Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (NBS) type from *Lens* species

M.W.F. Yaish, L.E. Sáenz de Miera, and M. Pérez de la Vega

Abstract: Most known plant disease-resistance genes (R genes) include in their encoded products domains such as a nucleotide-binding site (NBS) or leucine-rich repeats (LRRs). Sequences with unknown function, but encoding these conserved domains, have been defined as resistance gene analogues (RGAs). The conserved motifs within plant NBS domains make it possible to use degenerate primers and PCR to isolate RGAs. We used degenerate primers deduced from conserved motifs in the NBS domain of NBS–LRR resistance proteins to amplify genomic sequences from *Lens* species. Fragments from approximately 500–850 bp were obtained. The nucleotide sequence analysis of these fragments revealed 32 different RGA sequences in *Lens* species with a high similarity (up to 91%) to RGAs from other plants. The predicted amino acid sequences showed that lentil sequences contain all the conserved motifs (P-loop, kinase-2, kinase-3a, GLPL, and MHD) present in the majority of other known plant NBS–LRR resistance genes. Phylogenetic analyses grouped the *Lens* NBS sequences with the Toll and interleukin-1 receptor (TIR) subclass of NBS–LRR genes, as well as with RGA sequences isolated from other legume species. Using inverse PCR on one putative RGA of lentil, we were able to amplify the flanking regions of this sequence, which contained features found in R proteins.

Key words: disease resistance genes, comparative analysis, lentils, TIR, LRR.

Résumé : La majorité des gènes connus de résistance (gènes R) chez les plantes incluent des domaines tels que des sites de liaison de nucléotides (NBS) ou des régions riches en leucines (LRR). Des séquences de fonction inconnue, mais codant pour de tels domaines, ont été appelées analogues de gène de résistance (RGA). Les motifs conservés au sein des domaines NBS rendent possible l'utilisation d'amorces dégénérées pour isoler des RGA par PCR. Les auteurs ont employé des amorces dégénérées, basées sur les motifs conservés au sein du domaine NBS des protéines de résistance NBS–LRR, afin d'amplifier des séquences génomiques chez des espèces du genre *Lens*. Des fragments d'environ 500–850 pb ont été obtenus. Le séquençage de ces amplicons a révélé 32 RGA chez ces espèces et ces séquences montraient une grande similarité (jusqu'à 91 %) avec des RGA provenant d'autres espèces. Les séquences peptidiques prédites de ces gènes de lentilles contenaient tous les motifs conservés (boucle P, kinase-2, kinase-3a, GLPL et MHD) présents chez la majorité des autres gènes de résistance de type NBS–LRR connus chez les plantes. Des analyses phylogénétiques groupent les séquences NBS de lentilles au sein de la sous-classe TIR (« Toll and interleukin-1 receptor ») des gènes NBS–LRR de même qu'avec des séquences RGA isolées chez d'autres espèces de légumineuses. À l'aide de la PCR inverse, les auteurs ont réussi à amplifier les séquences adjacentes pour un RGA putatif de lentille et celles-ci montraient des caractéristiques typiques retrouvées dans les protéines R.

Mots clés : gènes de résistance, analyse comparée, lentilles, TIR, LRR.

[Traduit par la Rédaction]

Introduction

Over the past few years, many plant disease-resistance genes (R genes) and related genes have been sequenced. Most of the functionally defined R genes cloned to date encode products with similar amino acid sequences and structural domains irrespective of whether they confer resistance to viruses, bacteria, fungi, nematodes, or insects (Baker et al. 1997; Bent 1996; Ellis et al. 2000; Hammond-Kosack and Jones 1997). These sequences include nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains. These domains seem to be components of signal transduction sys-

Received 16 October 2003. Accepted 4 March 2004. Published on the NRC Research Press Web site at http://genome.nrc.ca on 23 July 2004.

Corresponding Editor: R. Appels.

M.W.F. Yaish,[1] L.E. Sáenz de Miera, and M. Pérez de la Vega.[2] Departamento de Genética, Facultad de Biología, Universidad de León, 24071 León, Spain.

[1]Present address: Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.
[2]Corresponding author (e-mail: degmpv@unileon.es).

 doi: 10.1139/G04-027

tems involved in plant resistance against pathogens (Michelmore and Meyers 1998); so far, the only role demonstrated for NBS–LRR-encoding plant genes is either in disease or pest resistance (Michelmore 2000). Therefore, the majority of plant resistance genes are members of large gene families that encode NBS and LRR motifs. Many are present in complex gene clusters (Ferrier-Cana et al. 2003; Meyers et al. 1998; Noël et al. 1999). It is estimated that the *Arabidopsis* genome contains approximately 200 NBS-encoding genes (Meyers et al. 1999). The list of known resistance genes with NBS and LRR includes, among others, *N* from tobacco; *RPS2*, *RPP5*, and *RPM1* from *Arabidopsis*; *L6* and *M* from flax; *Mi* from tomato; and *Xa1* from rice (Pan et al. 2000). In several higher plant taxa, such as legume species, functional resistance genes of this type have not yet been identified; although many sequences coding for internal domains conserved in known resistance genes have been isolated from numerous plant species. These sequences have been named resistance gene analogues (RGAs) or resistance gene homologues (RGHs) (Cannon et al. 2002). RGAs have been isolated from several legumes such as soybean, common bean, cowpea, pea, chickpea, and alfalfa and related species (Cannon et al. 2002; Chida et al. 2000; Cordero and Skinner 2002; Creusot et al. 1999; Ferrier-Cana et al. 2003; Geffroy et al. 1998; Graham et al. 2000; Kanazin et al. 1996; Rivkin et al. 1999; Yu et al. 1996; Zhu et al. 2002), as well as from many non-leguminous species.

NBS plant resistance genes are subdivided into two families based on the presence or absence of an N-terminal region with homology to *Drosophila* Toll and the human interleukin-1 receptor (TIR region) (Meyers et al. 1999; Pan et al. 2000; Young 2000), namely TIR and non-TIR genes. The non-TIR type commonly has a predicted leucine-zipper (LZ) domain or a putative coiled-coil (CC) domain at the N-terminal region (Pan et al. 2000). The non-TIR family is widely distributed in both monocotyledonous and dicotyledonous species, whereas the TIR family appears to be restricted to dicotyledonous species, since TIR-class genes have not been detected in genomic or expressed sequences from any monocot species (Meyers et al. 1999; Pan et al. 2000; Cannon et al. 2002). In *Arabidopsis*, approximately 150 sequences fall into the TIR group and 50 into the non-TIR group (Meyers et al. 1999).

Most of the functionally known resistance genes have been cloned and identified using transposon tagging and map-based cloning technologies, but the use of these approaches for the identification of new resistance genes is still laborious and time-consuming. On the other hand, there are eight major conserved amino acid motifs in the NBS (Meyers et al. 1999) that can be used to design degenerate primers to identify RGAs by means of PCR amplification and sequencing. This approach to identify RGAs has been used in a variety of plant species, such as soybean (Kanazin et al. 1996; Yu et al. 1996), common bean (Ferrier-Cana et al. 2003; Rivkin et al. 1999), alfalfa (Cordero and Skinner 2002; Zhu et al. 2002), lettuce (Shen et al. 1998), coffee (Noir et al. 2001), and grape (Di Gaspero and Cipriani 2002).

