Theor Appl Genet (2003) 108:53–61
DOI 10.1007/s00122-003-1410-1

ORIGINAL PAPER

Z. Deng · F. G. Gmitter

Cloning and characterization of receptor kinase class disease resistance gene candidates in Citrus

Received: 19 December 2002 / Accepted: 15 May 2003 / Published online: 6 September 2003
© Springer-Verlag 2003

Abstract The rice gene *Xa21* represents a unique class of plant disease resistance (*R*) genes with distinct protein structure and broad-spectrum specificity; few sequences or genes of this class have been cloned and characterized in other plant species. Degenerate primers were designed from the conserved motifs in the kinase domains of *Xa21* and tomato *Pto*, and used in PCR amplification to identify this class of resistance gene candidate (RGC) sequences from citrus for future evaluation of possible association with citrus canker resistance. Twenty-nine RGC sequences highly similar to the kinase domain of *Xa21* (55%–60% amino-acid identity) were cloned and characterized. To facilitate recovery of full-length gene structures and to overcome RGC mapping limitations, large-insert genomic clones (BACs) were identified, fingerprinted and assembled into contigs. Southern hybridization revealed the presence of 1–3 copies of receptor-like kinase sequences (i.e., clustering) in each BAC. Some of these sequences were sampled by PCR amplification and direct sequencing. Twenty-three sequences were thus obtained and classified into five groups and eight subgroups, which indicates the possibility of enhancing RGC sequence diversity from BACs. A primer-walking strategy was employed to derive full-length gene structures from two BAC clones; both sequences 17o6RLK and 26m19RLK contained all the features of the rice Xa21 protein, including a signal peptide, the same number of leucine-rich-repeats, and transmembrane and kinase domains. These results demonstrate that PCR amplification with appropriately designed degenerate primers is an efficient approach for cloning receptor-like kinase class RGCs. Utilization of BAC clones can facilitate this approach in multiple ways by improving sequence diversity, providing full-length genes, and assisting in understanding gene structures and distribution.

Keywords Citrus · Resistance gene candidates · Receptor-like kinases · Xa21-like genes · Citrus canker

Communicated by P. Langridge

F. G. Gmitter Jr. (✉)
University of Florida,
Citrus Research and Education Center,
700 Experiment Station Road, Lake Alfred, FL, 33850, USA

Z. Deng
University of Florida,
Gulf Coast Research and Education Center,
5007 60th Street East, Bradenton, FL, 34203, USA

Introduction

Enhancing disease resistance is one of the top priorities in plant breeding and genetic improvement programs. Hundreds of resistance (*R*) genes play essential roles in defending agricultural and horticultural crop plants against pathogen attacks (Michelmore 1996). For effective utilization of these natural *R* genes and for engineering novel resistance, it is important to clone and characterize these *R* genes.

At least two-dozen *R* genes have been cloned from several plant species, using map-based (positional) or transposon-tagging approaches. Characterization of these cloned genes has revealed interesting features of plant *R* genes: only a few classes of proteins are involved in plant disease resistance (Staskawicz et al. 2001) and the proteins tend to contain modular structures (Hammond-Kosack and Jones 1997). Particularly, it has been found that the NBS (nucleotide-binding site) domains from many *R* genes of different plant origin and responsible for resistance to diverse pathogens, share highly conserved amino-acid sequences (Hammond-Kosack and Jones 1997). These findings have led to the development of PCR-based methods using degenerate primers as new approaches for cloning *R* genes or *R*-gene candidates (RGCs; Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). Many of these RGCs are tightly linked or co-segregate with known disease resistance loci (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996; Deng et al. 2000); as such, they may represent the ultimate molecular markers for plant disease resistance (Michelemore 1996). Several RGCs have facilitated cloning of full-length

functional *R* genes, including lettuce *Dm3* that confers resistance to downy mildew (*Bremia lactucae*) (Meyers et al. 1998), and the *Arabidopsis Rpp8* that is responsible for downy mildew (*Pernospora parasitica*) resistance (Aarts et al. 1998; McDowell et al. 1998). Therefore, RGCs provide valuable sequence resources for using a candidate gene approach to clone plant *R* genes (Pflieger et al. 2001). In addition to their immediate utilization in genetics and breeding selection, RGCs also provide opportunities and tools to answer some fundamental questions about disease resistance genes, such as *R* gene organization, distribution and evolution (Michelemore 1996).

Hundreds of RGCs have been obtained from many plant species in the last few years and deposited in DNA sequence databanks such as GenBank (http://www.ncbi.nlm.nih.gov/). Nevertheless, most reported sequences are partial and short, containing only a small portion of full-length gene structures when compared to cloned functional *R* genes. Consequently, they are not ready to be tested in genetic transformation for their functions. In addition, most of these sequences were amplified with a limited number of NBS domain-based degenerate primers, and as a group they may not actually represent the breadth of plant *R*-gene repertoires. Among the several classes of cloned *R* genes, rice gene *Xa21* is unique in structure and function, thus representing a distinct class (a receptor-like kinase, RLK) of *R* genes (Song et al. 1995; Ronald 1997). *Xa21*, either in its natural or isolated form, confers resistance to more than 30 different *Xanthomonas oryzae* pv *oryzae* (*Xoo*) isolates. Its predicted gene product is a transmembrane protein containing extracellular LRRs (leucine-rich repeats) and a cytoplasmic serine-threonine kinase, i.e., one part similar to the LRR class *R* genes and the other part similar to kinase class *R* genes (Song et al. 1995). Few receptor-like kinase RGC sequences have been cloned or characterized in other plants. This may be partly due to the fact that the LRR domains are poorly conserved. Further, the kinase domains are too common among genes responsible for diverse functions, and suitable distinctive degenerate primers are not available.

