Comparison of $\textit{avrD}$ alleles from $\textit{Pseudomonas syringae pv. glycinea}$

Lisa Wolfson Keith  
\textit{University of Nebraska-Lincoln}  

Carol Boyd  
\textit{University of California}  

Noel T. Keen  
\textit{University of California}  

J. E. Patridge  
\textit{University of Nebraska-Lincoln}, james@unlinfo.unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/plantpathpapers

Part of the Other Plant Sciences Commons, Plant Biology Commons, and the Plant Pathology Commons

http://digitalcommons.unl.edu/plantpathpapers/333

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Plant Pathology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Avirulence gene D alleles reside on indigenous plasmids in races 0, 2, 3, 4, 5, and 6 of *Pseudomonas syringae pv. glycinea* (Psg), but the allele in race 1 appeared to be chromosomal. These were all nonfunctional avirulence genes because they neither induced the avirulence phenotype on *Rpg*4 soybean cultivars nor directed the production of syringolide elicitors when expressed in *Escherichia coli* cells. The predicted proteins encoded by the seven *Psg* avrD genes were very similar to that of a functional class II allele from *P. syringae pv. phaseolicola* G50 race 2, but contained mutations collectively affecting only nine amino acid positions. Despite these relatively small amino acid differences and the location of avrD from each isolate on a 5.6-kb HindIII restriction fragment, the flanking regions varied considerably among the *Psg* isolates. The presence of avrD alleles with few alterations but different locational contexts in all tested *Psg* races argues that they provide an important selected function in the bacteria but have been modified to escape defense surveillance in *Rpg*4 soybean plants.

Additional keywords: avr, hypersensitive response, phylogenetic, pulsed-field gel electrophoresis.

Many bacterial avirulence (avr) genes have been cloned and characterized (Keen 1990) and some of them appear to be important in virulence (Lorang and Keen 1995; Ritter and Dangl 1995; Swarup et al. 1992). *Pseudomonas syringae pv. glycinea* (Psg), the causal agent of bacterial blight of soybean (*Glycine max* (L.) Merr.) contains three avr genes—avrA, avrB, and avrC (Staskawicz et al. 1984, 1987)—that determine race phenotype and account for the resistance of soybean cultivars carrying complementary resistance genes (Keen and Buzzell 1991). A fourth avirulence gene (avrD) was first cloned from *P. syringae pv. tomato* (Pst) (Kobayashi et al. 1990a, 1990b). The avrD gene in *P. syringae pv. tomato* encoded a 34-kDa protein product of 311 amino acids. This gene directed an avirulence phenotype when introduced into *Psg* race 4 and the bacteria were inoculated into soybean plants carrying the matching disease resistance gene, *Rpg*4 (Keen and Buzzell 1991). Further work showed that *Escherichia coli* and other gram-negative bacteria expressing avrD produced low-molecular-weight products that behaved as elicitors of the hypersensitive response only in soybean plants carrying *Rpg*4 (Keen et al. 1990). These non-proteinaceous, low-molecular-weight C-glycosides were named syringolides (Midland et al. 1993; Smith et al. 1993).

Active alleles of avrD from several *P. syringae* pathogens have been cloned and sequenced (Yucel et al. 1994c; C. Boyd, unpublished data). Some of these alleles were highly homologous to *Pst avrD (>95%)*, but others were only ca. 85% homologous. Yucel et al. (1994c) discovered that the syringolides directed by this second group of alleles were also distinct from those directed by the avrD gene from *Pst* and related highly homologous genes. The different avrD genes were accordingly grouped as class I and class II. Class I alleles direct the production of syringolides with 8 or 10 carbon alkyl chains, while class II alleles lead to syringolides with 6 or 8 carbon alkyl chains.