Lentil (*Lens culinaris* Medik.) is a grain legume grown in many temperate areas mainly for human food. In spite of this, lentils have been less researched than other legume crops. The genus *Lens* Miller is included in the tribe Vicieae (Papilionacea, Leguminosae), which contains three other genera with cultivated species, namely, *Lathyrus* L., *Pisum* L., and *Vicia* L. *Lens culinaris* is the only cultivated species of *Lens*. In this study, we describe the isolation of a family of NBS sequences from *Lens* species using degenerate PCR primers, and the analysis of their structural motifs. The phylogenetic relationships of the NBS sequences from *Lens* species and other legume species are also analyzed.

Materials and methods

Plant material, DNA extraction, and quantification

Genomic DNA was isolated from dry seed from each *Lens* species (*L. culinaris* Medik., *Lens orientalis* Boiss, *Lens odemensis* Godr., *Lens ervoides* (Bring.) Grande, and *Lens nigricans* (Bieb.) Godr.; all diploid species with $2n = 14$). In the case of *L. culinaris*, the ILL5588 (Reg. No. GP-6, PI 592998) lentil accession was used because it was registered as being resistant to Vascular Wilt and *Ascochyta* Blight (Erskine et al. 1996). Each dry seed was crushed in a micromill and the flour was transferred to a 1.5-mL Eppendorf tube. DNA was extracted as described by Dellaporta et al. (1983) with minor modifications. DNA samples were quantified using a spectrophotometer at wavelength 260 nm and checked by electrophoresis in 1% *w*/*v* agarose gels.

Amplification and product analysis

Primers and PCR conditions

Oligonucleotide primers used in this study were either designed on the basis of conserved motifs in the NBS of known resistance proteins (RPP5 and N) using OLIGOTEST version 2.0 (Beroud et al. 1990) or obtained from previously published works (Table 1). Two hundred nanograms of DNA were used as template in 50-μL PCRs containing 1 μM of each primer, 10 mM Tris–HCl (pH 8), 100 mM KCl, 0.05% *w*/*v* gelatin, 1.5 mM $MgCl_2$, and 2 U *Taq* DNA polymerase (Promega, Madison, Wis.). The PCR program comprised a denaturation step of 95 °C for 1 min, followed by 38 cycles of 94 °C for 40 s, 45 °C for 60 s, and 72 °C for 120 s. The size of the amplified fragments was determined using 1% *w*/*v* agarose gel electrophoresis.

Cloning and sequencing of the PCR products

PCR amplified bands were recovered from agarose gels using the Geneclean method (Boyle and Lew 1995). Each of these bands represented a heterogeneous population of PCR products as determined by restriction endonucleases. PCR products were individualized by cloning into the pGEM-T vector system (Promega) and transformed in *Escherichia coli* DH5α competent cells. The positive clones were selected and recombinant plasmids isolated using the Holmes and Quigley (1981) miniprep method. One hundred clones of each transformation population were randomly chosen and grouped according to their restriction patterns using *Hae*III, *Rsa*I, *Msp*I, and *Taq*I. Three clones, representative of each group identified, were sequenced by the dideoxynucleotide chain termination method (Sanger et al.

Table 1. Primer sequences used for amplification from genomic DNA.

Primer	Motif	Oligonucleotide sequence[a] (5′→3′)	Combination	Product (bp)	Reference
B1	GGVGKT	GGIGGIRTIGGIAARACIAC			Rivkin et al. 1999
B2	GLPLAL	WTIARIGYIARIGGIARICC	B1/B2	K (512–515)	Rivkin et al. 1999
Q1[b]	FLDIACF	RAARCAIGCSATRTCIARRAA	B1/Q1	Q (669–684)	Pan et al. 2000
L2	MHD	CCKNSAGYMNRTCRTGCAT	B1/L2	C (817–832)	This study
INV1	—	CACATCATCAAGAACAAGAAGA			This study
INV2	—	GGAAGTAGAGTAATCATCACAAG	INV1/INV2	Variable	This study

[a]Codes for degenerate positions: I, inosine; R, A/G; Y, C/T; K, G/T; N, A/C/G/T; D, A/G/T; M, A/C; S, G/C; H, A/T/C; W, A/T and V, G/A/C.
[b]Q1 was named in Pan et al. (2000) as 17696.

1977). DNA sequencing was carried out using universal and reverse primers, the Thermo Sequenase Fluorescent Labelled Primer Cycle Sequence Kit (Amersham Pharmacia Biotech, Piscataway, N.J.) following the manufacturer's instructions; automatic sequencing; and ALF^{TM} Manager version 2.6 (Amersham Pharmacia Biotech). Each NBS clone was sequenced at least three times.

Inverse PCR

Genomic DNA was independently digested with 20 restriction enzymes, 4 of which, *Dra*I, *Eco*RI, *Hae*I, and *Hin*dIII, were selected because they produced large inverse PCR products. After the restriction, DNA was self ligated and extracted once with phenol–chloroform and once with chloroform – isoamyl alcohol. The sample was ethanol precipitated and resuspended in H_2O. Circularized DNA was linearized with *Xho*I (at 37 °C for 16 h), since previous sequencing data of the putative NBS RGAs of lentil indicated that there was a single *Xho*I site inside these sequences. The digested DNA was re-extracted with the same procedure, precipitated, and resuspended in H_2O. Inverse PCR (IPCR) was performed according to Silver and Keerikatte (1989) from 30-ng aliquots of DNA using Platinum® *Taq* DNA polymerase (GIBCO-BRL, Carlsbad, Calif.) with an annealing temperature of 50 °C and the INV1 and INV2 primers (Table 1).

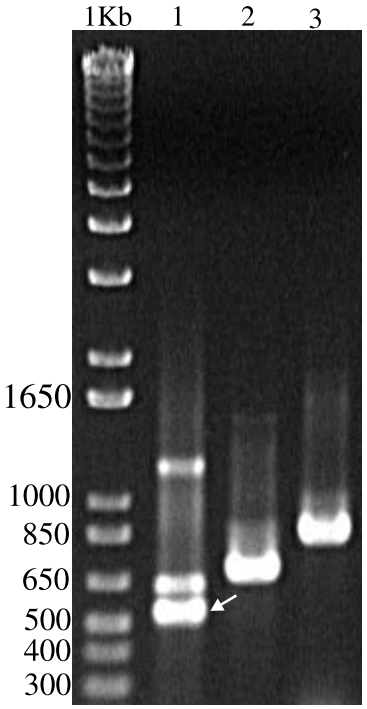
Database search and molecular modeling

The initial database search was carried out using the BLAST Web page (http://www.ncbi.nlm.nih.gov/blast) of the National Center of Biotechnology Information (NCBI, Bethesda, Md.). To analyze the protein domains, we applied ProDom-CG (http://www.toulouse.inra.fr/prodomCG.html). LRR motifs were identified using the BLOCKS database (Henikoff and Henikoff 1994) (http://bioweb.pasteur.fr). In addition to the NBS sequences of *Lens* species, NBS-encoding R gene sequences and other RGA sequences from legume were used for comparative phylogenetic analysis.