Citrus is the most important fruit crop worldwide, with an annual global production of more than 100 million metric tons (FAO, http://www.fao.org/). This production scale puts citrus well ahead of banana, grape and apple. A wide spectrum of pathogens including viruses, bacteria, fungi and nematodes cause huge losses of production. Some of these pathogens, when epidemic, have destroyed local or regional citrus industries in several countries. Examples include citrus tristeza virus (CTV) and *Xanthomonas axonopodis* pv *citri* (*Xac*). A single dominant gene (*Ctv*) has been found in a citrus relative (*Poncirus trifoliata*, Gmitter et al. 1996) that can confer durable broad-spectrum resistance to CTV. It is estimated that it will take more than 10 years from initial genetic mapping to final cloning of the *Ctv* gene, using a map-based strategy. *Xac* causes citrus canker, a very devastating disease in many citrus growing areas including Florida, Brazil and China (Gottwald et al. 2002). Resistance to citrus canker has been found in kumquats (*Fortunella* spp.) (Gottwald et al. 2002). However, the inheritance of the resistance trait is unclear; the number of gene loci controlling the resistance and the biochemical mechanism involved have yet to be determined. One of the current attempts to develop canker resistance is to express the rice gene *Xa21* in susceptible citrus (Jude Grosser, UF/IFAS-CREC, personal communication). Another approach under consideration is cloning *R* gene candidates, which involves cloning and screening various classes of RGCs, followed by functional tests of co-segregating candidates in transgenic plants. Previously, screening of NBS-LRR class RGCs allowed us to identify molecular markers that co-segregated with the *Ctv* locus and led to chromosome landing on *Ctv*-containing BACs (bacterial artificial chromosomes; Deng et al. 2001b), and to develop one marker that was linked to *Tyr1*, a major gene locus for citrus nematode resistance in *Poncirus trifoliata* (Deng et al. 2000; Ling et al. 2000).

The objectives of this study were: (1) to demonstrate the feasibility of cloning receptor-like kinase class *R*-gene sequences using appropriate degenerate primers and PCR amplification; (2) to obtain citrus RGCs for future evaluation of their association with citrus canker resistance; and (3) to evaluate the effectiveness of cloning RGCs and full-length genes by incorporating genomic resources, particularly BAC clone libraries.

Materials and methods

Degenerate primers

The amino-acid sequences of *Xa21* (Song et al. 1995) and tomato *R* gene *Pto* (Martin et al. 1993) were aligned using the program ClustalX (Thompson et al. 1997) to identify potential conserved regions and amino acids. Two conserved regions were chosen for primer designing. From the amino acids FG(K/S)VYKG and GY(A/I)(A/D)PEY, sense and antisense degenerate primers (kindF1 and kindR1) were designed, respectively; their sequences were: (kindF1) *5′-TT(T/C) GGI (A/T)II GTI TA(T/C) (A/T)(A/C/G)I GG-3′*, and (kindR1) *5′-(A/G)TA (T/C)TC IGG I(T/G)(A/C) I(G/A)(T/C) (A/G)TA ICC-3′*

PCR amplification, DNA cloning, and sequencing

The genomic DNA of USDA 17-47, an intergeneric hybrid of *Citrus grandis* and *P. trifoliata,* was used as the DNA template in initial PCR reactions. Amplifications were performed on a GeneAmp PCR System 9700 thermal cycler (PE Applied Biosystem) in a 50-μl reaction volume; each reaction contained 50 m*M* Tris-HCl (pH 8.3), 2 m*M* $MgCl_2$, 800 μ*M* dNTPs, 100 μ*M* primer kindF1 and 100 μ*M* primer kindR1, 300 ng of genomic DNA, and 2.5 units of *Taq* polymerase. The thermal cycling parameters were: initial denaturation at 94°C for 3 min, 35 cycles of 30 s at 94°C, 30 s at 45°C and 2 min at 72°C, followed by a final extension at 72°C for 8 min. PCR products were separated in a 1.8% agarose gel, and visualized by staining with ethidium bromide and UV illumination. DNA fragments similar to the expected size (540 bp) were excised from the gel and used as templates in PCR re-amplification. Subsequently, the re-amplified products were collected from a 1% agarose gel and cloned using the pGEM T Easy

system (Promega, Wis., USA). Recombinant clones were picked and maintained in 96-well microtiter plates.

To characterize these clones, their inserts were amplified by PCR using two primers (T7 and BAC4) designed from the insert-flanking vector regions, and bacterial cell culture suspension as PCR templates. PCR products were digested with restriction enzymes *Rsa*I and *Taq*I separately, and their digests were analyzed on agarose gels. Restriction fragment patterns of each clone were analyzed, and clones producing unique restriction patterns were chosen for sequencing. Double-stranded plasmid sequencing was performed at the University of Florida's DNA Sequencing Core Laboratory (UF-DSCL, Gainesville, Fla.), using the BigDye terminator system and an ABI 377 DNA Sequencer.

PCR amplification of BACs was performed using a protocol similar to the above genomic amplification. Primers A2UP and A2LW were designed from the insert sequence of plasmid clone A2. Similar components were included in PCR reactions, except that 10 μ*M* of each primer, 1 μl of BAC DNA preps, and a trace amount of RNase were used. PCR products were separated on agarose gels, excised from the gel, and eluted into 0.1 × TE buffer (0.1 m*M* Tris-HCl, pH 8.3, and 0.01 m*M* EDTA, pH 8.0). The purified products were sequenced using primer A2LW at the UF-DSCL.

