Novel genes involved in the regulation of pathogenicity factor production within the *rpf* gene cluster of *Xanthomonas campestris*

J. Maxwell Dow, Jia-Xun Feng,† Christine E. Barber, Ji-Liang Tang† and Michael J. Daniels

The synthesis of extracellular enzymes and extracellular polysaccharide (EPS) in *Xanthomonas campestris* pathovar *campestris* (*Xcc*) is subject to co-ordinate regulation by a cluster of genes called *rpf* (for regulation of pathogenicity factors). These genes are located within a 21 kb region of the chromosome isolated as the cosmid clone pIJ3020. The genes in the left-hand section of this region of the chromosome have previously been characterized. This paper reports on the genes in the right-hand section and on the phenotypes of mutants with transposon insertions in these genes. Sequence analysis identified eight genes or ORFs with the gene order *rpfD–orf1–orf2–orf3–orf4–recJ–rpfE–greA*. RecJ and GreA have established functions in recombination and transcriptional elongation, respectively. *rpfD* encoded a protein with some amino acid sequence relatedness to a hypothetical protein from *Caulobacter crescentus* and an autolysin response regulator in *Bacillus subtilis*. The predicted protein products of *orf1*, *2* and *3* were related to each other and had substantial amino acid sequence relatedness to hypothetical proteins from *C. crescentus*. Transposon insertions in *orf1*, *2* and *3* had no effect on the synthesis of extracellular enzymes or EPS. The predicted proteins RpfE and Orf4 showed the highest amino acid sequence relatedness to hypothetical proteins from *Bordetella pertussis* and *Klebsiella pneumoniae*, respectively. Transposon insertions in *rpfE* led to reduced levels of some extracellular enzymes (endoglucanase and protease) and increased levels of others (polygalacturonate lyase). Transposon insertions in *orf4* had no effect on polygalacturonate lyase but led to reduced levels of protease and endoglucanase. Levels of EPS were reduced in both *rpfE* and *orf4* mutants. These alterations in the levels of extracellular enzymes, which were relatively modest (between two- and threefold), did not affect the pathogenicity of *Xcc* on turnip. It is proposed that the gene designation should be *rpfI* for *orf4*.

**Keywords:** *Xanthomonas campestris*, regulation of pathogenicity factors, *rpf* gene cluster

**INTRODUCTION**

*Xanthomonas campestris* pathovar *campestris* (*Xcc*) is the causal agent of bacterial black rot of crucifers (Onsando, 1992). The organism produces extracellular polysaccharide (EPS) and a range of extracellular enzymes that contribute collectively to pathogenesis (Dow & Daniels, 1994). One of the aims of our work on *Xcc* is to understand how the synthesis of these pathogenicity factors is regulated. The synthesis of extracellular enzymes and EPS is subject to co-ordinate regulation by a cluster of genes called the *rpf* cluster, for regulation of pathogenicity factors (Dow & Daniels,
1994). These genes are located within a 21.9 kb region of the chromosome which has been isolated as the cosmid clone pIJ3020 (Tang et al., 1991). Analysis of pIJ3020 by transposon mutagenesis has demonstrated the existence of at least eight genes (rpfA–H) insertions in which lead to a co-ordinate reduction in the levels of all extracellular enzymes tested and in EPS (Tang et al., 1991). Some of the rpf genes have been characterized in detail. rpfA encodes the major aconitase of Xcc and is implicated in iron homeostasis (Wilson et al., 1998); rpfB and rpfF are involved in the synthesis of a small diffusible signal molecule (Barber et al., 1997; Slater, 1999); rpfG, C and H encode elements of a two-component sensory-transduction system (Tang et al., 1991; Slater, 1999). Adjacent to rpfG are two genes with broader regulatory functions; prfB encodes peptide release factor 2, and lysU encodes lysyl-tRNA synthetase (Dow & Daniels, 1994). The genes rpfA, B, F, C, H and G together with prfB and lysU comprise the left-hand part of the Xcc DNA within pIJ3020 (Fig. 1). Here we report the characterization of the genes in the right-hand section, which includes rpfD and rpfE. This analysis has identified a further gene (orf4) involved in the regulation of extracellular enzyme and EPS synthesis, together with three ORFs (orf1, 2 and 3) of unknown function. In addition, two genes were identified whose products are highly related to RecJ and GreA, which in other bacteria have established functions in recombination and transcriptional elongation, respectively. The predicted protein products of orf1, orf2, orf3, orf4, rpfD and rpfE show the most significant amino acid sequence similarity to hypothetical proteins predicted from (as yet unfinished) microbial genome sequencing projects.

