Phylogenetic evidence for extensive horizontal gene transfer of type III secretion system genes among enterobacterial plant pathogens

Marianna Naum,1,2 Eric W. Brown2 and Roberta J. Mason-Gamer1

1Department of Biological Sciences, University of Illinois at Chicago, Chicago, IL 60611, USA
2Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration (FDA), College Park, MD 20740, USA

This study uses sequences from four genes, which are involved in the formation of the type III secretion apparatus, to determine the role of horizontal gene transfer in the evolution of virulence genes for the enterobacterial plant pathogens. Sequences of *Erwinia*, *Brenneria*, *Pectobacterium*, *Dickeya* and *Pantoaea* were compared (a) with one another, (b) with sequences of enterobacterial animal pathogens, and (c) with sequences of plant pathogenic γ and β proteobacteria, to evaluate probable paths of lateral exchange leading to the current distribution of virulence determinants among these micro-organisms. Phylogenies were reconstructed based on *hrcC*, *hrcR*, *hrcJ* and *hrcV* gene sequences using parsimony and maximum-likelihood algorithms. Virulence gene phylogenies were also compared with several housekeeping gene loci in order to evaluate patterns of lateral versus vertical acquisition. The resulting phylogenies suggest that multiple horizontal gene transfer events have occurred both within and among the enterobacterial plant pathogens and plant pathogenic γ and β proteobacteria. *hrcJ* sequences are the most similar, exhibiting anywhere from 2 to 50 % variation at the nucleotide level, with the highest degree of variation present between plant and animal pathogen sequences. *hrcV* sequences are conserved among plant and animal pathogens at the N terminus. The C-terminal domain is conserved only among the enterobacterial plant pathogens, as are the *hrcC* and *hrcR* sequences. Additionally, *hrcJ* and *hrcV* sequence phylogenies suggest that at least some type III secretion system virulence genes from enterobacterial plant pathogens are related more closely to those of the genus *Pseudomonas*, a conclusion neither supported nor refuted by *hrcC* or *hrcR*.

INTRODUCTION

Type III secretion systems (T3SSs) are used by a wide array of plant and animal pathogenic Gram-negative bacteria to secrete and inject virulence factors directly into the cytosol of eukaryotic host cells (Galan & Collmer, 1999; Madden et al., 2001). The T3SSs of *Yersinia*, *Escherichia*, *Salmonella* and *Shigella* have been extensively studied largely due to the potentially devastating health effects that these pathogens pose to human life (Galan & Collmer, 1999; Groisman, 2001; Tampakaki et al., 2004). Interestingly, these same pathogenicity systems are also used by plant pathogens (e.g. *Erwinia*, *Pseudomonas*, *Xanthomonas* and *Ralstonia*) to infect numerous agriculturally important plants (Chen, 2006; Shumann, 1991; van der Zet & Beer, 1995), sometimes leading to millions of US dollars worth of losses each year among certain agricultural commodities.

T3SSs encode approximately 20–25 gene products, nine of which are conserved among plant and animal pathogens and are used in the formation of the needle complex, a unique T3SS structure (Fig. 1a) that is characteristic of this contact-dependent secretion mechanism (Bogdanove et al., 1996; Gophna et al., 2003). Of these nine conserved gene products, one is an outer membrane protein (*hrcC*), one is periplasmic (*hrcF*), likely serving as a connector of the inner and outer membrane components of the apparatus, four are integral inner membrane proteins (*hrcR*, *S*, *T* and *U*) with periplasmic extensions, participating in the rod formation of the apparatus, and three are peripheral cytoplasmic proteins (*hrcV*, *Q* and *N*) involved in initiating secretion of effectors from the cytoplasm (Tampakaki et al., 2004). These nine structural components are so conserved among plant and animal pathogens that plant pathogenic...
T3SS proteins of *Xanthomonas* have been shown to secrete virulence effectors (Yops) of *Yersinia* species *in vitro* (Rossier et al., 1999). The T3SSs of both plant and animal pathogens are located in general on pathogenicity islands (PAIs), which appear to be horizontally transferred among disparate micro-organisms (Groisman & Ochman, 1996). These PAIs are found mostly on plasmids among the enterobacterial animal pathogens. Conversely, they appear to have been integrated into the bacterial genomes of their enterobacterial plant pathogenic counterparts (Hacker et al., 1997; Oh et al., 2005; Toth et al., 2006). The pathogenicity island of *Erwinia amylovora*, which has been recently characterized (Oh et al., 2005), consists of a 62 kb region with 60 ORFs. A region of approximately 22 kb encodes the proteins of the T3SS in three operons: *hrpA*, *hrpC* and *hrpJ* (Fig. 1b; Oh et al., 2005).

