Theor Appl Genet (2003) 106:251–261
DOI 10.1007/s00122-002-1032-z

E. Ferrier-Cana · V. Geffroy · C. Macadré
F. Creusot · P. Imbert-Bolloré · M. Sévignac
T. Langin

Characterization of expressed NBS-LRR resistance gene candidates from common bean

Received: 6 December 2001 / Accepted: 18 April 2002 / Published online: 22 August 2002

Abstract A complex ancestral resistance (R) gene cluster, localized at the end of linkage group B4, and referred to as the *B4* R gene cluster, has been previously genetically characterized. The *B4* R gene cluster existed prior to the separation of the two major gene pools of cultivated common bean and contains several resistance specificities effective against the fungus *Colletotrichum lindemuthianum*. In this paper we report the molecular analysis of four expressed resistance gene candidates (RGCs) that map at the *B4* R-cluster and co-localize with R-specificities or R-QTLs effective against *C. lindemuthianum*. These RGCs have been isolated from two genotypes that are representative of the two major gene pools of common bean: the *BA8* and *BA11* RGCs originating from the Mesoamerican BAT93 genotype, and the *JA71* and *JA78* RGCs originating from the Andean JaloEEP558 genotype. These RGCs encode NBS-LRR resistance-like proteins that are closely similar to the tomato I2 R-protein. Based upon sequence comparisons and genetic localization, we established that these four bean RGCs belong to two different subfamilies of R-sequences independently of their gene pool of origin. No feature discriminating the four RGCs according to their gene pool of origin has been observed yet. Comparative sequence analyses of the full-length RGCs and their flanking genomic sequences confirmed the ancestral origin of the *B4* R-cluster.

Keywords *Phaseolus vulgaris* · *Colletotrichum lindemuthianum* · Molecular evolution · R-cluster

Communicated by C. Möllers

E. Ferrier-Cana · V. Geffroy (✉) · C. Macadré · F. Creusot
P. Imbert-Bolloré · M. Sévignac · T. Langin
Institut de Biotechnologie des Plantes, UMR-CNRS 8618,
LPPM, bât. 630, Université Paris-Sud, 91405 Orsay, France
e-mail: geffroy@ibp.u-psud.fr
Tel.: +33-1-6915-3370, Fax: +33-1-6915-3424

Introduction

The interaction between a plant and a pathogen often complies with the gene-for-gene model involving plant resistance genes (R-genes) and corresponding pathogen avirulence genes (Avr-genes) (Flor 1971). In this gene-for-gene system, the direct or indirect interaction of the products of these two genes is believed to lead to the establishment of a hypersensitive response (Innes 1998). About 30 R-genes have been isolated so far from diverse plant species and are effective against various pathogens including bacteria, nematodes, viruses and fungi (reviewed in Ellis and Jones 1998). Most of these genes are members of multigene families and are organized in complex clusters (Ellis et al. 1995; Parniske et al. 1997; Song et al. 1997; Meyers et al. 1998a; Noël et al. 1999).

To-date, most characterized R-genes encode proteins containing a leucine-rich repeat (LRR) region (Jones and Jones 1997; Ellis and Jones 1998). In mammalian systems, this region has been demonstrated to be involved in protein-protein interactions and signal transduction processes (Kobe and Deisenhofer 1995). Similarly, in plant systems, sequence comparisons of functional R-genes (Thomas et al. 1997; Dodds et al. 2001), domain swapping experiments (Ellis et al. 1999; Van der Hoorn et al. 2001; Wulff et al. 2001) and mutant analyses of R-genes (Anderson et al. 1997) revealed that the LRR region of plant R-genes is involved in R-specificity. Moreover, two cases of physical interaction between a plant nucleotide-binding site (NBS)-LRR R-protein and its cognate AVR-protein have been reported (Jia et al. 2000; Leister and Katagiri 2000).

Based on cloned R-gene sequences, R-gene analogs (RGAs) encoding NBS and LRR motifs have been identified in several species, and many of these RGAs are grouped in clusters and co-localized with known R-specificities (Kanazin et al. 1996; Leister et al. 1996, 1998; Yu et al. 1996; Bottela et al. 1997; Aarts et al. 1998; Shen et al. 1998; Speulman et al. 1998; Geffroy et al. 1999). The study of R-gene sequences and RGA clusters revealed the probable importance of duplication events

in generating new R-specificities. Indeed, classical genetic analyses and sequencing of R-gene clusters in maize (Collins et al. 1999), lettuce (Meyers et al. 1998a, b), tomato (Parniske et al. 1997), rice (Song et al. 1997) and flax (Ellis et al. 1999) revealed that duplicated RGA sequences are subject to unequal crossing-overs and gene conversions, in combination with other mechanisms such as transposition and point mutation. These mechanisms may explain the high diversity observed at R-gene clusters. However, the effect of natural polymorphism on R-gene functionality has only been studied in a few cases (Caicedo et al. 1999; Stahl et al. 1999) and no molecular analyses have yet been done on plants for which the evolutionary history and the pathogen selection pressures are known. The interaction between the common bean, *Phaseolus vulgaris*, and the fungus *Colletotrichum lindemuthianum* offers this opportunity.

The interaction between *P. vulgaris* and *C. lindemuthianum* complies with the gene-for-gene model. *C. lindemuthianum* is the causal agent of anthracnose, one of the most serious diseases of common bean in tropical areas (Pastor-Corrales and Tu 1989). The leguminous plant species *P. vulgaris* originates from Latin America and natural populations are found from Mexico to northwestern Argentina. Based on morphological traits (Delgado-Salinas et al. 1988), biochemical and molecular markers (Gepts 1988; Koenig and Gepts 1989; Singh et al. 1991; Debouck et al. 1993; Becerra-Velasquez and Gepts 1994; Sonnante et al. 1994; Kami et al. 1995) three centers of diversity of wild common bean have been described: Mesoamerican, South-Andean and North-Andean (Gepts 1998). Comparative studies of wild and cultivated beans have suggested that independent domestication events have occurred in the Mesoamerican and the Andean centers of diversity, generating two major gene pools of cultivated beans: the Andean and the Mesoamerican gene pools (Gepts 1998). Cross-inoculations between *P. vulgaris* and *C. lindemuthianum* have further revealed that coevolution processes between the two partners led to a differentiation for host resistance specificities and pathogen virulence between the three centers of diversity (Geffroy et al. 1999). These results revealed that Andean and Mesoamerican plant populations harbor distinct phenotypes for resistance and are subject to different pathogen selection pressures. Similar results have been obtained with two other fungal pathogens of common bean, *Phaeoisariopsis griseola* (Guzman et al. 1995) and *Uromyces appendiculatus* (Steadman et al. 1995).

At the molecular level, a complex R-cluster, referred to as the *B4* R-cluster, has been characterized, using a population of recombinant inbred lines (RILs) constructed from a cross between two genotypes that are representative of the Mesoamerican and the Andean gene pools of cultivated common bean. This complex R-cluster has been mapped on linkage group B4 (Geffroy et al. 1999) of the integrated linkage map of common bean (Freyre et al. 1998). It contains three R-specificities against *C. lindemuthianum* (*Co-y*, *Co-z* and *Co-9*), a NBS RGA-family, referred to as the *PRLJ1* family (Geffroy et al. 1999), and two resistance quantitative trait loci (R-QTLs) of Andean and Mesoamerican origins (Geffroy et al. 2000). The co-localization of Andean and Mesoamerican R-specificities suggests that the complex *B4* R-cluster existed prior to the geographic separation of the two major gene pools of cultivated common bean (Geffroy et al. 1999). Therefore, the model between *P. vulgaris* and *C. lindemuthianum* offers the opportunity to study at the molecular level, the parallel evolution of an ancestral R-cluster, following the separation of two gene pools of the same plant species.