**METHODS**

**Strains and culture conditions.** The Xcc strains used in this study were 8004, a rifampcin-resistant mutant of a wild-type isolate (Daniels et al., 1984), and mutant strains carrying Tn5 or Tn5lac within the rpf cluster of genes (Tang et al., 1991). Strain 8478 carries a Tn5 insertion in rpfD; strain 8535 carries a Tn5lac insertion in orf4. All strains were normally grown in peptone/yeast extract/glycerol medium (NYGB) at 28 °C as described by Daniels et al. (1984).

**Enzyme and EPS assays.** For measurements of endoglucanase activities, Xcc strains were grown in NYGB medium to an OD of 0.6 at 30 °C. Enzyme activities in cell-free culture supernatants were measured by radial diffusion assays into substrate-containing agar plates as described by Gough et al. (1988). For induction of polygalacturonate lyase (PGL), strains were grown in MMXC medium supplemented with 0.25% (w/v) polygalacturonate (Dow et al., 1987) to an OD of 0.5. PGL activity was measured spectrophotometrically at 235 nm as described by Dow et al. (1987). One unit of activity was defined as a change in optical density of 1 unit min⁻¹ corresponding to the production of 0.5 μmol unsaturated product. For measurements of EPS production, strains were grown in NYGB medium supplemented with 2% (w/v) glucose for 48 h. EPS was precipitated from culture supernatants by ethanol, dried and weighed as described by Vojnov et al. (1998). Indicator plates for protease production contained NYGB supplemented with 1% (w/v) skimmed milk and solidified with 2% (w/v) agar (Tang et al., 1987). Cultures of bacteria were washed and resuspended in water to an OD of 0.2. Aliquots (5 μl) of these suspensions were plated and allowed to dry before growth at 30 °C. Methods used to assay the β-galactosidase reporter activity are given in Barber et al. (1997). Levels of the diffusible signal molecule DSF were assessed by a bioassay as described by Slater (1999).

**DNA sequencing and analysis.** The insert DNA from the cosmid clone pIJ3020 which carries the full rpf cluster (Tang et al., 1991) was subcloned into the pBluescript BS(−) vector (Stratagene) using convenient restriction sites for EcoRI and BamHI prior to sequencing. To complete the sequence of the right border of the cluster, a cosmid clone (pIJ3092) carrying Xcc DNA which overlapped the right-hand border of pIJ3020 was identified from a library of Xcc DNA cut with BamHI and constructed in pLAFR3 (Staskawicz et al., 1987). A 50 kb EcoRV fragment from pIJ3092, which spanned the EcoRI site at the right-hand side of pIJ3020, was cloned in pBluescript KS(−). A 90 bp PstI fragment spanning the EcoRI site was subcloned into pGEM-3zf(+) (Promega) for sequencing. Determination of the DNA sequence was done by automated sequencing using Applied Biosystems DNA sequencers (373 and 377) in conjunction with the PRISM Ready Reaction DyeDeoxy Terminator Cycle sequencing kit (Applied Biosystems). Double-stranded DNA templates were prepared as recommended by the manufacturer. Plasmids were sequenced in both directions using the forward and reverse universal primers and by walking, using a series of synthetic primers (Genosys) based on the determined sequence. The sequence was analysed and compiled using the xbase program of Staden (1989). The GGCG suite of programs was used to analyse the data and the predicted amino acid sequence. Homology searches of the databases were performed using BLAST with BLAST with BLASTOMATIC (Altschul et al., 1990). The Genemark program (Borodovsky & McIninch, 1993) and FRAME (Bibb et al., 1984) were used to predict potential coding regions. Protein structural predictions were made using the PredictProtein program at EMBL (Rost, 1996). BLAST searches against unfinished
microbial genome sequences were performed using the websites at The Institute for Genomic Research (TIGR) at http://www.tigr.org and The Sanger Centre at http://www.sanger.ac.uk/.