Sequence alignment, construction of phylogenetic trees, and nucleotide polymorphism

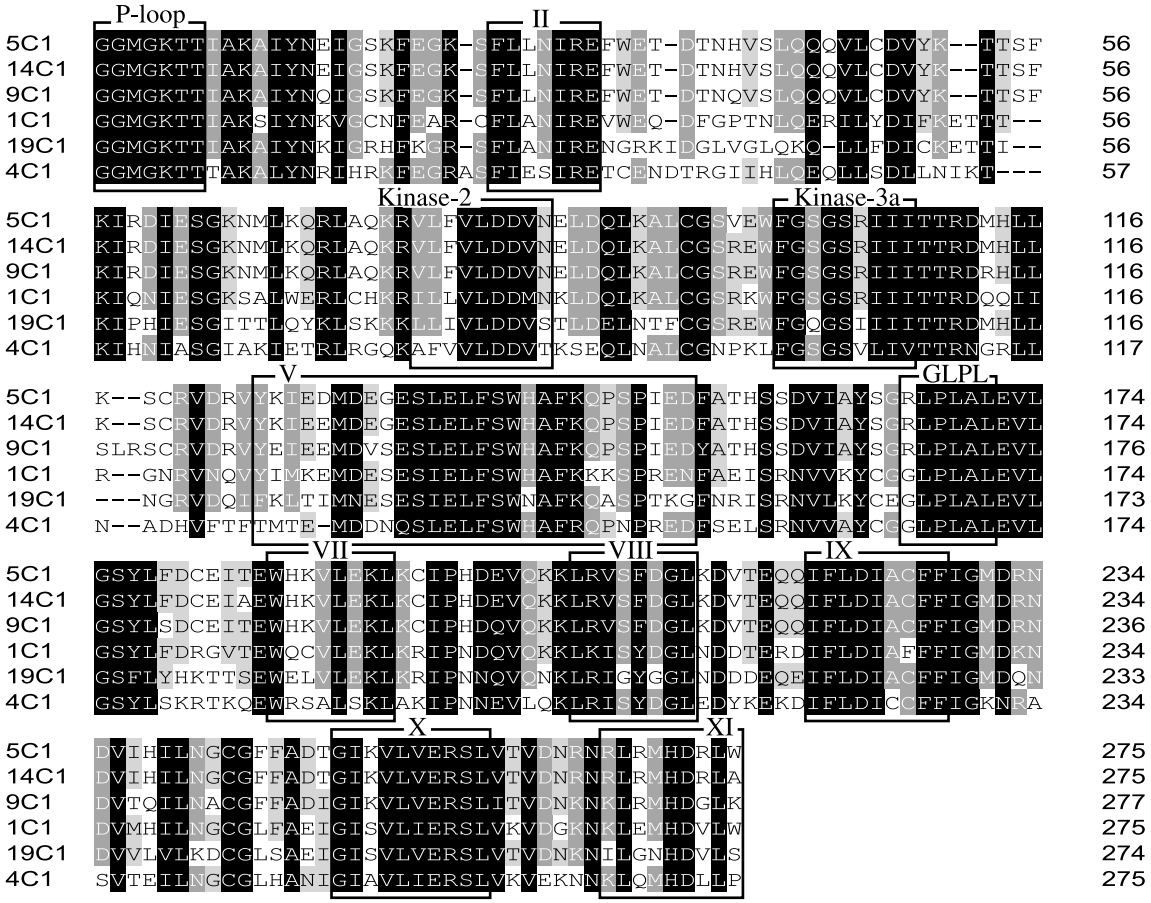
The *Lens* DNA sequences obtained in this study were deposited in the EMBL Database under Accession numbers AJ516060 to AJ516091. Phylogenetic analyses were performed using the putative lentil RGA nucleotide sequences as well as with their deduced protein sequences. Alignment of the nucleotide sequences and of the putative amino acid sequences (from *Lens* and other plant NBS sequences from EMBL and GeneBank databases) were performed using

Fig. 1. PCR products amplified with the degenerate primers, B1/B2, B1/Q1 and B1/L2 are represented in lanes 1, 2 and 3 respectively. The "1Kb" ladder (GIBCO-BRL) was used as size markers. The arrow indicates the band cloned with the first primer combination.



CLUSTAL X (Thompson et al. 1997) with the standard parameters; for the protein sequences, the PAM 250 matrix of Dayhoff (1979) was used. For comparative analyses, only the segments common to all sequences after trimming the primer annealing sites and deleting sequence gaps were considered. Phylogenetic analyses were carried out using of the PHYLIP 3.5 (Felsenstein 1993) software package. Distances were calculated using Kimura's two-parameter (Kimura 1980) model and phylogenetic trees were obtained using the Fitch and Margoliash (1967) method with DNADIST and FITCH programs. For protein sequences the substitution rates were corrected for multiple hits according to Dayhoff's PAM matrix (Dayhoff 1979) using the PROTDIST program from the PHYLIP package. The number of synonymous (K_s) and nonsynonymous (K_a) substitutions were calculated using the Nei and Gojobori (1986) method with the MEGA 2.1 software package (Kumar et al. 2001). Nucleotide diversity, P_i (Nei 1987), of the NBS of lentil sequences was calculated

Fig. 2. Sequence alignment of the deduced amino acid sequences of the six *Lens culinaris* NBS sequences in cluster J (Fig. 4). Conserved residues are shaded and conserved motifs are numbered as in Pan et al. (2000); P-loop; Kinase-2, Kinase-3a and GLPL motifs correspond to sequences I, III, IV and VI. Sequences I, VI, IX and XI were used to design degenerate primers for the PCR amplification of NBS sequences.

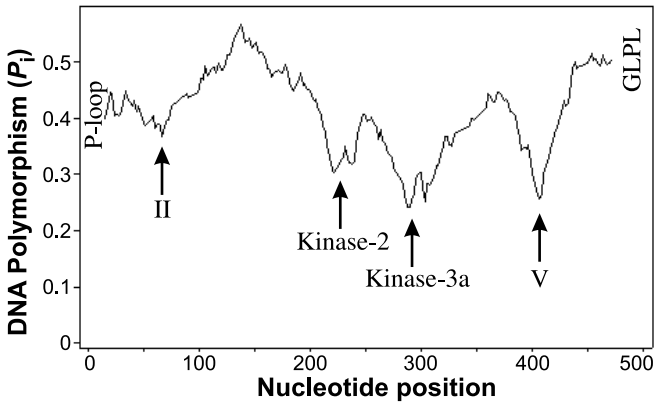


in shifting windows of 30 nucleotides with a step size of 2, considering all sequences as haplotypes of the same locus. This parameter was calculated using the DNASP 3.51 software program (Rozas and Rozas 1999). For phylogenetic analyses, the nomenclature and some sequences used by Cannon et al. (2002) were used.