In this study, a preliminary analysis of the polymorphism occurring between expressed resistance-gene candidate (RGC) sequences mapped at the *B4* R-cluster was conducted. Four expressed RGCs of either Andean or Mesoamerican origin have been isolated that co-localize with R-specificities or R-QTLs effective against *C. lindemuthianum* at the *B4* R-cluster. These RGCs encode NBS-LRR R-like proteins close to the tomato I2 R-protein (Simons et al. 1998). These four RGCs belong to the same family of R-like gene sequences. However, they could be separated into two subfamilies based both on their sequence polymorphism and their genetic localization, independently of their gene pool of origin. No feature specific for each gene pool could be identified from the four sequences that were analyzed so far.

Materials and methods

Development of cDNA and genomic libraries

The bean genotypes BAT93 and JaloEEP558 were used to construct cDNA and genomic libraries. For the cDNA libraries, 1-week-old plants of each genotype were infected with the *C. lindemuthianum* M38 strain. Cotyledonary leaves were harvested before inoculation and 24, 48 and 72 h after inoculation. Total RNA corresponding to each sampling time was isolated via the procedure described in Vallelian-Bindschedler et al. (1998), except for the use of 2 ml of phenol instead of 1 ml. Poly(A)+ RNA was purified using an oligo (dt) magnetic bead system (Dynal, Oslo, Norway) and the resulting DNA was pooled in equal quantities. The ZAP express cDNA synthesis kit and the ZAP express cDNA gigapack III Gold packaging kit (Stratagene, La Jolla, Calif.) were used according to the manufacturer's recommendations to construct the cDNA libraries. Genomic libraries were constructed from 1-week-old non-infected cotyledonary leaves of BAT93 and JaloEEP558 genotypes. DNA was extracted as described in Geffroy et al. (1998), partially digested with *Sau*3A and ligated in the lambda FIX II/*Xho*I partial fill-in vector kit (Stratagene, La Jolla, Calif.) according to the manufacturer's recommendations.

Library screening

Twelve copies of the BAT93 genome and 12 copies of the JaloEEP558 genome, represented each by 5×10^5 recombinant plaques, 2×10^6 recombinant plaques of the BAT93 cDNA library and 4×10^6 recombinant plaques of the JaloEEP558 cDNA library, were screened with the *PRLJ1* probe (Geffroy et al. 1999) under the following conditions: over-night hybridization at 65 °C in 5 × SSC (3 M NaCl and 0.3 M sodium citrate), 5% SDS and 5 × Denhardt's (0.1% Ficoll, 0.1% PVP, 0.1% BSA) followed by three washing steps of 15 min in 0.5 × SSC and 0.5% SDS at 65 °C. A primary screen of positive cDNA clones was done by sizing and sequencing of both cDNA ends. Long clones (more than 3 kb)

with ends displaying homology with reported R-genes were further sequenced and led to the characterization of the *cBA8*, *cBA11*, *cJA71* and *cJA78* cDNAs. Positive genomic clones isolated from the BAT93 and the JaloEEP558 genomic libraries respectively, using *PRLJ1* as a probe, were secondly screened using specific primers for each cDNA in order to select genomic clones corresponding to each cDNA. Specific primers were designed in the 3′-UTR for each JaloEEP558 gene: jalo2 (5′-GATACGATGAGTCACAAG-3′) and jalo10 (5′-GTGTTGCAAAGGTTATCTT-3′) for the *JA71* RGC, primers jalo22 (5′-GGTTCAGAAATACACTTTCC-3′) and jalo23 (5′-TGATCATAAGATGAGTCAGAA-3′) for the *JA78* RGC. For the 3′-UTRs of *BA8* and *BA11* RGCs, it was not possible to define specific primers. Consequently, specific primers obat30SV (5′-ATCGCATACTGATGGCATCTT-3′) and obat32 (5′-GTGCCAGACAAATTCAAACATAC-3′) for the *BA8* RGC, obat34 (5′-GTATTTTTAGTTGCCCTCAG-3′) and obat33 (5′-AAATAGGTGCACGCCTAA-3′) for the *BA11* RGC, were designed in the variable LRR region.

5′-RACE experiments

5′-RACE analyses (Frohman et al. 1988) were performed to identify the 5′-UTR sequences of the *BA8*, *BA11*, *JA71* and *JA78* transcripts. The Life Technologies 5′-RACE kit (version 2, GIBCO BRL, Cergy Pontoise, France) was used as specified by the manufacturer. First-strand cDNA synthesis used 0.5 µg of the composite poly(A)+ RNA pools with the Race3bis nested primer (5′-GGTGAGTTCATAGTCTATTTCACCC-3′) which hybridized with the four cDNAs. After adapter ligation, amplifications were performed with a second nested primer Race1Spe (5′-GGACTAGTCCGCTTTAGTTCAGCATCATCAGCGAG-3′) and the AAP primer provided by the manufacturer, using the Expand high fidelity PCR system (Boehringer Mannheim, Meylan, France). Each amplification reaction yielded amplification products of approximately 200 bp which were cloned into BlueScript II KS(+), (Stratagene, La Jolla, Calif.) for sequence analyses.

Sequencing and analyses of cDNA and genomic clones

The sequencing of cDNA and genomic clones was done using an automated 373A DNA sequencer and the Big Dye dideoxy chain Terminator cycle sequencing Kit (Perkin Elmer, Applied Biosystems, Roissy, France). Sequence analyses and multiple alignments were performed with the MacMolly program (version 3.5.1, Soft gene GmbH) and the Gap, Olddistances and Overlap programs of the Genetics Computer Group (GCG) package (Madison, Wis., USA). Sequence similarities were established using the Blast Algorithm (Altschul et al. 1997). Nucleotide substitution ratios were calculated with the Diverge program of the GCG software package that uses the method published by Li et al. (1985), modified by Li (1993) and Pamilo and Bianchi (1993). Exon-intron boundaries were determined by comparison between the cDNA and genomic sequences. GenBank accession numbers of the four cDNAs and their corresponding genomic sequences are AF306499 to AF306506.

Mapping of cDNA and genomic clones

A PCR approach was used to map the four cDNAs and their corresponding genomic clones on the common bean integrated linkage map (Freye et al. 1998). The specific oligonucleotide primers designed for each cDNA (jalo2/jalo10 for *JA71*, jalo22/jalo23 for *JA78*; obat30SV/obat32 for *BA8* and obat33/obat34 for *BA11*) were used to amplify genomic DNA of the 77 BAT93x JaloEEP558 F9 recombinant inbred lines (RILs) obtained from the University of California, Davis, USA. PCR reactions were done in a final volume of 25 µl containing 50-ng of template DNA, 1 × PCR reaction buffer, 3 pmol of each primer, 50 µM of each dNTP and 0.5 units of Red *Taq* Goldstar polymerase (Eurogentec, Seraing, Belgium). Amplifications were performed in a GeneAmp PCR system 9600 (Perking Elmer, Norwalk, USA). The MAPMAKER software version 3.0 was used to map the cDNA and genomic clones (Lander et al. 1987). Linkage groups were established with a LOD threshold of 3.0 and a recombination fraction of 0.3. Marker order was estimated with a LOD threshold of 2.0 based on multipoint "Compare", "Order" and "Ripple" analyses. Genetic distances between markers were estimated with the Kosambi mapping function (Kosambi 1944).

