A Cluster of Four Receptor-Like Genes Resides in the *Vf* Locus That Confers Resistance to Apple Scab Disease

Mingliang Xu and Schuyler S. Korban[1]

Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, Illinois 61801

Manuscript received July 12, 2002
Accepted for publication September 25, 2002

ABSTRACT

The *Vf* locus, derived from the crabapple species *Malus floribunda* 821, confers resistance to five races of the fungal pathogen *Venturia inaequalis*, the causal agent of apple scab disease. In our previous research, the *Vf* locus was restricted to a BAC contig of ~290 kb covered by five overlapping BAC clones. Here, we report on cloning of the resistance gene(s) present in the *Vf* BAC contig using a highly reliable and straightforward approach. This approach relies on hybridization of labeled cDNAs to amplified inserts of subclones derived from BAC inserts, followed by recovery of full-size transcripts by rapid amplification of cDNA ends (RACE). A cluster of four resistance paralogs (*Vfa1*, *Vfa2*, *Vfa3*, and *Vfa4*) was identified in the *Vf* locus. *Vfa1*, *Vfa2* and *Vfa4* had no introns and are predicted to encode proteins characterized with extracellular leucine-rich repeats (LRRs) and transmembrane (TM) domains. However, *Vfa3* contains an insertion of 780 bp at the end of the LRR motif, resulting in multiple truncated transcripts. Comparison of *Vfa1*, *Vfa2*, and *Vfa4* paralogs revealed a high degree of overall homology in their deduced amino acid sequences, while divergences were mainly restricted within LRR domains, including variable LRR units, numerous amino acid substitutions, and several residue deletions/duplications. Differential expression profiles among the four paralogs were observed during leaf development. *Vfa1*, *Vfa2*, and *Vfa3* were active in immature leaves, but slightly expressed in mature leaves, while *Vfa4* was active in immature leaves and was highly expressed in mature leaves.

MORE than 30 disease resistance (R) genes have been isolated and characterized from a wide range of plant species (Hulbert *et al.* 2001). Numerous R genes are predicted to encode receptors that function in the recognition of corresponding "ligands" controlled by specific dominant avirulence (*Avr*) genes present in pathogens (Hammond-Kosack and Jones 1997). This leads to an active plant defense response, previously described by Flor (1956) as gene-for-gene interaction. Most R genes are organized as complex gene families clustered within particular chromosomal regions (Wise 2000). Although isolated R genes confer resistance to a wide range of pathogens (fungi, viruses, bacteria, and nematodes), they share various conserved functional domains and fall into several distinct classes (Richter and Ronald 2000; Wise 2000). The most prevalent class is the NBS-LRR family that contains a nucleotide-binding site (NBS) and multiple leucine-rich repeats (LRRs) (Young 2000). The tobacco *N* gene; the Arabidopsis *RPS2*, *RPM1*, and *RPP5* genes; and the flax *L6* and *M* genes belong to this class (Hammond-Kosack and Jones 1997). The tomato *Cf* genes represent the second class, which is characterized by extracytoplasmic LRRs and a C-terminal membrane anchor (Dixon *et al.* 1996; Parniske *et al.* 1997). The tomato *Pto* gene, conferring resistance to a bacterial pathogen, falls into a third class that encodes a serine/threonine protein kinase, but lacks LRRs (Martin *et al.* 1993) and requires an NBS-LRR gene (*Prf*) to function (Salmeron *et al.* 1996). The rice *Xa21* gene, conferring resistance to a bacterial pathogen, is regarded as the fourth class, featuring both extracellular LRRs and a transmembrane protein kinase (Song *et al.* 1995).

On the basis of the LRR protein structure model, a consensus sequence, xxLxLxx, within the LRR is predicted to form a β-strand/β-turn structure (Jones and Jones 1997). The aliphatic residues (L) project into the hydrophobic core, whereas the other residues (x) form a solvent-exposed surface that presumably makes direct contact with elicitors of a specific pathogen, leading to recognition of that pathogen (Kobe and Deisenhofer 1995). Comparison of nucleotide sequences among closely related gene homologs has revealed that solvent-exposed residues are hypervariable (Parniske *et al.* 1997). The ratio of nonsynonymous to synonymous nucleotide substitutions (K_A:K_S ratio) of solvent-exposed residues is greater than one, thus indicating that these residues must have undergone adaptive evolution (Michelmore and Meyers 1998). The ever-changing population of a pathogen imposes a selection pressure to continually alter recognition specificity (Bergelson *et al.* 2001). In flax, domain swaps between alleles of the *L6* gene have demonstrated that the LRR is an important deter-

[1]*Corresponding author:* Department of Natural Resources and Environmental Sciences, University of Illinois, 310 Madigan Bldg., 1201 W. Gregory Dr., Urbana, IL 61801. E-mail: s-korban@uiuc.edu

minant of specificity (Ellis *et al.* 1997). Likewise, domain swaps between *Pto* and a closely related paralog, *Fen*, have identified a few amino acids in the ligand-binding domain that are critical for pathogen specificity (Scofield *et al.* 1996). Recent experiments with the rice resistance gene *Pi-ta* and its *Avr* gene show specific binding between the LRR domain of the Pi-ta protein and the cognate *Avr-Pi-ta* protein (Jia *et al.* 2000). Active R genes with differential expression are involved in durable resistance of the *Cf-9* locus to *Cladosporium fulvum*, whereby *Cf-4* and *Cf-9* genes are active from the seedling stage onward, while *Hcr9s*, *-9E*, and *-9A* or *-9B* are detectable in adult plants only (Parniske *et al.* 1997).

Apple scab, incited by the fungal pathogen *Venturia inaequalis* (Cke.) Wint., is one of the most serious diseases of apple (Bénaouf and Parisi 2000). Almost all apple cultivars grown commercially around the world are susceptible to this disease. Resistance to this disease was derived from a wild species of apple, *Malus floribunda* 821, and the chromosomal region containing a scab-resistance locus, *Vf*, has been widely introgressed into susceptible commercial apple cultivars (Korban 1998). A bacterial artificial chromosome (BAC) contig covering the *Vf* locus has been constructed and consists of a minimal tilling path of three BAC clones (Xu and Korban 2002).

Identifying putative coding sequences in a BAC contig containing gene(s) of interest is critical in a map-based cloning strategy. Traditional available methods in pursuing this goal include the following: (1) direct screening of a cDNA library (Elvin *et al.* 1990), (2) cDNA selection by hybridization (Lovett *et al.* 1991; Parimoo *et al.* 1991), (3) the CpG island approach (Cross *et al.* 1999), and (4) the exon-trapping method (Buckler *et al.* 1991). More recently, sequencing an entire BAC insert and predicting open reading frames are becoming more common (Zhang 1997). However, no single method is capable of identifying all transcribed sequences from a BAC contig. Frequently, a combination of methods is required to construct transcript maps for a defined chromosomal region (Sood *et al.* 2001). We present herein an efficient and straightforward approach for identifying putative transcribed sequences contained within BAC inserts. This approach is based on hybridization of a labeled cDNA population to an array of amplified inserts of subclones constructed from BAC inserts containing *Vf* gene(s). A high signal-to-noise ratio in hybridization allows pinpointing positive subclones containing putative coding sequences. Sequencing inserts of positive subclones and comparing deduced amino acid sequences with those available in the GenBank database have revealed the existence of putative resistance genes within the *Vf* locus. A full-length cDNA sequence is then obtained, using rapid amplification of cDNA ends (RACE). This sequence information allows further cloning of genomic gene(s) from BAC clones containing *Vf* gene(s).