High-density colony screening and characterization of BAC clones

Sixteen high-density colony filters of citrus BAC clones were prepared and screened as described previously (Deng et al. 2001a). The filters contained approximately 24,000 clones of the USDA 17-47 *Bam*HI BAC library (Deng et al. 2001a). BAC DNA of positive clones was prepared using a modified alkaline procedure described by Zhang et al. (1996). DNA preps were digested with *Hin*dIII and *Bam*HI at 37°C for 4 h, and run on 1% agarose gels in 1 × TAE buffer. The gels were stained with ethidium bromide and documented. DNA fragments in agarose gels were transferred onto nylon membranes under alkaline conditions. Southern hybridizations were conducted using DIG (digoxigenin)-labeled probes. Manufacturer's recommendations were followed in probe labeling, hybridization, and detection (Roche Molecular Biochemicals, Mannheim, Germany).

Primer walking-based sequencing of BACs

Pure BAC DNA was prepared from 500 ml of cell culture, using the Qiagen plasmid maxi kit (Qiagen, Valencia, Calif., USA). Sequencing was performed by the UF-DSCL, using the BigDye terminator system and an ABI 377 DNA Sequencer. Initial sequences were obtained with primer A2LW. From these and subsequent new sequences, primers were designed with the assistance of computer program OLIGO 6 and used to walk on BACs for sequencing in both directions. At the end of primer walking, numerous new primers were designed to sequence the other strand of the target region. From the obtained sequence it reads from both strands, and a contigous sequence of the target region was generated in DNASIS.

Sequence analysis

DNA sequence editing and analysis were performed with DNASIS or DNASIS MAX v1.0 (MiraiBio Inc.), and the Genetics Computer Group (GCG) SeqWeb software programs. Database searches were conducted using Gapped-BLAST (Altschul et al. 1997). Phylogenetic analysis was performed using the CLUSTALX package (Thompson et al. 1997). Ten iterations of sequence alignment and tree construction were conducted. In each iteration, a neighbor-joining (NJ) tree (Saitou and Nei 1987) was generated and used as the guide tree for the next cycle of alignment. Bootstrap analysis was performed in CLUSTALX to evaluate the reliability of the nodes of the phylogenetic trees.

Results

Amplification, cloning, and sequencing of RLK class RGCs from genomic DNA

Sequence alignment of the kinase domains of Xa21 and Pto revealed two well-conserved regions in the subdomains I and VIII, respectively (Fig. 1). Together, the two regions contained three of the 15 invariant amino acids (Val, Pro and Glu) identified in eukaryotic protein kinases (Hanks et al. 1988). When two degenerate primers (kindF1 and kindR1) designed from these two regions were used to amplify the genomic DNA of USDA 17-47, two major DNA bands and three minor bands were observed, from the PCR amplification, on an argarose gel (Fig. 2). The approximate 520-bp major band and the approximate 650-bp minor band displayed fragment sizes close to the expected (540 bp), and therefore, the two bands were excised from the gel and re-amplified by PCR. DNA fragments in the bands were cloned. Many of the recombinant clones were light-blue on selective argar plates containing IPTG (isopropylthio-*β*-D-galactoside) and X-Gal (5-bromo-4-chloro-3-indolyl-*β*-D-galactoside), suggesting possible encoding capacities of the DNA inserts. One hundred of the recombinant clones

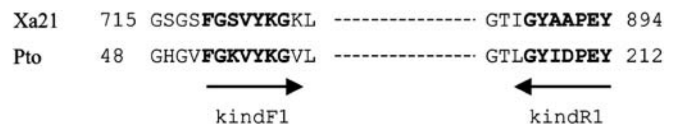

Fig. 1 Two highly conserved regions in the rice blight resistance gene protein Xa21 (Song et al. 1995) and tomato bacterial speck resistance gene protein Pto (Martin et al. 1993), and location of degenerate primers designed in this study

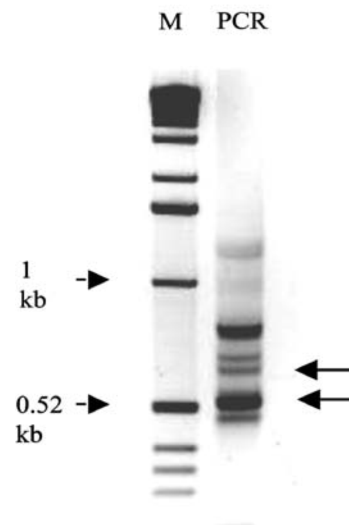

Fig. 2 PCR amplification of USDA 17-47 genomic DNA with primers kindF1 and kindR1 (reversed image). M: 1-kb ladder DNA size-markers. Lane marked PCR contained the PCR products. Arrows point to the bands excised, re-amplified and cloned

Table 1 Receptor-like kinase class RGCs cloned from citrus genomic DNA (prefixed with letters) or amplified from citrus BACs (prefixed with numbers). The RGCs were classified into groups and subgroups based on multiple sequence alignments and pairwise amino-acid identity values. RGCs in boldface and italics were representatives of each subgroups and they were used in phylogenetic relationship analysis. Percent amino-acid identity and similarity (in parenthesis) values against the rice Xa21 kinase domain (aa 708–888; Song et al. 1995) were calculated using the GCG 'GAP' program, with gap weight = 8 and length weight = 2

