environmental microbiology

Environmental Microbiology (2010) **12**(1), 147–155 doi:10.1111/j.1462-2920.2009.02056.x

A single genetic locus in the phytopathogen *Pantoea stewartii* enables gut colonization and pathogenicity in an insect host

John Stavrinides,[1]* Alexander No[2] and Howard Ochman[2]

[1]*Department of Biology, University of Regina, Regina, Saskatchewan, S4S0A2, Canada.*

[2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.*

Summary

Aphids are typically exposed to a variety of epiphytic and phytopathogenic bacteria, many of which have entomopathogenic potential. Here we describe the interaction between *Pantoea stewartii* ssp. stewartii DC283 (DC283), an enteric phytopathogen and causal agent of Stewart's wilt, and the pea aphid, *Acyrthosiphon pisum*. When ingested by aphids, DC283 establishes and aggregates in the crop and gut, preventing honeydew flow and excretion, resulting in aphid death in 72 h. A mutagenesis screen identified a single locus, termed *ucp1* (you cannot pass), whose disruption abolishes aphid pathogenicity. Moreover, the expression of *ucp1* in *Escherichia coli* is sufficient to mediate the hindgut aggregation phenotype by this normally avirulent species. Ucp1 is related to six other proteins in the DC283 genome, each having a common N-terminal region and a divergent C-terminus, but only *ucp1 has* a role in pathogenicity. Based on predicted motifs and secondary structure, Ucp1 is a membrane-bound protein that functions in bacterial adhesion and promotes the formation of aggregates that are lethal to the insect host. These results illustrate that the enteric plant pathogenic bacteria have the capacity to exploit alternative non-plant hosts, and retain genetic determinants for colonizing the gut.

Introduction

The plant-dependent lifestyle of many insects predisposes them to frequent encounters with epiphytic and plant pathogenic bacteria. For the agriculturally and economically important aphid, many bacterial plant pathogens have been identified as constituents of its gut microbiome (Harada *et al.*, 1996). These associations are of ecological consequence in that several phytopathogens, including *Erwinia aphidicola* (Harada and Ishikawa, 1997; Harada *et al.*, 1997; Santos *et al.*, 2009) *Dickeya dadantii* (Grenier *et al.*, 2006) and *Pseudomonas syringae* pv. syringae B728a (Stavrinides *et al.*, 2009), can exploit the aphid as an alternative host, suggesting that these bacteria retain genetic determinants for insect colonization in addition to their regular complement of plant pathogenicity factors.

Little is known about the nature and specificity of the genetic determinants that mediate the interaction between bacterial pathogen and aphid. Exploration of the *D. dadantii*–pea aphid interaction implicated several bacterial Cyt-like toxin genes in aphid pathogenesis (Grenier *et al.*, 2006), and although the Cyt gene homologues of *Bacillus thuringiensis* confer measurable insecticidal activity (Bravo *et al.*, 2007), the *D. dadantii* Cyt homologues are not wholly responsible for its pathogenic potential. In fact, disruption of the *cytA-D* in *D. dadantii* only slightly attenuates virulence, indicating the requirement of additional virulence factors. In contrast, a *fliL* mutant of the phytopathogen *Pseudomonas syringae* was rendered avirulent towards the pea aphid, suggesting a virulence mechanism linked to bacterial motility (Stavrinides *et al.*, 2009). Other insect-specific virulence factors are likely present in the genome of *P. syringae* (Buell *et al.* 2003; Feil *et al.*, 2005; Lindeberg *et al.*, 2008), but the identity of these genes is still unknown.

Several studies have implicated members of RHS/YD repeat family in insect–pathogen interactions (Lin *et al.*, 1984; Sadosky *et al.*, 1989; Zhao *et al.*, 1993; Hill *et al.*, 1994; Zhao and Hill, 1995; Minet *et al.*, 1999; Vallet-Gely *et al.*, 2008). The *rhs* (rearrangement hotspot) genes were first described in *Escherichia coli* and characterized as mosaics whose common core mediates chromosomal rearrangements through homologous exchange (Lin *et al.*, 1984). The RHS family in *E. coli* consists of eight proteins classified into three subfamilies, all having a highly conserved 3.7 kb N-terminal domain but variable C-termini (Wang *et al.*, 1998). There is some suggestion

Received 27 May, 2009; accepted 10 August, 2009 *For correspondence. E-mail john.stavrinides@uregina.ca; Tel. (+1) 306 585 5635; Fax (+1) 306 337 2410.

that these *E. coli* proteins associate with the bacterial cell surface and bind specific ligands (Wang *et al.*, 1998; Hill, 1999), but this has not been demonstrated experimentally leaving their specific function unknown. RHS/YD repeat proteins of several bacteria have been shown to mediate pathogen–insect interactions, such as the Tcc proteins of *Photorhabdus luminescens*, which have been implicated in toxicity to insect hosts (Waterfield *et al.*, 2001), and the Sep proteins of *Serratia entomophila*, which are responsible for the development of amber disease in the grass grub *Costelytra zealandica* (Hurst *et al.* 2000). More recently, the RHS-like toxin encoded on the APSE phage in the aphid endosymbiont *Candidatus* Hamiltonella defensa has been suggested to antagonize development of parasitoid wasp ova and larvae growing within the aphid (Degnan and Moran, 2008). Ultimately, the role and importance of RHS genes in host–bacterial interactions remain unclear.

