIVFSENYASSSWCLDGLVRILDFTEDNH     100
RPVIPVFFDVEPSHVRHQKGIYGEALAMHERRLNPESYKVMKWRNALRQA     150
ANLSGYAFKHGDGYEYKLIEKIVEDISNKIKISRPVVDRPVGLEYRMLEV     200
DWLLDATSLAGVHMIGICGIGGIGKTTLARAVYHSAAGHFDTSCFLGNVR     250
ENAMKHGLVHLQQTLLAEIFRENNIRLTSVEQGISLIKKMLPRKRLLLVL     300
DDVCELDDLRALVGSPDWFGPGSRVIITTRDRHLLKAHGVDKVYEVEVLA     350
NGEALELLCWKAFRTDRVHPDFINKLNRAITFASGIPLALELIGSSLYGR     400
GIEEWESTLDQYEKNPPRDIHMALKISFDALGYLEKEVFLDIACFFNGFE     450
LAEIEHILGAHHGCCLKFHIGALVEKSLIMIDEHGRVQMHDLIQQMGREI     500
VRQESPEHPGKRSRLWSTEDIVHVLEDNTGTCKIQSIILDFSKSEKVVQW     550
DGMAFVKMISLRTLIIRKMFSKGPKNFQILKMLEWWGCPSKSLPSDFKPE     600
KLAILKLPYSGFMSLELPNFLHMRVLNFDRCEFLTRTPD                639
```

B

```
L sg F pi L ke L f F vf C en L ve IH ds      663
V gf L dk L ei M n F eg C sk L et FP p       686
I -k L ts L es I n L sh C ss L vs FP ei      709
L gk M en I th L s L -e Y ta I sk LP ns      732
I re L vr L qs L e L hn C -g M vq LP ss      755
I vt L re L ev L s I cq C eg L -r FS         776
kqdedvknksllmpssylkqvnlwscsisdefidtg         812
L aw F an V ks L d L s- A nn F ti LP sc      835
I qe C rl L rk L y L dy C th L he IR g       859
I ppn L et L sa I r C ts L kd L dla VP le    885
T ke G cc L rq L i L dd C en L qe IR         907
α xx L xx L xx L x L xx C xx L xx αP   L20a consensus

          L xx L x I xx C xx α xx αP   I2-C
L xx L xx L xx L x L   xxxxx   L xx LP   N
α xx L xx L xx L x α   xxxxx   α xx αP   Rps2
α xx L xx L xx L x L xx C                L6
```

C

```
GIPPSIEFLSATNCRSLTASCRRMLLKQELHEAGNKRYSLPGT      950
RIPEWFEHCSRGQSISFWFRNKFPVISLCLAGLMHKHPFGLKPIVSINGN     1000
KMKTEFQRRWFYFEFPVLTDHILIFGERQIKFEDNVDEVVSENDWNHVVV     1050
SVDVDFKWNPTEPLVVRTGLHVIKPKSSVEDIRFIDPYKPTFL           1093
```

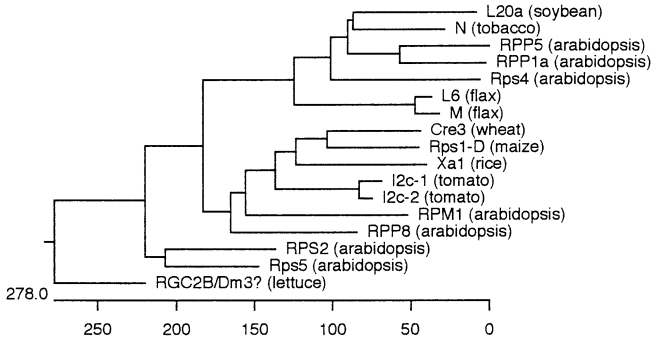

Fig. 3 Phylogenetic tree showing relatedness of the various NBS/LRR disease resistance genes with the F-linked soybean gene represented by L20a. The tree was constructed after alignment of the deduced L20a amino acid sequence with other NBS/LRR gene products using the Clustral method

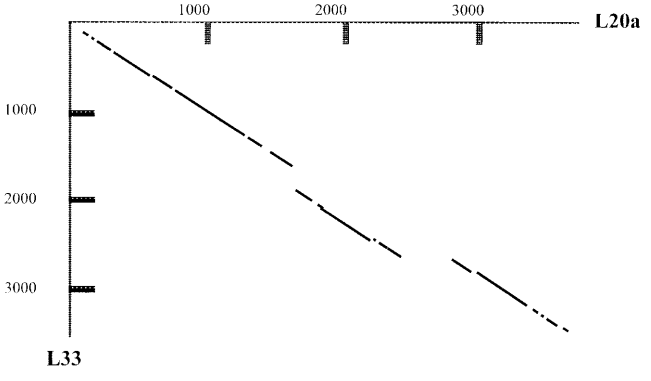

Fig. 4 Dotplot diagram comparing the nucleotide sequence of the F-linked gene, represented by L20a, with the E-linked gene, represented by L33. *Lines* in the diagram signify stretches of nucleotide sequence which share greater than 75% identity. The *first break* in the *line* represents an intron spliced out of the L20a cDNA which is retained in the L33 sequence. The *second break* shows an alternately spliced intron which is retained in L20a. This sequence is not present in the L33 cDNA

Discussion

Using the soybean NBS5 clone, we have identified a full-length disease resistance candidate gene. After completely sequencing this gene, as represented by the L20a cDNA clone, we conducted sequence comparisons to determine the relatedness of this clone to other cloned disease resistance genes (Fig. 3). This gene clearly belongs to the class of plant disease resistance genes that contain an NBS and LRR. The L20a clone most closely resembles the *N* gene cloned from tobacco that confers resistance to tobacco mosaic virus. Sequence similarity between these two genes is 26%. This gene is in a subgroup of NBS-LRR resistance genes, which includes *N*, *L*6 (flax), *M* (flax), *Rpp*1 (*Arabidopsis*), and *Rpp*5 (*Arabidopsis*). All of these gene products contain a N-terminal motif which is similar to the *Drosophila* Toll and interleukin-1 receptor proteins. The similarities of these disease resistance genes to L20a range from 20.4% to 26%. A second subgroup of NBS-LRR genes which includes *Rps*2 (*Arabidopsis*), *Rpm*1 (*Arabidopsis*), *Rpp*8 (*Arabidopsis*), *I*2 (tomato), *Xa*1 (rice), and the putative *Cre*3 gene (wheat) and *Dm*3 gene (lettuce) are similar to L20a in the range of 11.6% to 14.6%.