Results

Molecular characterization of four expressed bean R-gene candidates

Molecular analyses of RGAs in several species have allowed the identification of either expressed R-genes or pseudogenes (Thomas et al. 1997; Meyers et al. 1998a; Michelmore and Meyers 1998). In order to select only expressed R-genes, we have chosen to screen cDNA libraries constructed from the representative Mesoamerican BAT93 and the Andean JaloEEP558 bean genotypes. The screening of the two cDNA libraries with the NBS *PRLJ1* probe (Geffroy et al. 1999) led to the isolation of four full-length cDNAs with high similarities to R-genes encoding NBS-LRR R-proteins (Jones and Jones 1997). Two cDNAs, *cBA8* and *cBA11*, were isolated from the BAT93 genotype (*cBA* stands for cDNA isolated from the BAT93 genotype) and two other cDNAs, *cJA71* and *cJA78*, were isolated from the JaloEEP558 genotype (*cJA* stands for cDNA isolated from the JaloEEP558 genotype) (Fig. 1A). The sequence of the NBS *PRLJ1* probe shows an average nucleotide identity of 89% with the four cDNAs, but none of them corresponds to the *PRLJ1* RGA sequence.

The nucleotide environment of the candidate translation start of each ORF was studied in order to test whether the nucleotides in the vicinity of each predicted start codon follow the general consensus observed for the majority of functional dicot plant genes. The predicted start codon of the four RGCs is preceded by an A and a C in the –3 and –2 positions and is followed by a G and a C in the +4 and +5 positions, respectively. Consequently, each of the four genes contains a start-codon surrounding region that fits well with those observed for functional dicot genes (Joshi et al. 1997).

Complementary 5′-RACE analyses were performed to characterize the 5′ untranslated regions (5′-UTRs) of the four bean cDNAs and to locate more precisely the translation start codon position. Twenty two distinct and 18 distinct 5′-RACE products have been isolated from the BAT93 and JaloEEP558 genotypes, respectively. Nucleotide identities among the 22 BAT93 5′-RACE products and among the 18 5′-RACE JaloEEP558 products range from 74 to 99.5% and from 67.9 to 99.3%, respectively. Moreover, between 67.9 and 100% of nucleotide identity has also been observed upon comparison of the BAT93 and the JaloEEP558 5′-RACE products with each other. In their overlapping sequences (approximately 150 bp), the 5′-RACE products and the 5′-end of the four cDNAs

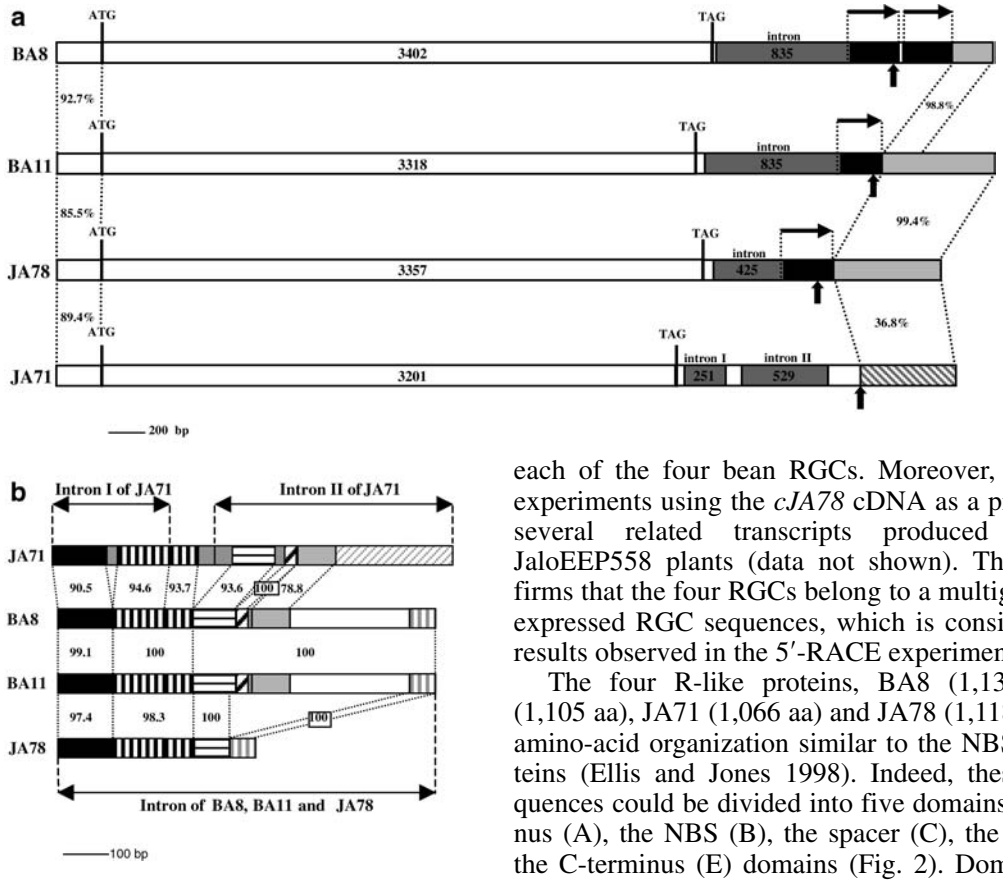

Fig. 1A, B Schematic representation of the four bean R-gene candidate sequences. **A** Alignment of the four bean ORF sequences and their available genomic flanking sequences. ATG and TAG indicate the start and the stop codons, respectively. The ORF lengths (bp) are indicated inside *open boxes*. *Dark gray boxes* correspond to the introns and their lengths (bp) are indicated inside *boxes*. The conserved sequence observed in the *BA8*, *BA11* and *JA78* RGCs is indicated by a *black box* and with an *horizontal black arrow* above the drawing. The *vertical arrows* indicate the beginning of the polyA tail in cDNA sequences. The percentage of nucleotide identity between the 5′ and 3′ flanking genomic sequences are mentioned between drawings. **B** Schematic representation of the intron sequences of the *JA71*, *BA8*, *BA11* and *JA78* RGCs. The intron sequences have been separated in several blocks, indicating different motifs, according to their homology between the four sequences. The percentages of nucleotide identity between these distinct blocks are indicated between sequences. *Dark gray boxes* correspond to small additional sequences that are specific for, or absent in, the *JA71* RGC sequence. The boundaries of these sequences are indicated by *hyphens*. The boundaries of the *JA71* RGC introns I and II and the *BA8*, *BA11* and *JA78* RGC introns are indicated above and below the drawing by *arrows*. The *diagonal gray hatched box* corresponds to the sequence of the *JA71* RGC intron II that could not be aligned with the three other intron sequences

have nucleotide identities that range from 67.1 to 97.9%. Consequently, none of the recovered sequences correspond to the 5′-UTR of one of the four cDNAs. Moreover, analyses of the 5′-RACE products indicate an average size of 80 bp and 70 bp for the 5′-UTR of all transcripts identified from the BAT93 and the JaloEEP558 genotypes, respectively. Therefore, all these results identify the most-likely position of the translation start for each of the four bean RGCs. Moreover, Northern-blot experiments using the *cJA78* cDNA as a probe, revealed several related transcripts produced in infected JaloEEP558 plants (data not shown). This result confirms that the four RGCs belong to a multigene family of expressed RGC sequences, which is consistent with the results observed in the 5′-RACE experiments.