A gene homologous to functional avrD alleles was found in an isolate of *Psg* race 4 (Kobayashi et al. 1990b), but this allele did not elicit the hypersensitive response in soybean cultivars carrying the *Rpg*4 disease resistance gene. The race 4 gene was therefore considered “nonfunctional” but its occurrence as a complete open reading frame raised several questions: Do avrD alleles occur in all isolates of *Psg* and are they all “nonfunctional”? Do the avrD alleles in a range of *Psg* isolates have wholesale mutations, including nonsense and deletion mutations, or do a limited number of specific missense mutations recur? Was a functional avrD allele evolutionarily introduced into *Psg* and subsequently mutated? Was avrD introduced into *Psg* once or multiple times? To provide evidence germane to these questions, we examined the avrD genes occurring in two additional race 4 isolates of *Psg* as well as those occurring in isolates from six other races. DNA flanking the genes was examined by restriction fragment length polymorphism (RFLP) analysis and the sequences were compared with that of a highly homologous but functional class II avrD allele cloned from *P. syringae pv. phaseolicola* G50 race 2.

Corresponding author: J. E. Partridge; E-mail: james@unlinfo.unl.edu

Joint contribution of University of Nebraska Agricultural Research Division and The University of California - Riverside. Published Journal Series 11626, Agricultural Research Division, University of Nebraska, Lincoln.

Nucleotide and/or amino acid sequence data are to be found at GenBank as accession number J03682.
For RFLP analysis, total plasmid preparations were digested with restriction enzymes under conditions specified by the manufacturer (New England Biolabs, Beverly, MA), and resolved on 9-cm-long 0.7% agarose gels with E buffer (40 mM Tris, 2 mM Na$_2$EDTA, pH 7.9), run at 70 V for 2.5 h at 4°C. DNA was transferred to 0.45-µm, Hybond N nylon membranes for Southern analysis as described above.

To take advantage of the highly conserved 5' and 3' ends of the avrD genes, Psg alleles were amplified from genomic DNA by polymerase chain reaction with restriction sites incorporated into the primers (Life Technologies, Gaithersburg, MD) to facilitate cloning [5' primer with an NsiI site TAC CTG ATG GAA CCA AAT CC or an EcoRI site (Kobayashi et al. 1990a, 1990b), and a complementary 3' primer with a BamHI site TCG AAA GGAT CCT AAA GGG GTA ATA]). After amplification, fragments were prepared for cloning following the method of Crowe et al. (1991). Vent DNA polymerase (New England Biolabs) with 3' to 5' exonuclease activity was used for amplification to reduce the likelihood of polymerase errors. The avrD genes from each race of Psg were individually cloned and sequenced on at least two separate occasions to ensure fidelity and prevent cross-contamination errors from the other races.

Alleles of avrD were harvested by all of the newly examined isolates, representing seven races of Psg, and an avrD allele was also amplified from genomic DNA of P. syringae pv. phaseolicola G50. When cloned into the expression plasmid pNIII A-2 and expressed in E. coli DH5α (Yucel et al. 1994c), the G50 allele produced large amounts of syringolide, but none of the Psg alleles resulted in detectable production as measured by high-pressure liquid chromatography or elicitor activity in Rpe4 soybean leaves. In common with the previously investigated Psg race 4 avrD allele, all of the new Psg avrD alleles therefore appeared to be inactive as avirulence genes.

The DNA sequences of all the new Psg avrD alleles and that of P. syringae pv. phaseolicola G50 race 2 were very similar and gave predicted amino acid sequences (see Table 1). In view of the substantial number of avrD alleles now cloned, we propose institution of systematic naming. Thus, the initially cloned allele from P. syringae pv. tomato is now called avrD$^4$, the P. syringae pv. phaseolicola G50 race 2 allele is referred to as avrD$^{2}$ (the previously cloned allele from