This study focuses on species of the genus *Erwinia* and the closely related genera *Brenneria*, *Pectobacterium*, *Dickeya* and *Pantoea*. This economically important group of enterobacterial plant pathogens is distributed worldwide, and they vary widely in their ecology, host range tropism, and pathogenicity (Krieg & Holt, 1985). The strains used in this study have been well characterized taxonomically and their evolutionary history has been described based on numerous housekeeping genes (Brown et al., 2000; Hauben et al., 1998; Naum et al., 2008; Young & Park, 2007). For example, Fig. 2(a) depicts a de novo evolutionary reconstruction of enterobacterial plant pathogenic species based on 16S rDNA gene sequences (Naum et al., 2008). To understand the role of horizontal gene transfer (HGT) of T3SS genes among the enterobacterial plant pathogens, we analysed sequences from four genes, *hrpC*, *hrpR*, *hrpJ* and *hrpV*. These four genes represent each of the three characterized operons (Oh et al., 2005), and encode proteins which serve different functional purposes (Fig. 1b).

**METHODS**

**Sampling.** A total of 29 strains were included in the phylogenetic analysis of the *hrpC* sequences: 28 isolates were included in *hrpR* and *hrpV* analysis, and 52 samples were included in the *hrpJ* analysis (Table 1). Gene sequences for the enterobacterial plant pathogens were generated in our laboratory with the exception of the two *Erwinia pyrifoliae* sequences, which along with the remaining samples were downloaded from GenBank. The *hrpJ* dataset included the broadest sample of taxa because sequences for this gene were alignable across plant and animal pathogenic representatives. Strains of *Erwinia*, *Brenneria*, *Pectobacterium*, *Dickeya* and *Pantoea* were obtained from the American Type Culture Collection (ATCC), United States Department of Agriculture (USDA) (Appalachian Fruit Research Center, Kearneysville, WV), and from Dr Amy Charkowski (University of Wisconsin, Madison) as frozen cultures and Luria–Bertani (LB) slants respectively. Sample authentication was confirmed based on 16S rDNA sequence analysis prior to their use (Naum et al., 2008).

**Bacterial DNA extraction and PCR amplification.** Total genomic DNA was extracted using a silica-based matrix (Instagene, Bio-Rad) as described in the manufacturer’s directions, which allows for total chromosomal isolation from Gram-negative bacteria. Briefly, bacterial cells from 24 h cultures grown on LB agar at room temperature were resuspended in Instagene matrix and heated at 56 °C for 30 min.