In this study, we examined the interaction between the enteric Stewart's wilt pathogen *Pantoea stewartii* ssp. stewartii DC283 (DC283) (Coplin *et al.*, 1986) and the pea aphid, *Acyrthosiphon pisum*. Previous work identified two *Pantoea* species that are aggressive pathogens of aphids (Harada and Ishikawa, 1997; Grenier *et al.*, 2006), and due to the association of *P. stewartii* with an insect vector, the corn flea beetle (*Chaetocnema pulicaria*), and its ability to establish in the plant vasculature during plant pathogenesis (Leigh and Coplin, 1992; von Bodman *et al.*, 2003), we hypothesized that our pathosystem may have potentially relevant ecological implications. A genetic screen in aphids designed to identify hypovirulent DC283 mutants recovered an *rhs*-like gene that enables bacteria to form aggregates in the gut of the pea aphid, ultimately reducing honeydew excretion and causing death. Disruption of this gene in DC283 results in complete abolition of aphid pathogenicity, and transfer of the gene to *E. coli* is sufficient to convert this normally avirulent strain into a pathogen.

Results

The ucp1 *mutant of DC283 is avirulent*

Virulence assays that tested the oral toxicity of DC283 towards the aphid revealed that DC283 is an aggressive pathogen, reaching titres of 5×10^8 cfu and killing the aphid within 72 h (Fig. 1). Transposon mutagenesis of DC283 yielded a hypovirulent mutant that was disrupted in a sequence resembling an *rhs* gene, which we called *ucp1* (you cannot pass). Growth of the mutant line, DC283 (*ucp1::Tn5*), in the aphid was substantially lower than wild-type, reaching titres of only 1×10^8 cfu/aphid by 96 h. Survival of aphids fed the mutant line DC283 (*ucp1::Tn5*) was the same as that for the no-bacteria control.

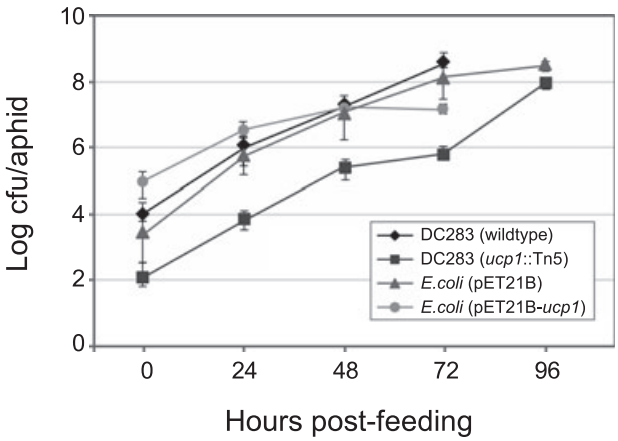

Fig. 1. Growth of *ucp1*-proficient and *ucp1*-deficient strains in the aphid. By 72 h, DC283r grows to densities > 10^8 cfu, whereas *E. coli* (pET21b-*ucp1*) grows to 10^7 cfu. Aphids feeding on either of these strains do not survive beyond 72 h. In contrast, aphids fed the *ucp1* mutant of DC283r or the *E. coli* empty vector control (pET21b) survive for more than 96 h, equivalent to aphids fed on no-bacteria control (not shown). Each data point represents the average of nine aphids, three aphids from each of three independent feeding plates. Error bars show standard deviation.

Functional features of Ucp

Ucp contains a predicted kinase domain, similar to those in the phosphofructokinase B family (PS00583), as well as three transmembrane (TM) domains. Despite its current annotation as a member of the RHS family of proteins (COG3209), it does not possess the 22-residue repeat containing the YD motif that characterizes this family, but can instead be placed in the DUF808 protein family (domain of unknown function). Similarity searches of *ucp1*-like genes in the DC283 genome yielded six potential homologues, ranging from 1200 to 1300 bp in length, which we have named *ulg1–6* (*ucp1*-like gene). The seven genes are similar at their 5′ ends (i.e. the initial 663 bp) but differ from one another in that they have unique 3′ regions. A particularly striking feature was the variation in GC contents across each of the seven genes: the conserved N-terminus has a GC content of 64% while the 3′ end averages only 42% GC (Fig. 2). A similar chimeric pattern of compositional variation is apparent in all members of this gene family. Like Ucp, each of the Ulg proteins contains a positionally conserved predicted kinase domain, but they are variable in the number and location of their predicted TM domains. Most have TM regions between amino acids 137–159 (TM1), 165–185 (TM2) and 206–225 (TM3), except for Ulg1, which has an additional predicted TM domain (TM0) between amino acids 98–118, and whose TM2 lies between amino acids 174–194. Ulg3 lacks TM3, and Ulg4 has a predicted F-box in its C-terminus, a domain with potential role in protein–protein interactions.

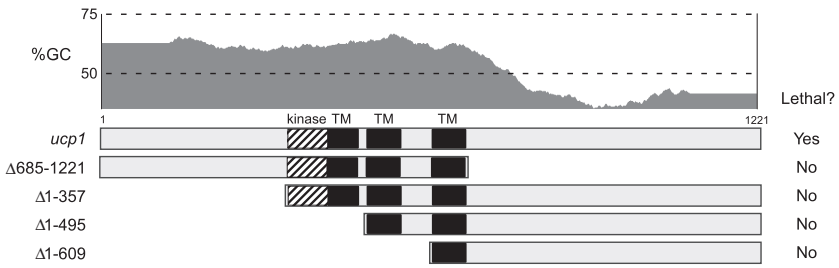

Fig. 2. Genetic structure and predicted motifs of *ucp1*. *ucp1* possesses a putative kinase domain, and three transmembrane (TM) domains, all within the GC-rich 5′ end of the gene. Portions of *ucp1* encompassing key domains of the genes do not confer pathogenicity when expressed in *E. coli*, suggesting that the entire gene is necessary for proper protein structure or function.