\begin{table}
\centering
\begin{tabular}{|l|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{Isolate} & \textbf{Allele designation} & \multicolumn{12}{|c|}{\textbf{Amino acid at position number}} \\
\hline
\multicolumn{1}{|c|}{\textbf{Race 0}} & \textbf{avrD} & 19 & A & A & R & H & S & M & E & V & N & \ldots & \ldots \\
\hline
\textbf{Race 1} & \textbf{avrD} & 1 & R & A & A & H & Y & S & M & E & V & N & F & S \\
\hline
\textbf{Race 2} & \textbf{avrD} & 2 & C & A & R & H & H & L & M & E & A & K & L & S \\
\hline
\textbf{Race 3} & \textbf{avrD} & 3 & H & A & R & H & H & S & M & E & A & K & L & S \\
\hline
\textbf{Race 4} & \textbf{avrD} & 4 & C & A & R & H & H & L & M & E & A & K & L & S \\
\hline
\textbf{Race 5} & \textbf{avrD} & 5 & C & A & R & H & H & L & M & E & A & K & L & S \\
\hline
\textbf{Race 6} & \textbf{avrD} & 6 & H & A & R & H & H & S & M & E & A & K & L & S \\
\hline
\textbf{P. s. phas} & \textbf{avrD} & 2 & R & E & R & H & H & S & V & E & V & K & F & S \\
\hline
\textbf{P. s. lach} & \textbf{avrD} & 2 & R & E & R & H & H & S & M & A & V & K & F & S \\
\hline
\end{tabular}
\end{table
The derived amino acid sequences from the seven \textit{Psg} \textit{avrd} alleles were compared with those amino acids that varied from \textit{avrd}0 (\textit{P. syringae pv. phaseolicola} G50) race 2 which was designated as the outgroup and analyzed phylogenetically with the Protpars program of Phylip (Felsenstein 1989) (Fig. 1). This program infers an unrooted phylogeny from protein sequences, using a new method intermediate between the approaches of Eck and Dayhoff (1966) and Fitch (1971). Eck and Dayhoff (1966) allowed any amino acid change to any other, and counted the number of such changes needed to evolve the protein sequences on each given phylogeny. This has the problem that it allows replacements that are not consistent. Fitch (1971), on the other hand, counted the minimum number of nucleotide substitutions that would be needed to achieve the given protein sequences (Felsenstein 1988). As a result, the Fitch method counts silent changes equally with those that change the amino acid.

The Protpars method “insists that amino acid changes must be consistent with the genetic code so that, for example, lysine is allowed to change to methionine but not to proline. However, changes between two amino acids via a third are allowed and counted as two changes if each of the two replacements is individually allowed” (Felsenstein 1988). For these reasons, Protpars is appropriate for this analysis of very closely related genes. The amino acid sequences were analyzed in all possible paired combinations. This resulted in three trees that varied only in the relative positions of races 2, 4, and 5 to each other. Because races 2, 4, and 5 have identical amino acid sequences and the program cannot represent unresolved trichotomies, the apparent alternative trees are thus resolved.

Amino acid differences are shown as unique characters on the cladogram in Figure 1. The radial presentation permits examination of differences as well as assessment of transitional and transversional changes between the races without declaring either to be evolved or primitive characters.

Nucleotide sequence submitted to phylogenetic analysis produced the same relationships that are shown in Figure 1. There are two silent mutations at bases 102 and 498, coding for amino acids 34 and 166, respectively. In both cases all of the \textit{Psg} \textit{avrd} alleles have the same base and \textit{P. syringae pv. phaseolicola} and \textit{P. syringae pv. lachrymans} \textit{avrd} alleles the other base. A third silent mutation occurs at base 198, amino acid 66, wherein \textit{avrd}2/4/5 and \textit{avrd}3/6 differ from \textit{avrd}0 in their coding for arginine.