**Plant tests.** For growth assays, bacterial suspensions in water at 10⁶ c.f.u. ml⁻¹ were introduced into the leaf lamina of mature leaves of turnip cv. Just Right using a syringe with no needle, and plants were maintained as described by Collinge et al. (1997). Bacterial numbers in plants were measured as described by Parker et al. (1993). For pathogenicity tests, bacterial suspensions (10⁶ c.f.u. ml⁻¹) were introduced into the vascular system of turnip through nicks at the vein endings and the plants were maintained in high humidity conditions as described by Dow et al. (1990). The ability of bacterial strains to induce the hypersensitive response was tested in pepper cv. ECW10R as described by Newman et al. (1997).

**RESULTS**

**Organization of genes in the right-hand section of the Xcc DNA in pIJ3020**

A physical map of the Xcc DNA in pIJ3020, with the restriction sites for BamHI and EcoRI and the location of the region analysed here, is shown in Fig. 1. The DNA sequence of the right-hand section of the Xcc DNA in pIJ3020 comprised 8163 nt. Analysis of this sequence showed that Orf1 and Orf2 have amino acid sequence similarity to each other. An alignment of the three predicted proteins (Fig. 2) showed that Orf1 and Orf2 have amino acid sequence relatedness to the N-terminal and C-terminal parts of

**Table 1. Genes or ORFs predicted by sequence analysis of the right-hand section of the Xcc DNA cloned in pIJ3020**

The size of the predicted gene products, the most significantly related database entries as revealed by BLAST searches, and the percentage similarity and identity scores are given.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Size of gene product (aa)</th>
<th>Amino acid sequence relatedness with BLAST score</th>
<th>Accession no. of homologue</th>
<th>Size of homologue (aa)</th>
<th>Percentage identical (I) and similar (S) amino acids</th>
<th>No. of positions compared</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpfD</td>
<td>305</td>
<td>Hypothetical protein from <em>C. crescentus</em> (10⁻¹⁸)</td>
<td>*</td>
<td>*</td>
<td>31 % I, 48 % S</td>
<td>235</td>
</tr>
<tr>
<td>orf1</td>
<td>108</td>
<td>Hypothetical protein from <em>C. crescentus</em> (10⁻³¹); homologous to Orf3 (9 x 10⁻³¹)</td>
<td>*</td>
<td>*</td>
<td>61 % I, 78 % S</td>
<td>70</td>
</tr>
<tr>
<td>orf2</td>
<td>300</td>
<td>Hypothetical protein from <em>C. crescentus</em> (10⁻³⁵); homologous to Orf3 (9 x 10⁻³⁵)</td>
<td>*</td>
<td>*</td>
<td>42 % I, 56 % S</td>
<td>246</td>
</tr>
<tr>
<td>orf3</td>
<td>454† (420)</td>
<td>Hypothetical protein from <em>C. crescentus</em> (10⁻⁴⁳)</td>
<td>*</td>
<td>*</td>
<td>37 % I, 55 % S</td>
<td>312</td>
</tr>
<tr>
<td>orf4</td>
<td>368</td>
<td>Hypothetical protein from <em>Klebsiella pneumoniae</em> (10⁻⁵)</td>
<td>*</td>
<td>*</td>
<td>45 % I, 55 % S</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hypothetical protein YPFG of <em>E. coli</em> (10⁻¹⁸)</td>
<td>P76559</td>
<td>347</td>
<td>29 % I, 40 % S</td>
<td>337</td>
</tr>
<tr>
<td>recJ</td>
<td>582</td>
<td>Single-stranded DNA specific exonuclease RecJ of <em>Erwinia chrysanthemi</em> and <em>E. coli</em> (10⁻¹²³)</td>
<td>P39693</td>
<td>575</td>
<td>47 % I, 60 % S</td>
<td>578</td>
</tr>
<tr>
<td>rpfE</td>
<td>306</td>
<td>Hypothetical protein predicted from genome sequencing of <em>Bordetella pertussis</em> (5.2 x 10⁻¹⁵)</td>
<td>*</td>
<td>*</td>
<td>33 % I, 67 % S</td>
<td>151</td>
</tr>
<tr>
<td>greA</td>
<td>158</td>
<td>Transcriptional elongation factor GreA of <em>E. coli</em> (10⁻³⁹)</td>
<td>P21346</td>
<td>158</td>
<td>61 % I, 79 % S</td>
<td>154</td>
</tr>
</tbody>
</table>