The four R-like proteins, BA8 (1,133 aa), BA11 (1,105 aa), JA71 (1,066 aa) and JA78 (1,118 aa), share an amino-acid organization similar to the NBS-LRR R-proteins (Ellis and Jones 1998). Indeed, these protein sequences could be divided into five domains: the N-terminus (A), the NBS (B), the spacer (C), the LRR (D) and the C-terminus (E) domains (Fig. 2). Domain A has no characteristic motif such as a leucine zipper or a TIR domain found in some NBS-LRR R-proteins (Ellis and Jones 1998). The NBS (domain B) is composed of three highly conserved sequences: Ploop, Kinase 2 and Kinase 3. Four additional regions, conserved to a variable extent, are also present in these proteins (domains IV, V, VI and VII) as reported for the tomato I2 (Simons et al. 1998), the *Arabidopsis* RPM1 (Mindrinos et al. 1994; Grant et al. 1995) and RPS2 (Bent et al. 1994) R-proteins. The LRR region (D) could be structured in imperfect LRRs of variable length (13 to 33 amino acids) with an "axxaxaxxCxxax" amino-acid consensus sequence, where (a) and (x) refer to an aliphatic residue and to any residue, respectively (Fig. 2). Moreover, comparisons of the four full-length BA8, BA11, JA71 and JA78 R-like proteins with the full-length I2 R-protein (the most closely related R-proteins to the four bean R-like proteins), revealed a 38.3/47.9% amino-acid identity/similarity. The amino-acid sequence structures of the I2 R-protein (Simons et al. 1998) and the BA8, BA11, JA71 and JA78 R-like proteins are highly similar: the position of the start codon for translation, the position of the conserved domains in the N-terminus region and the beginning of the LRR region are conserved among these sequences.

Sequence homology among known bean R-like proteins

The BA8, BA11, JA71 and JA78 R-like proteins were compared with the previously published bean NBS-like

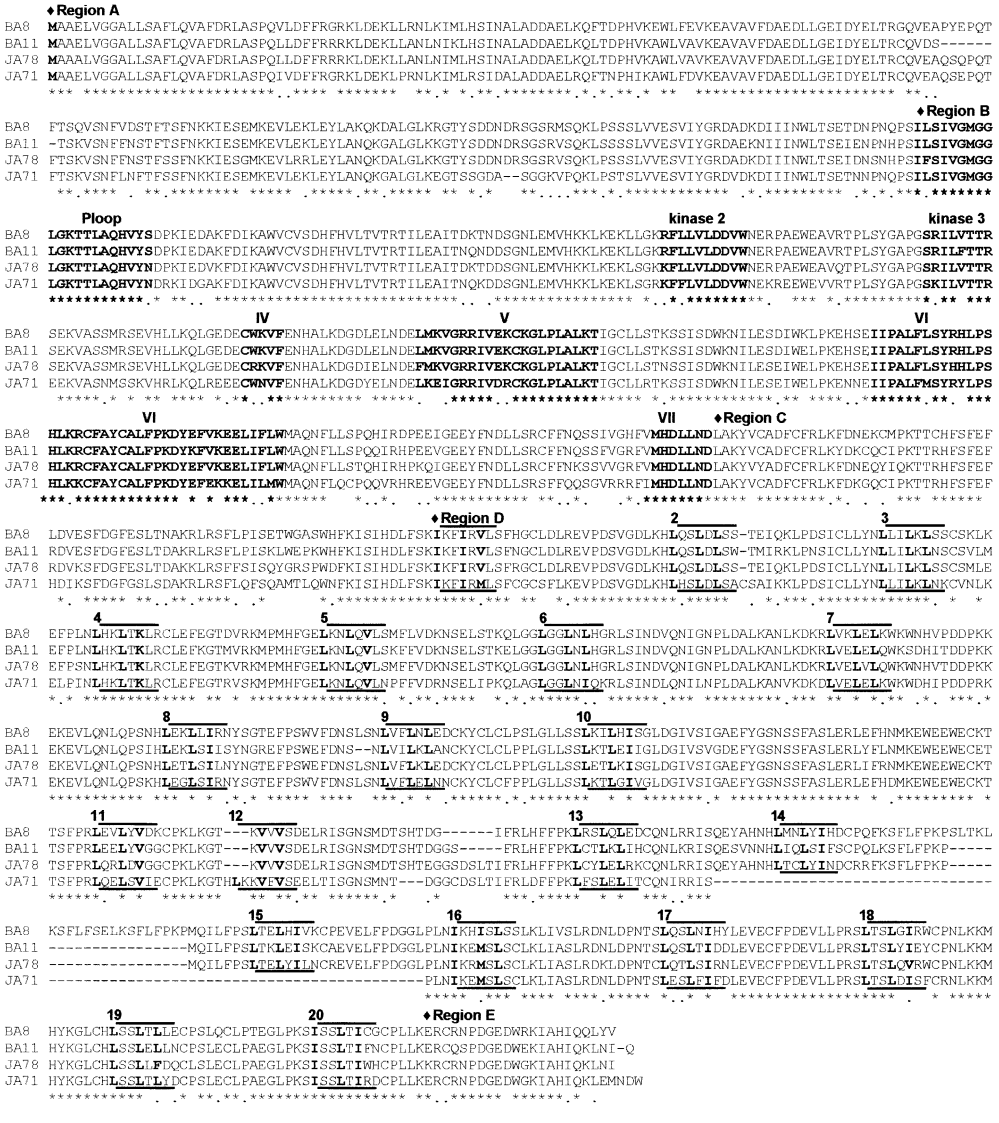

Fig. 2 Alignment of the four R-like proteins encoded by the four bean RGCs. The amino-acid sequences of the four proteins have been divided in five regions: the N-terminus (*A*), the NBS (*B*), the spacer (*C*), the LRR (*D*) and the C-terminus (*E*) regions. Diamonds (black) above the four sequences indicate the beginning of each region. The conserved domains of the NBS region, as well as domains IV to VII that are conserved among the NBS-LRR R-like sequences, are indicated in *boldface characters*. The first amino acid of each LRR is indicated in *bold* and the *number* of each LRR is indicated above the sequence. The LRR consensus sequence is indicated at the bottom of the figure: a corresponds to an aliphatic residue (*I*, *F*, *L*, *M* or *V*) and *x* corresponds to any residue. The conserved aliphatic residues of the β-strand/β-turn domains (xxaxaxx) are shown between *horizontal lines* and aliphatic amino acids are highlighted in *bold*. *Asterisks and dots* indicate identical and similar amino acids in the consensus sequence, respectively

sequences: the Ploop-HD OB/SB sequences (OB1, OB2, OB3, SB1, SB3, SB4, SB5 and SB8; Rivkin et al. 1999), and the Ploop-HD region of the ORF1 RGC sequence that belongs to the *Co-2* NBS-LRR RGC family (Creusot et al. 1999). Amino-acid sequence comparisons revealed that the B4 RGCs are more similar to the ORF1 RGC than to the OB/SB RGCs (Table 1). The BA8, BA11, JA71 and JA78 R-like proteins were also compared to the full protein sequences obtained from two partial cDNAs which belong to the *Co-2* RGC family: cD7 that begins just after the conserved domain IV, and cD8 that begins at the conserved kinase 3 sequence (Creusot et al. 1999). An average percentage of 55.1% of amino-acid identity was observed and the position of the N-terminus domains V to VII, as well as the beginning of the LRR regions, are conserved. However, in Southern analyses, no cross hybridization was observed between the *Co-2* and the *B4* RGC families as demonstrated by the mapping of the members of the two families on two distinct linkage groups of the integrated linkage map of common bean (Geffroy et al. 1998, 1999). These results indicate that the *Co-2* sequences, the *OB*/*SB* RGA sequences and our four RGCs belong to distinct families of bean RGCs, despite the similar amino-acid sequence organization of the corresponding R-like protein products.