\textbf{Fig. 1.} Radial cladogram resulting from a Protpars program of Phylip analysis of seven races of \textit{Pseudomonas syringae} pv. \textit{glycinea} based on amino acid dissimilarities with \textit{P. syringae pv. phaseolicola} (G50) race 2 as the outgroup. The letters indicate the amino acids change(s) (transition or transversion) at the nearest branch point. Lettering begins arbitrarily with the outgroup \textit{P. syringae pv. phaseolicola}. \textbf{A}, Transition at amino acid 250 from M to V. \textbf{B}, Transition at amino acid 41 from E to A and transition at amino acid 44 from R to A. \textbf{C}, Transition at amino acid 283 from K to N. \textbf{D}, Transition at amino acid 66 from H to R and transversion to termination at amino acid 294. \textbf{E}, Transition at amino acid 266 from H to Y. \textbf{F}, Transition at amino acid 280 from V to A and transition at amino acid 301 from F to L. \textbf{G}, Transition at amino acid 19 from R to C, transition at amino acid 245 from S to L, and transition at amino acid 304 from S to L. \textbf{H}, Transition at amino acid 19 from R to H.

\textbf{Fig. 2.} Plasmids from the seven races of \textit{Pseudomonas syringae} pv. \textit{glycinea} (\textit{Psg}). A variation of the methods of Kado and Liu (1981) and Birnboim and Doly (1979) as described by Maniatis et al. (1982) was used to isolate the indigenous plasmids of \textit{Psg}. The plasmids were resolved by pulsed-field gel electrophoresis with the CHEF-DR II system. Agarose gels (1%) in 0.5× Tris-borate-EDTA buffer (Maniatis et al. 1982) were run at 15 Vcm⁻¹, at 4°C, with ramped pulse times of 3 to 10 s for 24 h. \textbf{A}, Indigenous plasmids revealed by ethidium bromide. \textbf{B}, Southern blot of \textit{A} probed with a full-length \textit{avrd}0 orf 1 digoxigenin-labeled probe.
Sequence analysis suggests that the alleles found in races 2, 3, 4, 5, and 6 may have resulted from a single evolutionary introduction, with the selection of two subsequent transitions causing amino acid changes at positions 280 and 301 (Table 1). Yucel et al. (1994b) showed that the change at position 280 is crucial for activity and the change at position 301 significantly reduces syringolide production, depending on the context of residues at positions 41 through 44. The next mutations were two independent transitions in the same amino acid causing a divergence of the alleles occurring at position 19 (R to C in avrD<sub>g2/4/5</sub> and R to H in avrD<sub>g3/6</sub>). avrD<sub>g2/4/5</sub> further accumulated two more transitions (both S to L at positions 245 and 304) not found in avrD<sub>g3/6</sub>.

The race 0 and race 1 alleles have the transversion K to N at position 283 not found in any other alleles (Table 1). These alleles then diverged from each other, avrD<sup>5</sup> having an H to Y transition at position 226 and avrD<sup>6</sup> having an H to R transition at position 294. The avrD alleles in races 0 and 1 appear to represent an independent introduction of avrD into Psg since none of the mutations in these alleles occur in those from races 2, 3, 4, 5, or 6, and vice versa. Indeed, the alleles from races 0 and 1 have greater homology to the active class II <i>P. syringae</i> pv. <i>lachrymans</i> (avrD<sup>5</sup>) allele than they do to the alleles found in races 2, 3, 4, 5, and 6.

In order to further study the avrD alleles occurring in the seven Psg races, the location of these genes was examined. Each Psg race contained a unique plasmid profile (Fig. 2A) and hybridization with avrD<sup>at</sup> orf1 was only observed to the largest indigenous plasmid in all races tested except race 1 (Fig. 2B). Murillo et al. (1994) showed that avrD in <i>Pst</i> occurred on an indigenous plasmid and plasmids have commonly been detected in <i>Psg</i> isolates (Curiale and Mills 1977, 1983; Ullrich et al. 1993). Plasmid sizes were estimated as somewhat less than 90 kb in Psg races 0, 2, 3, and 5 and 95 kb in races 4 and 6 (Fig. 2). It appears from the difference in hybridization pattern from the CHEF gel Southern blot that avrD is chromosomally located in race 1.