Groups	Subgroups	Name of sequences	Percent amino-acid identity and similarity to rice Xa21	BLASTP Expect value
CRK1	a	***A2***, A3, A5, B4, B9, C2, C8, C9, E1, E5, E8, E9, F4, F6, F9, D8, F8, G12, G1, 17o6, 19h7	56.8 (68.5)	9e-46
	b	***B10***	55.3 (67.3)	5e-57
	c	***E3***	56.6 (68.7)	1e-54
CRK2	a	***26m19***, 47F11, 6m9, 42h8, 43j5, 28b7, 49m131, 31a4	59.7 (70.8)	1e-44
	b	***A1***, A4, B5, C6, C7, D6, D9, F12, G3	57.6 (69.4)	1e-56
	c	***15n13***	55.6 (66.7)	3e-41
CRK3	a	***2j4***, 8L10, 29a13, 32m16, 41L8, 50n111, 49j23	56.5 (69.4)	2e-41
	b	***25F17***	55.6 (66.7)	1e-43
	c	***29b4***	54.9 (65.5)	3e-41
CRK4	a	***19h16***, 26o12	57.0 (68.3)	2e-41
CRK5	a	***16b111***	58.0 (71.3)	2e-43

were characterized further. Inserts of the clones were amplified by PCR with primers T7 and BAC4, and digested with *Rsa*I and *Taq*I, respectively. Fifty three clones that displayed unique restriction patterns on agarose gels were chosen for DNA sequencing.

Sequencing data revealed that most of the selected clones contained inserts from 490 bp to 650 bp in length; there was considerable heterogeneity among the clones, and therefore, in the original DNA bands excised from the agarose gel. BLASTX searches (Altschul et al. 1997) of the GenBank showed that 33 of the clones possessed strong similarities to numerous protein kinases, but the rice blight resistance gene *Xa21* (Song et al. 1995) and a number of *Xa21*-like genes identified in large-scale genome sequencing projects were their best hits in the database, with the expected value less than 9e-46 (Table 1). The majority of these sequences (29 out of 33) could be translated to polypeptides without any stop codons. These sequences fell into two different groups in clustering analysis (see below). Thirteen clones showed the greatest similarities to other classes of plant-kinase genes, including receptor protein kinases, cytokinin-regulated kinases, lectin-like protein kinase, S-receptor kinases and brassinosteroid-receptor kinases. Several clones were found similar to hypothetical proteins, or transposases. These clones were excluded from further analysis in this study.

Identification of BAC clones and RLK class RGCs from BACs

To facilitate recovery of more diverse sequences and cloning of full-length gene structures, BAC clones were screened with the plasmid clone A2 obtained above. This clone represented the most-abundant type among the 29 *Xa21*-like clones. The BAC library used was constructed from *Bam*HI partially digested high-molecular-weight genomic DNA of USDA 17-47 and consisted of more than 24,000 clones with an average insert size of 115 kb (Deng et al. 2001a). Therefore, the library represented approximately seven equivalents of the *Poncirus* and *Citrus* genomes. After moderately high stringency post-hybridization washes (2 times of 20-min washes with 0.2 × SSC + 0.1% SDS solution at 65°C), 79 positive BAC clones were identified. The insert sizes of these clones ranged from 45 kb to 175 kb. To identify overlapping clones for contig development, all the positive BACs were characterized by fingerprinting analysis and Southern blotting. Fingerprinting was conducted by simultaneous digestion of BACs with two restriction enzymes (*Bam*HI and *Hin*dIII) followed by electrophoresis on agarose gels. Southern blots from these gels then were probed with the A2. Based on these analyses, the 79 clones were grouped into 35 contigs: 21 of them consisted of 2–6 BAC clones each, and the remaining 14 of the contigs were singletons with one BAC clone each.

To examine the distribution of RLK-like sequences in the above-identified BAC clones, BAC DNA from different contigs were restricted with *Bam*HI and *Hin*dIII, separated on agarose gels, and blotted onto nylon membranes. The blots were probed with a pool of eight PCR fragments that represented the five groups and eight subgroups of *Xa21*-like sequences sampled from these BAC clones (see below). One to three hybridizing DNA fragments (an average of 1.6 fragments) were observed in each BAC, indicating a possible clustering of kinase domain-encoding DNA sequences (Fig. 3).

PCR amplification was attempted with primers A2UP and A2LW that were designed from the A2 DNA sequence to sample some of RLK sequences in the BACs. Twenty three of the 79 clones yielded PCR products of apparently similar sizes (about 450 bp). These 23 clones were from 13 of the contigs described above. The PCR products were gel-purified and then sequenced with the primer A2UP. Sequence information from these

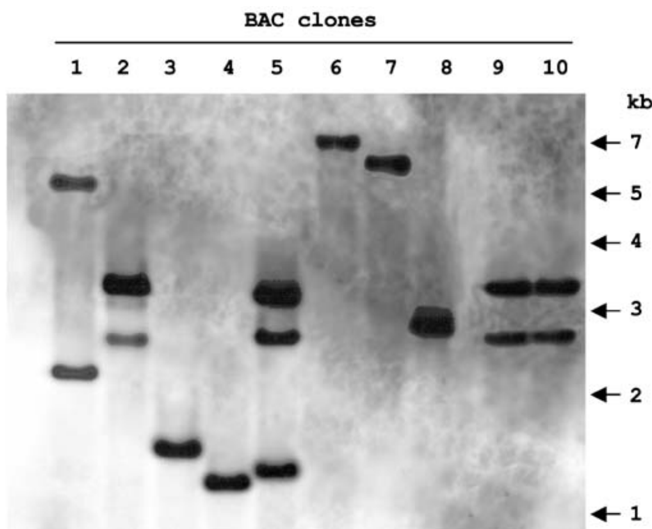

Fig. 3 Clustering of receptor-like kinase-encoding DNA fragments in BAC clones. Ten BAC clones from different contigs were restricted with enzymes *Bam*HI and *Hin*dIII; and the DNA blot was probed with a pool of eight representative PCR fragments that are similar to the kinase domain-encoding region of rice resistance gene *Xa21* (Table 1). The eight fragments were 2j4, 15n13, 16b111, 17o6, 19h16, 25F17, 26m19 and 29b4, and they were labeled with DIG (digoxigenin), using Roche's DIG-High Prime kit. Post-hybridization wash conditions consisted of two washes at 65°C with 0.5×SSC+0.1% SDS (15 min/each time), and one wash at 65°C with 0.2×SSC+0.1% SDS for 15 min. Chemiluminescent signal detection was performed according to the manufacturer's recommendations (Roche Molecular Biochemicals). DNA size markers (in kilobases) are indicated to the right of the image

PCR products supported the above grouping of BAC clones based on hybridization and fingerprints, i.e., PCR products from clones in the same contigs contained the same sequence reads (data not shown).