The four bean RGCs belong to two distinct subfamilies of R-sequences

Sequence analyses of the four bean cDNA sequences revealed two features. The four cDNAs display a high percentage of nucleotide identity (at least 89.2%) in their coding region, whereas based upon their 3′-UTR sequences the *cJA71* cDNA appears to be distinct from the three other cDNAs (Table 2). At the protein level, amino-acid identities range from 83.8 to 89.6% between the four bean R-like proteins. The length and the structure of

Table 1 Amino-acid comparison of the Ploop-HD regions of the putative BA8, BA11, JA78 and JA71 R-like proteins with the Ploop-HD regions of the putative proteins encoded by *OB/SB* and *ORF1* bean RGAs isolated from bean. The Ploop-HD amino-acid sequences of the four bean R-like proteins, BA8, BA11, JA71 and JA78, have been compared with the Ploop-HD sequences of the putative proteins encoded by the RGAs isolated from bean: OB/SB (OB1, OB2, OB3, SB1, SB3, SB4, SB5 and SB8; Rivkin et al. 1999) and ORF1 (Creusot et al. 1999). The Ploop-HD amino-acid sequences of the four bean R-like proteins BA8, BA11, JA71 and JA78 have been also compared with the corresponding region of the I2 R-protein (Simons et al. 1998) which is the most similar R-protein to the four bean R-like proteins. The last column corresponds to the average percentage obtained over the individual analyses. I and S indicate the percentages of amino-acid identity and similarity, respectively

RGAs	BA8 I/S	BA11 I/S	JA78 I/S	JA71 I/S	Average percentage I/S
ORF1	58.2/65.3	57.6/64.7	58.8/65.2	60.5/67.6	58.7/65.7
OB1	29.2/42.6	28.6/40.8	30.4/42.6	29.7/42.4	29.4/42.1
OB2	27.1/38.5	27.7/39.1	30.5/40.8	30.5/42	28.9/40.1
OB3	29.5/42.9	32/45.1	29.5/43.6	29.5/43.6	30.1/43.8
SB1	30/42.3	29.4/41.1	28.7/38.3	29.8/41.4	29.4/40.7
SB3	31.9/46.2	30.2/44.4	31.2/45	32.5/46.8	31.4/45.6
SB4	43.2/53.2	43.7/53.8	42.6/50.8	44.9/53.2	43.6/52.7
SB5	29.5/37.3	22.8/30	28.9/37.4	28.9/41.5	27.5/36.5
SB8	31.6/40	31/40.3	30.5/44.5	29.4/40.5	30.6/41.3
I2	47/56.4	46.4/55.8	47.6/56.4	48.2/57.6	47.3/56.5

Table 2 Percentages of nucleotide identity between the four cDNA sequences. Percentages of identity between the cDNA ORFs and the 3′-UTRs are presented above and below the diagonal, respectively

	cBA8	*cBA11*	*cJA78*	*cJA71*
cBA8		93.5	94.1	89.6
cBA11	98.8		93.6	89.2
cJA78	98.5	98.7		89.8
cJA71	42.1	48.4	53.8	

the four proteins are highly similar in regions A, B, C and E, while most of the differences are localized in the LRR region (Fig. 2). The BA8, BA11 and JA78 R-like proteins harbor 20 LRRs whereas the JA71 R-like protein lacks LRR14 and LRR15 that are present in the three other R-like proteins. In addition, length polymorphism is observed for the LRR14 (between the BA8, BA11 and JA78 R-like proteins) and for the LRR12 (between the four R-like proteins). Therefore, from these comparisons, the JA71 R-like protein appears to be the most different among the four R-like proteins.

Comparisons of the four RGC cDNA sequences with their corresponding genomic clones revealed the presence of intron sequences in the 3′-UTRs of the four RGCs (Fig. 1A and B). The *BA8*, *BA11* and *JA78* intron sequences are highly similar (more than 97.4% of nucleotide identity) and the major difference is the presence of an additional central sequence of 410 bp in both the *BA8* and *BA11* introns that is absent in the *JA78* intron (Fig. 1B). Moreover, the *BA8* and *BA11* RGCs introns differ by only one base substitution. In the *JA71* RGC two introns are present. In the same way, the full intron I sequence and the 5′-end sequence of intron II of the *JA71* RGC are more than 90% identical with the 5′-end of the *BA8*, *BA11* and *JA78* RGC introns, whereas the 3′-end of the *JA71* intron II (260 bp) is not homologous with the *BA8*, *BA11* and *JA78* intron sequences (Fig. 1B).

High nucleotide sequence identity (85.5 to 99.4%) was also observed for the 5′ and 3′ flanking genomic regions of the four bean RGCs, independently of the genotype, except for the *JA71* RGC 3′ flanking genomic sequence which is only 36.8% identical to the three other corresponding genomic regions (Fig. 1A). Furthermore, the *BA8*, *BA11* and *JA78* RGCs exhibit a conserved sequence that is present in two copies in the *BA8* RGC and in one copy in the *BA11* and *JA78* RGC (Fig. 1A). All copies of this conserved sequence are highly similar (between 98.6 and 100% of identity). We conclude from all those sequence analyses that the four bean RGCs belong to the same NBS-LRR family of RGCs. However, two subfamilies can be defined based upon the percentages of nucleotide identity and specific features of the sequences: *JA71* RGC is representative of a first subfamily whereas the *BA8*, *BA11* and *JA78* RGCs are representative of a second subfamily.

Probes specific for each subfamily of RGCs have been used in Southern hybridization experiments in order to confirm that the four RGCs belong to distinct multigene R-subfamilies. When the conserved sequence present in the *BA8*, *BA11* and *JA78* RGCs (Fig. 1A) is used as a probe, a small multigene subfamily comprising at least eight and five members are revealed in the BAT93 and JaloEEP558 genotypes, respectively. In the same way, a probe corresponding to a part of intron II, that is specific for the *JA71* RGC, has revealed that this sequence also belongs to a small multigene subfamily of at least five members in both the BAT93 and JaloEEP558 genotypes, respectively. Moreover, two subfamily members are also observed following hybridization with the NBS *PRLJ1* probe (for the *BA8*, *BA11*, *JA78* R-subfamily) or with the *JA71* ORF sequence (for the *JA71* R-subfamily). These results confirm that the *BA8*, *BA11*, *JA78* and *JA71* RGCs belong to two small and distinct multigene subfamilies, of which members harbor either the conserved sequence of the *BA8*, *BA11* and *JA78* RGCs or the part of the *JA71* intron II, respectively.

Table 3 Ka/Ks ratios calculated on different domains of the bean R-gene candidates. The Ka/Ks ratios have been determined for each pair of genes for the NBS, the spacer regions and distinct parts of the LRR region. Ka/Ks above 1 (diversifying selection) are indicated in bold characters. Na: not applicable; in this comparison Ks is equal to zero preventing the estimation of the Ka/Ks ratio

Gene pair	NBS region	Spacer	Full LRR region LRR 1 to 20		N-terminus part of the LRR region LRR 1 to 15		C-terminus part of the LRR region LRR 16 to 20	
			Intervening residues	Hypervariable residues	Intervening residues	Hypervariable residues	Intervening residues	Hypervariable residues
BA8/BA11	0.334	7.177	1.004	2.939	0.934	4.505	2.074	2.189
BA8/JA78	0.732	3.998	1.242	3.851	0.911	3.677	4.806	4.245
BA11/JA78	0.404	9.384	1.331	2.743	1.318	2.046	1.296	10.534
BA8/JA71	0.772	0.535	0.633	1.651	0.559	1.275	2.538	4.432
BA11/JA71	0.587	0.549	0.917	2.211	0.880	1.590	Na	6.069
JA71/JA78	0.834	0.6	0.834	2.45	0.771	1.592	1.525	16.385

Nucleotide substitution-ratio comparisons suggest that the bean RGC sequences are under diversifying selection

We calculated Ka/Ks ratios, where Ka and Ks correspond to the frequencies of non-synonymous and synonymous nucleotide substitutions, respectively (Li et al. 1985; Li 1993; Pamilo and Bianchi 1993), for the NBS (B), the spacer (C) and the LRR (D) regions of the four bean RGCs (Table 3). These ratios reveal whether sequence polymorphism is in agreement with diversifying (Ka/Ks > 1) or purifying (Ka/Ks < 1) selection. The NBS encoding sequences of the four bean R-like proteins show a Ka/Ks below 1, suggesting that these regions are under purifying selection, a result described for all R-gene families (Michelmore and Meyers 1998). Interestingly, some Ka/Ks ratios calculated on the spacer regions (C) are above 1, a feature that has not been reported before in studies of other plant R-gene families. For example, Ka/Ks ratios range from 3.998 to 9.384 for the *BA8*, *BA11* and *JA78* spacer regions, whereas these ratios drop below 1 when pair comparisons include the *JA71* spacer region (Table 3).