The *ucp1*-like genes are clustered in the DC283 genome, with *ucp1* and *ulg1* being arranged convergently and separated by two genes commonly associated with mobile elements (Fig. 3). Disruption of *ucp1* is therefore less likely to have polar effects. *ulg2*, *ulg3* and *ulg4* are organized co-linearly, with the latter two ORFs overlapping and possibly co-transcribed from a single regulatory promoter. Upstream of these three genes is a putative nematicidal gene belonging to the RHS/YD repeat family. *ulg5* and *ulg6* are adjacent but separated by 625 bp, suggesting that they are controlled by distinct promoters.

The majority of the *ucp1* coding sequence is contained within genes annotated as RHS-like, nematicidal, or hypothetical in the enteric bacteria *Yersinia intermedia* (ZP_00832839), *Erwinia tasmiensis* (YP_001906495, YP_001908938), *Hamiltonella defensa* (ACJ10121), *Providencia stuartii* (ZP_02960661) and *Xenorhabdus bovienii* (CAC19493). These genes are usually much longer than *ucp1*, and their products often bear characteristics of the RHS/YD repeat family of proteins. Similarly, *ulg1* is similar over its entire length to the C-terminus of the *Erwinia tasmaniensis* nematicidal gene 1 ($E = 1e-79$, YP_001908938), whereas both *ulg3* and *ulg6* show full-length similarity to the C-terminus of the *Xenorhabdus bovienii* nematicidal gene 2 ($E < 8e-54$, CAC19493).

ucp1 *is sufficient for aphid pathogenicity*

To determine whether *ucp1* was sufficient for entomopathogenicity, aphids were allowed to feed on a recombinant strain of *E. coli* that constitutively expressed *ucp1*. Identical to the observed effect of DC283 on aphids, the *E. coli* expressing *ucp1* was rendered toxic to aphids, reaching bacterial titres of 10^7 cfu/aphid and resulting in death by 72 h (Fig. 1). In contrast, *E. coli* vector control (pET-21b) surpassed 10^8 cfu, with aphids surviving beyond 96 h. To determine the pathogenic potential of each of the *ulg* alleles, we tested the toxicity of *E. coli* expressing each of these genes in similar lethality assays. In each case, these lines behaved like the vector control and were avirulent to aphid hosts. We also assayed the pathogenicity of these constructs in the monocot aphids *Sitobion avenae* and *Schizaphis graminum*, and as with the pea aphid, *E. coli*-expressing *ucp1* are lethal to both species, with death occurring in 3 days.

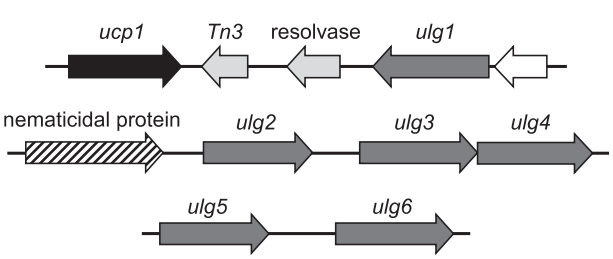

Fig. 3. Genomic context and organization of *ucp1* and related *ulg* family members. Genes are clustered and confined to three regions of the DC283 genome. With the exception of *ucp1* and *ulg1*, all other clusters are co-linear. *ucp1* and *ulg1* are separated by mobile genetic elements, and a putative nematicidal protein flanks *ulg2*.

Given the chimeric nature of *ucp1* and the *ulg* family, and the fact that expression of only *ucp1* is lethal to aphids, we attempted to determine if the C-terminus alone was responsible for Ucp1 functionality. Aphids feeding on *E. coli* expressing different portions of *ucp1* (Fig. 2) survived beyond those aphids feeding on *E. coli* (*ucp1*), indicating that none of the regions tested was sufficient for entomopathogenicity.

Pathology and Ucp1 mode of action

By observing aphids during the pathogenicity assays, it became apparent that honeydew production was abated drastically in aphids feeding on DC283 (Fig. S1). Although originally attributed to the cessation of feeding due to infection, aphids were immobile and tapped into their food source continually over the 72 h period, as if feeding actively during this time. Microscopic examination of the aphids after 48 h revealed a non-turgid body and a sunken abdomen characteristic of starved aphids. A similar pathology was observed for aphids feeding on *E. coli* expressing *ucp1* (although honeydew production was not affected as dramatically). In contrast, DC283 (*ucp1::Tn5*)-fed aphids produced substantially more honeydew and remained fully engorged over the same 72 h period.

Microscopic examination of the gut from aphids feeding on DC283 revealed the formation of solid aggregates in the crop and hindgut (Fig. 4, 1A and 1B). Bacterial aggre-