In order to assess for differences among the plasmid environments of the avrD alleles in each race, RFLP analyses were performed. As shown in Figure 3, all of the new Psg...
isolates contained the 5.6-HindIII restriction fragment observed for all avrD genes thus far cloned (Kobayashi et al. 1990a). Several conserved internal restriction sites in the 5.6-kb fragment are also shown in Figure 3. These sites were used to “anchor” restriction fragments. Thus, fragments obtained with restriction enzymes EcoRI, ClaI, BamHI, and PstI were positioned based on the known sites within the 5.6-kb fragment. In the case of BglII, SacI, and BstXI, the long fragments were necessarily placed upstream of the known restriction sitebecause shorter fragments would indicate second internal restriction sites within the 5.6-kb fragment.

Figures 4A, B, C, D and 5A show cleavage fragments obtained with restriction endonucleases HindIII, ClaI, BamHI, EcoRI, and PstI, respectively. These enzymes recognize a single known site within the avrD 5.6-kb HindIII fragment but not within the 1.3-kb avrD probe fragment. Accordingly, the restriction fragment sizes identify the position of restriction endonuclease sites upstream from the 5′ ends of the various avrD sequences. These data indicate that there are no large deletions or insertions in the 5.6-kb HindIII fragment in the Psg isolates. The BamHI digest revealed RFLPs among the races in the upstream region (Fig. 4C). Most races exhibited a single BamHI fragment greater than 21 kb, while race 5 showed a band at 4.7 kb.

Restriction endonucleases ClaI (Fig. 4B) and EcoRI (Fig. 4D) have two cleavage sites at 3.6 and 4.1, and 4.6 and 5.2 kb, respectively, within the 5.6-kb HindIII fragment, but only the sites close to avrD were germane to this analysis. Again, all races showed single RFLP bands with these enzymes and RFLPs among the races.

In the 5.6-kb HindIII fragment from PstI, there is a single restriction endonuclease cleavage site for PstI at base 5182. All races, except race 1, show single RFLP bands at about 19 kb while race 1 shows two bands at 5.2 and 6.5 kb (Fig. 5A).

Figures 5 and 6 present bands resulting from cleavage by restriction endonucleases (PstI, BglII, BstXI, and SacI) digestions of plasmid preparations of each of the seven races of Pseudomonas syringae pv. glycinea (Psg) probed by the orf 1 probe. Following single-enzyme digestion of total plasmid preparations with restriction enzymes (PstI, BglII, BstXI, and SacI), restriction fragment length polymorphism (RFLP) bands were resolved on 9-cm-long 0.7% agarose gels with E buffer (40 mM Tris, 2 mM Na2EDTA, pH 7.9) run at 70 V for 2.5 h at 4°C. DNA was transferred to 0.45-µm Hybond N nylon membranes for Southern analysis. A, B, C, D, RFLP bands after digestion with restriction endonucleases PstI, BglII, BstXI, and SacI, respectively. Lanes 0, 1, 2, 3, 4, 5, and 6 correspond to Psg races 0, 1, 2, 3, 4, 5, and 6, respectively. In A, owing to inadequate loading, lane 0 presents a faint band equivalently positioned with bands in lanes 2, 3, 4, 5, and 6.
restriction endonucleases BglII, BstXI, and SacI with recognition sites at bases 467, 298, and 345, respectively. Because the probe used was a full-length copy of avrD, two bands were expected from each enzyme. Except for race 1, this result was obtained in each case. Again there are RFLPs for each enzyme among the races.

Figure 6A, B, C, and D present bands obtained with restriction endonucleases KpnI, SacI, SmaI, and XhoI. These enzymes do not cleave within the 5.6-kb fragment and, therefore, only single bands were anticipated. This result was obtained, with RFLPs apparent among the races with enzymes SacI and XhoI.

Some races showed more similarity than others in RFLP patterns, and consequently we were unable to separate races 0 and 2 from races 4 and 6 based on RFLPs. Because there are no race-specific restriction sites within avrD and the number of restriction endonucleases used was small, it is not appropriate to attempt a phylogenetic analysis based on RFLP patterns.