MATERIALS AND METHODS

Subcloning of BAC inserts into the plasmid vector pBlueScriptII-KS: A BAC contig spanning the *Vf* locus contained five overlapping BAC clones (*M4-P-11*, *M61-M-16*, *G11-J-23*, *G7-C-18*, and *G53-N-7*; Figure 1). The BAC clone *M61-M-16*, located in the middle of the BAC contig, was selected to construct a subclone library. BAC DNA was extracted from a 2-liter culture using the Plasmid Maxi kit (QIAGEN, Valencia, CA), followed by purification using CsCl-gradient centrifugation (Sambrook *et al.* 1989). A total of 4 μg of purified BAC DNA was partially digested with the restriction enzyme *Sau*3AI. The reaction was performed at 37° for 15 min in a total volume of 150 μl, using 3 units of *Sau*3AI. Digested fragments of ~2 kb in size were harvested from a 1% agarose gel using a QIAEX II gel extraction kit (QIAGEN) and then ligated into a *Bam*HI-digested pBlueScriptII-KS vector. A total of 1 μl of ligation product was used to transform competent *Escherichia coli* cells by electroporation using a Bio-Rad (Richmond, CA) gene pulser. Transformed cells were grown onto Luria broth (LB) plates containing ampicillin (100 μg/ml), X-gal, and isopropyl thiogalactoside. White colonies were picked and placed in a grid onto fresh LB plates. Following an overnight incubation at 37°, each colony was dissolved in 30 μl TE buffer, heated at 100° for 10 min, and centrifuged at 4000 rpm for 20 min. A pair of PCR primers (T3 and T7 promoters), flanking the multiple cloning site of pBlueScriptII-KS, was introduced to amplify inserts of subclones using their DNA-containing supernatants as template DNAs. A total of 1 μl of the PCR product was denatured, using an equal volume of 1 N NaOH, and placed in a grid onto a nylon membrane for hybridization. A total of 96 (12 × 8) dots were arrayed on a single membrane.

cDNA synthesis from apple leaf tissues: Fresh leaf tissues of a *Vf*-containing apple cultivar, "GoldRush," were sprayed with 2 mM salicylic acid and ddH_2O, respectively. Salicylic acid is known to activate resistance gene(s) and its downstream defense system in plant tissues. Therefore, it is used in this study to induce transcripts for disease resistance in apple leaf tissues of GoldRush. Following a 24-hr incubation at room temperature, leaves from both treatments were harvested and stored at −80° in an ultra-freezer. Total RNA was extracted from 2 g of leaf tissue using the LiCl method (Sambrook *et al.* 1989), and mRNA was isolated using a Poly(A) Tract mRNA Isolation System III (Promega, Madison, WI). The isolated mRNA was then used to synthesize cDNA, using a Marathon cDNA amplification kit (CLONTECH, Palo Alto, CA).

Identification of subclones containing putative transcribed sequences: cDNAs from both salicylic acid- and ddH_2O-treated leaf tissues were mixed and labeled with a ^{32}P isotope, using a Random Primers labeling system (GIBCO BRL, Gaithersburg, MD). Prehybridization, hybridization, and membrane washing were performed according to Sambrook *et al.* (1989). Membranes were exposed to a hyperfilm (Amersham, Buckinghamshire, England) for 3 days at −80°. Positive subclones were then spotted onto the film, and their corresponding recombinant plasmids were extracted using the Mini Preparation kit (QIAGEN). Inserts of positive subclones were sequenced from both ends, using T3 and T7 promoter primers. Analysis of insert sequences was performed for all positive subclones to identify putative transcribed sequences present within the BAC insert. Resistance gene homologs were predicted after comparing deduced amino acid sequences with the GenBank database using BLAST.

PCR-based search for additional *Vf* homologous sequences: A two-step PCR-based strategy was pursued. First, a pair of common primers was designed on the basis of conserved regions among the *Vf* gene homologs found in the BAC clone *M61-M-16*. The common primers were then used

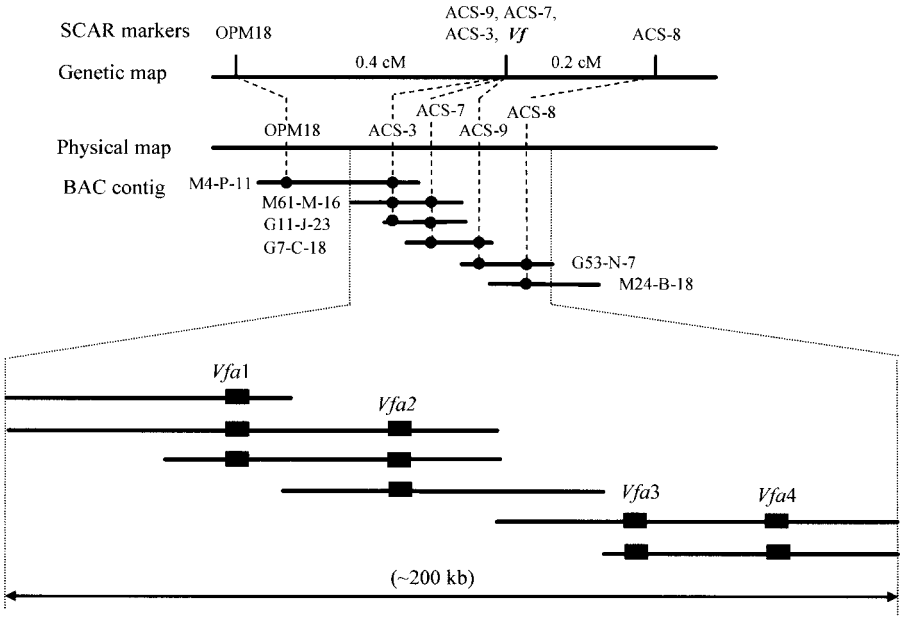

FIGURE 1.—A 200-kb BAC contig of the *Vf* locus and schematic representation of the four *Vf* gene paralogs.

to amplify all five BAC clones spanning the *Vf* locus, along with their neighboring BAC clones, to search for additional *Vf* homologous sequences present in the *Vf* region. PCR products were cloned into pBlueScriptII-KS, followed by extensive investigation of fingerprints of cloned PCR products using an array of restriction enzymes, including *Eco*O109I, *Msp*I, *Mob*I, *Hin*fI, *Hin*cII, *Sau*3AI, and *Taq*I. Several fingerprinting patterns were observed, and for each fingerprinting pattern, at least two clones were selected for sequencing. Second, a pair of nested common primers was designed on the basis of the newly identified conserved sequences and used to further amplify all BAC clones in the *Vf* region. Cloning and analysis of PCR products were performed as described in the first step. Using this two-step PCR-based search approach, all putative *Vf* homologous sequences in the *Vf* region were recovered.

It is highly likely that PCR artifacts are generated and cloned during the PCR-based search strategy. Thus, clones harboring PCR artifacts must be ruled out; otherwise, they will interfere with identifying authentic *Vf* homologous sequences. This can be achieved by a simple PCR reaction. A pair of specific primers for each putative *Vf* homologous sequence is designed on the basis of its polymorphic nucleotides and used to amplify three sources of template DNAs: the cognate PCR product clone, the cognate BAC clone, and the genomic DNA of GoldRush. A putative *Vf* gene homolog is deemed authentic only if uniform PCR products are present in all three sources of template DNAs, using its specific primers. In contrast, a putative *Vf* gene homolog is considered a PCR artifact when specific primers can amplify only the cognate PCR product clone, but not the cognate BAC clone and the genomic DNA of GoldRush.

Recovery of full-length cDNA using RACE: The RACE method was utilized to construct and characterize cDNAs transcribed by *Vf* gene paralogs, including putative start and termination points for both transcription and translation, sizes of the open reading frames, and leader and trailer sequences. Sequence information from RACE analysis was further used to clone the full-length cDNA and genomic gene for each *Vf* paralog. For RACE analysis, both 3′ and 5′ primers were designed on the basis of conserved regions of the four *Vf* paralogs and with an overlapping region of ~1.5 kb. A mixture of cDNAs from both immature and mature leaves of GoldRush was used as template DNA to amplify both 3′ and 5′ cDNA ends. RACE products of expected sizes were cloned into a pT-Adv vector. Therefore, both 3′ and 5′ RACE clones were a mixture of RACE clones including all active *Vf* gene paralogs. Identifying a RACE clone containing an individual *Vf* gene paralog was accomplished using *Vfa*-specific primers.

Expression profiles of *Vf* gene paralogs revealed by reverse transcription-PCR: Three sources of cDNAs were synthesized from respective mature leaves of a scab-susceptible apple cultivar "Red Delicious" and immature and mature leaves of the *Vf*-containing resistance apple cultivar GoldRush. The nested common primers as well as *Vfa*-specific primers were used to amplify the three sources of cDNAs to determine expression profiles for each of the four *Vf* gene paralogs. Negative controls were also set up separately by using corresponding mRNAs (100 ng each) from these tissues in PCR reactions, utilizing both common and *Vfa*-specific primers, to rule out any genomic DNA contamination prior to running reverse transcription (RT)-PCR. Each of the RT-PCR analyses was repeated a second time to confirm repeatability of results.