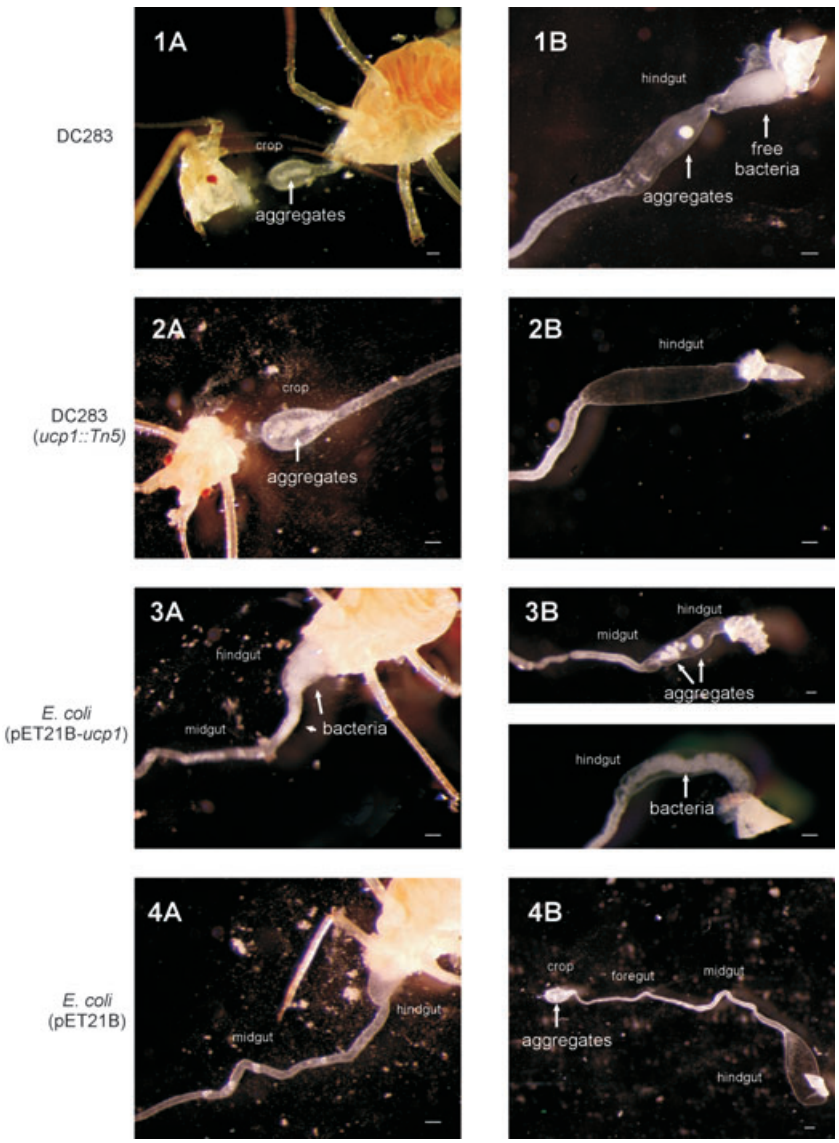

Fig. 4. Histological examination of infected aphids. Bacterial aggregates form in the crop of aphids fed DC283 (1A), with spherical aggregates and free bacteria accumulating in the hindgut (1B). Bacterial aggregates appear in the crop of aphids fed on the *ucp1* mutant of DC283 (2A), but the hindgut remains clear (2B). The hindgut of aphids fed on *E. coli* expressing *ucp1* is filled with bacterial aggregates (3A and 3B). As in aphids fed the *ucp1* mutant of DC283, the hindgut of aphids fed on the *E. coli* vector control remains clear (4A and 4B), although bacterial aggregates sometimes develop in the crop (4B). Scale bar is 200 μm in each panel. The crop and hindgut of aphids fed on a no-bacteria control contain no aggregates and are entirely clear (not shown).

gates began to form in the crop by 24 h, and by 48 h, our dissections revealed large spherical aggregates in the hindgut that were often suspended in a cloudy haze of bacteria. During dissection, aggregates often flowed out of the crop, moving through the foregut and midgut, and settling in the hindgut, indicating that they were not permanently bound to either the crop or gut epithelium. Such aggregates were also observed in the crop of aphids feeding on DC283 (*ucp1::Tn5*) (Fig. 4, 2A) but were never present in the hindgut (Fig. 4, 2B). In some preparations, the crop was completely filled with bacteria, which could be released by puncturing the crop (Fig. 5).

Because *E. coli* (*ucp1*) was also lethal to aphids, we looked for commonalities between the gut of aphids fed on *E. coli* (*ucp1*) with those fed DC283. Both produced bacterial aggregates in the crop and the hindgut, with many preparations being filled from crop to hindgut with an opaque tubular formation of bacteria (Fig. 4, 3A and 3B). Much like that of DC283 (*ucp1::Tn5*)-fed aphids, the gut of aphids fed the *E. coli* vector control contained few aggregates in the crop but none in the hindgut (Fig. 4, 4A and 4B). Aphids fed on diet without bacteria lacked aggregates in their crop and gut (data not shown).

Discussion

We identified a single gene, *ucp1*, responsible for DC283 pathogenicity in aphids. Disruption of *ucp1* completely abolishes the virulence attributes of DC283, and the heterologous expression of this one gene is sufficient to render a non-pathogenic strain of *E. coli* virulent. The formation of prominent aggregates in the crop and hindgut of aphids fed *ucp1*-containing bacteria suggests a primary role for Ucp1 in bacterial adhesion and aggregation. This