The N-terminal region of the translated F-linked gene suquence contains numerous conserved motifs, both defined and undefined, that are shared by the members of the NBS/LRR class of disease resistance genes. These include the p-loop, kinase-2, and putative kinase-3a regions as defined by Lawrence et al. (1995), who also described other undefined regions of amino and conservation. All the potentially significant conserved motifs shared between L20a and other cloned resistance genes that are present in the N-terminal half of the protein are listed in Table 1.

In addition to N-terminal conservation in the putative protein product, there are conserved characteristics in the COOH-terminal region, which further suggest a disease resistance role for this gene product. Like all genes in this class, this soybean gene contains a region that resembles the common LRR found on all NBS-LRR resistance genes in this portion of the protein product (Fig. 2). While no two LRRs in this superfamily of genes are identical, all share commonalities of structure. LRRs are hypothesized to play a role in protein-protein interactions, which suggests that this portion of the gene is involved in pathogen recognition (Baker et al. 1997). It is perhaps not surprising then that all of these resistance genes show slightly different LRR patterns, lending to their various specificities.

It is interesting to note that the F-linked gene actually contains an alternate splice site in the LRR region. Gene products like L20a whose translation includes this sequence contain two similar LRRs of six and four repeats, consecutively. Other gene products like the partial cDNA, L28, have this region spliced out and thus contain only a four-repeat LRR. It is tantalizing to speculate at the role this may play in disease resistance. Alternate splicing has been investigated in the flax *L*6 gene and the tobacco *N* gene (Ayliffe et al. 1999). With *L*6, the alternate splicing results in truncated products that do not contain an LRR. It is of interest to note that in the L20a gene the reading frame is retained in the intron region such that two transcripts, both of which contain an LRR region, are produced. Several studies have implicated the LRR region as being involved in pathogen recognition (Ellis et al. 1997; Jones and Jones 1997).

Soybean is thought to be derived from an allotetraploid progenitor (Hymowitz and Singh 1987), therefore it is not surprising that we identified a paralogous coding sequence on MLG E. Shoemaker et al. (1996) have previously reported an extensive region of homoeology between the soybean E and F linkage groups. It is apparent from the sequence data that this paralogous gene on E is not a complete resistance gene as NBS-LRR disease resistance genes are understood. Instead, it seems to represent the genetic degradation of a resistance locus. That is to say, a number of mutations are evident in the E-linked gene, as represented by L33, which result in frame shifts and stop codons, as compared to the F-linked gene.

The entire LRR portion of the E-linked gene product is presumably not translated, even though it is retained in the transcribed mRNA. It is interesting to note, however, that the N-terminal region of this E-linked gene remains intact. There is an uninterrupted ORF through the entire portion of the gene that is highly conserved in other NBS-LRR resistance genes. This portion of the gene codes for the NBS region that is postulated to be involved in signal transduction following triggering of the defense response. This conservation of ORF sequence could indicate some potential role for this truncated protein product. In the *L*6 and *N* resistance genes, transcription of truncated genes which only contain the NBS region have been observed (Dinesh-Kumar et al. 1995; Lawrence et al. 1995). These truncated products have been postulated to be involved in the regulation of transcription. Ayliffe et al. (1999) further speculate that with *L*6, these truncated products could act in *trans* from homoeologous regions based on transformation experiments. The observation of truncated products which are transcribed in soybean could add credence to this theory.