Sequencing genomic *Vf* gene paralogs: Genomic genes of four *Vf* paralogs were isolated from two BAC clones, *M61-M-16* and *G7-C-18*. A total of 4 μg of purified BAC DNA was partially digested at 37° for 15 min with 0.5 units of *Sau*3AI in a total volume of 150 μl. The digested fragments of ~10 kb in size were harvested from a 1% agarose gel using the QIAEX II gel extraction kit (QIAGEN) and then ligated into a *Bam*HI-digested pCAMBIA2301 vector. A total of 1 μl of ligation product was used to transform competent *E. coli* cells by electroporation, using a Bio-Rad gene pulser. Both the nested common and *Vfa*-specific primers were used to screen all subclones to find those containing *Vf* gene paralogs, each with a full-length coding sequence, at least 2-kb promoter, and 3′-flanking region. Promoter and coding regions were sequenced for each of the four *Vf* gene paralogs.

RESULTS

Identification of positive subclones containing putative transcribed sequences: PCR amplification of subclone inserts showed that >90% of subclones derived

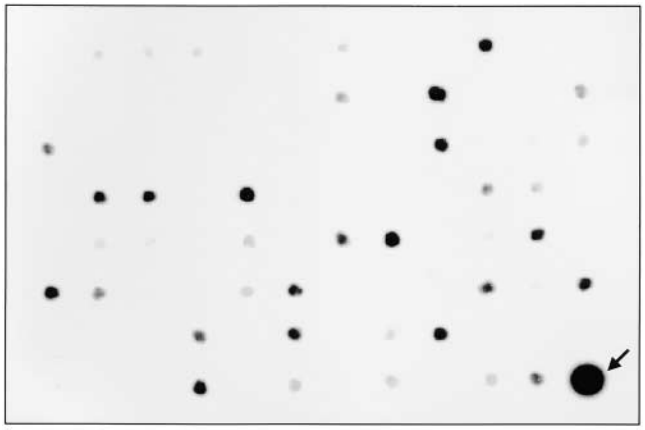

FIGURE 2.—Dot-blot hybridization to pinpoint subclones harboring putative coding sequences. A subclone library is constructed from a *Vf*-containing BAC clone, *M61-M-16*. Inserts of the subclone are amplified using common primers flanking the cloning sites of the pBluescript II-KS. PCR products are denatured and then dot blotted onto membranes (12 × 8 dots per membrane). Hybridization is carried out using labeled cDNAs derived from the *Vf*-containing apple cultivar GoldRush. Positive subclones with very high signal-to-noise ratios have been obtained, and inserts of positive subclones are sequenced to identify putative coding regions. The large dot (indicated by an arrow) corresponds to *M61-M-16* BAC DNA used as control.

from the BAC clone *M61-M-16* contained inserts of ~2 kb in size. A total of 190 amplified inserts were denatured and arrayed onto two hybridization membranes. The insert of the BAC clone *M61-M-16* was estimated to be ~115 kb in size (XU and KORBAN 2002). Therefore, all 190 subclones, with an average insert size of ~2 kb, represented more than three times the BAC insert, thus providing a 95% probability of covering the entire BAC insert. Hybridization of dot-blot membranes with labeled cDNA revealed a total of 30 positive subclones with strong signals (Figure 2). Of these 30 positive subclones, 15 were randomly selected for sequencing. Analysis of insert sequences revealed at least four open reading frames. The deduced amino acid sequences were compared to those found in the GenBank database using BLAST. Two amino acid sequences highly matched some resistance genes. They showed 29% sequence identity and 31% sequence similarity to *Cf2/Cf5* resistance genes in tomato and 31% sequence identity and 35% sequence similarity to some putative disease resistance genes in Arabidopsis, respectively. These two open reading frames were deemed likely *Vf* gene homologs and were designated as *Vfa1* and *Vfa2*. The third open reading frame was similar to two unknown proteins in Arabidopsis, while the fourth open reading frame was likely to be a transposase. Only the two resistance-like genes were subjected to further analysis.

Recovery of all putative resistance gene homologs from the *Vf* region: Sequence alignment between *Vfa1* and *Vfa2* revealed a deletion of 162 bp in the *Vfa2*, a number of polymorphic single nucleotides, and several conserved regions flanking the deletion. A pair of PCR primers was designed on the basis of flanking conserved regions to search for additional putative *Vf* gene homologs along the *Vf* locus (Table 1). Five overlapping BAC clones in the *Vf* locus, along with their neighboring BAC clones, were subjected to PCR amplification. As a result, all five BAC clones and one neighboring BAC clone, *M24-B-18*, gave rise to PCR products. Interestingly, PCR products from different BAC clones varied in both band numbers and sizes. From left to right, *M4-P-11* had a single band of ~1.2 kb; both *M61-M-16* and *G11-J-23* had two bands of ~1.2 and ~1.04 kb in size; *G7-C-18* had a single band of ~1.04 kb; and both *G53-N-7* and *M24-B-18* gave rise to two bands of ~1.04 and 0.8 kb in size. The PCR products from different BAC clones were individually cloned into pBlueScriptII-KS, and restriction fingerprints of cloned PCR products were analyzed with a series of restriction enzymes. A total of five banding patterns were observed, indicating that at least five different PCR products had been generated. At least two clones for each banding pattern were selected for sequencing. Accordingly, five putative *Vf* gene homologous sequences were obtained. Sequence alignment of these five *Vf* homologs revealed some conserved regions that were available for designing a pair of nested common primers (Table 1). The nested common primers were used to amplify the same set of BAC clones. The resulting PCR products were cloned into pBlueScriptII-KS and analyzed using a series of restriction enzymes. Following sequencing and sequence alignment, three new putative *Vf* homologous sequences were identified. In total, eight putative *Vf* gene homologs (types 1–8) were recovered from the two-step PCR-based search strategy.

Sequence alignment showed some interesting features present in these eight putative *Vf* homologous sequences (Figure 3). Types 1 and 3 were derived from *Vfa1* and *Vfa2*, respectively, while type 2 was a chimeric sequence between types 1 and 3. The remaining five types were all derived from BAC clones *G53-N-7* and *M24-B-18*. Of these five putative *Vf* homologous sequences, types 4 and 7 were deemed novel. Types 5 and 6 were chimeric sequences between types 4 and 7, while type 8 shared the same sequence as type 7, except for a single-base-pair difference. Considering that only a small fraction of PCR product clones harbored chimeric sequences, and that all chimeric sequences were derived from those BAC clones containing at least two putative *Vf* homologous sequences (*M61-M-16*, *G11-J-23*, *G7-C-18*, and *M24-B-18*), it was speculated that such chimeric sequences were derived from PCR artifacts. To confirm this, a pair of specific primers was designed for each *Vf* homolog on the basis of its polymorphic nucleotides and used to amplify its cognate PCR product clones, cognate BAC clone, and genomic DNA of Gold-

TABLE 1

Primers used in the discovery of apple scab-resistance *Vf* gene(s)

Primers	Sequence
Vf common primers	
Common primers	For: 5′-ATGGAGAGAACCATGAGAGTTGT-3′
	Rev: 5′-TGTGTCCGCAACCACATTGGCCA-3′
Nested common primers	For: 5′-GGTTCTATGACAAGTTTAACACACC-3′
	Rev: 5′-AAAGGAGGAACCCAATCTCG-3′
Common primers in RACE	For: 5′-CACATCCACGAGCTGCACCTTAAT-3′
	Rev: 5′-AAGGAGGAACCCAATCTCGACTAG-3′
Vfa-specific primers	
Vfa1-specific primers	For: 5′-TCTATCTCAGTAGTTTCTATAATTCC-3′
	Rev: 5′-GTAGTTACTCTCAAGATTAAGAACTT-3′
Vfa2-specific primers	For: 5′-CTCAATCTCAGTAGTTTCTATGGA-3′
	Rev: 5′-CCCCCGAGATTAAGAGTTG-3′
Vfa3-specific primers	For: 5′-ATATTAGTAGTTTCTATAATCTGAAGG-3′
	Rev: 5′-CCCCCGAGATTAAGAGATG-3′
Vfa4-specific primers	For: 5′-TATCTCAATCTCAGTAGTAATAGTATC-3′
	Rev: 5′-GACCTTGGAAACCACAATC-3′

For, forward; Rev, reverse.