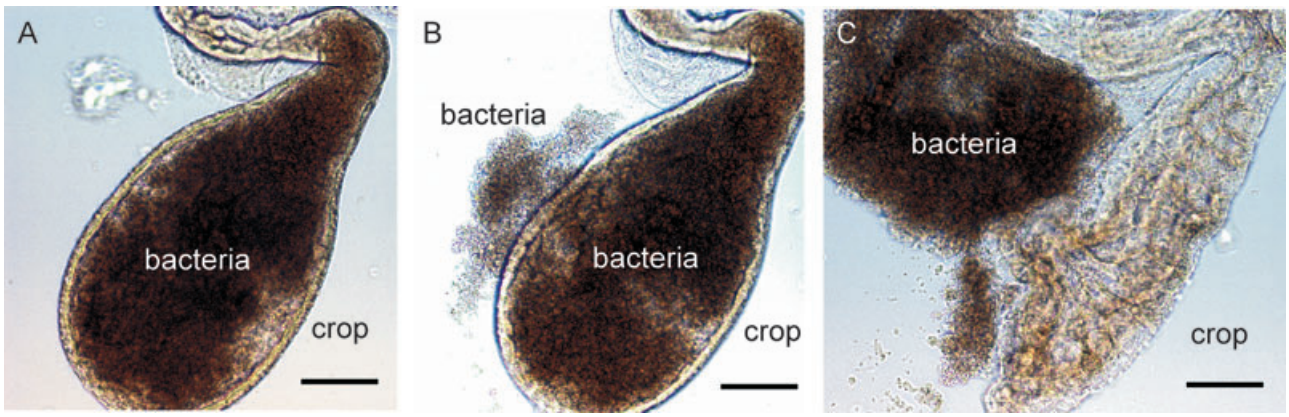

Fig. 5. Crop from an aphid infected with *E. coli* (pET21b-*ucp1*).
A. Crop filled with ingested bacteria.
B. After puncturing, bacteria seep out of the crop.
C. Crop is eventually purged of bacteria.

activity is supported by the structural and functional features of *ucp1* predicted from its sequence: Ucp1 contains several putative TM domains, indicating membrane localization and a possible role in substrate or matrix binding.

Based on its sequence similarity to portions of full-length RHS family members, *ucp1* was originally believed to encode a putative RHS family protein; however, it is substantially smaller than most genes encoding RHS/YD repeat proteins and does not possess the ligand-binding YD repeat motif characteristic of this family (Lin *et al*., 1984; Sadosky *et al*., 1989; Minet *et al*., 1999). Still, Ucp1 and the Ulg relatives retain several structural and organizational features ascribed to the original RHS family of proteins from *E. coli*, including mosaicism marked by conserved 5′-cores with divergent and seemingly non-homologous 3′ ends, and drastically different base compositions at their 5′ and 3′ ends (Lin *et al*., 1984; Sadosky *et al*., 1989; Zhao *et al*., 1993; Hill *et al*., 1994; Zhao and Hill, 1995; Hill, 1999). Although the functional implications of the disparate GC composition between 5′ and 3′ ends of *ucp1* are unclear, the evolution of seven related proteins with non-homologous 3′ ends may be significant from the perspective of host–pathogen interactions.

Based on its predicted secondary structure, we propose that Ucp1 and its relatives are membrane bound, with the hypervariable C-terminus being exposed to the extracellular environment (Fig. 6). Our proposed model is complicated by the predicted N-terminal kinase domain, which would be localized to the periplasm instead of the cytoplasm (Fig. 6). Nonetheless, if our model is correct, *ucp1* could function in manner analogous to the Microbial Surface Component Recognizing Adhesive Matrix Molecules (MSCRAMM) family of adhesin proteins used by animal pathogens, which bind to key proteinaceous components of the eukaryotic host cell to facilitate pathogenesis (Patti *et al*., 1994). The unique Ucp C-terminus may bind components of the aphid gut epithelial cells, and because only *ucp1* can confer pathogenicity towards the aphid, the related Ulg proteins could function to bind alternative matrix molecules in different hosts. Under this scenario, the variability of the exposed C-terminus may be the result of selection imposed by host immunity (Korotkova *et al*., 2007). Similar patterns are seen in type III effectors of pathogenic bacteria, which are shuffled into new variants by the process of terminal re-assortment (Stavrinides *et al*., 2006; 2008; McCann and Guttman, 2008), and by the *Neisseria gonorrheae* pilin gene, which is shuffled into new variants via recombination with inactive pilin gene variants in the genome (Seifert *et al*., 1988). Alternatively, the C-terminus of Ucp1 does not bind eukaryotic proteins, but rather other exposed Ucp C-termini of nearby cells, linking them together into bacterial matrices.

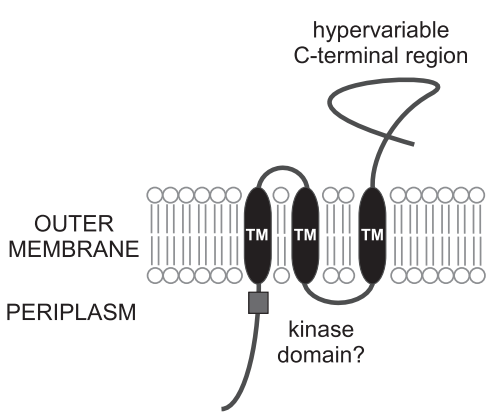

Fig. 6. Proposed model of Ucp structure and localization. Three transmembrane (TM) domains loop through the outer membrane of the Gram-negative cell, with the C-terminal hypervariable region exposed to the extracellular environment and the putative N-terminal kinase domain localized to the periplasm.

Because the RHS members share highly similar N-terminal regions, the distinct phenotype conferred by Ucp1 is specified by its C-terminus; however, the C-terminus alone does not advance aphid lethality and the aggregation phenotype, as it was insufficient for protein functionality (Fig. 2). Even fragments that incorporated all three TM domains and the predicted kinase domain did not retain functionality (Fig. 2), suggesting the N-terminus possesses key regions for protein function or conformational folding. Nonetheless, it is clear that Ucp promotes the formation of spherical bacterial aggregates in the aphid alimentary tract, as well as the accumulation of free bacteria, obstructing the normal flow through the aphid gut and reducing honeydew expulsion.