Results

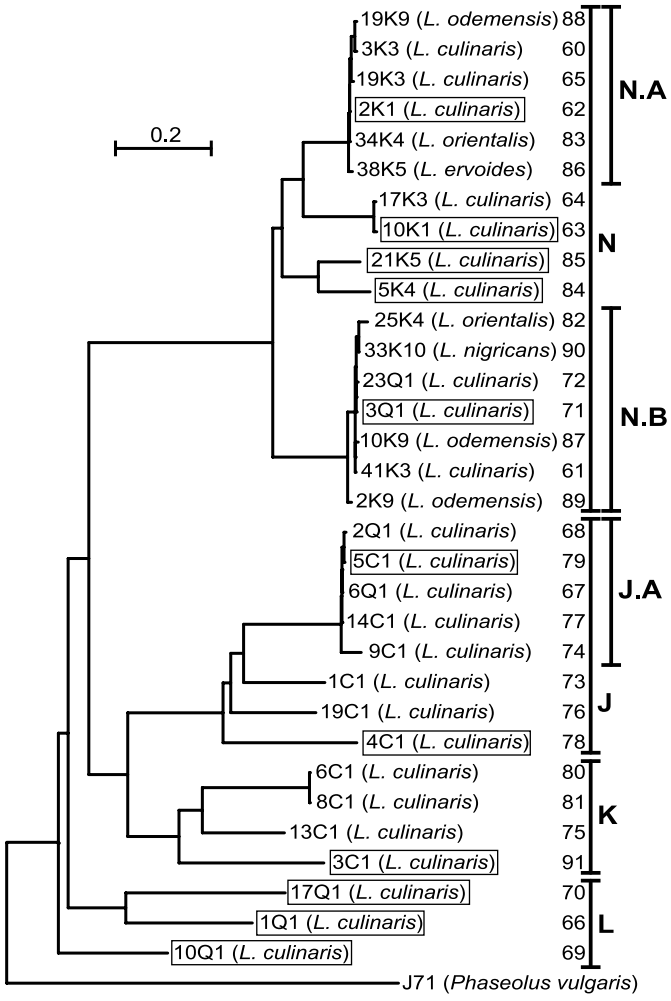
The use of degenerate oligo-primed PCR is a powerful method to amplify and clone new or uncharacterized sequences or genes that belong to a known gene family (Compton 1990). Thus, we chose this approach as a first step in the characterization of putative resistance genes in *Lens*. PCR amplification of lentil DNA using B1/Q1 and B1/L2 degenerate primer sets (Table 1) resulted in main products of approximately 670 and 820 bp, respectively. However, primer combination B1/B2 produced three bands, from which the band of approximately 520 bp was selected for further analyses, as it had the expected size of the NBS-coding region without introns (Fig. 1). Digestion of each of these products with restriction enzymes (*Rsa*I, *Msp*I, *Hae*I, and *Taq*I) resulted in several fragments whose sum was greater than the length of the original PCR product. The presence of heterogeneous PCR products suggested that, as

Fig. 3. Nucleotide polymorphism of NBS of *Lens* species. Lower polymorphism rates indicates the conserved regions of NBS.



in other plant species, a gene family encoded NBS proteins in *Lens*. Sequences obtained with B1/B2, B1/Q1, and B1/L2 primers were named K, Q, and C, respectively, (Table 1). Sequence analysis indicated that even sequences with the same restriction pattern had different nucleotide sequences. In the wild *Lens* species tested, the primer combinations produced amplification products of the same length as in the

Fig. 4. Phylogenetic tree for NBS sequences of *Lens* species. The tree was constructed using the Fich-Margoliash method on the basis of Kimura's two-parameter distances, and the nucleotide sequences of the NBS region amplified between B1 and B2 primers. Common bean RGA sequence J71 was used as the outgroup. Sequences inside rectangles were used to construct the phylogenetic tree in Fig. 5. Numbers on the right side of sequences correspond to the two last digits of the respective accession number in the EMBL Database (from AJ516060 to AJ516091).



cultivated lentil, but only some of the shortest products, a result of the B1/B2 combination, were selected for sequencing. All of the *Lens* sequences differed at nucleotide and amino acid levels; although it cannot be ruled out that some differences could be due to *Taq* polymerase errors. Degenerate primers directed to other conserved motifs (LKRCFLY, YCALFPE, YDVF, and MAASS) failed to amplify products; they also failed in combination with B1. Database searches confirmed that the sequences amplified from *Lens* were related to NBS sequences of cloned RGAs from other legume species and R genes from other plant species. For instance, the deduced amino acid sequence of the NBS domain of the 3C1 sequence shared a 62% similarity with the *Arabidopsis* RPP1 protein, whereas 4C1 showed 60% similarity with the flax L6 protein.

Sequence analysis of the NBS-clones showed that all 32 *Lens* sequences had an uninterrupted ORF. Their deduced amino acid sequences showed the internal motifs characteristic of the NBS–LRR gene class (Traut 1994). Kinase-2 (LLVLDDV), and kinase-3a (FGPGSR) motifs were present in all NBS sequences from lentil, although in some sequences with minor amino acid substitutions. The hydrophobic motif sequence (GLPLAL), conserved in this class of proteins, was present in lentil NBS domains. However, in some cases, this conserved motif underwent some amino acid substitutions, such as in clones 6Q1, 5C1, and 14C1 where the GLPLAL sequence was substituted by GRLPLAL, and in clones 1Q1 and 10Q1 where it was substituted by GHPLAL. This motif was not studied in K clones, since one of the primers encoded this sequence (Fig. 2) in K clones. Some conserved amino acids in the NBS domain sequences deduced in *Lens* species suggested that all the amplified sequences were of the TIR type, as indicated by Meyers et al. (1999) and Pan et al. (2000). The analysis of the nucleotide polymorphism and diversity of the lentil sequences showed that they were less polymorphic and more conserved at the kinase-2 and kinase-3a motifs (two functionally defined NBS motifs) than at other parts of the sequence, although other motifs in the NBS domain, such as V, also showed a low polymorphism (Fig. 3).

Phylogenetic analysis and evolution of NBS in *Lens* species

Phylogenetic analyses of lentil NBS were carried out in two steps: firstly, considering only lentil sequences and, secondly, including some sequences of other leguminous and non-leguminous species used by Cannon et al. (2002) to identify the different groups of RGA sequences as described by these authors. In the first set of analyses, nucleotide sequences and deduced amino acids sequences were used. In both cases, similar tree topologies were obtained. Analyses showed the presence of four groups, J, K, L, and N (with two subgroups, NA and NB), of RGA sequences in lentil (Fig. 4, based on DNA sequences). The nucleotide variation between closely related sequences of the same class was mainly due to point mutations. For instance, *L. culinaris* sequences 6C1 and 8C1 (Group K) differed in a single nucleotide substitution and an amino acid replacement, while the differences between *L. culinaris* 3K3 and *L. odemensis* 19K9 (group NA) were four nucleotide substitutions but a single amino acid replacement.

In the second set of analyses, when sequences from other species were included in analyses, only amino acid sequences were considered. To present a simpler tree, Fig. 5 shows the phylogenetic tree obtained using only some *Lens* sequences and some additional NBS sequences of other species. Lentil sequences were chosen to represent groups and subgroups as described in Fig. 4, and additional R and RGA sequences from legumes and other plant species were chosen as representatives of the different sequence groups described by Cannon et al. (2002). In legumes, there are no functionally characterized R genes, thus only legume RGAs could be included to determine phylogenetic relationship. Two major branches, designated TIR and non-TIR, were observed in the phylogenetic tree (Fig. 5), indicating that lentil RGAs sequenced so far belonged to the TIR type, supporting the pre-

Table 2. Synonymous substitutions per synonymous site (K_s) between sequences within each of three groups of NBS lentil sequences (above the diagonal) and nonsynonymous to synonymous ratios (K_a/K_s) (below the diagonal).