Rush. As a result, specific primers for types 1, 3, 4, and 7 amplified their corresponding three sources of template DNAs. This indicated that these four putative *Vf* homologous sequences were present in both BAC clones and genomic DNA of GoldRush. However, for the three chimeric sequences, PCR products were observed only in the cognate PCR product clones and not in cognate BAC clones or in genomic DNA of GoldRush. This indicated that these were derived from PCR artifacts. A single polymorphic nucleotide present in type 8 was likely introduced during the PCR reaction, as only one PCR product clone that harbored the sequence of type 8 was obtained.

Following this extensive PCR-based search and validation strategy, only four *Vf* gene homologous sequences were found clustered within the *Vf* locus. The two novel *Vf* homologs were designated *Vfa3* and *Vfa4*. Since all four *Vf* gene homologous sequences were derived from the *Vf* locus, these were then deemed as *Vf* gene paralogs. The left-most BAC clone *M4-P-11* and the subsequent two clones, *M61-M-16* and *G11-J-*23, contained *Vfa1*, suggesting that *Vfa1* was the left-most *Vf* paralog. This was followed by *Vfa2*, which was present in three BAC clones, *M61-M-16*, *G11-J-23*, and *G7-C-18*. The flanking region of *Vfa3* shared some restriction fingerprints with the BAC clone *G7-C-18*, thus suggesting *Vfa3* was the third *Vf* paralog (data not shown). The *Vfa4* was then located as the right-most *Vf* paralog. Both *Vfa3* and *Vfa4* were present in *G7-C-18* and *M24-B-18* (Figure 1).

Characterization of cDNA sequences of the four *Vf* paralogs using RACE: With partial sequences of the four *Vf* paralogs in hand, 5′ and 3′ *Vf* gene primers, with an overlapping region of ~1.5 kb, were designed for RACE analysis (Table 1). Each primer was 25 bp in length and contained a high G + C content to enhance PCR specificity. Screening of RACE product clones was first carried out using the nested common primers and then using each of the four pairs of *Vf* paralog-specific primers (Table 1). Since the nested common primers and all *Vfa*-specific primers were located within the overlapping region, these primers were available for screening both 5′ and 3′ RACE clones. In this RACE analysis, both 5′ and 3′ RACE clones for each of the four *Vf* paralogs were obtained, suggesting that all four *Vf* paralogs were transcribed. For each *Vf* paralog, both 5′ and 3′ RACE products were sequenced. Putative start codons for the four *Vf* paralogs were identified following sequence analysis of 5′-end cDNA fragments having the largest sizes. Leader sequences, upstream of the putative start codon for the four *Vf* paralogs, were estimated to be ~25 bp. All 3′-end cDNA fragments had typical polyadenylation signals, indicating the fidelity of the RACE reaction. Sizes of 3′ untranslated regions (UTRs) varied not only among different *Vf* paralogs, but also among different 3′ RACE clones of the same *Vf* paralog. For instance, we sequenced seven 3′ RACE clones of *Vfa4*, and sizes of UTRs were found to be 84 bp for one clone, 250 bp for two clones, and 354 bp for the remaining four clones. Generally, sizes of UTRs ranged from ~100 to ~400 bp.

A full-length cDNA sequence was reconstructed for each of the four *Vf* paralogs by merging sequences of 5′- and 3′-end cDNA fragments on the basis of their overlapping regions. The sizes of coding sequences were estimated to be 3048 bp for *Vfa1*, 2943 bp for *Vfa2*, and 2748 bp for *Vfa4*, respectively. *Vfa3* had multiple truncated transcripts resulting from premature termina-

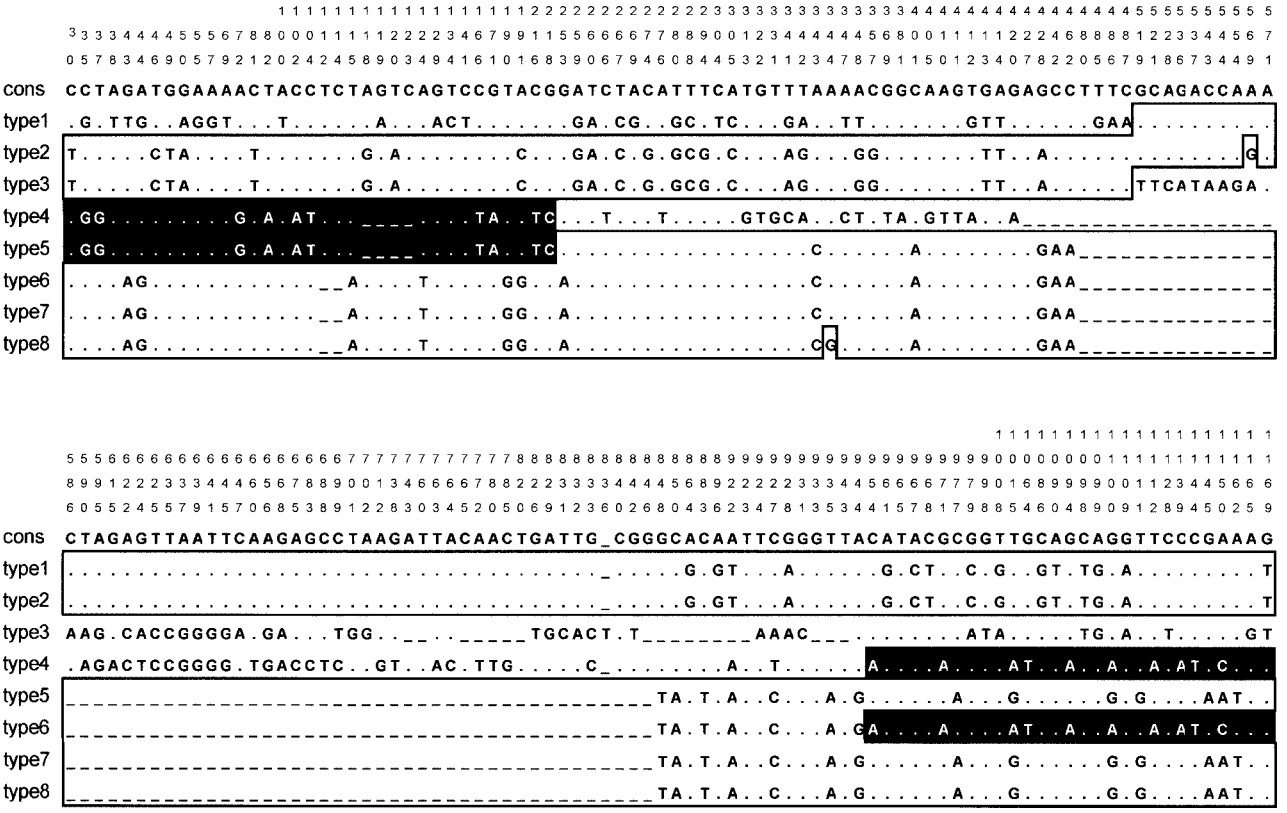

FIGURE 3.—Alignment of eight types of *Vf* gene homologous sequences that have been recovered from the *Vf* locus by using a two-step PCR-based search strategy. Only polymorphic nucleotides are shown. Types 1, 3, 4, and 7 are confirmed to be derived from the *Vf* locus, whereas types 2, 5, 6 (chimeric), and 8 (point mutation) are deemed PCR artifacts. Numbers written vertically along the top of conserved sequences correspond to nucleotide positions. Dots (. . .) indicate identical nucleotides with the consensus sequence. Dashes (- - -) correspond to missing nucleotides. Solid and open rectangles show identical sections among different *Vf* homologous sequences.

tion of transcription. The shortest truncated transcript was only 1220 bp. Moreover, a nonsense mutation was present at 228 bp downstream of the start codon in the *Vfa3* transcripts.