Similarity searches of the GenBank with these PCR product sequences as queries indicated that all the 23 sequences were most similar to rice gene *Xa21* (Song et al. 1995) and many other *Xa21*-like genes from *Arabidopsis* and rice genome projects (expect values: lower than 3e-41) (Table 1). The sequences were translated into polypeptides without stop codons. These sequences fell into five groups in ClustalX analysis (see below) and they seem to represent a much greater diversity than the 29 sequences amplified and cloned from the USDA 17-47 genomic DNA.

Sequence alignment and characterization

Multiple rounds of sequence comparisons and alignments were conducted using the deduced amino-acid sequences of the 29 *Xa21*-like clones from genomic DNA and the 23 PCR products from BACs. Before the comparisons and alignments, sequences corresponding to the degenerate primers (kindF1 and kindR1) in plasmid clones, or primer A2LW in PCR products, were removed. Computer programs GCG 'PileUP' and ClustalX were used, and

Table 2 Percent amino-acid identities and similarities among five groups (CRK1–5) of citrus receptor-like kinase class RGCs. The values were calculated as described in Table 1

Group	CRK1	CRK2	CRK3	CRK4
CRK2	77.7 (83.1)			
CRK3	76.0 (81.9)	85.4 (86.0)		
CRK4	78.1 (83.0)	77.8 (82.2)	74.3 (79.3)	
CRK5	73.8 (79.6)	75.5 (79.7)	76.7 (82.3)	77.1 (82.7)

both produced similar results. The 53 sequences were grouped into five clusters (data not shown). These clusters were designated as groups CRK (citrus receptor-like kinase) 1–5 (Table 2). The percent amino-acid identity values between the groups were generally lower than 79% (Table 2), except for that between groups CRK 2 and CRK3, which shared more than 85% amino-acid identity. Within each group, 2–4 subgroups were recognized (Fig. 4 and Table 1). The exception to this was groups CRK4 and 5, which contained one member each (Table 1). The percent amino-acid identity values within subgroups were generally greater than 95%, whereas the identity values between subgroups within each cluster were around 84%–91%. Some of the sequences within the same subgroups were identical or highly similar (more than 98% amino-acid identity).

Multiple sequence alignments and clustering analyses put the 29 *Xa21*-like sequences amplified from the USDA 17-47 genomic DNA into two groups and four subgroups; A2 and the other 18 clones in CRK1a, B10 in CRK1b, E3 in CRK1c, and A1 and the other eight clones in CRK2b. Clones A2 and A1 were the representatives of the most-abundant types identified from genomic amplification. The 23 sequences amplified from BACs were clustered into five groups and eight subgroups. Except for subgroup CRK1a that was shared, the other subgroups were uniquely represented by BAC-derived sequences. This distribution seems to indicate the presence of a much-greater diversity among the BACs that produced PCR products with primers A2UP and A2LW.

Phylogenetic analysis was performed to evaluate further the relationship of citrus sequences to rice resistance gene *Xa21* and two other closely related plant receptor-like kinases (RLKs). Xa21 belongs to one of the five plant RLK classes that possess extracellular leucine-rich repeats (Song et al. 1995; Pastuglia et al. 1997). This class of RLK also includes *Arabidopsis TMK1* for transmembrane kinase 1, for which functions remain to be known (Chang et al. 1992), and the petunia *PRK1* for pollen receptor kinase 1 (Mu et al. 1994). Eleven representative citrus sequences and the corresponding regions of the kinase domains of Xa21, TMK1 and RPK1 were included in the analysis; Fig. 5 shows one typical NJ (neighbor-joining) tree derived from multiple sequence alignment. Most of the nodes were supported by more than 600 bootstrap values. It is evident that the citrus sequences had greatest similarities to Xa21 among the three closely related RLKs. The percent amino-acid identity and similarity values of citrus sequences to