We have shown that Psg races 0, 2, 3, 5, and 6 contain plasmid-borne avrD homologues and that race 1 appears to contain a chromosomal avrD homologue. DNA sequence data (Table 1) established that these Psg alleles fall in the class II family of avrD alleles, but all of them have accumulated a limited number of discrete mutations that account for their lack of the avirulence phenotype. Thus, as shown by Yucel et al. (1994b), avrD would be expected to be inactive because of the nonsense mutation at position 294 that eliminates essential amino acid residues at positions 301 and 304. avrD would also be predicted to be inactive because it contains mutations at positions 41 and 44 that do not function with phenylalanine at position 301 in class II alleles. avrD contains cysteine at position 19, alanine at position 280, leucine at position 301, and leucine at position 304, all of which are sufficient to eliminate the avirulence phenotype (Yucel et al. 1994b). avrD contains histidine at position 19, alanine at position 280, and leucine at position 301, all of which are sufficient to eliminate avirulence activity. The race 0 allele is of particular interest because it was the only one containing a nonsense mutation, converting the codon for amino acid 295 from GGA (glycine) to TGA (termination codon). The premature termination results in a protein lacking 17 amino acids, including residues 301 and 304, which are required for avrD activity, as discussed above.

Despite lacking the avirulence phenotype, all examined isolates of Psg contained complete or nearly complete and intact avrD genes as well as the 5.6-kb HindIII DNA fragment on which they occur. Indeed, very few mutations have accumulated in Psg avrD genes and, excepting the nonsense mutation in avrD, these are generally limited to missense mutations that have been shown previously to abolish the avirulence function (Yucel et al. 1994b). The avrD alleles currently present in Psg have, excepting avrD, therefore been strongly selected to contain only mutations yielding amino acid substitutions that eliminate the avirulence phenotype but neither wholesale missense or nonsense mutations nor deletions/insertions were observed.

The results permit answers to some of the questions posed at the outset regarding Psg avrD alleles. First, these alleles were present in all tested Psg isolates, representing seven physiologic races, and all of them are nonfunctional as avirulence genes. However, the clustering of mutations in only nine amino acid codons and one C-terminal nonsense mutation clearly shows that mutations have not accumulated at random in an ancestral avrD allele. Further, sequence comparisons of the various Psg avrD alleles strongly suggest that avrD was independently introduced at least twice into Psg. Had only one introduction occurred, then avrD would have been expected to share at least one common mutation with avrD or avrD.

In contrast to DNA sequence analysis, which reveals the limited number of mutations that have occurred in Psg avrD alleles and the strict conservation of the 5.6-kb HindIII fragment on which they occur, the RFLP analysis estimates considerable variation in the flanking environment among the various Psg isolates. Thus, while avrD occurs on indigenous plasmids in all isolates except race 1, which appears to harbor a chromosomal copy, RFLP analyses (Fig. 3) established that DNA flanking the 5.6-kb HindIII fragment carrying avrD is quite polymorphic.

Finally, our results offer insight into the selection pressure exerted on pathogen avirulence genes by cognate plant disease resistance genes. The Rpg resistance gene is distributed in many but not all contemporary soybean cultivars (Keen and Buzzell 1991). Correspondingly, the presence of an avirulence phenotype in the parasite is by its very nature detrimental. Our data clearly suggest that the occurrence of Rpg in the soybean germ plasm selected only certain avrD mutations after its introduction into Psg in ways that specifically eliminated the avirulence phenotype but retained a complete or nearly complete open reading frame. These observations all foster the speculation that the gene named avrD may have a naturally selected function in Psg independent of the avirulence phenotype.

ACKNOWLEDGMENTS

We thank C. Bender and T. Denny for helpful comments on this work, and A. K. Vidaver and S. Patil for supplying the bacterial cultures. This research was supported in part by a special grant from the Agricultural Research Division (J.E.P.), and by grant no. 9501093 from the USDA Competitive grants (N.T.K.).

LITERATURE CITED


Keen, N. T., and Buzzell, R. I. 1991. New disease resistance genes in