When compared with the database in the GenBank, the deduced amino acid sequences of *Vfa1*, *Vfa2*, and *Vfa4* show high homology to *Cf* resistance genes in tomato and can be subdivided into similar functional domains (Figure 4). Domain A is a putative signal peptide of 23 amino acids and is likely predicted to be removed in the mature protein. Domain B is the presumed NH_2 terminus of the mature protein. Domain C contains multiple N-terminal LRRs. Domain D is of unknown function. Domain F is composed of hydrophobic residues and is likely predicted to be the transmembrane (TM) domain. Domains E and G are rich in acidic and basic residues, respectively, and presumably function in anchoring and directing the transmembrane domain to the cell membrane. Thus, the *Vf* paralogs belong to a LRR-TM resistance gene class and encode proteins characterized with extracellular LRRs and TM domains.

At the amino acid level, a high degree of overall sequence homology was observed among *Vfa1*, *Vfa2*, and *Vfa4*. Sequence identity was 84% for the pair *Vfa1/Vfa2*, 78% for the pair *Vfa1/Vfa4*, and 82% for the pair *Vfa2/Vfa3*. Just like other resistance genes identified in various plant species, amino acid sequence divergence was mainly present in the LRR motifs, including LRR copy numbers, amino acid substitutions, deletions, and duplications (Figure 4). The LRR copy numbers for *Vfa1*, *Vfa2*, and *Vfa4* were 30, 29, and 26, respectively. Several amino acid deletions/duplications were scattered along the LRR motifs. Pairwise comparisons of the three *Vf* paralogs within the LRR motifs revealed 26 sites with three different amino acid residues and 106 positions with two variable amino acid residues. Compared to LRR motifs, C-terminal domains (D, E, and F) were highly homologous among *Vfa1*, *Vfa2*, and *Vfa4*, and only a few amino acid substitutions were observed.

Upon closer examination of the amino acid sequences of the three full-length *Vfa* paralogs, *Vfa1*, *Vfa2*, and *Vfa4*, the 6 amino acids on either side of the first four core sequences (xxLxLxx) show distinct differences (Figure 4). Those amino acids on the N-terminal side are highly conserved (only 1 of 24 amino acids is polymorphic), whereas those on the C-terminal side are highly polymorphic (18 of 24 amino acids are polymor-

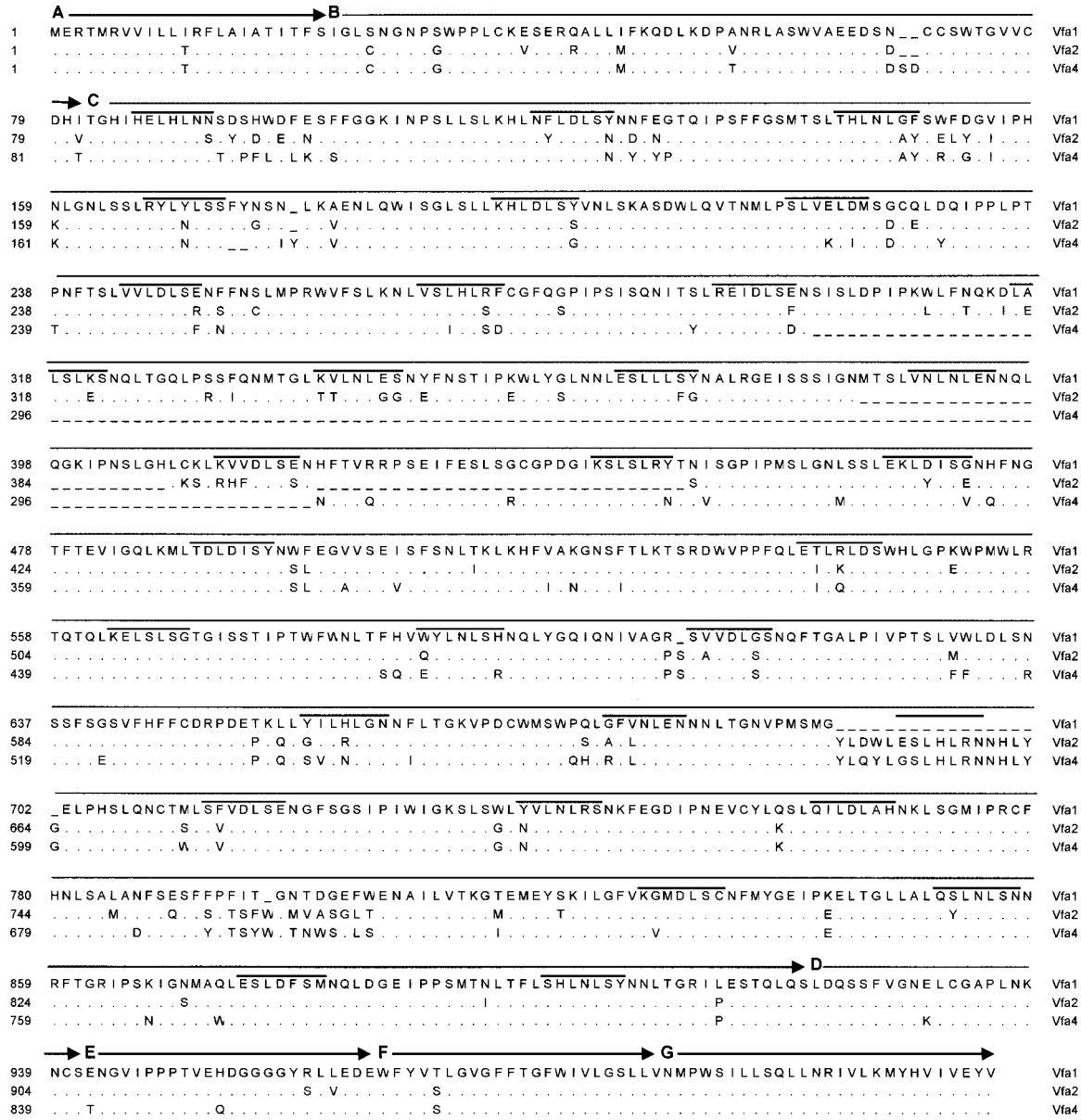

FIGURE 4.—Comparison of amino acid sequences of *Vfa1*, *Vfa2*, and *Vfa4*. The amino acid sequences of the three *Vf* paralogs are divided into seven domains, similar to those of *Cf* genes in tomato. Domain A is a signal peptide; B is the NH_2 terminus of a mature protein; C is an LRR domain, and the core structure of xxLxLxx is highlighted; D is of unknown function; E is rich in acid residues; F is a transmembrane domain; and G is rich in basic residues. A high rate of divergence is present in the LRR domain, including deletion of LRR units, residue deletions/duplications, and amino acid substitutions.

phic). By contrast, when considering the 3 amino acids on either side of core sequences 20–27 (counting from the N-terminal end), it is revealed that those on the N-terminal side contain 12 polymorphic amino acids of 24 residues while those on the C-terminal side contain only 1 polymorphic amino acid among 24 residues (Figure 4). Whether the high level of polymorphism in particular locations in the amino acid sequences of the *Vf* paralogs is the result of adaptive evolution (because the residues are solvent exposed) or is a result of mutations that survived by chance and have a neutral effect (because the altered residues are not solvent exposed) should be further explored.

Expression profiles of the four *Vf* paralogs: RT-PCR was used to investigate expression patterns of the four *Vf* gene paralogs during leaf development. The nested common primer and *Vf* paralog-specific primers were used to amplify cDNAs and their corresponding mRNAs derived from mature leaves of a scab-susceptible cultivar, Red Delicious, as well as from immature and mature leaves of a scab-resistant cultivar, GoldRush. No PCR products were observed in standard PCR reactions (negative controls) using mRNAs as templates, thus indicating no genomic DNA contamination occurred during mRNA preparations.