In the LRR region, Ka/Ks ratios were calculated separately on nucleotides predicted to encode the solvent-exposed hypervariable (x) amino-acid residues and on nucleotides predicted to encode the aliphatic residues (a) of the β-strand/β-turn motif (xxaxaxx), plus nucleotides encoding all the remaining residues of the LRR region. The latter are designated as "intervening" residues (Meyers et al. 1998b). The Ka/Ks ratios calculated on the full-length LRR regions (LRR1 to LRR20) show that the hypervariable residues seem to be under diversifying selection (all Ka/Ks > 1), whereas the intervening residues seem to be either under purifying or diversifying selection (Table 3).

In order to test if the bean LRR regions present a variable pattern of sequence diversity, as observed for the lettuce RGC2 sequences (Meyers et al. 1998b), the Ka/Ks ratio was calculated for several domains of the LRR region. All comparisons revealed a heterogeneous pattern of sequence diversity illustrated by the analysis of the LRR N-terminus part corresponding to LRR1 to LRR15 and the LRR C-terminus part corresponding to LRR16 to LRR20. For all comparisons, Ka/Ks ratios calculated on the intervening residues are smaller than Ka/Ks ratios calculated on the hypervariable (x) residues, except in the C-terminus part for the *BA8/JA78* pair (Table 3). Moreover, for intervening residues in the N-terminus part, Ka/Ks are below 1 in five combinations out of six, suggesting a purifying selection. Finally, Ka/Ks ratios for the intervening residues are smaller in the N-terminus region than in the C-terminus region, except for the *BA11/JA78* pair where intervening residues seem to be under diversifying selection (Table 3).

For the hypervariable (x) residues, all Ka/Ks ratios are above 1 suggesting these residues are under diversifying selection. Particularly, in the C-terminus region, some of the Ka/Ks ratios observed for (x) residues are the highest reported so far (i.e.: 10.534 and 16.385). Therefore, two regions seem to be under diversifying selection in the *B4* RGC family: the spacer and the LRR regions.

Co-localization of the bean R-gene candidates with qualitative and quantitative R-traits

A polymerase chain reaction (PCR) approach, using specific primers of the four bean RGCs, was used to map the four RGCs on the integrated linkage map of common bean (Freyre et al. 1998). The four genes map on linkage group B4 at the complex R-cluster *B4* (Fig. 3). The *BA8* and *BA11* RGCs co-segregate with the Mesoamerican *Co-9* R-specificity which confer resistance against *C. lindemuthianum*, and with a R-QTL against the same pathogen for which the BAT93 allele increases the resistance in leaf, petiole and stem (Geffroy et al. 2000). Similarly, the *JA78* RGC co-segregates with the Andean *Co-y* R-specificity conferring resistance against *C. lindemuthianum*. Interestingly, the *JA71* RGC maps 2 cM from the three other RGCs and co-segregates with a R-QTL for which the JaloEEP558 allele increases resistance in the leaf (Geffroy et al. 2000). These results might reflect a relationship between the genetic localization of RGC sequences and their membership in a defined subfamily of R-sequences.

The polymorphic bands detected in the Andean (JaloEEP558) and Mesoamerican (BAT93) genotypes,

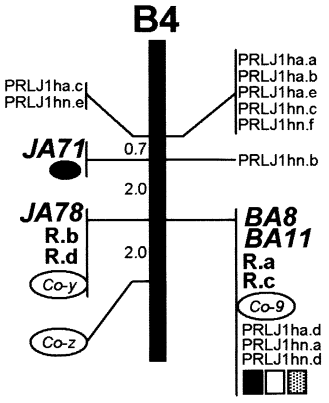

Fig. 3 Detailed genetic localization of the four bean RGCs and their related sequences on the *B4* R-cluster. The *BA8*, *BA11*, *JA71* and *JA78* RGCs map at the *B4* R-cluster at the end of the B4 linkage group of the common bean integrated linkage map (Freyre et al. 1998). The Andean (*Co-y*, *Co-z*) and the Mesoamerican (*Co-9*) R-specificities are indicated in *ovals* (Geffroy et al. 1999). The *black oval* corresponds to the Andean R-QTL effective on leaf. *Black*, *white and hatched rectangles* correspond to the Mesoamerican R-QTLs effective on the leaf, petiole and stem, respectively (Geffroy et al. 2000). R. a, b, c and d correspond to the polymorphic members of the *BA8*, *BA11*, *JA78* subfamily revealed upon hybridization with the conserved sequence observed in the *BA8*, *BA11* and *JA78* RGCs (Fig. 1A). PRLJ1ha.a to PRLJ1ha.e and PRLJ1hn.a, b, c, d, e and f correspond to the members of the *PRLJ1* RGA family previously described and mapped in Geffroy et al. (1999). Genetic distances between each marker are indicated on the left in Kosambi cM

after hybridization with the probe corresponding to the conserved sequence present in the *BA8*, *BA11* and *JA78* RGCs (Fig. 1A), were also mapped (Fig. 3). These bands are localized at the *B4* R-cluster and co-segregate with the *Co-y* and the *Co-9* R-specificities (Fig. 3). This suggests that the *Co-y/Co-9* R-specificities might be members of the RGC subfamily corresponding to the *BA8*, *BA11* and *JA78* RGCs.

Discussion

Molecular and genetic data on the four expressed RGCs are consistent with their putative biological function as R-genes

This paper reports on the first characterization of four full-length bean RGCs encoding NBS-LRR R-like proteins with several features suggesting that they may encode functional R-genes. First, the four RGCs are co-localized either with R-specificities (Geffroy et al. 1999) or R-QTLs that are effective against the fungal pathogen *C. lindemuthianum* (Geffroy et al. 2000). Second, the four RGCs belong to a family of expressed RGCs. Third, the Ka/Ks ratios determined on the NBS and LRR regions of the four RGCs are in agreement with previous results observed for functional R-genes and their analogues (reviewed in Michelmore and Meyers 1998). The NBS regions seem to be under purifying selection, which is consistent with their proposed effector role (Michelmore and Meyers 1998). Hypervariable (x) residues of the β-strand/β-turn motifs in the LRR regions seem to be under high diversifying selection, especially in the 3′ LRR encoding part, where ratios up to 10-fold higher than the previous studies were observed (Parniske et al. 1997; Thomas et al. 1997; Mc Dowell et al. 1998; Meyers et al. 1998b; Michelmore and Meyers 1998; Noël et al. 1999). Such patterns of variability have been observed for the *RGC2* R-genes where diversifying selection occurs similarly on (x) residues of the 3′ encoded LRR region (Meyers et al. 1998b) whereas diversifying selection occurs in the 5′ encoded LRR region for *Cf* R-genes (Parniske et al. 1997; Thomas et al. 1997). As hypervariable residues are hypothesized to be involved in the ligand binding specificity (Kobe and Deisenhofer 1994; Jones and Jones 1997) we could hypothesize that the 3′ encoded LRR region of the bean RGCs may be involved in the recognition of molecules from pathogen origin.