J.A		2Q1	5C1	6Q1	14C1	9C1		
	2Q1		0.011	0.011	0.011	0.107		
	5C1	0.273		0.000	0.000	0.107		
	6Q1	0.818	**n.s.**		0.000	0.107		
	14C1	0.545	**n.s.**	**n.s.**		0.107		
	9C1	0.243	0.271	0.215	0.187			
N.A		19K9	3K3	19K3	2K1	34K4	38K5	
	19K9		0.030	0.030	0.020	0.030	0.020	
	3K3	0.100		0.040	0.030	0.040	0.030	
	19K3	0.500	0.300		0.010	0.020	0.010	
	2K1	0.850	0.500	**1.200**		0.010	0.000	
	34K4	0.500	0.300	0.450	0.300		0.010	
	38K5	**1.300**	0.767	**2.000**	**n.s.**	**1.200**		
N.B		25K4	33K10	23Q1	3Q1	10K9	41K3	2K9
	25K4		0.015	0.031	0.025	0.024	0.024	0.052
	33K10	**1.467**		0.026	0.020	0.020	0.020	0.047
	23Q1	**1.226**	0.615		0.007	0.007	0.007	0.034
	3Q1	**1.120**	0.300	**2.286**		0.000	0.000	0.028
	10K9	**1.542**	0.750	**3.429**	**n.s.**		0.000	0.028
	41K3	**1.292**	0.450	**2.571**	**n.s.**	**n.s.**		0.028
	2K9	**2.173**	**1.936**	**2.971**	**3.250**	**3.25**	**3.143**	

Note: Comparisons in which nonsynonymous substitutions were higher than synonymous are indicated in bold. See Fig. 4 for groups of sequences.

vious hypothesis that all the amplified sequences from *Lens* were included in the TIR type of RGA sequences.

Although sequences of lentil NBS domains showed a relatively low similarity with NBS domains of R gene products from non-legume species such as *Arabidopsis* and flax, they shared higher similarity with RGA products of legume species such as alfalfa, chickpea, common bean, and soybean. For instance, group N sequences were similar to putative disease resistance proteins from alfalfa (up to 91% in the case of 3Q1 and 33K10 with regard to 171 amino acids of sequence AF487955 of *Medicago sativa*), groups K and L showed up to 90% similarity with a family of soybean NBS sequences, and group J of lentil NBS shared up to 88% sequence similarity with an RGA family isolated from chickpea and pea.

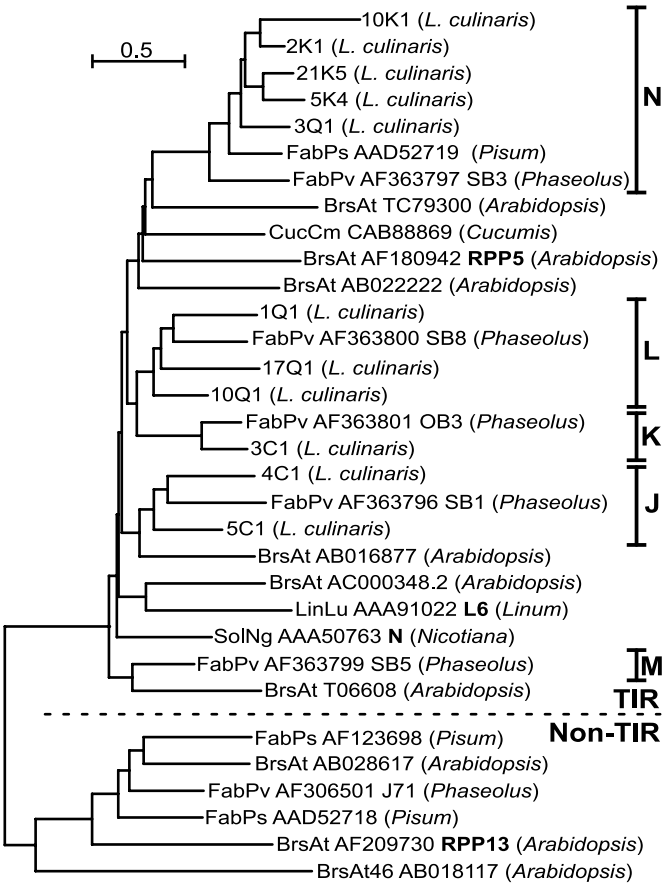
Pattern of nucleotide substitution in the NBS encoding sequences from *Lens*

Table 2 shows the frequencies of synonymous nucleotide substitutions (K_s) and the ratio K_a/K_s (K_a, non-synonymous substitutions) within the three clusters (NA, NB, and JA) with a higher number of sequences. The average values within NA, NB, and JA groups were, respectively, 0.014, 0.041, and 0.014 for K_a and 0.022, 0.021, and 0.046 for K_s, and the respective average K_a:K_s values were 0.636, 1.952, and 0.304. The K_a:K_s reveals whether sequence polymorphism supports either diversifying (K_a:K_s > 1) or purifying (K_a:K_s < 1) selection. Within groups, the compared lentil NBS domains showed several ratio values higher than 1, and in particular group NB showed an average value of 1.952, which indicated that diversifying selection was predominant between the sequences included in that cluster. Comparisons between pairs of sequences from different groups and subgroups always yielded K_a:K_s values much lower than 1.

Isolation of 3C1 flanking regions

Inverse PCR (IPCR) was used to amplify the flanking regions of the initially cloned 3C1 sequence. 3C1 was selected because it showed a relatively high identity to RGA sequences isolated from common bean and soybean. DNA restriction with *Dra*I, *Eco*RI, *Hae*I, or *Hin*dIII, self-annealing, *Xho*I linearization, and PCR amplification yield PCR products of 2000, 1400, 2700, and 1700 bp, respectively, which were further sequenced. The position of IPCR primers (INV) is shown in Fig. 6*a*. The final DNA sequence, a result of the alignment of the four sequences, was 3511 bp long including 982 bp of the 5′ untranslated region (5′ UTR), two exons coding for a sequence of 810 amino acids, and a 99-bp intron (Fig. 6*a*). The first exon coded for 584 amino acids and included the predicted TIR and NBS domains, while the second coded for 226 amino acids and included a predicted LRR region. The lentil deduced amino acid sequence showed good overall similarity to other resistance gene products. For example, it shares 53% similarity with the *N* gene product of tobacco, and 52% with RPP1 of *Arabidopsis*. But the similarity was higher for the TIR domain, reaching an amino acid sequence similarity of 70% and 68% with the TIR domain of N and L6 R proteins, respectively. The putative 5′ UTR was rich in A+T pairs (67%), and several possible TATA boxes were detected in this region using the Hamming–Clustering method for eukaryotic genes. Furthermore, the 5′ UTR included some sequences similar to the so-called G box, which is a common *cis*-element involved in responses to environmental

Fig. 5. Phylogenetic tree constructed with representative *Lens* NBS amino acid sequences and NBS sequences of RGAs from other legume species and some known R genes (in bold) from other plant species. Note that all *Lens* sequences amplified belonged to the TIR-NBS family of sequences.