The nested common primer amplified all three

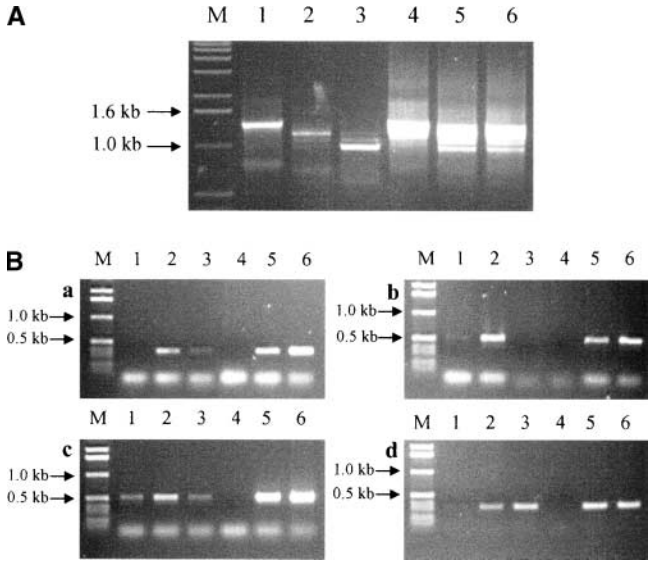

FIGURE 5.—Analysis of expression profiles of the four *Vfa* paralogs by RT-PCR. PCR was performed using either cDNAs (lanes 1, immature leaves of a scab-susceptible apple cultivar Red Delicious; lanes 2, immature leaves of a *Vf*-containing scab-resistant cultivar GoldRush; and lanes 3, mature leaves of GoldRush) or genomic DNA (lanes 4, Red Delicious; lanes 5, GoldRush; and lanes 6, *M. floribunda* 821, the original source of *Vf*). M, 1-kb DNA ladder marker. (A) PCR amplification using the nested common primers, designed on the basis of conserved regions of LRR domains. The sizes of PCR products were expected to be 1199 bp for *Vfa1*, 1037 bp for *Vfa2*, 1049 bp for *Vfa3*, and 836 bp for *Vfa4*. The bulk of RT-PCR products in immature leaf tissue of GoldRush corresponded to *Vfa1* (lanes 2), while most RT-PCR products in mature leaf tissue of GoldRush corresponded to *Vfa4* (lanes 3). RT-PCR products of large sizes were also observed in cDNAs from the scab-susceptible cultivar Red Delicious (lanes 1). (B) PCR amplification using *Vfa*-specific primers. (a) *Vfa1*-specific primers; (b) *Vfa2*-specific primers; (c) *Vfa3*-specific primers; and (d) *Vfa4*-specific primers. *Vfa1*, *Vfa2*, and *Vfa3* are strongly expressed in immature leaf tissues (a–c, lanes 2), but are weakly expressed in mature leaf tissue (a–c, lanes 3) of GoldRush, whereas *Vfa4* is expressed in immature leaf tissue and is strongly expressed in mature leaf tissues of GoldRush (d, lane 3).

sources of cDNAs (Figure 5). As cDNA of Red Delicious gave rise to the largest PCR band, this indicated the presence of some active *Vf* homologous sequences of large sizes in the susceptible cultivar. Interestingly, cDNAs from both immature and mature leaves of GoldRush gave rise to variable PCR products. The bulk of RT-PCR products from immature leaves of GoldRush were estimated to be 1.2 kb in size and corresponded to *Vfa1*, whereas most RT-PCR products from mature leaves of GoldRush were ~0.8 kb in size and corresponded to *Vfa4*. The four *Vf* paralog-specific primers further revealed expression profiles of the four *Vf* gene paralogs. Three *Vf* paralogs, *Vfa1*, *Vfa2*, and *Vfa3*, were expressed in immature leaves, but were only slightly detectable in mature leaves of GoldRush, whereas *Vfa4* was active in immature leaves and was highly expressed in mature leaves (Figure 5). This indicated that *Vf* gene paralogs were differentially expressed during leaf development.

Characterization of four *Vf* gene paralogs: The four genomic *Vf* paralogs were cloned from their corresponding BAC clones, each including at least a 2-kb promoter region, the full-length coding sequence, and 5–12 kb of the 3′-flanking region. Sequencing of the coding and the promoter regions was performed for each *Vf* paralog. Alignment of coding regions between cDNA and genomic sequences revealed that *Vfa1*, *Vfa2*, and *Vfa4* were intronless genes and exhibited a 100% match between cDNAs and genomic genes. Furthermore, putative start and termination codons as well as leader and trailer sequences derived from cDNA analysis were further confirmed using genomic sequences, whereas *Vfa3* had a 780-bp insert with a high A + T content at the end of the LRR motif, and the insert was assumed to cause multiple premature terminations of RNA transcription. The insertion produced 16-bp (5′-AATTTATCGAATAATC-3′) direct repeats of the target sequences flanking the insert, suggesting that the insert was a transposable-like element. No sequences were found to be homologous with this insert following BLAST search with the database in GenBank.

The promoter regions of the four *Vf* paralogs are high homologous for ~300 bp upstream of the putative start codon, but beyond that they show wide divergence (Figure 6). A TATA box is clearly identified 63 bp upstream of the putative start codon. Several nucleotide deletions, duplications, and substitutions are scattered along the promoter regions. The *Vfa1, Vfa2*, and *Vfa3* promoters are very similar in expression profiles, but exhibit a certain degree of nucleotide divergence. Two deletions in *Vfa1* (216 and 147 bp upstream of the start codon, respectively), one nucleotide duplication in *Vfa2* (186 bp upstream of the start codon), and several nucleotide substitutions for all *Vfa1, Vfa2*, and *Vfa3* have been observed. It can be concluded that these sequence divergences in the promoter regions have no significant impacts on their functions. The most striking divergence is a 5′-TCCCT-3′ direct duplication immediately upstream of the TATA box in *Vfa4*. This duplication may be strongly related to the high expression level of the *Vfa4* promoter in mature leaves of GoldRush.

DISCUSSION

Search for putative coding sequences from BAC clones: To pinpoint putative coding sequences within a BAC contig spanning the target gene(s) is a crucial step for map-based gene cloning. Traditional methods often fall short of this expectation and frequently result in delays in cloning of the target gene(s). Identification of positive cDNA clones from a cDNA library usually requires screening of a large number of cDNA clones using labeled BAC inserts. Often, it is difficult to obtain

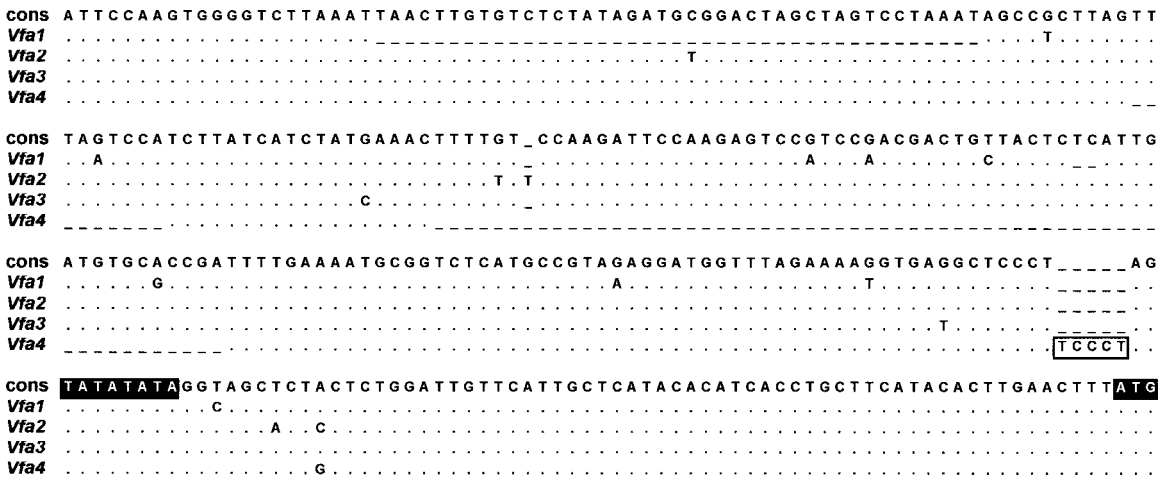

FIGURE 6.—Alignment of promoter sequences (~300 bp upstream of the start codon) for the four *Vfa* gene paralogs. Start codon and TATA box are indicated by solid rectangles. A direct repeat of 5′-TCCCT-3′ is present immediately upstream of the TATA box in *Vfa4* and is presumed to enhance the expression level of *Vfa4* at the mature stage of leaf development.

true positive cDNA clones due to low abundance of target cDNA clones or interference of homologous sequences (ELVIN *et al.* 1990; VINATZER *et al.* 2001). The exon-trapping approach is useful only when a gene contains introns that can be recognized by the host cell (BUCKLER *et al.* 1991). Gene discovery by isolating CpG islands is mainly used in conjunction with other methods as CpG islands are linked to ~60% of human genes (CROSS *et al.* 1999). cDNA selection by hybridization is an expression-dependent method (LOVETT *et al.* 1991; PARIMOO *et al.* 1991) and therefore falls short in constructing a transcript map of a BAC insert. While sequencing the entire BAC insert to predict putative coding regions is an efficient method, it is costly and offers an indirect approach for gene discovery.