* Data obtained from incomplete sequencing project; accession numbers and full length of homologue not available.
† GTG start codon, all other predicted ORFs start with an ATG codon.
**Fig. 2.** Amino acid sequence similarities between Orf1, 2 and 3, which are predicted to be associated with the cytoplasmic membrane. The sequences were aligned using the algorithm PILEUP (Devereux et al., 1984). Residues found in at least two of the sequences are highlighted in bold. The positions of the ten predicted membrane-spanning domains of Orf3 are indicated by bars under the alignment and are numbered. Orf1 has three predicted membrane-spanning domains, equivalent in position to domains 1–3 of Orf3. Orf2 has seven predicted membrane-spanning domains, equivalent in position to domains 4–10 of Orf3.

Orf3, respectively. There were two possible start codons for Orf3: a GTG, giving a product of 454 amino acids in length, and an ATG, giving a product of 420 amino acids in length. Optimal alignments between Orf1 and Orf3 were seen with the shorter protein product. In database searches, Orf1, 2 and 3 were related to incomplete hypothetical proteins predicted from the (unfinished) genome sequence of *Caulobacter crescentus* (Table 1). Orf2 and the C-terminal part of Orf3 were both highly related to the same hypothetical protein of *C. crescentus*; Orf1 (and the N-terminal part of Orf3) was related to a protein predicted from a different DNA sequence. Whether these incomplete hypothetical proteins represent different domains of the same protein will only be revealed after assembly of the full *C. crescentus* sequence. Using the PredictProtein program, Orf1, 2 and 3 were suggested to be localized in the cytoplasmic membrane. Orf3 is predicted to have ten membrane-spanning helices, whereas Orf2 has seven and Orf1 has three (Fig. 2). The termination codon of orf1 overlaps the predicted start codon of orf2. Expression studies with Tn5lac show that orf1, 2 and 3 are only weakly expressed (50–150 Miller units) whereas orf4 is more highly expressed (550–1000 Miller units). Although it is possible that orf1, 2 and 3 are transcribed as an operon, the higher expression of orf4 suggests that it must have its own independent promoter. Orf4 is distinct from Orf1, 2 and 3, and has amino acid sequence relatedness to hypothetical proteins from *Klebsiella pneumoniae* and *Escherichia coli* (Table 1). Flanking orf1 is rpfD, which is transcribed in the opposite direction and encodes a protein with amino acid sequence relatedness to a hypothetical protein from *C. crescentus* (Table 1) and also, to a lesser extent, the autolysin response regulator LytT of *Bacillus subtilis* (BLAST score 10^-6"). Interestingly, analysis of the preliminary *C. crescentus* nucleotide sequences suggests that the genes encoding the RpfD and Orf1 homologues are also adjacent in the chromosome and divergently transcribed. Using the PredictProtein program, RpfD, Orf4 and RpfE were suggested to be globular soluble proteins.