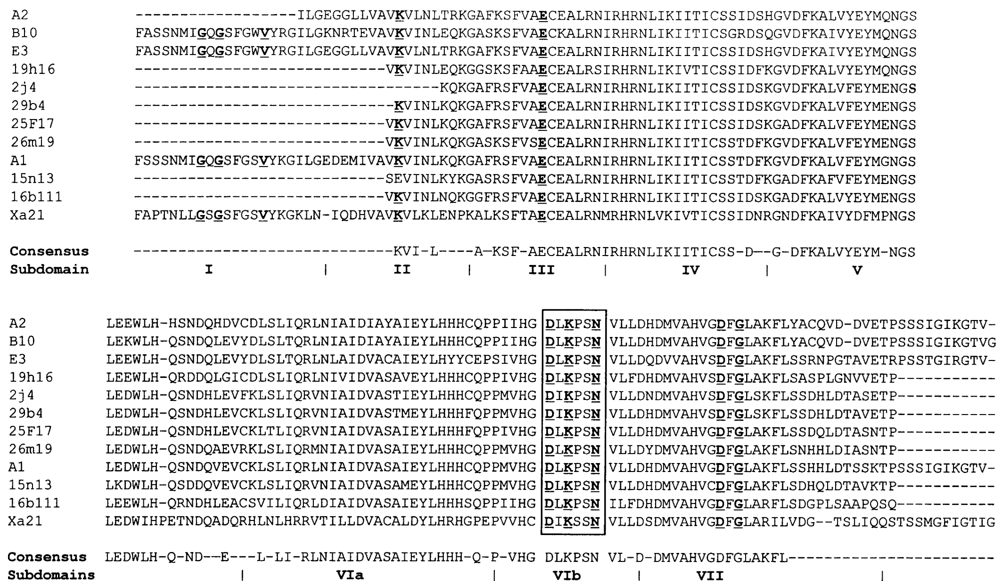

Fig. 4 Alignment of the deduced amino-acid sequences of 11 citrus receptor-like kinase class RGCs and the kinase domain of rice Xa21 (amino acids 708–888) (Song et al. 1995). The computer program CLUSTALX (Thompson et al. 1997) was used in the alignment analysis. Below the alignments is the consensus sequence conserved among rice Xa21 protein and the citrus RGCs. Six to seven subdomains could be recognized based on the available sequences and are indicated in *Roman numerals*. The invariant amino acids present in all eukaryotic protein kinases (Hanks et al. 1988) are indicated in *bold*, *underlined letters*. The *boxed portions* in subdomain VIb is one of the two motifs for serine-threonine specificity

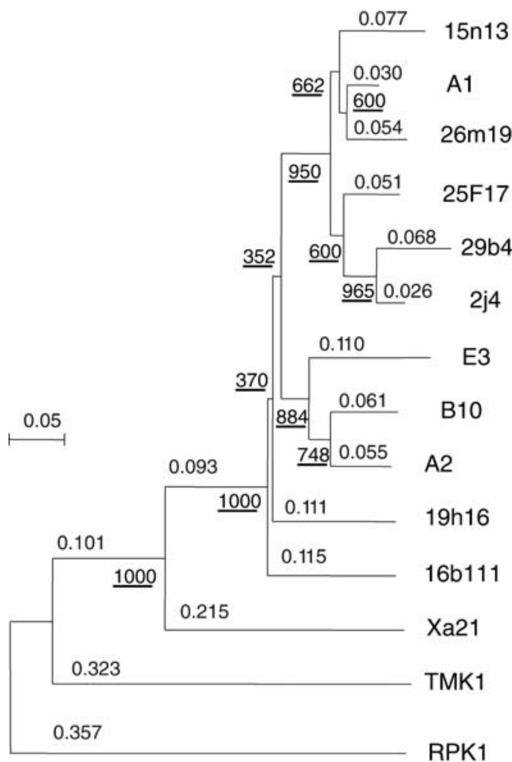

Fig. 5 Phylogenetic tree based on alignment of the deduced amino-acid sequences of 11 *Xa21*-like citrus RGCs and the kinase domains of rice *Xa21* (Song et al. 1995), *Arabidopsis TMK1* (Chang et al. 1992), and petunia *PRK1* (Mu et al. 1994). The tree was constructed using the neighbor-joining method in the program CLUSTERALX. Branch lengths and bootsrap values (*in boldface and underlined*) are indicated

Xa21, calculated using the GCG 'GAP' program, ranged from 54.9% to 59.7%, and from 65.5% to 71.3%, respectively (Table 1).

Sequence alignments in Fig. 4 revealed that citrus sequences contained all the conserved residues (invariant amino acids) found in the corresponding subdomains II, III, VIb and VII of eukaryotic protein kinases (Hanks et al. 1988). The consensus citrus sequence, generated from the alignment, shared with Xa21 the same patterns of alternating high and low conservation and highly conserved residues in available subdomains (II to VII). The serine-threonine kinase specificity of Xa21 is determined by 15 conserved residues in subdomain VIb and VIII (Song et al. 1995). Though the current citrus sequences did not extend to subdomain VIII, they did contain the amino-acid sequences in subdomain VIb (DLKPSN) that are indicative of serine-threonine specificity.

Full-length gene-structure comparison

To demonstrate the feasibility of cloning receptor-like kinase class *R* gene candidates and to obtain full-length gene structures, sequences downstream and upstream of the kinase domains were obtained from BAC clones 17o6 and 26m19 that were identified from the library screening. Approximately 6-kb of double-stranded sequence was obtained from each of the clones, based on primer walking on the large BAC insert DNA. Computational analysis, using a number of gene identification programs such as GenScanW (http://genes.mit.edu/; Burge and Karlin 1997), revealed one large open reading frame

Table 3 Comparison of two RLK class citrus sequences to the rice resistance gene *Xa21* at the DNA and protein levels

Type	Xa21	17o6RLKP	26m19RLKP
DNA structure	One intron	One intron[a]	One intron[a]
Peptide length	1,025 aa	1,019 aa	1,020 aa
Amino-acid identity			
Overall		44.2 (53.1)	43.7 (52.9)
LRRs		45.2 (52.6)	43.2 (50.5)
Kinase		51.9 (63.7)	49.5 (63.1)
Domains			
Signal peptide	P [b]	P	P
Unknown function	P	P	P
LRRs	544 aa 23 repeats	544 aa 23 repeats	553 aa 23 repeats
Charged	P	P	P
Transmembrane	P	P	P
Charged	P	P	P
Juxtamembrane	P	P	P
Kinase	297 aa 12 subdomains	300 aa 12 subdomains	300 aa 12 subdomains
COOH-terminal tail	S/T kinase P	S/T kinase P	S/T kinase P

[a] Introns were predicted by gene identification computer programs
[b] P: Presence of the corresponding domain