The strategy presented in this study is reliable, direct, and quite powerful to pinpoint all potential coding sequences within a BAC insert. A key factor in this strategy is the use of amplified inserts of subclones rather than colonies in hybridization. This modification dramatically reduces the strong background that is often detected in colony hybridization and results in positive clones with much higher signal-to-noise ratios. Sequencing inserts of positive subclones and comparing their deduced amino acid sequences with those in the GenBank database allow rapid identification of transcripts contained within BAC inserts. Like other expression-dependent methods, hybridization using labeled cDNAs involves similar challenges as gene discovery is strongly dependent on the sources of cDNA populations (LOVETT *et al.* 1991; PARIMOO *et al.* 1991). As our aim is to search for *Vf* resistance gene(s) that are expressed in leaf tissue, only cDNAs from the leaf tissue of a *Vf*-containing apple cultivar, GoldRush, are used. For the purpose of constructing a transcription map, all sources of cDNAs or normalized cDNA populations are required. Positive subclones with strong signals have indicated the presence of a large portion of a gene that is highly expressed in the leaf tissue. In contrast, subclones with faint signals contain either a short fragment of a strongly expressed gene or a large fragment of a weakly expressed gene.

To recover all potential *Vf* homologous sequences from the *Vf* locus, a two-step PCR-based screening approach is introduced. This PCR-based strategy is independent of gene expression and therefore provides a complementary approach for revealing all *Vf* homologous sequences. As demonstrated in this study, the PCR-based search strategy has proved to be quite powerful. Not only eight putative *Vf* homologous sequences are identified within the *Vf* locus, but also three other *Vf* haplotypes are recovered in the *M. floribunda* 821 genome, one of which is very close to the *Vf* locus (data not shown). One problem encountered in this PCR-based search is the occurrence of PCR artifacts, as these interfere with identification of authentic *Vf* homologous sequences. This is also a major problem encountered in a resistance gene analog (RGA) search strategy whereby degenerate PCR primers are utilized to identify all possible RGAs spread throughout a whole genome (LEISTER *et al.* 1996). Therefore, it is necessary to validate the authenticity of the *Vf* homologous sequences in the PCR-based search strategy to rule out all PCR artifacts. Our approach of using specific primers to amplify different sources of template DNAs has proved to be effective and highly reliable.

The RACE technique rapidly enables construction of a full-length cDNA sequence based on a partial transcribed sequence. In addition, the usefulness of this technique can be expanded to search for all active members of a gene family and to identify related genes on the basis of conserved sequence information. During cloning of 5′ RACE products, multiple clones of different sizes have been recovered due to premature termination of the first-strand cDNA synthesis. This is an

intrinsic limitation of the RACE technique (Chenchik *et al.* 1995). Therefore, several clones having large 5′ RACE products must be sequenced before the putative start codon can be identified. All 3′ RACE products have typical poly(A) signals; however, 3′ UTRs are variable among the different *Vf* paralogs as well as among clones in the same paralog. Highly homologous sequences downstream of stop codons are present among the *Vf* paralogs, and several sequence blocks are rich in A + T content. It is likely that RNA terminal riboadenylate transferase binds to different A + T-rich blocks, thus resulting in the recovery of variable sizes of UTRs. The sequence information revealed by RACE analysis is then used to clone the full-length cDNA and its corresponding genomic gene.

Apart from resistance-like genes involved in the *Vf* locus, other genes are also present in this region. Whether or not these closely linked genes are related to the plant defense system is yet unknown and must be investigated. Our results indicate that the *Vf* locus is a gene-rich region. This is consistent with the finding of Vinatzer *et al.* (2001) where >50 expressed sequences have been recovered during screening of positive cDNA clones using labeled BAC inserts.

Common and specific PCR primers: We have taken full advantage of the sequence information of the four *Vf* paralogs to design both common and specific primers. In the first half of the LRR domain, deletion of complete LRR units is observed whereby both *Vfa2* and *Vfa3* lack two LRRs, while *Vfa4* lacks five LRRs. Conserved sequences flanking the deletion region have been used to design three common primers (Table 1). These common primers not only have amplified the four *Vf* paralogs and other *Vf* homologous sequences, but also on the basis of the sizes of the PCR products have allowed determination of the identities of the *Vf* homologs. The polymorphic sequences of each *Vf* paralog have been used to design *Vfa*-specific primers, and this has allowed pinpointing of individual *Vf* paralogs. As demonstrated in this study, utilization of both common and *Vfa*-specific primers in PCR-based search, RACE, RT-PCR, and screening of genomic genes has dramatically accelerated our efforts to clone and characterize the *Vf* paralogs. Moreover, additional information has been unveiled by using these primers. During RACE analysis, a total of 48 positive RACE clones have been identified using the nested common primers. Of these, 38 clones are confirmed to be derived from the *Vf* locus by using the four *Vfa*-specific primers. This suggests that the *Vf* locus is very active in the resistance to apple scab. The remaining 10 positive RACE clones that could not be confirmed with *Vfa*-specific primers must then be derived from other chromosomal regions containing *Vf* homologous sequences. Likewise, during RT-PCR analysis, RT-PCR products have been generated from the susceptible apple cultivar Red Delicious by using the nested common primers. All these findings suggest that some other *Vf* homologous sequences, apart from the *Vf* paralogs, are also being expressed in the leaf tissue. Whether or not these expressed sequences also confer resistance to apple scab or to some other diseases must be investigated.

Four *Vf* paralogs are the only resistance genes present in the *Vf* locus: After sequencing of 15 positive clones followed by an intensive PCR-based search, a total of four *Vf* paralogs were identified in the *Vf* locus. These four *Vf* paralogs belong to the LRR-TM resistance gene class and share many common features with the *Cf* genes found in tomato (Dixon *et al.* 1996; Parniske *et al.* 1997). Is there any other class of resistance gene(s) interspersed within the *Vf* locus? To answer this question, first we used degenerate primers derived from the major resistance gene class of NBS-LRR to amplify BAC clones covering the *Vf* locus. However, no NBS-LRR resistance-like sequences have been obtained. Second, the four *Vfa*-specific primers have been used to screen an apple-mapping population consisting of 468 individuals (Xu and Korban 2000), and no recombinant was detected between *Vf* resistance and the four *Vf* paralogs. Third, for all R gene families reported thus far, only members of the same resistance gene class tend to cluster within the same locus (Dixon *et al.* 1996, 1998; Parniske *et al.* 1997; Meyers *et al.* 1998; Wise 2000). All the above findings strongly support the fact that no genes from other resistance classes are interspersed within the *Vf* locus, and the four *Vf* paralogs are the only resistance genes clustered within the *Vf* locus. The coding sequences of *Vfa1* and *Vfa2* are identical with those of *HcVf1* and *HcVf2*, previously reported by Vinatzer *et al.* (2001). The *Vfa3* and *Vfa4* paralogs are novel and have not been previously reported.

Possible roles of the *Vf* paralogs within the *Vf* locus: A resistance gene belonging to the LRR-TM class first recognizes the elicitor of a specific pathogen and then passes down this signal to such elements as protein kinases to induce a plant's defense system. The observed high homology in the C-terminal region among *Vfa1*, *Vfa2*, and *Vfa4* suggests that these three *Vf* paralogs activate the same defense system in apple.

Three active *Vf* paralogs, *Vfa1*, *Vfa2*, and *Vfa4*, in the *Vf* locus exhibited differential expression profiles during leaf development. This suggested that a natural pyramiding of genes must have occurred within the *Vf* locus. Transferring these *Vf* paralogs individually into apple via genetic transformation, followed by analysis of scab resistance profiles of transgenic lines, will elucidate the functionality of these *Vf* paralogs and assess their role(s) in the durability of scab resistance in apple leaf tissues.