Although we had originally predicted that no aggregates would form in *ucp1*-deficient bacteria, some aggregates were observed in aphids fed either *E. coli* empty-vector control or DC283 (*ucp1::Tn5*). In these cases, however, the aggregates were confined to the crop, and the hindgut was always entirely clear, unlike the situation in aphids fed on bacteria expressing *ucp1*. This suggests that the aggregates of *ucp1*-deficient bacteria may erode in the crop such that they do not impede the flow of honeydew. Aggregates of *ucp1*-expressing bacteria may therefore be more resilient, accumulating until they begin to block the natural flow of honeydew. It appears that the aggregation phenotype of *ucp1*-proficient strains is not simply due to *ucp1* increasing bacterial growth, since bacteria lacking the gene can grow to higher densities in the aphid (Fig. 1).

The ecological significance of the DC283–aphid interaction remains uncertain since ingested DC283 are trapped within a dead-end host and are not readily transmitted to new hosts. In contrast, the interaction between *P. syringae* and the pea aphid described previously appears ecologically relevant, since bacteria acquired from plant surfaces by the aphid can replicate within the aphid host, and eventually escape to the plant surface via the aphid honeydew prior to host death (Stavrinides *et al.*, 2009). In the DC283–aphid interaction, the abatement of honeydew production coupled with aphid lethality effectively prevents escape into the environment, hindering bacterial dissemination and reducing overall fitness. Still, the recurrent isolation of enteric phytopathogens from aphids suggests that colonization of this insect host is frequent. Recently, the aphid pathogen *E. aphidicola* was identified as a plant pathogen that can infect both snap bean (*Phaselous vulgaris*) and fava bean (*Vicia faba*) (Santos *et al.*, 2009), plants that commonly serve as hosts to pea aphids (Sohi and Swenson, 1964; Blackman and Eastop, 2000; Nault *et al.*, 2004). In addition, *E. aphidicola*, was shown originally to group phylogenetically with phytopathogenic *Erwinia ananas* and *Erwinia herbicola* (Harada *et al.*, 1997), both of which have since been re-assigned to the genus *Pantoea* (Mergaert *et al.*, 1993), suggesting species of *Pantoea* are commonly associated with aphids.

The ability of *P. stewartii* to colonize and aggregate in the aphid gut so effectively lends additional support to the idea that enteric plant pathogens were originally associated with insect hosts, but may have evolved the potential to colonize plants after their frequent deposition into the phyllosphere by insects (Stavrinides, 2009). The results of this study illustrate that the enteric phytopathogens, although typically considered restricted to plant hosts, may retain the genetic means for establishing and colonizing the intestinal environment of alternative non-plant hosts with which they share a common niche and are likely to encounter frequently.

Experimental procedures

Aphid lines, bacterial strains and growth conditions

Cultures of DC283r, a spontaneous rifampicin-resistant mutant of DC283 that was used for all experiments in this study, were grown in Luria–Bertani (LB) medium at 25°C, and *Escherichia coli* BL21 (DE3) was grown in LB medium at 37°C (Table S1). Where applicable, antibiotics were supplemented at the following concentrations: ampicillin (150 μg ml^{-1}), chloramphenicol (64 μg ml^{-1}), rifampicin (50 μg ml^{-1}) and kanamycin (50 μg ml^{-1}). The pea aphid, *Acyrthosiphon pisum* strain 5 A, was maintained at 20°C on fava bean plants (*Vicia faba* L.). For screening and growth assays, aphids were maintained on an artificial diet containing essential amino acids, vitamins and sucrose (Febvay *et al.*, 1988). Fava bean plants were grown by sowing fava bean seeds in 1:1 ratio of Sunshine Mix to vermiculite.

Transposon mutagenesis and screening

Transposon mutagenesis of DC283 was performed using the mini-Tn5 vector pBSL118 in *E. coli* VPE42 (carrying kanamycin and ampicillin resistance), which was mobilized into the recipient DC283r using the helper plasmid RK600 (chloramphenicol resistance) (Table S1). To facilitate this triparental mating, the three parental strains were combined in a 1:1:1 ratio, spotted on LB agar plates (without antibiotics) and incubated overnight at 28–30°C. Traces of bacterial lawns were resuspended in 10 mM $MgSO_4$, and serial dilutions plated on LB medium containing rifampicin and kanamycin. Screening of single transconjugants was performed in 96-well microtitre plates as described previously (Stavrinides *et al.*, 2009). Plates were kept at ambient temperature and scored after 3–4 days. Transconjugants recovered from wells of surviving aphids were confirmed to be DC283 via PCR and re-assayed eight times against aphids. A total of over 1100 mutants were screened.

Mutated genes were identified by inverse PCR (Ochman *et al.*, 1988) as follows: genomic DNA from each mutant was extracted using the Gentra Puregene Kit (Qiagen, CA, USA), ~1 μg of gDNA was digested using either HincII and EcoRI in

a 20 μl reaction volume, and 10 μl of the digest used in a 200 μl unimolecular ligation with 15 units of T4 DNA ligase (Fermentas, MD, USA). After incubation at 16°C overnight, the ligation was filtered with the Qiagen PCR purification kit (Qiagen) and eluted to a final volume of 30 μl. Individual PCRs were performed on 2, 5 and 10 μl of the purified ligation, using primers npt-41 and npt+772 to test for circularized molecules (Table S2).

Cloning and transformations

The *ucp1* and *ulg* genes, along with all *ucp1* deletion derivatives (Δ685-1221), Δ1-357, Δ1-495 and Δ1-609), were amplified with *Pfu*-Turbo (Stratagene, California, USA) using gene-specific primers (Table S2) and purified with the Qiagen PCR purification kit. PCR products and the pET-21b(+) expression vector (Novagen, New Jersey, USA) (Table S1) were each subjected to 16 h restriction digests with EcoRI and HindIII using 5–10 units of enzyme in a 40 μl reaction volume, and purified with the Qiagen PCR purification kit. PCR products were ligated into the pET-21b(+) vector, and the ligation mix transformed directly into competent *E. coli* BL21 (DE3) (Chung *et al*., 1989). Transformants were selected on LB containing ampicillin, and positives were confirmed via PCR in a set of reactions employing two vector primers, one vector and one insert-specific primer, and two insert-specific primers.