---

**Fig. 1.** (a) Schematic representation of the type III secretion needle complex used by Gram-negative plant and animal pathogenic micro-organisms. The four genes used to construct the phylogenies presented in this study are integral components of the needle complex, interacting with both the bacterial cell membranes and the chaperone–effector molecules secreted into the host cells. (b) Schematic representation of the characterized *hrp/hrpC* operons as described for *Erwinia amylovora* (Oh et al., 2005). For visual clarity, oblique lines were added to the figure to denote the removal of T3SS chromosomal segments not related to the four *hrp* genes analysed here.
The preparations were then vortexed vigorously and boiled for 10 min. Suspensions were pelleted and the supernatant containing the DNA was decanted into a clean microtube. The optimum amount of DNA to be used in PCR amplification reactions was determined to be 2 μl, after an optimization PCR was carried out using 20, 10, 5, 2 and 1 μl of DNA. Oligonucleotide primers were designed for amplification of the \( hrcJ \), \( hrcV \), \( hrcC \) and \( hrcR \) genes from conserved sequences flanking variable regions of the genes, using the few available \( Erwinia \), \( Brenneria \) and \( Pectobacterium \) sequences from GenBank. The primers used were: \( hrcCF \) (5'-CTG GAG CAC CAT TTC GAG TGG T-3') and \( hrcCR \) (5'-CCT AGC GCC TGG TAT CCG G-3'); \( hrcRF \) (5'-CCA ATA TGG CGC TCT ACG GCA-3') and \( hrcRR \) (5'-GCC GTC GAG CAC CGT GGT GCA-3'); \( hrcEF \) (5'-AAC GAT GCC AAG GAA G-3') and \( hrcFR \) (5'-GGG ACA AAG ACA ATC GCC A-3'); and \( hrcVF \) (5'-CAT GAT CAT CAT TCG GCC GCA-3') and \( hrcVR \) (5'-AGC TCT TGG GCC AGC TCT GGT CAT CT-3').