All apple cultivars, regardless of their resistance or susceptibility to apple scab, show resistance to apple scab in older mature leaves, commonly referred to as ontogenic resistance (Gessler and Stumm 1984). Thus,

it can be speculated that some factors may contribute to scab resistance in older mature leaves. One such factor is the physical barrier, such as the strength of cell walls and/or cuticle layers of mature leaves. This will either prevent the pathogen from penetrating into cells of the cuticle layer or lead to accumulation of cutinase inhibitors, thus blocking development of subcuticular stromata of the fungal pathogen and preventing its spread (Valsangiacomo and Gessler 1988; Köller *et al.* 1991). Another factor may involve the presence of some unknown *Vf*-like resistance genes that are active during late stages of leaf development. In this study, RACE and RT-PCR have demonstrated the presence of some *Vf* orthologs that are outside the *Vf* locus and are transcribed in mature leaves. It is likely that these *Vf* orthologs may contribute to resistance to apple scab in mature leaves of susceptible apple cultivars. However, this may be different in the wild crabapple species *M. floribunda* 821, the original source of the *Vf* locus. As the *Vf* locus confers sustained resistance to apple scab, it seems unlikely for *M. floribunda* 821 to carry other highly functional *Vf* orthologs. The *Vf* haplotypes detected in the PCR-based search may mainly serve as reservoirs for generating novel resistance specificities to counterattack the ever-changing populations of the pathogen. After the *Vf* locus is incorporated into susceptible apple cultivars via genetic transformation, it is anticipated that a more durable resistance will be present in both immature and mature leaves.

To date, no single recombination event among the three active *Vf* paralogs has been observed. This indicates that the *Vf* locus is usually inherited intact. However, individuals containing the *Vf* locus have exhibited a wide range of symptoms (Chevalier *et al.* 1991), implying that genetic backgrounds may play a significant role in scab resistance. This may be achieved through either modification in transcription and translation of the *Vf* paralogs or alteration in the expression of genes involved in the defense system in the apple genome.

This work has been supported by funds received from the Illinois Council for Food and Agricultural Research external program.

LITERATURE CITED

Bénaouf, G., and G. Parisi, 2000 Genetics of host-pathogen relationships between *Venturia inaequalis* races 6 and 7 and *Malus* species. Phytopathology **90:** 236–242.

Bergelson, J., M. Kreitman, E. A. Stahl and D. Tian, 2001 Evolutionary dynamics of plant R-genes. Science **292:** 2281–2285.

Buckler, A. J., D. D. Chang, S. L. Graw, J. D. Brook, D. A. Haber *et al.*, 1991 Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA **88:** 4005–4009.

Chenchik, A., F. Moqadam and P. Siebert, 1995 Marathon cDNA amplification: a new method for cloning full-length cDNAs. CLONTECHniques **10:** 5–8.

Chevalier, M., Y. Lespinasse and S. Renaudin, 1991 A microscopic study of different classes of symptoms coded by the *Vf* gene in apple for resistance to scab (*Venturia ineaqualis*). Plant Pathol. **40:** 249–256.

Cross, S. H., V. H. Clark and A. P. Bird, 1999 Isolation of CpG islands from large genomic clones. Nucleic Acids Res. **27:** 2099–2107.

Dixon, M. S., D. A. Jones, J. S. Keddie, C. M. Thomas, K. Harrison *et al.*, 1996 The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell **84:** 451–459.

Dixon, M. S., K. Hatzixanthis, D. A. Jones, K. Harrison and J. D. G. Jones, 1998 The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. Plant Cell **10:** 1915–1925.

Ellis, J. G., G. J. Lawrence, M. Ayliffe, P. Anderson, N. Collins *et al.*, 1997 Advances in the molecular genetic analysis of the flax-flax rust interaction. Annu. Rev. Phytopathol. **35:** 271–291.

Elvin, P., G. Slynn, D. Black, A. Graham, R. Butler *et al.*, 1990 Isolation of cDNA clones using yeast artificial chromosome probes. Nucleic Acids Res. **18:** 3913–3917.

Flor, H. H., 1956 The complementary systems in flax and flax rust. Adv. Genet. **8:** 29–54.

Gessler, C., and D. Stumm, 1984 Infection and stroma formation by *Venturia inaequalis* on apple leaves with different degrees of susceptibility to scab. Phytopathol. Z. **110:** 119–126.

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

Hulbert, S. H., C. A. Webb, S. M. Smith and Q. Sun, 2001 Resistance gene complexes: evolution and utilization. Annu. Rev. Phytopathol. **39:** 285–312.

Jia, Y., S. A. McAdams, G. T. Bryan, H. P. Hershey and B. Valent, 2000 Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. EMBO J. **19:** 4004–4014.

Jones, D. A., and J. D. G. Jones, 1997 The role of leucine-rich repeat proteins in plant defenses. Adv. Bot. Res. Adv. Plant Pathol. **24:** 89–167.

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

Köller, W., D. M. Parker and C. M. Becker, 1991 Role of cutinase in the penetration of apple leaves by *Venturia inaequalis*. Phytopathology **81:** 1375–1379.

Korban, S. S., 1998 What's new with disease-resistant apple cultivars. Proc. Trans. Ill. Hort. Soc. **131:** 74–76.

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

Lovett, M., J. Kere and L. M. Hinton, 1991 Direct selection: method for the isolation of cDNAs encoded by large genomic regions. Proc. Natl. Acad. Sci. USA **80:** 9628–9632.

Martin, G. B., S. H. Brommonschenkel, J. Chunwongse, A. Frary, M. W. Ganal *et al.*, 1993 Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science **262:** 1432–1436.

Meyers, B. C., K. A. Shen, P. Rohani, B. S. Gaut and R. W. Michelmore, 1998 Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. Plant Cell **10:** 1833–1846.

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

Parimoo, S., S. R. Patanjali, H. Shukla, D. D. Chaplin and S. M. Weissman, 1991 cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc. Natl. Acad. Sci. USA **80:** 9623–9627.

Parniske, M., K. E. Hammond-Kosack, C. Golstein, C. M. Thomas, D. A. Jones *et al.*, 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.

Richter, T. E., and P. C. Ronald, 2000 The evolution of disease resistance genes. Plant Mol. Biol. **42:** 195–204.

Salmeron, J. M., G. E. D. Oldroyd, C. M. T. Rommens, S. R. Scofield, H. S. Kim *et al.*, 1996 Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. Cell **86:** 123–133.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scofield, S. R., C. M. Tobias, J. P. Rathjen, J. H. Chang, D. T. Lavelle *et al.*, 1996 Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science **274:** 2063–2065.

Song, W. Y., G. L. Wang, L. L. Chen, H. S. Kim, L. Y. Pi *et al.*, 1995 A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science **270:** 1804–1806.

Sood, R., T. I. Bonner, I. Makalowska, D. A. Stephan, C. M. Robbins *et al.*, 2001 Cloning and characterization of 13 novel transcripts and the human *RGS8* gene from the *1q25* region encompassing the heredity prostate cancer (*HPC1*) locus. Genomics **73:** 211–222.

Valsangiacomo, C., and C. Gessler, 1988 Role of the cuticular membrane in ontogenic and *Vf*-resistance of apple leaves against *Venturia inaequalis*. Phytopathology **78:** 1066–1069.

Vinatzer, B. A., A. Patocchi, L. Gianfranceschi, S. Tartarini, H. B. Zhang *et al.*, 2001 Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with *Vf* apple scab resistance. Mol. Plant-Microbe Interact. **14:** 508–515.

Wise, R. P., 2000 Disease resistance: what's brewing in barley genomics. Plant Dis. **84:** 1160–1170.

Xu, M. L., and S. S. Korban, 2000 Saturation mapping of the apple scab resistance gene *Vf* using AFLP markers. Theor. Appl. Genet. **101:** 844–851.

Xu, M. L., and S. S. Korban, 2002 AFLP-derived SCARs facilitate construction of a 1.1 Mb sequence-ready map of a region that spans the *Vf* locus in the apple genome. Plant Mol. Biol. **50:** 803–818.

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

Zhang, M. Q., 1997 Identification of protein coding regions in the human genome by quadratic discrimination analysis. Proc. Natl. Acad. Sci. USA **94:** 565–568.

Communicating editor: A. H. D. Brown