### Analysis of effects of transposon insertion on the production of extracellular enzymes and EPS

We have previously reported the mutagenesis of pIJ3020 with Tn5 and Tn5lac and the effects that some of these insertions have on production of extracellular enzymes, EPS and pathogenicity (Tang et al., 1991). The positions of transposon insertion in the right-hand section of the Xcc DNA in pIJ3020 are shown in Fig. 1. The effect of these transposon insertions on the production of certain enzymes and EPS are listed in Table 1.
extracellular enzymes and EPS was assessed. Tn5 insertions within orf1, orf2 and recJ and Tn5lac insertions within recJ and orf3 had no effect on the production of extracellular enzymes or EPS. Tn5lac insertion into rpfD had only minor effects (Tang et al., 1991; and data not shown). In contrast, mutation of orf4 and rpfE with Tn5lac and Tn5, respectively, gave significant alterations in the levels of pathogenicity factors. Mutation of orf4 led to a reduction in endoglucanase and protease levels and in EPS (Table 2 and Fig. 3), although the levels of PGL were unaltered. Mutation of rpfE had more complex effects than originally reported by Tang et al. (1991) leading to a reduction in levels of endoglucanase, protease and EPS production than those seen after mutation of rpfC, rpfG, rpfH, rpfB and rpfE. More significantly, mutation of rpfE led to an elevated level of PGL, an effect that is not seen as a consequence of mutation of any other gene within the rpf cluster. Mutations in an unlinked gene, rpfN, do lead to an increase in PGL, but the levels of other enzymes also increase co-ordinately (Tang et al., 1990). Divergent effects on the levels of different extracellular enzymes are seen after mutation of clp, a gene encoding a catabolite-activation-factor-like protein in X. campestris (De Crecy-Lagard et al., 1990). A clp mutant has elevated levels of protease but reduced levels of endoglucanase and PGL. Mutation of orf4 had no effect on the production of PGL although levels of EPS were fivefold lower. These less pronounced effects on enzymes and EPS compared to other rpf mutants may explain the lack of a detectable effect of mutation of either rpfE or orf4 on pathogenesis or on growth in plants. It is possible that regulation of extracellular enzyme and EPS production by orf4 and rpfE is important in other phases of the disease cycle of Xcc or may be more pronounced under other culture conditions.

The two genes whose functions could be unequivocally assigned were recJ and greA. No detectable change in phenotype was associated with mutation of recJ, a situation which is found in other bacteria where RecJ is thought to have functions which are redundant with other single-stranded DNA exonucleases (Viswanathan & Lovett, 1998; Harris et al., 1998). We did not obtain any transposon insertions in greA. The significance (if any) of the organization of greA within an operon containing recJ and rpfE is not known. greA is known to occur in different (genomic) contexts in different bacteria. In Pseudomonas aeruginosa, greA is the most downstream gene in an operon encoding enzymes involved in pyrimidine biosynthesis (Lu et al., 1997) whereas in Rickettsia prowazekii it is the most upstream assessed over a period of 5 d. No significant difference in the growth rate or final populations was observed (data not shown). Pathogenicity tests in which bacteria were introduced into the vascular system of the plant through cuts at the leaf vein endings showed no loss of virulence associated with mutation of rpfE, orf3 or orf4. In addition, rpfE, orf3 and orf4 mutants, as well as the wild-type, were all able to induce a hypersensitive response in the non-host plant pepper (Capsicum annuum) cv. ECW10R.

**DISCUSSION**

Our analysis of the right-hand section of the 21.9 kb Xcc DNA fragment cloned in pJ3020 has identified two novel genes, orf4 and rpfE, which appear to regulate the production of extracellular enzymes and EPS. The mechanism by which the gene products effect changes in the level of these pathogenicity factors is obscure. The effects of mutation of orf4 and rpfE are considerably different from those seen after mutation of other rpf genes. Mutation of rpfE led to a less pronounced reduction of protease, endoglucanase and EPS production than seen after mutation of rpfC, rpfG, rpfH, rpfB and rpfE. More significantly, mutation of rpfE led to an elevated level of PGL, an effect that is not seen as a consequence of mutation of any other gene within the rpf cluster. Mutations in an unlinked gene, rpfN, do lead to an increase in PGL, but the levels of other enzymes also increase co-ordinately (Tang et al., 1990). Divergent effects on the levels of different extracellular enzymes are seen after mutation of clp, a gene encoding a catabolite-activation-factor-like protein in X. campestris (De Crecy-Lagard et al., 1990). A clp mutant has elevated levels of protease but reduced levels of endoglucanase and PGL. Mutation of orf4 had no effect on the production of PGL although levels of EPS were fivefold lower. These less pronounced effects on enzymes and EPS compared to other rpf mutants may explain the lack of a detectable effect of mutation of either rpfE or orf4 on pathogenesis or on growth in plants. It is possible that regulation of extracellular enzyme and EPS production by orf4 and rpfE is important in other phases of the disease cycle of Xcc or may be more pronounced under other culture conditions.