**Fig. 2.** (a) A de novo reconstructed model of a 16S rDNA gene phylogeny encompassing the same enterobacterial plant pathogenic strains as those used in this study. To simplify this schematic representation we indicate only the genus and not the species names. (b) Redrawing of previously published relationships of the plant pathogenic enterobacteria based on a comparative analysis of concatenated alignments of \( atpD \), \( carA \) and \( recA \) sequences using the likelihood algorithm (Young & Park, 2007). (c) ML tree resulting from the phylogenetic analysis of the concatenated \( gapA \)/\( gyrA \) sequence dataset consisting of 899 nucleotide positions, produced in GARLI and viewed in PAUP*. ML bootstrap nodal support was generated in PAUP*, and subsequent values are reported for each node in parentheses. Abbreviations for the taxa used are as follows: \( Erwinia \), \( Er \); \( Brenneria \), \( Br \); \( Pectobacterium \), \( Pe \); \( Dickeya \), \( Di \); \( Pantoea \), \( Pa \); \( Citrobacter \), \( Ci \); \( Enterobacter \), \( En \); \( Escherichia \), \( Es \); \( Ewingella \), \( Ew \); \( Hafnia \), \( Ha \); \( Klebsiella \), \( Kl \); \( Photobacterium \), \( Ph \); \( Proteus \), \( Pr \); \( Rhanella \), \( Rh \); \( Salmonella \), \( Sa \); \( Serratia \), \( Se \); \( Shigella \), \( Sh \); \( Xenorhabdus \), \( Xe \); \( Yersinia \), \( Ye \); \( Haemophilus \), \( H \). Multiple strains are denoted as 1 or 2 in the order that they have been listed in Table 3. For legibility purposes we have not italicized species names. Additionally, the plant pathogenic genera have been colour-coded for visual clarity, with \( Erwinia \) in red, \( Brenneria \) in blue, \( Pectobacterium \) in green, \( Pantoea \) in yellow and \( Dickeya \) in dark green.
<table>
<thead>
<tr>
<th>Taxon</th>
<th>Strain</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>hrcJ</td>
</tr>
<tr>
<td><strong>Erwinia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Er. amylovora</em></td>
<td>178</td>
<td>EU561269</td>
</tr>
<tr>
<td><em>Er. amylovora</em></td>
<td>552</td>
<td></td>
</tr>
<tr>
<td><em>Er. amylovora</em></td>
<td>8645</td>
<td>EU561270</td>
</tr>
<tr>
<td><em>Er. amylovora</em></td>
<td>8646</td>
<td>EU561271</td>
</tr>
<tr>
<td><em>Er. rhapontici</em></td>
<td>1</td>
<td>EU561272</td>
</tr>
<tr>
<td><em>Er. rhapontici</em></td>
<td>29283*</td>
<td>EU561273</td>
</tr>
<tr>
<td><em>Er. lapiniola</em></td>
<td>wl31</td>
<td>EU561274</td>
</tr>
<tr>
<td><em>Er. lapiniola</em></td>
<td>348</td>
<td>EU561275</td>
</tr>
<tr>
<td><em>Er. psidii</em></td>
<td>8423</td>
<td>EU561276</td>
</tr>
<tr>
<td><em>Er. psidii</em></td>
<td>8429</td>
<td>EU561277</td>
</tr>
<tr>
<td><em>Er. tracheiphila</em></td>
<td>5845</td>
<td>EU561253</td>
</tr>
<tr>
<td><em>Er. tracheiphila</em></td>
<td>33245*</td>
<td></td>
</tr>
<tr>
<td><em>Er. pyrifoliae</em></td>
<td>DSM12163</td>
<td>DQ180962</td>
</tr>
<tr>
<td><strong>Brenneria</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Br. nigrifluens</em></td>
<td>1391</td>
<td>EU561279</td>
</tr>
<tr>
<td><em>Br. nigrifluens</em></td>
<td>4789</td>
<td>EU561280</td>
</tr>
<tr>
<td><em>Br. alni</em></td>
<td>Pf20*</td>
<td>EU561281</td>
</tr>
<tr>
<td><em>Br. salicis</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Br. rubrifaciens</em></td>
<td>9136</td>
<td>EU561256</td>
</tr>
<tr>
<td><em>Br. rubrifaciens</em></td>
<td>5950</td>
<td>EU561313</td>
</tr>
<tr>
<td><em>Br. quercina</em></td>
<td>29291*</td>
<td>EU561249</td>
</tr>
<tr>
<td><em>Br. quercina</em></td>
<td>1846</td>
<td>EU561250</td>
</tr>
<tr>
<td><em>Br. quercina</em></td>
<td>1895</td>
<td>EU561302</td>
</tr>
<tr>
<td><strong>Pectobacterium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>subsp. carotovorum</td>
<td>4</td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>subsp. carotovorum</td>