(ORF) in the DNA sequences derived from 17o6 and 26m19. Each of the ORFs was interrupted by one putative intron, similar to what was observed for the *Xa21* gene in rice (Song et al. 1995). Amino-acid sequences deduced from the ORFs were designated as 17o6RLKP and 26m19RLKP, respectively. BLAST similarity searches showed that they were most-similar to rice Xa21 protein, with expected values = 0. Sequence comparison revealed that both putative protein sequences contained all the nine domains of rice Xa21, including a signal peptide, extracellular LRRs, transmembrane and kinase domains (Table 3). Overall, 17o6RLKP and 26m19RLKP shared approximately 44% amino-acid identity and 51% similarity with rice Xa21. The similarity was particularly high at the kinase domain: approximately 50% identity and 63.5% similarity. They contained all the 12 subdomains and the serine-threonine-specifying motifs of rice Xa21 and the 15 invariant amino acids of protein kinases (data not shown). The deduced LRR domains in 17o6RLKP and 26m19RLKP consisted of 23 copies of imperfect repeats (the same as Xa21), and they showed approximately 44% amino-acid identity and 51% similarity to the corresponding domain of Xa21 (Table 3).

Discussion

A PCR approach to clone RLK class-resistance gene candidates

Recent studies show that plants seem to use just a few classes of gene products for resistance to various bacterial, fungal and viral pathogens, and nematodes (Hammond-Kosack and Jones 1997; Staskawicz et al. 2001). Rice *Xa21* represents a unique class (a receptor-like kinase) of plant *R* genes (Song et al. 1995; Ronald 1997). Few of this class of *R* genes have been cloned and characterized in other plants, except in cultivated rice and its relatives, and *Arabidopsis*. This hinders our further understanding of this class of *R* genes in plant disease resistance.

Homology-based PCR approaches have been used to clone hundreds of NBS-LRR class *R*-gene sequences and to identify molecular markers for resistance gene loci in many different plant species (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). These sequences are valuable resources for using the candidate gene approach to clone functional plant *R* genes (Pflieger et al. 2001). It will be useful to develop a similar PCR-based approach to clone RLK-class RGC sequences. By aligning the kinase domains of rice *Xa21* and tomato *Pto*, we identified two well-conserved regions. Degenerate primers designed from these regions amplified citrus DNA fragments that are most similar to the rice *Xa21* gene. A majority of the amplified fragments contained open reading frames and could be translated into polypeptides without stop codons. The deduced peptide sequences contained all the features in the corresponding region of rice Xa21 protein, including kinase subdomains, the serine-threonine specificity motif and the invariant amino acids. Phylogenetic analyses show that these sequences have the greatest relatedness with Xa21 among related receptor-like kinases. Sequence information upstream and downstream of the kinase domains from BAC clones demonstrated further that they contained all the features of rice Xa21, including a signal peptide, LRRs, transmembrane and kinase domains. These lines of evidence seem to indicate that the PCR-based homology approach using the two degenerate primers could allow effective targeting onto *Xa21*-like sequences in complex plant genomes that may contain several hundreds of protein kinases (Stone and Walker 1995).

These primers and the PCR approach described above are expected to be applicable to many other plants due to the conservation of *R*-gene sequences and structures among plants (Hammond-Kosack and Jones 1997). Each plant genome may contain hundreds of potential *R*-gene sequences (Michelmore 1996). Isolating these sequences one by one can be time-consuming. Quick cloning of numerous *R*-gene-like sequences can provide important resources for tagging and understanding plant *R* genes (Michelemore 1996). This seems particularly important in several groups of plants, such as woody perennial species (many fruit crops and forest species) and polyploid species (such as cultivated strawberry and banana). Because of the long generation times, large tree sizes, or complex genetic behaviors, genetic mapping and map-based cloning of disease resistance genes could not be readily performed in these species. Genome-wide sequencing of such plants would not be conducted likewise.

BAC resources facilitate cloning and characterization of *R*-gene sequences

The majority of the reported NBS-LRR class RGCs are short sequences (about 500 bp). Several problems are encountered in using these sequences for marker development, genetic mapping, or structural and/or functional analysis. It was difficult designing specific PCR primers for many of the short cloned RGC sequences and identifying polymorphism that could be used for mapping them in relation to resistance traits or gene loci. To overcome these problems, we made extensive use of large-insert genomic resources such as BAC clones and libraries that have been developed recently. Specifically, BACs facilitated *R*-gene cloning and characterization in a number of ways. For example, BACs 17o6 and 26m19 allowed recovery of full-length gene structures and sequence information upstream and downstream of coding regions. In other cases, we have used BAC clones to identify promoters and other regulatory sequences for *R*-genes (unpublished data). By fingerprinting positive BAC clones and assembling BAC contigs, estimations of gene copy number could be made. Sequencing and mapping BAC insert ends allowed localization of *R*-gene sequences. Our experience indicates that integration of BAC and other genomic resources can greatly facilitate not only cloning of *R*-gene sequences, but also characterization of them. In several plants, physical maps and assembled BAC or TAC contigs are available (http://hbz.tamu.edu/; http://genome.clemson.edu). They provide excellent resources to overcome the problems associated with short RGC sequences.