The spacer region and the intervening residues of the LRR region of the four bean RGCs seem to be under diversifying selection

The Ka/Ks calculations also revealed two specific features for these four bean RGC sequences that have not been reported for other RGCs. First, the spacer-encoding regions seem to be under diversifying selection. This result suggests that, like the LRR region, the spacer region could also participate in pathogen recognition as observed for the flax *L* R-genes (Ellis et al. 1999; Luck et al. 2000). Second, the intervening residues seem to be under diversifying selection in the 3′ encoded LRR region, which is different from previous observations on other RGC families from various species (Parniske et al. 1997; Thomas et al. 1997; Mc Dowell et al. 1998; Meyers et al. 1998b; Dodds et al. 2001). As aliphatic residues of the β-strand/β-turn motifs participate in the LRR backbone structure, their variability could be associated with conformational changes in the three-dimensional structure of the LRR (Kobe and Deisenhofer 1995; Jones and Jones 1997), and thus could increase the number of binding opportunities for a pathogen-derived ligand.

Bean RGC families have been generated both in a dicot plant ancestor and in a *Phaseolus* ancestor

Comparison of the four bean BA8, BA11, JA78 and JA71 R-like proteins with several R-proteins isolated from bean and other plant species, revealed that there is no relationship between the percentage of amino-acid identities among R-protein sequences and their species of origin. For example, the BA8, BA11, JA78 and JA71 R-like proteins share a higher score of identity with the tomato I2 R-protein (Simons et al. 1998) than with the bean OB/SB sequences (Rivkin et al. 1999). These re-

sults obtained upon analyses of full-length RGCs (including the NBS and LRR encoding sequence) confirm those obtained upon comparison of the NBS encoding sequence of several RGAs and R-genes (Rivkin et al. 1999; Meyers et al. 1999; Pan et al. 2000a, b). Moreover, we observed a strictly conserved sequence organization between our RGCs and the I2 R-gene, despite the evolutionary distance between the two species of origin, bean (Fabaceae) and tomato (Solanaceae). Such results have also been observed for the common bean *COK-4* RGC which encodes a kinase R-like protein that presents high similarity with the tomato *Pto* R-gene (Melotto and Kelly 2001).

The percentage of amino-acid identity observed between the *B4*, *Co-2* and *OB/SB* bean RGC families confirms that distinct evolutionary histories have occurred within *P. vulgaris*, as previously suggested by Rivkin et al. (1999). Members of the *B4* RGC family are less similar to members of the *OB*/*SB* RGC family than to the tomato I2 R-protein, which could indicate that the *OB*/*SB* and *B4* R-clusters could have arisen from an ancient duplication event preceding the separation of botanical families. Conversely, the BA8, BA11, JA78 and JA71 R-like proteins are more similar to the bean cD7 and cD8 R-like proteins of the *Co-2* RGC family than to the tomato I2 R-protein. Therefore, we could hypothesize that the *Co-2* and *B4* R-clusters resulted from a duplication event that had occurred after separation of botanical families in a *Phaseolus* ancestor. These results argue in favor of the theory that suggests that the number of R-clusters is not fixed and could change at any step during botanical evolution (separation of botanical families or separation of plant species) (Meyers et al. 1999; Pan et al. 2000b).

Generation of sequence diversity and evolution of R-specificities at the *B4* R-cluster

Molecular analyses of many R-gene sequences and their genomic environment in different plant species have increased our knowledge of the structure of complex R-clusters and on the possible mechanisms involved in their evolution. Based on genetic analysis, a common ancestral origin of the *B4* R-cluster in both Andean and Mesoamerican genotypes has been proposed (Geffroy et al. 1999). The molecular results presented in this paper confirm this ancestral origin. Indeed, the four RGCs are highly similar, belong to the same family of RGC sequences and present no feature related to their gene pool origin, namely: (1) the 5′ and 3′ flanking genomic regions share identical sequences independently of the genotype, (2) the structure of the first intron is conserved in the four RGCs, and (3) a conserved sequence is present in the RGCs of both Andean (*JA78*) and Mesoamerican (*BA8* and *BA11*) origins. Similar observations have been reported in the *Lycopersicon* genus where comparisons of the *Cf-4*/*Cf-9*/*Cf-0* locus in genotypes from diverse *Lycopersicon* species have revealed that flanking regions of the *Cf* R-genes and of their homologues are partially identical, suggesting that these loci have arisen by duplication of an ancestral locus (Parniske et al. 1997; Thomas et al. 1997).

Based upon population, genetic and molecular analyses of the *PRLJ1* RGA family, Geffroy et al. (1999) have suggested that different sets of R-specificities have been randomly generated in each center of diversity of bean, and that different R-specificities were selected after separation of the different bean gene pools. Because the similarity among the putative Mesoamerican RGCs paralogues (*BA8*/*BA11*) is not higher than the similarity between the putative Andean and Mesoamerican RGCs orthologues (*BA8*/*JA78* and *BA11*/*JA78*), our results argue against concerted evolution of the RGC sequences that belong to the *B4* R-cluster. Moreover, as the four RGCs display major polymorphism in the LRR region (i.e.: amino-acid polymorphism and variation of the LRR number) and high Ka/Ks ratios were found in the 3′ part of the LRR region, we could hypothesize that putative R-specificity relies on variation in this region of the protein. Similarly, previous results reported that variations in the LRR region of related R-genes are associated with a modification of R-specificity (Anderson et al. 1997; Thomas et al. 1997; Mc Dowell et al. 1998; Michelmore and Meyers 1998; Meyers et al. 1998b; Ellis et al. 1999; Dodds et al. 2001; Van der Hoorn et al. 2001; Wulff et al. 2001).

Interestingly, the distinct structure of the LRR region between the *JA71* and the three other RGCs could also reflect that these RGCs confer distinct types of R-traits as both qualitative and quantitative resistance are localized at the *B4* R-locus (Fig. 3). Transient expression experiments using *Agrobacterium tumefasciens* transformation are underway in our laboratory in order to test the functionality of the four R-gene candidates and establish whether one of them corresponds to a previously characterized R-specificity (Geffroy et al. 1999).

The molecular analysis of the present study also provides clues about the putative mechanisms that have contributed to the evolution of the ancestral *B4* R-cluster in bean. At least, two distinct duplication events seem to have occurred: a first ancient one that led to the creation of the two RGC subfamilies and a more recent one that led to the creation of the *BA8* and *BA11* RGCs. Thus, the *B4* R-cluster has apparently evolved through duplication and point mutation like the *Xa21* R-locus in rice (Song et al. 1997), the *Cf* R-loci in tomato (Parniske et al. 1997) and the *Dm3* locus in lettuce (Meyers et al. 1998a).

Finally, as the *JA71* RGC maps 2 cM from the three other RGCs (*BA8*, *BA11* and *JA78*), a relationship is so far observed between the genetic localization of the four RGCs and the fact that they belong to each subfamily of R-sequences. Such a result has been observed for the members of the *Xa21* R-cluster (Song et al. 1997) and for the *Hcr9* members of the *Cf* R-clusters (Parniske and Jones 1999), but not for the members of the *Dm3* R-cluster (Meyers et al. 1998a). The isolation of RGCs that co-localize with the Andean *Co-z* R-specificity will reveal whether such genes belong to one of the two previously

defined subfamilies or if they belong to a new subfamily of RGC with their own features. Isolation and characterization of numerous paralogue and orthologue R-genes, present at distinct genomic positions within the *B4* R-cluster, are underway. This should provide us with information on whether several other RGC subfamilies exist at the *B4* R-cluster in the BAT93 and JaloEEP558 genotypes, and on the molecular events that have been involved in the evolution of the *B4* R-cluster in the two gene pools.