<td>15713*</td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>subsp. brasiliensis</td>
<td>1692</td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>subsp. brasiliensis</td>
<td>1695</td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>subsp. odoriferum</td>
<td>SCR1 482</td>
</tr>
<tr>
<td><em>Pe. betavasculorum</em></td>
<td>Ecb1</td>
<td>EU561259</td>
</tr>
<tr>
<td><em>Pe. betavasculorum</em></td>
<td>Ecb2</td>
<td>EU561260</td>
</tr>
<tr>
<td><em>Pe. betavasculorum</em></td>
<td>SCR1 1043</td>
<td>EU561287</td>
</tr>
<tr>
<td><strong>Dickeya</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Di. dadantii</em></td>
<td>11663*</td>
<td>EU561335</td>
</tr>
<tr>
<td><em>Di. dadantii</em></td>
<td>EC16</td>
<td>EU561266</td>
</tr>
<tr>
<td><strong>Pantoea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pa. agglomerans</em></td>
<td>33243*</td>
<td>EU561336</td>
</tr>
<tr>
<td><em>Pa. stewartii</em></td>
<td>8199</td>
<td>EU561288</td>
</tr>
<tr>
<td><strong>Pseudomonas</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Ps. syringae</em></td>
<td>B728a</td>
<td>NC_007005</td>
</tr>
<tr>
<td><em>Ps. syringae</em></td>
<td>1448A</td>
<td>NC_005773</td>
</tr>
<tr>
<td><em>Ps. syringae</em></td>
<td>2</td>
<td>DC3000</td>
</tr>
<tr>
<td><em>Ps. syringae</em></td>
<td>3</td>
<td>LMG5090</td>
</tr>
<tr>
<td><em>Ps. syringae</em></td>
<td>4</td>
<td>ATCC115288</td>
</tr>
<tr>
<td><em>Ps. mendocina</em></td>
<td>ymp</td>
<td>CP000680</td>
</tr>
<tr>
<td><em>Ps. cichorii</em></td>
<td>83-1</td>
<td>DQ16884</td>
</tr>
<tr>
<td><em>Ps. viridiflava</em></td>
<td>RMX3.1b</td>
<td>AY597283</td>
</tr>
<tr>
<td><em>Ps. viridiflava</em></td>
<td>RMX23.1a</td>
<td>AY597282</td>
</tr>
<tr>
<td><em>Ps. viridiflava</em></td>
<td>ME3.1b</td>
<td>AY597281</td>
</tr>
<tr>
<td><strong>Xanthomonas</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Xa. campestris</em></td>
<td>pv. vescatora</td>
<td>AM039952</td>
</tr>
<tr>
<td><em>Xa. campestris</em></td>
<td>85-10</td>
<td>NC_007508</td>
</tr>
<tr>
<td><em>Xa. campestris</em></td>
<td>80004</td>
<td>NC_007086</td>
</tr>
</tbody>
</table>
Table 1. cont.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Strain</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>hrcJ</td>
</tr>
<tr>
<td>Xa. campestris 3</td>
<td>ATCC 33913</td>
<td>NC_003902</td>
</tr>
<tr>
<td>Xa. axonopodis</td>
<td>306</td>
<td>NC_003919</td>
</tr>
<tr>
<td>Xa. oryzae 1</td>
<td>BLS256</td>
<td>AY875714</td>
</tr>
<tr>
<td>Xa. oryzae 2</td>
<td>MAFF311018</td>
<td>AP008229</td>
</tr>
<tr>
<td>Xa. oryzae 3</td>
<td>KACC10331</td>
<td>AB115081</td>
</tr>
<tr>
<td><strong>Ralstonia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ra. solanacearum</td>
<td>GMI1000</td>
<td>AL646053</td>
</tr>
<tr>
<td>Ra. solanacearum 1</td>
<td>IPO1609</td>
<td>AJ245811</td>
</tr>
<tr>
<td>Ra. solanacearum 2</td>
<td></td>
<td>NC_003296</td>
</tr>
<tr>
<td><strong>Escherichia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex. coli</td>
<td>042</td>
<td>AF200363</td>
</tr>
<tr>
<td>Ex. coli 1</td>
<td>O157:H7</td>
<td>NC_002695</td>
</tr>
<tr>
<td>Ex. coli 2</td>
<td>E2348/69</td>
<td>AF022236</td>
</tr>
<tr>
<td>Ex. coli 3</td>
<td>101-1</td>
<td>AJ303141</td>
</tr>
<tr>
<td>Ex. coli 4</td>
<td>0181-6/86</td>
<td>AJ633129</td>
</tr>
<tr>
<td><strong>Salmonella</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sa. enterica</td>
<td>SC-B67</td>
<td>NC_006905</td>
</tr>
<tr>
<td>Sa. enterica 1</td>
<td>CT18</td>
<td>AL627271</td>
</tr>
<tr>
<td>Sa. typhimurium</td>
<td>LT2</td>
<td>NC_003197</td>
</tr>
<tr>
<td><strong>Yersinia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ye. pestis</td>
<td>CO92</td>
<td>AL117189</td>
</tr>
<tr>
<td>Ye. enterocolitica</td>
<td>8081</td>
<td>AM286415</td>
</tr>
</tbody>
</table>