Bacterial growth and virulence assays

Second and third instar aphids were allowed to feed for 17–18 h on artificial diet containing 10^7 cfu ml^{-1} of bacteria, after which, aphids were transferred to fresh diet that lacked bacteria. Three aphids were sampled at 24 h intervals, macerated together in 100 μl of 10 mM $MgSO_4$ and dilutions plated on LB containing rifampicin. Three independent growth assays were performed.

Qualitative virulence assays were conducted on second and third instar aphids, which were allowed to feed on artificial diet containing 10^7 cfu ml^{-1} of bacteria. For assays examining the contribution of *ulg*1–6 to pathogenicity, and determining the functional domain of *ucp1*, IPTG was supplemented in all treatments (including the no-bacteria control) to a final concentration of 0.5 mM. Aphids were maintained on assay plates at ambient temperature for the duration of the assay. Five independent assays were performed.

Genomics and sequence analysis

Identification of *ucp*1 and related sequences within the DC283 genome was achieved by BLASTn and tBLASTx against a customized database constructed from sequence data obtained from Baylor College of Medicine Sequencing Center (http://www.hgsc.bcm.tmc.edu). The *ucp1* and *ucp1*-like homologues in other species were recovered by BLASTn and BLASTp searches against the nr database. The DNA sequence of *Pantoea stewartii* ssp. stewartii DC283 was supported by 2004-35600-14174 from the USDA to George Weinstock (BCM-HGSC), Susanne von Bodman (University of Connecticut, CT, USA), Nicole Perna (University of Wisconsin, Madison, WI, USA) and David L. Coplin (The Ohio State University, Columbus, OH, USA). Motif and domain architecture of all proteins were analysed using both InterproScan (Zdobnov and Apweiler, 2001) and PredictProtein (Rost *et al*., 2004).

Microscopy

Aphid dissections were conducted under a Leica dissecting microscope retrofitted with a Nikon camera. Aphid guts were removed and dissected using two fine-tipped tweezers by pinching and pulling either the head or the last segment on the abdomen directly in phosphate buffered saline. Higher magnification images were taken on a Nikon light/fluorescence compound microscope.

Acknowledgements

We acknowledge gratefully the technical and intellectual contributions of Ryan Wong and Nancy Moran. We would also like to thank Dave Coplin for generously providing DC283, and the members of the *Pantoea stewartii* ssp. stewartii DC283 genome project for making the genome sequence available prior to publication. This work was supported by NIH Grant GM56120 to Howard Ochman. John Stavrinides was supported by a Postdoctoral Fellowship from the National Sciences and Engineering Research Council of Canada.

References

Blackman, R.L., and Eastop, V.F. (2000) *Aphids on the World's Crops: An Identification and Information Guide*. London, UK: J. Wiley.

von Bodman, S.B., Bauer, W.D., and Coplin, D.L. (2003) Quorum sensing in plant-pathogenic bacteria. *Annu Rev Phytopathol* **41:** 455–482.

Bravo, A., Gill, S.S., and Soberon, M. (2007) Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* **49:** 423–435.

Buell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T., Gwinn, M.L., *et al.* (2003) The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci USA* **100:** 10181–10186.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli* – transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* **86:** 2172–2175.

Coplin, D.L., Frederick, R.D., Majerczak, D.R., and Haas, E.S. (1986) Molecular cloning of virulence genes from *Erwinia stewartii*. *J Bacteriol* **168:** 619–623.

Degnan, P.H., and Moran, N.A. (2008) Diverse phage-encoded toxins in a protective insect endosymbiont. *Appl Environ Microbiol* **74:** 6782–6791.

Febvay, G., Delobel, B., and Rhabe, Y. (1988) Influence of amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (*Homoptera*: *Aphididae*). *Can J Zool* **66:** 2449–2453.

Feil, H., Feil, W.S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., *et al.* (2005) Comparison of the complete

genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *Proc Natl Acad Sci USA* **102:** 11064–11069.

Grenier, A.M., Duport, G., Pages, S., Condemine, G., and Rahbe, Y. (2006) The phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) is a pathogen of the pea aphid. *Appl Environ Microbiol* **72:** 1956–1965.

Harada, H., and Ishikawa, H. (1997) Experimental pathogenicity of *Erwinia aphidicola* to pea aphid, *Acyrthosiphon pisum*. *J Gen Appl Microbiol* **43:** 363–367.

Harada, H., Oyaizu, H., and Ishikawa, H. (1996) A consideration about the origin of aphid intracellular symbiont in connection with gut bacterial flora. *J Gen Appl Microbiol* **42:** 17–26.

Harada, H., Oyaizu, H., Kosako, Y., and Ishikawa, H. (1997) *Erwinia aphidicola*, a new species isolated from pea aphid, *Acyrthosiphon pisum*. *J Gen Appl Microbiol* **43:** 349–354.

Hill, C.W. (1999) Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. *Res Microbiol* **150:** 665–674.

Hill, C.W., Sandt, C.H., and Vlazny, D.A. (1994) Rhs elements of *Escherichia coli* – a family of genetic composites each encoding a large mosaic protein. *Mol Microbiol* **12:** 865–871.