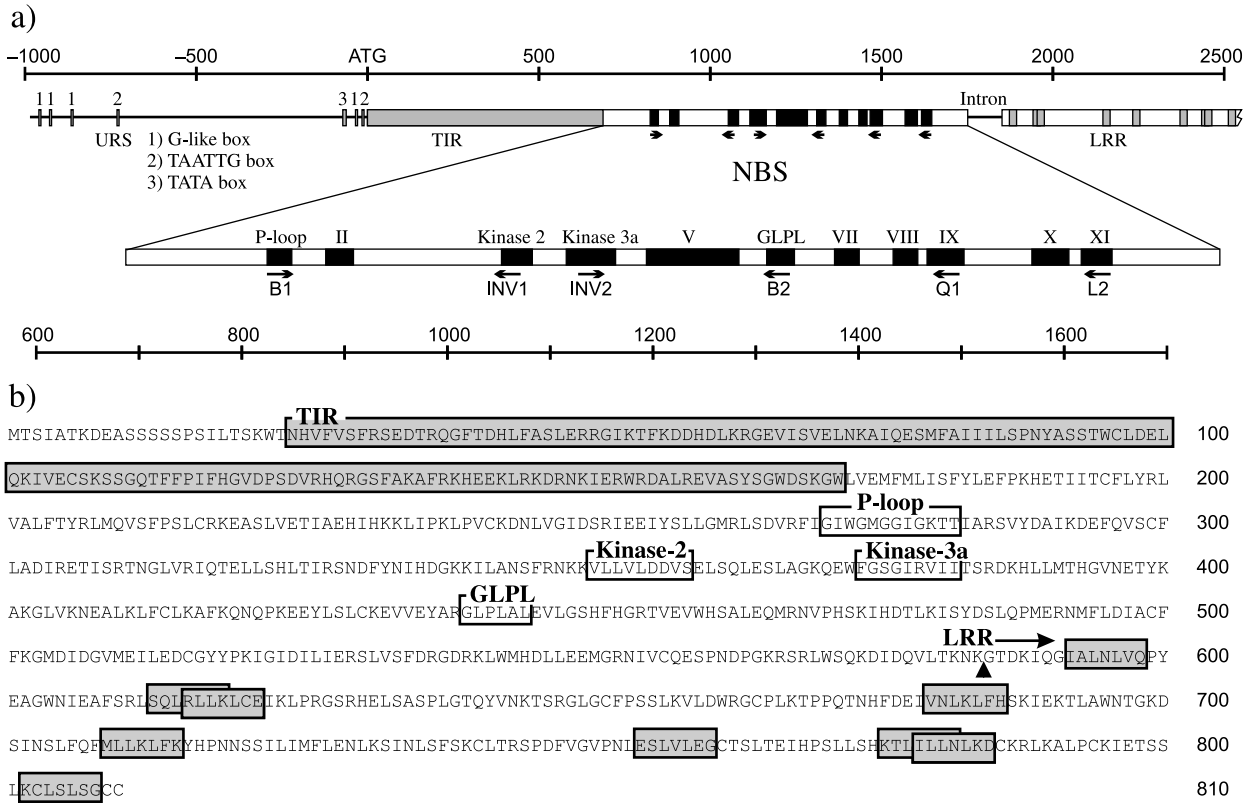
stresses (Kawaoka et al. 1994), and two TAATTG boxes (one in direct and another in inverse orientation), a *cis*-element of the pathogenesis-related protein 2 gene (*pr2*) of parsley (Korfhage et al. 1994). The reading frame included approximately 1.0 kb of sequence downstream from the NBS domain. This region was interrupted by a small (99 bp) noncoding region, which is supposed to be an intron, followed by a region encoding LRRs. The putative intron located between the NBS and LRR domain was flanked by characteristic borders (5′-GTAT...ATAG-3′) and its omission restituted the single in-frame sequence, generating a protein domain with LRR characteristics. Nine repeated motifs of the β-strand–β-turn structural motif, XXLXLXX, where X stands for any amino acid (Kobe and Deisenhofer 1994), were located in the 226 amino acids of the LRR region. Unfortunately, the gene segment sequenced did not include the 3′ end.

Discussion

The two most critical factors in degenerate oligo-primed PCR are the design of the primers and the PCR conditions. The primers should be designed to an amino acid region with minimal degeneracy in codon usage (Compton 1990) and PCR conditions should be optimized to give a balance between efficiency and specificity. Two types of sequence were not expected to be recovered in our study: (*i*) NBS sequences that did not contain the conserved motifs used for priming or that contained motifs different enough to prevent the annealing of primers directed to them; (*ii*) genes that contain large introns in this region. Although the presence of introns has been reported in some NBS–LRR sequences (Meyers et al. 2003), lentil NBS sequences with an intron of a few hundred base pairs could have been amplified under the PCR conditions used. Inspection of available sequences from *A. thaliana* and other species (both, mono and dicotyledonous) indicates that all of the NBS–LRR-type genes contain the motifs primed; therefore, sequence changes in these motifs (LKRCFLY, YCALFPE, YDVF, and MAASS) were the most probable cause of the amplification failure in lentil. Furthermore, variants in the motif GLPLAL are the most probable reason why we were unable to amplify some NBS sequence subfamilies by PCR using the B1/B2 primer set, since the degenerate primer B2 was deduced from GLPLAL. The consensus XGXGKTT of the P-loop motif is present in almost all NBS–LRR-type plant resistance genes and is also conserved in the human *APAF-1* sequence. In fact, the B1 primer directed to this motif always generated amplification products with other primers. The most common P-loop motif of lentil sequences isolated in this study was GGMGKTT, although they can be spurious owing to the primer effect. This hypothesis is supported by the fact that the sequence obtained in 3C1 after inverse PCR was GGIGKTT; and it was GGLGKTT in four common bean genes (Ferrier-Cana et al. 2003). The kinase-2 motif (LLVLDDVX) in the non-TIR-type sequences has a characteristic tryptophan residue (LLVLDDVW) that was absent in the TIR sequences (Meyers et al. 1999). This fact could be used to predict with 95% accuracy whether an NBS region would fall in the TIR-NBS or into the non-TIR-NBS group of RGA. Instead of the tryptophan residue (W), the kinase-2 motifs found in lentil NBS sequences had aspartic acid (D), asparagine (N), serine (S), or threonine (T) residues at this last position. On the other hand, four bean sequences having a LLVLDDVW kinase-2 consensus sequence (Ferrier-Cana et al. 2003) fall into the non-TIR group.

It can be seen that there was an association between the primer combination and the type of sequence amplified. All the sequences amplified using the B1/B2 combination were clustered in group N, while those amplified with B1/L2 were in groups J and K, and those with B1/Q were distributed between NB, JA, and L (Fig. 4). This result may be an indication of specific differences between conserved motifs depending upon the sequence group. If this is so, slight changes in primers could be used to amplify specific sets of RGA sequences in lentil and other species. Conserved motifs in the NBS domain seem to play an important role in the function of resistance genes, since several loss-of-function alleles in R genes are due to point mutations within conserved blocks inside the NBS domain (Pan et al. 2000). For example, a spontaneous flax mutant of *P2* susceptible to rust was caused by a G→E amino acid substitution in the GLPLAL motif (Dodds et al. 2001). Together with the amino acid motifs previously mentioned, many other small similar stretches of amino acids were found in lentil se-

Fig. 6. (*a*) Schematic representation of the PCR and IPCR strategy used to isolate NBS sequences from lentil, and the flanking regions of 3C1 sequence. The different domains, motifs and the putative 5′UTR transcription signals are indicated. Arrows indicate the position and orientation of primers. Scale indicates length in base pairs. (*b*) Amino acid sequence of the 3C1. The predicted TIR region is inside the longest shaded box, NBS conserved motifs are inside rectangles, and residues corresponding to the β-strand/β-turn structural motifs (XXLXLXX) of the LRR domain are inside the short shaded rectangles. The arrowhead marks the position of the intron.



quences (Fig. 2). Small pockets of similarity in amino acid sequences have also been described between R gene products (i.e., *Arabidopsis* RPS2 and RPM1, tobacco N, and flax L6) from different species (Grant et al. 1995; Staskawicz et al. 1995). These poorly understood regions of conservation are likely to represent functionally relevant sites and useful landmarks in the isolation of *R* gene homologues (Bent 1996).