Acknowledgements We are grateful for R. Laugé and D. Sicard for helpful discussions and critical reading of the manuscript. We thank B.C. Meyers (Dept. Of Vegetable Crops, UC Davis, USA) and B. Gaut (Dept. of Ecology and Evolutionary Biology, UC Davis, USA) for helpful discussions on Ka/Ks calculations. E.F.C. was supported by a grant from the Ministère de l'Education Nationale (contract number: 96302). The research was supported by the European community (INCO-CT980317 project), CNRS and INRA. GenBank accession numbers. cDNA sequences: *cBA8* (AF306499), *cBA11* (AF306500), *cJA71* (AF306501) and *cJA78* (AF306502). Genomic sequences: *BA8* (AF306503), *BA11* (AF306504), *JA71* (AF306505) and *JA78* (AF306506)

References

Aarts MG, 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

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9:641–651

Becerra-Velasquez VL, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centers of origin. Genome 37:256–263

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawicz BJ (1994) *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856–1860

Bottela MA, Coleman MJ, Hughes DE, Nishimura MT, Jones JDG, Sommerville SC (1997) Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant J 12:1197–1211

Caicedo AL, Schaal BA, Kunkel BN (1999) Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 96:302–306

Collins N, Drake J, Ayliffe M, Sun Q, Ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize *Rp1-D* rust resistance haplotype and its mutants. Plant Cell 11:1365–1376

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

Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* (*Fabaceae*) in Northern South America. Econ Bot 47:408–423

Delgado-Salinas A, Bonet A, Gepts P (1988) The wild relative of *Phaseolus vulgaris* in Middle America. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer Academic Publishers, London, pp 163–184

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

Ellis J, Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant Biol 1:288–293

Ellis JG, Lawrence GJ, Finnegan EJ, Anderson PA (1995) Contrasting complexity of two rust resistance loci in flax. Proc Natl Acad Sci USA 92:4185–4188

Ellis JG, Lawrence GJ, Luck JE, Dodds N (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. Plant Cell 11:495–506

Flor HH (1971) Current status of the gene-for-gene concept. Anu Rev Phytopathol 9:275–296

Freyre R, Skroch P, Geffroy V, Adam-Blondon A-F, Shirmohamadali A, Johnson WC, Llca V, Nodari RO, Pereira PA, Tsai SM, Tohme J, Dron M, Nienhuis J, Vallejos CE, Gepts P (1998) Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. Theor Appl Genet 97:847–856

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85:8998–9002

Geffroy V, Creusot F, Falquet J, Sévignac M, Adam-Blondon A-F, Bannerot H, Gepts P, Dron 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

Geffroy V, Sicard D, De Oliveira JC, Sévignac M, Cohen S, Gepts P, Neema C, Langin T, Dron M (1999) Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. Mol Plant-Microbe Interact 12:774–784

Geffroy V, Sévignac M, De Oliveira JCF, Fouilloux G, Skroch P, Toquet P, Gepts P, Langin T, Dron M (2000) Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of QTLs with genes involved in specific resistance. Mol Plant-Microbe Interact 13:287–296

Gepts P (1988) Phaseolin as an evolutionary marker. In: Gepts P (ed) Genetic resources of *Phaseolus* beans. Kluwer Academic Publishers London, pp 215–241

Gepts P (1998) Origin and evolution of common bean: past events and recent trends. HortScience 33:1124–1130

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995) Structure of the *Arabidopsis Rpm1* gene enabling dual specificity disease resistance. Science 269:843–846

Guzman P, Gilbertson RL, Nodari R, Johnson WC, Temple SR, Mandala D, Mkandawire ABC, Gepts P (1995) Characterization of variability in the fungus *Phaeoisariopsis griseola* suggests coevolution with the common bean (*Phaseolus vulgaris*). Phytopathology 85:600–607

Innes RW (1998) Genetic dissection of R gene signal transduction pathways. Curr Opin Plant Biol 1:229–304

Jia Y, McAdams SA, Bryan GT, Hershey HP, Valent B (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J 19:4004–4014

Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defenses. In: Andrews JH, Tommerup IC (eds) Advances in botanical research. Academic Press Incorporated, San Diego, pp 89–167

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35: 993–1001

Kami J, Velasquez VB, Debouck DG, Gepts P (1995) Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc Natl Acad Sci USA 92:1101–1104

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11,746–11,750

Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. Trends Genet 19:415–421

Kobe B, Deisenhofer J (1995) A structural basis of the interaction between leucine-rich repeats and protein ligands. Nature 374: 183–186

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of genetic diversity. Theor Appl Genet 78:809–817

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

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

Leister RT, Katagiri F (2000) A resistance gene product of the nucleotide binding site-leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant J 22:345–354

Leister D, Ballvora A, Salamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genet 14:421–429

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

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Luck JE, Lawrence GJ, Dodds PN, Shepherd KW, Ellis JG (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. Plant Cell 12:1367–1377

Mc Dowell 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 *RPP8* locus of *Arabidopsis*. Plant Cell 10:1861–1874

Melotto M, Kelly JD (2001) Fine mapping of the *Co-4* locus of common bean reveals a resistance gene candidate, *COK-4*, that encodes for a protein kinase. Theor Appl Genet 103:508–517

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

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998b) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell 11:1833–1846

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrihnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Cell 20:317–332

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8:1113–1130

Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994) The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78:1089–1099

Noël L, Moores TL, van der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JDG (1999) Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. Plant Cell 11:2099–2111

Pamilo P, Bianchi NP (1993) Evolution of the *Zfx* and *Zfy* genes: rates and interdependance between the genes. Mol Biol Evol 10:271–281

Pan Q, Liu YS, Budai-Hadrian O, Sela M, Carmel-Goren L, Zamir D, Fluhr R (2000a) Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: tomato and Arabidopsis. Genetics 155:309–322

Pan Q, Wendel J, Fluhr R (2000b) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J Mol Evol 50:203–213

Parniske M, Jones JDG (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. Proc Natl Acad Sci USA 96:5850–5855

Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BBH, Jones JDG (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell 91:821–832

Pastor-Corrales MA, Tu JC (1989) Anthracnose. In: Schwartz HF, Pastor-Corrales MA (eds) Bean production problems in the tropics. Centro International de Agricultura Tropical, Cali, Colombia, pp 77–104

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

Shen KA, Meyers BC, Islam-Faridi MN, 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

Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, Van der Lee T, Bleeker M, Onstenk J, de Both M, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P (1998) Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. Plant Cell 10:1055–1068

Singh SP, Nodari R, Gepts P (1991) Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci 31:19–23

Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279–1287

Sonnante G, Stockton T, Nodari RO, Beccera-Velasquez VL, Gepts P (1994) Evolution of genetic diversity during the domestication of common bean *(Phaseolus vulgaris* L.). Theor Appl Genet 89:629–635

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

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400:667–671

Steadman JR, Beaver J, Boudreau M, Coyne D, Groth J, Kelly J, McMillan M, McMillan R, Miklas PMPC, Schwartz H, Stavely J (1995) Progress reported at the 2nd International Bean Rust Workshop. Annu Rep Bean Improv Coop 38:1–10

Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JDG (1997) Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. Plant Cell 9:2209–2224

Vallelian-Bindschedler L, Mösinger E, Métraux JP, Schweizer P (1998) Structure, expression and localization of a germini-like protein in barley (*Hordeum vulgare* L.) that is insolubilized in stressed leaves. Plant Mol Biol 37:297–308

Van der Hoorn RAL, Roth R, De Wit PJGM (2001) Identification of distinct specificity determinants in resistance protein *Cf-4* allows construction of a *Cf-9* mutant that confers recognition of avirulence protein AVR4. Plant Cell 13:273–285

Wulff BBH, Thomas CM, Smoker M, Grant M, Jones JDG (2001) Domain swapping and gene shuffling identity sequences required for induction of an Avr-dependent hypersensitive response by the tomato *Cf-4* and *Cf-9* proteins. Plant Cell 13: 255–272

Yu YG, Buss GR, Maroof MAS (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 93:11,751–11,756