An apparent disease resistance gene having been identified in soybean, the next major goal remaining is the identification of the pathogen interaction in which the L20a gene is involved, if any. There is the likelihood that the L20a sequence cloned from cv. Williams 82 represents a susceptible allele, in which case transformation experiments would be pointless. Pedigree information does not mention the presence of any resistance genes in Williams 82 that are known to be present on MLG F, with the exception of *Rpg*1, conferring resistance to bacterial blight. Preliminary studies by Ashfield et al. (1998) indicate that there is genetic distance between *Rpg*1 and NBS5. We have conducted sequence analysis on the F-linked gene from a number of lines in soybean which carry various resistance genes and alleles from the F linkage group cluster (data not reported). Evidence

from sequencing *Rps*2 from several *Arabidopsis* lines indicated that resistant alleles have highly conserved sequences, while susceptible alleles tend to be more divergent as a group (Caicedo et al. 1999). Our sequence comparisons show that the F-linked gene is highly conserved between lines and, thus, this sequence information from multiple soybean lines has not been informative with respect to the nature of this resistance gene. Further studies will have to be conducted in order to determine the nature of this apparent disease resistance gene in soybean.

Acknowledgments This study was supported in part by the U.S. Dept. of Agric., Natl. Research Initiative, Competitive Grants Program, Grant no. 96-35300-3648 and by the United Soybean Board.

References

Aarts MGM, Hekkert BL, Holub EB, Beynon JL, Stiekema WJ, Pereira A (1998) Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol Plant-Microbe Interact 11:251–258

Ashfield T, Danzer JR, Held D, Clayton K, Keim P, Saghai Maroof MA, Webb DM, Innes RM (1998) *Rpg*1, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theor Appl Genet 96:1013–1021

Ayliffe MA, Frost DV, Finnegan EJ, Lawrence GJ, Anderson PA, Ellis JG (1999) Analysis of alternative transcripts of the flax *L*6 rust resistance gene. Plant J 17:287–292

Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP (1997) Signaling in plant microbe interactions. Science 276:726–733

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz B (1994) *Rps*2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856–1860

Caicedo A, Schaal B, Kunkel B (1999) Diversity and molecular evolution of the *Rps*2 resistance gene in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96:302–306

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant-Microbe Interact 11:968–978

Dinesh-Kumar SP, Whitman S, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene N: similarity to Toll and the interleukin-1 receptor. Cell 78:1101–1115

Ellis J, Lawrence G, Ayliffe M, Anderson P, Collins N, Finnegan J, Frost D, Luck J, Pryor T (1997) Advances in the molecular genetic analysis of the flax-flax rust interaction. Annu Rev Phytopathol 35:271–291

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Labrouhe DTD, Nicolas P (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus*, L. Theor Appl Genet 96:519–525

Higgins DG, Sharp PM (1989) Fast and sensitive mutliple sequence alignment on a microcomputer. CABIOS 5:151–153

Hymowitz T, Singh RJ (1987) Taxonomy and speciation. Agronomy 16:23–48

Jones DA, Jones JDG (1997) The roles of leucine-rich-repeat proteins in plant defenses. Adv Bot Res 24:89–167

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11746–11750

Lander ES, Green P, Abrahamson J, Barlow A, Daly JM, Lincoln SE, Newberg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The *L*6 gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *Rps*2 and the tobacco viral resistance gene *N*. Plant Cell 7:1195–1206

Leister D, Ballvora A, Salamini F, Gephardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genet 14:421–428

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologous in cereal genomes. Proc Natl Acad Sci USA 95:370–375

Maughan PJ, Saghai Maroof MA, Buss GR (1996) Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor Appl Genet 93:574–579

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *Rpp*8 locus of *Arabidopsis*. Plant Cell 10:1861–1874

Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Michelmore RW (1998) The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell 10:1817–1832

Ming-Tsan S, Venkatesh TV, Bodmer R (1998) Large- and small-scale preparation of bacteriophage lambda lysate and DNA. BioTechniques 25:44–46

Ohmori T, Murata M, Motoyoshi F (1998) Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theor Appl Genet 96:331–338

Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42:41–47

Seah S, Sivasithamparam K, Karakousis A, Lagudah E (1998) Cloning and characterisation of a family of disease resistance gene analogs from wheat and barley. Theor Appl Genet 97:937–945

Shen KA, Meyers BC, Faridi MNI, Chin DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol Plant-Microbe Interact 11:815–823

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP (1996) Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329–338

Speulman E, Bouchez D, Holub EB, Beynon JL (1998) Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. Plant J 14:467–474

Whitman S, Dinesh-Kumar SP, Choi D, Hehl R, Corr C, Baker B (1994) The product of the tobacco mosaic virus resistance gene *N*: similarity to Toll and the Interleukin-1 receptor. Cell 78:1101–1115

Yu YG, Buss GR, Saghai Maroof MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 93:11751–11756