Phylogenetic analyses of lentil NBS sequences showed that some sequences from different species are more similar among each other than some sequences from the same species; this is true when taking only *Lens* into account (Fig. 4) or when also considering species from other plant genera (Fig. 5). This was an expected result according to the divergent gene duplication and further gene diversification hypothesis on the evolution of NBS–LRR-type disease resistance genes (Pan et al. 2000), which states that many of the gene duplications predated the divergence of main higher plant taxa. Similar trees have been obtained with legume and non-legume species (Cannon et al. 2002; Cordero and Skinner 2002; Di Gaspero and Cipriani 2002).

The values of the synonymous and non-synonymous substitutions in the NBS domain were calculated within groups of sequences separated by relatively short genetic distances to avoid saturation at synonymous sites. The average K_a:K_s in groups JA and NA are compatible with purifying selection, since the average ratios were lower than 1; however, within group NB, diversifying selection seemed to be predominant, since the average K_a:K_s was greater than 1. Diversifying selection could be the result of a differential selection pressure after a speciation process and (or), within a species, the result of the specialization of different paralogous sequences in the identification of or the response to different pathogens or pathovars. Although in some instances higher numbers of non-synonymous substitutions could be due to inter-specific comparisons (i.e., sequence 2K9 from *L. odemensis* in relation to all other sequences of group NB), in other cases there were more non-synonymous than synonymous substitutions within species (i.e., 23Q1 and 3Q1 within NB, or 5C1, 6Q1 and 14C1 within JA; all these sequences from *L. culinaris*). Thus, it seems possible that the specialization of different sequences in response to different pathogens could be one of the causes of diversifying selection in lentil NBS sequences. Ferrier-Cana et al. (2003) did not find any evidence of diversifying selection in the NBS region of seven common bean RGA sequences, but they did find diversifying selection in the LRR and spacer region of these sequences. It has been pointed out that the LRR domain is usually under diversifying selection (Michelmore and Meyers 1998). Many RGA sequences are grouped in clusters in plant genomes (Ferrier-Cana et al. 2003); however, concerted evolution does not seem to play an evolutionary role in these genes, and the results in *Lens* agree with this hypothesis, since sequences from different

species are more similar than some sequences from the same species.

Inverse PCR is a convenient and versatile method of cloning unknown sequences upstream or downstream of known sequences (Triglia et al. 1988). Inverse PCR circumvents the laborious procedures of producing and screening genomic libraries (Forster et al. 1994). Nucleotide sequences flanking the NBS sequence have been previously isolated using IPCR in lettuce (Shen et al. 1998). This procedure allowed us to expand the 3C1 sequence in both directions. The 3C1 predicted product had the TIR, NBS, LRR conserved domains (Fig. 6*a*) characteristics of TIR-NBS-LRR products of RGA and R genes (e.g., *L6, N, M, RPP5*, and *RPP1*). In particular, the highest amino acid similarity among 3C1 and the proteins encoded by R genes *N* and *L6* was observed in the TIR domain. Furthermore, the 5′ UTR of 3C1 showed some motifs similar to the G-box or to the TAATTG box, which have been found in genes involved in response to stresses, including pathogens (Kawaoka et al. 1994; Korfhage et al. 1994). Structural and sequence similarities with functionally characterized resistance proteins provide strong evidence that we have identified at least some lentil resistance gene candidates, and not only sequences with similarity to NBS domains.

In summation, the use of degenerate primers directed to conserved sequences has demonstrated to be a useful technique to isolate resistance gene candidates in lentil, as proven in other plant species (Cordero and Skinner 2002; Di Gaspero and Cipriani 2002; Ferrier-Cana et al. 2003; Kanazin et al. 1996; Noir et al. 2001; Rivkin et al. 1999; Shen et al. 1998; Yu et al. 1996; Zhu et al. 2002). But, according to the information from other plant species (Meyers et al. 1999), many other NBS sequences must be present in the lentil genome. In fact, no sequence of the non-TIR groups of NBS–LRR sequences has been amplified in lentil with the primers tested in this study, while many non-TIR sequences isolated from dicotyledonous species, including legumes (Ferrier-Cana et al. 2003), are available in data bases. New primers to be obtained with the help of the information deduced from NBS sequences of *Lens culinaris*, and the sequencing of new clones will allow for the isolation of most, if not all, of these NBS–LRR sequences in this genus, allowing for further studies such as the mapping of these sequences and further evolutionary analyses. Breeding lentil programs include, among other alternative techniques, wide crosses with wild *Lens* species. The use of wild species of the genus will serve to explore variants of R genes absent in the cultivated gene pool that could be useful for the introduction of new resistances by traditional or molecular breeding techniques.

Acknowledgements

This work was supported by grants AGF99-0192 and AGL2002-0023 from the Spanish Dirección General de Investigación, grant No. LE 21/01 from the Consejería de Educación y Cultura de la Junta de Castilla y León, and a personal Ph.D. grant to M.W.F.Y. from the Spanish Agency for International Cooperation.

References

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S.P. 1997. Signaling in plant-microbe interactions. Science (Washington, D.C.), **276**: 726–733.

Bent, A.F. 1996. Plant disease resistance genes, function meets structure. Plant Cell, **8**: 1757–1771.

Beroud, C., Antignac, C., Jeanpierre, C., and Junien, C. 1990. Un programme informatique pour la recherche d'amorces pour l'amplification par PCR. Med. Sci. **6**: 901–903.

Boyle, J.S., and Lew, A.M. 1995. An inexpensive alternative to glassmilk for DNA purification. Trends Genet. **11**: 8.

Cannon, S.B., Zhu, H., Baumgarten, A.M., Spangler, R., May, G., Cook, D.R., and Young, N.D. 2002. Diversity, distribution, and ancient taxonomic relationships within the TIR and Non-TIR NBS-LRR resistance gene subfamilies. J. Mol. Evol. **54**: 548–562.

Chida, Y., Okazaki, K., Karasawa, A., Akashi, K., Nakazawa, S.H., Takahashi, H., and Ehara, Y. 2000. Isolation of molecular markers linked to the *Cry* locus conferring resistance to cucumber mosaic cucumovirus infection in cowpea. J. Gen. Plant Pathol. **66**: 242–250.

Compton, T. 1990. Degenerate primers for DNA amplification. *In* PCR protocols, a guide to methods and applications. *Edited by* M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White. Academic Press, London, UK. pp. 39–45.

Cordero, J.C., and Skinner, D.Z. 2002. Isolation from alfalfa of resistance gene analogues containing nucleotide binding sites. Theor. Appl. Genet. **104**: 1283–1289.