Hurst, M.R.H., Glare, T.R., Jackson, T.A., and Ronson, C.W. (2000) Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. *J Bacteriol* **182:** 5127–5138.

Korotkova, N., Chattopadhyay, S., Tabata, T.A., Beskhlebnaya, V., Vigdorovich, V., Kaiser, B.K., *et al.* (2007) Selection for functional diversity drives accumulation of point mutations in Dr adhesins of *Escherichia coli*. *Mol Microbiol* **64:** 180–194.

Leigh, J.A., and Coplin, D.L. (1992) Exopolysaccharides in plant–bacterial interactions. *Annu Rev Microbiol* **46:** 307–346.

Lin, R.J., Capage, M., and Hill, C.W. (1984) A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K12 chromosome. *J Mol Biol* **177:** 1–18.

Lindeberg, M., Myers, C.R., Collmer, A., and Schneider, D.J. (2008) Roadmap to new virulence determinants in *Pseudomonas syringae*: insights from comparative genomics and genome organization. *Mol Plant-Microbe Interact* **21:** 685–700.

McCann, H.C., and Guttman, D.S. (2008) Evolution of the type III secretion system and its effectors in plant–microbe interactions. *New Phytol* **177:** 33–47.

Mergaert, J., Verdonck, L., and Kersters, K. (1993) Transfer of *Erwinia ananas* (synonym, *Erwinia uredovora*) and *Erwinia stewartii* to the genus *Pantoea* emend as *Pantoea ananas* (Serrano 1928) comb-nov and *Pantoea stewartii* (Smith 1898) comb-nov, respectively, and description of *Pantoea stewartii* subsp indologenes subsp nov. *Int J Syst Bacteriol* **43:** 162–173.

Minet, A.D., Rubin, B.P., Tucker, R.P., Baumgartner, S., and Chiquet-Ehrismann, R. (1999) Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene Ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci* **112:** 2019–2032.

Nault, B.A., Shah, D.A., Dillard, H.R., and McFaul, A.C. (2004) Seasonal and spatial dynamics of alate aphid dispersal in snap bean fields in proximity to alfalfa and implications for virus management. *Environ Entomol* **33:** 1593–1601.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120:** 621–623.

Patti, J.M., Allen, B.L., McGavin, M.J., and Hook, M. (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **48:** 585–617.

Rost, B., Yachdav, G., and Liu, J.F. (2004) The PredictProtein server. *Nucleic Acids Res* **32:** W321–W326.

Sadosky, A.B., Davidson, A., Lin, R.J., and Hill, C.W. (1989) *rhs* gene family of *Escherichia coli* K12. *J Bacteriol* **171:** 636–642.

Santos, M., Diánez, F., Miñano, J., Marín, F., Martínez, S., de Cara, M., and Tello, J.C. (2009) First report of *Erwinia aphidicola* from *Phaseolus vulgaris* and *Pisum sativum* in Spain. *New Dis Rep* **19:** 13–14.

Seifert, H.S., Ajioka, R.S., Marchal, C., Sparling, P.F., and So, M. (1988) DNA transformation leads to pilin antigenic variation in *Neisseria gonorrhoeae*. *Nature* **336:** 392–395.

Sohi, S.S., and Swenson, K.G. (1964) Pea aphid biotypes differing in bean yellow mosaic virus transmission. *Entomol Exp Appl* **7:** 9–14.

Stavrinides, J. (2009) Origin and evolution of phytopathogenic bacteria. In *Plant Pathogenic Bacteria: Genomics and Molecular Biology*. Jackson, R.W. (ed.). Norfolk, UK: Caister Academic Press, p. 330.

Stavrinides, J., Ma, W.B., and Guttman, D.S. (2006) Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens. *PLoS Pathog* **2:** 913–921.

Stavrinides, J., McCann, H.C., and Guttman, D.S. (2008) Host-pathogen interplay and the evolution of bacterial effectors. *Cell Microbiol* **10:** 285–292.

Stavrinides, J., McCloskey, J.K., and Ochman, H. (2009) Pea aphid as both host and vector for the phytopathogenic bacterium *Pseudomonas syringae*. *Appl Environ Microbiol* **75:** 2230–2235.

Vallet-Gely, I., Lemaitre, B., and Boccard, F. (2008) Bacterial strategies to overcome insect defences. *Nat Rev Microbiol* **6:** 302–313.

Wang, Y.D., Zhao, S., and Hill, C.W. (1998) Rhs elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*. *J Bacteriol* **180:** 4102–4110.

Waterfield, N.R., Bowen, D.J., Fetherston, J.D., Perry, R.D., and ffrench-Constant, R.H. (2001) The *tc* genes of *Photorhabdus*: a growing family. *Trends Microbiol* **9:** 185–191.

Zdobnov, E.M., and Apweiler, R. (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847–848.

Zhao, S., and Hill, C.W. (1995) Reshuffling of RHS components to create a new element. *J Bacteriol* **177:** 1393–1398.

Zhao, S., Sandt, C.H., Feulner, G., Vlazny, D.A., Gray, J.A., and Hill, C.W. (1993) *rhs* elements of *Escherichia coli* K12 – complex composites of shared and unique components that have different evolutionary histories. *J Bacteriol* **175:** 2799–2808.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Honeydew excretion by aphids ingesting *ucp1*-proficient and *ucp1*-deficient DC283. Aphids feeding on wild-type DC283 excrete less honeydew than those feeding on DC283 (*ucp1::Tn5*). Aphids feeding on DC283 (*ucp1::Tn5*) produce equivalent amounts of honeydew to those feeding on the no-bacteria control. Error bars show standard deviation.
Table S1. Strains and constructs.
Table S2. Primers.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.