Citrus RLK sequences and their features

Genome-wide sequencing in a number of model plants has shown that each plant may contain hundreds of *R*-gene sequences. Therefore, it is necessary to clone as many potential *R*-gene sequences as possible for use of reverse genetics approaches to isolate and characterize agriculturally important *R* genes, and for marker development, gene tagging, marker-assisted selection and application of a candidate gene approach. Cloning *R*-gene sequences was not possible in citrus (and many other plants) until recently (Fang et al. 1998; Deng et al. 2000, 2001a). Previously, we isolated ten classes of citrus sequences similar to the NBS-LRR class *R* genes and identified approximately 77 contigs that may contain 150 copies of NBS-LRR *R*-gene sequences (Deng et al. 2001a). In this study, we obtained 52 sequences that are most similar to rice *Xa21*. These sequences belong to five groups and eight subgroups. From 79 positive BAC clones, 35 contigs were assembled. Each contig may correspond to one genetic locus, and in general may contain 1–3 copies of *Xa21*-like sequences. Assuming an average of two copies per contig, the 35 contigs contain approximately 50–70 copies of *Xa21*-like sequences. One of the features observed in these BAC clones was the clustering of *Xa21*-like sequences in the BACs, or in the citrus genome. This phenomenon has been seen with NBS-LRR class of *R*-gene sequences in the citrus genome (Deng et al. 2001a). Further analysis of the *Xa21*-like sequences within these BAC contigs should provide important information regarding *R*-gene distribution and evolution in citrus. In addition, BAC end sequences are being obtained so that specific primers can be designed for each contig, and linkage of each contig with resistance traits can be examined. It will be particularly interesting to test any possible association of these *Xa21*-like sequences with resistance to citrus canker, one of the most important bacterial diseases in citrus (Gottwald et al. 2002). This disease is caused by *Xanthomonas axonopodis* pv *citri*, a bacterium closely related to *Xanthomonas oryzae* pv *oryzae* (*Xoo*), the pathogen in the rice *Xa21* gene and the *Xoo* avirulence gene-interaction system.

Acknowledgements The authors thank Mr. S. Huang, Ms. T. Guo, Ms. M. Wendell, Ms. Y. Wang and Ms. L. Wang for their excellent technical support and assistance with the research. This project was supported in part by grants from the USDA NRICGP (#9600748), USDA/National Citrus Research Council (#98012205), the Florida Citrus Production Research Advisory Council (#942–27), and a USDA-ARS/UF-IFAS Specific Cooperative Agreement. All experiments described in this paper comply with the current laws of the United States of America. Florida Agricultural Experiment Station Journal Series No R-09244.

References

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

Altschul SF, Thomas L, Madden A, Schäffer A, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Chang C, Schaller GE, Patterson SE, Kwok SF, Meyerowitz EM, Bleecker AB (1992) The TMK1 gene from *Arabidopsis* codes for a protein with structural and biochemical characteristics of a receptor protein kinase. Plant Cell 4:1263–1271

Deng Z, Huang S, Ling P, Chen C, Yu C, Weber CA, Moore GA, Gmitter FG Jr (2000) Cloning and characterization of NBS-LRR class resistance gene-candidate sequences in citrus. Theor Appl Genet 101:814–822

Deng Z, Tao Q, Chang YL, Huang S, Ling P, Yu C, Chen C, Gmitter FG Jr, Zhang H-B (2001a) Construction of a bacterial artificial chromosome (BAC) library for citrus and identification of BAC contigs containing resistance gene candidates. Theor Appl Genet 102:1177–1184

Deng Z, Huang S, Ling P, Yu C, Tao Q, Chen C, Wendell MK, Zhang H-B, Gmitter FG Jr. (2001b) Fine genetic mapping and BAC contig development for the citrus tristeza virus resistance gene locus in *Poncirus trifoliata* (Raf.). Mol Genet Genomics 265:739–747

Fang DQ, Federici CT, Roose ML (1998) A high-resolution linkage map of the citrus tristeza virus resistance gene region in *Poncirus trifoliata* (L.) Raf. Genetics 150:883–890

Gottwald TR, Graham JH, Schubert TS (2002) Citrus canker: the pathogen and its impact. Online. Plant Health Progress doi:10.1094/PHP-2002–0812–01-RV

Gmitter FG, Jr, Xiao S, Huang S, Hu XL, Garnsey SM, Deng Z (1996) A localized linkage map of the citrus tristeza virus resistance gene region. Theor Appl Genet 92:688–695

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48:575–607

Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42–52

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

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

Ling P, Duncan LW, Deng Z, Dunn D, Hu X, Huang S, Gmitter FG Jr. (2000) Inheritance of citrus nematode resistance and its linkage with molecular markers. Theor Appl Genet 100:1010–1017

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *Rpp8* locus of *Arabidopsis*. Plant Cell 10:1861–1874

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

Michelmore R (1996) Flood warning—resistance genes unleashed. Nature Genet 14:376–378

Mu JH, Lee HS, Kao T-H (1994) Characterization of a pollen-expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. Plant Cell 6:709–721

Pastuglia M, Roby D, Dumas C, Cock JM (1997) Rapid induction by wounding and bacterial infection of an *S* gene family receptor-like kinase gene in *Brassica oleracea*. Plant Cell 9:49–60

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291

Ronald PC (1997) The molecular basis of disease resistance in rice. Plant Mol Biol 35:179–196

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Gardner J, Wang B, Holsten TE, Zhai WX, Zhu LH, Fauquet C, Ronald PC (1995) A receptor kinase-like protein encoded by the rice disease resistance gene *Xa21*. Science 270:1804–1806

Staskawicz BJ, Mudgett MB, Dangl JL, Galan JE (2001) Common and contrasting themes of plant and animal diseases. Science 292:2285–2289

Stone JM, Walker JC (1995) Plant protein kinase families and signal transduction. Plant Physiol 108:451–457

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Wang GL, Ruan DL, Song WY, Sideris S, Chen L-L, Pi L-Y, Zhang S, Zhang Z, Fauquet C, Gaut B, Ronald P (1998) Xa21D encodes a receptor-like molecule with a leucine rice repeat domain that determines race specific recognition and is subject to adaptive evolution. Plant Cell 10:1–5

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

Zhang H-B, Choi SD, Woo SS, Li ZK, Wing RA (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombination inbred mapping population. Mol Breed 2:11–24