Creusot, F., Macadré, C., Ferrier Cana, E., Riou, C., Geffroy, V., Sévignac, M., Dron, M., and Langin, T. 1999. Cloning and molecular characterization of three members of NBS-LRR subfamily located in the vicinity of the Co-2 locus for anthracnose resistance in *Phaseolus vulgaris*. Genome, **42**: 254–264.

Dayhoff, M.O. 1979. Atlas of protein sequences and structure. Vol. 5. Suppl. 3. National Biomedical Research Foundation, Washington, D.C.

Dellaporta, S.L., Wood, J., and Hicks, J.B. 1983. A plant DNA minipreperation, version II. Plant Mol. Biol. **1**: 19–21.

Di Gaspero, G., and Cipriani, G. 2002. Resistance gene analogs are candidate markers for disease-resistance genes in grape (*Vitis* spp.). Theor. Appl. Genet. **106**: 163–172.

Dodds, P.N., Lawrence, G.J., and Ellis, J.G. 2001. Six amino acid changes confined to the leucine-rich repeat b-strand/b-turn motif determine the difference between the P and P2 rust resistance specificities in flax. Plant Cell, **13**: 163–178.

Ellis, J., Dodds, P., and Pryor, T. 2000. Structure, function and evolution of plant disease resistance genes. Curr. Opin. Plant Biol. **3**: 278–284.

Erskine, W., Bayaa, B., and Saxena, M.C. 1996. Registration of ILL5588 lentil germplasm resistant to vascular wilt and ascochyta blight. Crop Sci. **36**: 1080.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle, Wash.

Ferrier-Cana, E., Geoffroy, V., Macadré, C., Creusot, F., Imbert-Belloré, P., Sívingac, M., and Langing, T. 2003. Characterization of expressed NBS-LRR resistance gene candidates from common bean. Theor. Appl. Genet. **106**: 251–261.

Fitch, W.M., and Margoliash, E. 1967. Construction of phylogenetic trees. Science (Washington, D.C.), **155**: 279–284.

Forster, C., Arthur, E., Crespi, S., Hobbs, S.L.A., Mullineaux, P., and Casey, R. 1994. Isolation of a pea (*Pisum sativum*) seed

lipoxygenase promoter by inverse polymerase chain reaction and characterization of its expression in transgenic tobacco. Plant Mol. Biol. **26**: 235–248.

Geffroy, V., Creusot, F., Falquet, J., and Sévignac, M. 1998. A family of LRR sequences in the vicinity of the Co-2 locus for anthracnose resistance in *Phaseolus vulgaris* and its potential use in marker-assisted selection. Theor. Appl. Genet. **96**: 494–502.

Graham, M.A., Marek, L.F., Lohnes, D., Cregan, P., and Shoemaker, C. 2000. Expression and genome organization of resistance gene analogs in soybean. Genome, **43**: 96–93.

Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. 1995. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. Science (Washington, D.C.), **269**: 843–846.

Hammond-Kosack, K.E., and Jones, J.D.G. 1997. Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48**: 575–607.

Henikoff, S., and Henikoff, J. 1994. Protein family classification based on searching a database of blocks. Genomics, **19**: 97–107.

Holmes, D.S., and Quigley, M. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**: 193–197.

Kanazin, V., Fredrick Marek, L., and Shoemaker, R.C. 1996. Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. U.S.A. **93**: 11 746 – 11 750.

Kawaoka, A., Kawamoto, T., Sekine, M., Yoshida, K., Takano, M., and Shinmyo, A. 1994. A *cis*-acting element and trans-acting factor involved in two wound-induced expression of horseradish peroxidase gene. Plant J. **6**: 87–97.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**: 111–120.

Kobe, B., and Deisenhofer, J. 1994. The leucine-rich repeat, a versatile binding motif. Trends Biochem. Sci. **19**: 415–421.

Korfhage, U., Trezzini, G.F., Meier, I., Hahlbrock, K., and Somssich, E. 1994. Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene. Plant Cell, **6**: 695–708.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. 2001. MEGA2: molecular evolutionary genetics analysis software. Arizona State University, Tempe, Ariz.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z., and Michelmore, R.W. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. Plant Cell, **10**: 1817–1832.

Meyers, B.C., Dickerman, A.W., Michelmore, R.W., Sivaramakrishnan, S., Sobral, B.W., and Young, N.D. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J. **20**: 317–332.

Meyers, B.C., Kozik, A., Griego, A., Kuang, H.H., and Michelmore, R.W. 2003. Genome-wide analysis of NBS–LRR-encoding genes in *Arabidopsis*. Plant Cell, **15**: 809–834.

Michelmore, R.W. 2000. Genomic approaches to plant disease resistance. Curr. Opin. Plant Biol. **3**: 125–131.

Michelmore, R.W., and Meyers, B.C. 1998. Clusters of resistance genes in plants evolve by divergent selection and birth-and-death process. Genome Res. **8**: 1113–1130.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, N.Y.

Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**: 418–426.

Noël, L., Moorers, T.L., Van der Biezen, E.A., Parniske, M., Daniels, M.J., Parker, J.E., and Jones, J.D.G. 1999. Pronounced intraspecific haplotype divergence in the *RPP5* complex disease resistance locus of *Arabidopsis*. Plant Cell, **11**: 2099–2111.

Noir, S., Combes, M.-C., Anthony, F., and Lashermes, P. 2001. Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee tree (*Coffea* L.). Mol. Genet. Genomics, **265**: 654–662.

Pan, Q., Jonathan, W., and Fluhr, R. 2000. Divergent evolution of plant NBS–LRR resistance gene homologues in dicot and cereal genomes. J. Mol. Evol. **50**: 203–213.

Rivkin, M.I., Vallejos, C.E., and McClean, P.E. 1999. Disease-resistance related sequences in common bean. Genome, **42**: 41–47.

Rozas, J., and Rozas, R. 1999. DnaSP version 3, an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics, **15**: 174–175.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **179**: 1144–1147.

Shen, K.A., Meyers, B.C., Isalm-Faridi, M.N., Chin, D.B., Stelly, D.M., and Michelmore, R.W. 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.

Silver, J., and Keerikatte, V. 1989. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63**: 1924–1928.

Staskawicz, B.J., Ausubel, F.M., Baker, B.J, Ellis, J.G., and Jones, J.D.J. 1995. Molecular genetics of plant disease resistance. Science (Washington, D.C.), **268**: 661–667.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4782.

Traut, T.W. 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur. J. Biochem. **222**: 9–19.

Triglia, T., Peterson, M.G., and Kemp, D.J. 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16**: 8186.

Young, N.D. 2000. The genetic architecture of resistance. Curr. Opin. Plant Biol. **3**: 285–290.

Yu, Y.G., Buss, G.R., and Saghai Maroof, M.A. 1996. Isolation of a superfamily of candidate disease resistance genes in soybean based on a conserved nucleotide-binding site. Proc. Natl. Acad. Sci. U.S.A. **93**: 11 751 – 11 756.

Zhu, H., Cannon, S.B., Young, N.D., and Cook, D.R. 2002. Phylogeny and genomic organization of the TIR and Non-TIR NBS–LRR resistance gene family in *Medicago truncatula*. Mol. Plant–Microbe Interact. **15**: 529–539.