The two genes whose functions could be unequivocally assigned were recJ and greA. No detectable change in phenotype was associated with mutation of recJ, a situation which is found in other bacteria where RecJ is thought to have functions which are redundant with other single-stranded DNA exonucleases (Viswanathan & Lovett, 1998; Harris et al., 1998). We did not obtain any transposon insertions in greA. The significance (if any) of the organization of greA within an operon containing recJ and rpfE is not known. greA is known to occur in different (genomic) contexts in different bacteria. In Pseudomonas aeruginosa, greA is the most downstream gene in an operon encoding enzymes involved in pyrimidine biosynthesis (Lu et al., 1997) whereas in Rickettsia prowazekii it is the most upstream

**Table 2. Levels of extracellular enzymes and EPS in strains of Xcc**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Endoglucanase (U ml⁻¹)</th>
<th>PGL (mU ml⁻¹)</th>
<th>EPS (mg ml⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8004 (wild-type)</td>
<td>202 ± 30</td>
<td>10 ± 2</td>
<td>2.16 ± 0.05</td>
</tr>
<tr>
<td>8478 (rpfE)</td>
<td>66 ± 10</td>
<td>31 ± 7</td>
<td>0.98 ± 0.02</td>
</tr>
<tr>
<td>8535 (orf4)</td>
<td>78 ± 10</td>
<td>10 ± 2</td>
<td>0.39 ± 0.02</td>
</tr>
</tbody>
</table>

**Fig. 3.** Protease production by strains of Xcc as revealed by zones of clearing around colonies on skimmed milk indicator plates. The strains were 8004 (wild-type), 8478 (rpfE) and 8535 (orf4).

---

Xanthomonas pathogenicity gene regulation
gene in the major macromolecular synthesis operon (Shaw et al., 1997), which includes the gene encoding the major sigma factor RpoD.

Mutation of rpfE and orf4 had no effect on the production of the diffusible signal molecule DSF. DSF may act as a stress- or starvation-induced signal for the synthesis of all extracellular enzymes and EPS (Barber et al., 1997). Its synthesis requires RpfB and RpfF (Barber et al., 1997) and its perception may involve the two-component regulatory system encoded by rpfG/H/C (Slater, 1999). The phenotypes of rpfE and orf4 mutants suggest that RpfE and Orf4 are not essential components within the DSF regulatory system. Similar considerations apply to RpfA, which is implicated in iron homeostasis but not in DSF-dependent regulation (Wilson et al., 1998). The organization of the genes within the region of the Xcc chromosome cloned in pIJ3020 reflects the functional connection to the DSF-dependent regulation of extracellular enzyme and EPS synthesis; the genes rpfB, rpfE, rpfC, rpfH and rpfG are present in two convergently transcribed operons which are contiguous (Slater, 1999). This group of genes is flanked on the left by rpfA (Wilson et al., 1998). Transposon insertions in genes immediately to the left of rpfA have no effect on the synthesis of extracellular enzymes or EPS (T. J. G. Wilson & M. J. Daniels, unpublished). The other rpf genes (rpfD, rpfE and orf4) are to the right of rpfB–G and are grouped along with genes of diverse function including those with broad regulatory functions such as rpfB, lysU and greA, and those with, as yet, no defined function (orf1, 2 and 3).

The significance of the clustering of so many genes with regulatory functions within the region of the Xcc chromosome cloned as pIJ3020 is not understood.

In conclusion, our analysis has identified two genes (orf4 and rpfE) with roles in the regulation of pathogenicity factors in Xcc; we propose the gene designation rpfI for orf4. Although mutation of rpfE and orf4 had no detectable effect on the interaction of Xcc with plants, the role in pathogenesis of the related genes that occur in animal pathogens certainly warrants investigation.

ACKNOWLEDGEMENTS

The Sainsbury Laboratory is supported by a grant from the Gatsby Charitable Foundation. This work was carried out according to the provisions of the Ministry of Agriculture, Fisheries and Food, licence PHF 1185/8(48) issued under the Plant Health (Great Britain) Order 1987 (Statutory Instrument 1758).

REFERENCES


Received 15 September 1999; revised 11 November 1999; accepted 21 December 1999.