*Indicates the type culture for each of the species represented as originally obtained from ATCC.

volume of 25 µl PCR was performed in a PTC-200 thermal cycler (MJ Research) under the following conditions: initial denaturation at 94 °C for 2 min, 94 °C for 50 s, 58 °C for 50 s and 72 °C for 45 s (35 cycles), ending with incubation at 72 °C for 8 min. The amplified segments were ~850 bp long for hrcC, ~500 bp long for hrcR, ~500 bp long for hrcJ, and ~1500 bp long for hrcV.

**Cloning and DNA sequencing.** PCR products were purified using QIAquick gel extraction kits (Qiagen) according to the manufacturer’s protocol, except that DNA was eluted in 20 µl buffer. Cloning reactions were completed using pGEM T Easy cloning kits (Promega) according to the manufacturer’s protocol, except that the ligation and transformation reaction volumes were halved. Cloned inserts were amplified directly from colonies using the original PCR primers and amplification conditions. PCR products were cleaned using exonuclease I and shrimp alkaline phosphatase as in Mason-Gamer (2004). Purified PCR products were sequenced in both directions using an automated Sanger dye-deoxy-chain-termination method (Amplicon Express). DNA sequences were edited and combined in Sequencher v.4.1 (Gene Codes Corporation).

**Sequence alignment and phylogenetic analyses.** DNA sequences and resultant computer-translated amino acid sequences (hrcC and hrcR) were aligned using CLUSTAL W v.1.5 (Thompson et al., 1994), followed by manual adjustments in MacClade v.4.0 (Maddison & Maddison, 2000). Amino acid codon coloration was used in MacClade v.4.0 (Maddison & Maddison, 2000) on the nucleotide datasets to ensure the resulting alignments did not alter the protein sequence. Regions judged to have an ambiguous alignment were excluded (~400 bp for hrcC, 45 %; ~120 bp for hrcR, 21 %; and ~590 bp for hrcV, 40 %) prior to the phylogenetic analyses. All datasets were first analysed in PAUP* (Swofford, 2001) using the principle of maximum parsimony (MP) under heuristic search methods using tree bisection-reconnection (TBR) branch-swapping, and 100 random taxon addition replicates. For the nucleotide datasets, the shortest trees were then used as the starting topology for the evaluation of 16 nested models of sequence evolution (Felsenstein, 1985; Swofford et al., 1997; Sullivan et al., 1995; Swofford et al., 1996). The best-fit model of substitution for each dataset was determined by the likelihood score of each under a chi-squared approximation of the null distribution (Yang et al., 1995). Parameter space was searched for the best tree with simultaneous estimation for model parameters using a maximum likelihood (ML) search conducted using the Genetic Algorithm for Rapid Likelihood Inference (GARLI) software (Zwickl, 2006). As recommended by Zwickl (2006), multiple runs were performed (~20) to ensure that results were consistent, as the algorithm is inherently stochastic. The log-likelihood values of each run were retained in order to compare the individual runs, and the tree and model parameters corresponding to the best score were used.

Branch support was determined in GARLI by 100 ML bootstrap iterations (Felsenstein, 1985), and with Bayesian posterior probability (MrBayes v.3.1.2) approximation of 1 000 000 generations, discarding 25 % (2500) of the tree samples as recommended in the MrBayes v.3.1.2 manual (Huelsenbeck & Ronquist, 2001). For Bayesian character support methods, parameters of sequence evolution estimated from the final ML tree were used (Table 2). Amino acid character support was determined in PAUP* by 1000 bootstrap iterations and 1 000 000 generations of Bayesian (MrBayes v.3.1.2; Huelsenbeck & Ronquist, 2001) posterior probability approximation under the principle of MP, again discarding 25 % (2500) of the tree samples.
Nucleotide gene trees were rooted by midpoint rooting, rather than outgroup rooting (Schoettzelp & Hoot, 2006), because the homologous gene sequences from the closest available outgroups were too divergent to align reliably. The hrcC and hrcR amino acid sequence trees were rooted with Escherichia coli (GenBank accession numbers AJ303141 and AF022236) and Yersinia pestis (GenBank accession number NC_004839), respectively. The absence of suitable outgroup taxa impeded the detection of the direction of certain HGT events between and among plant and animal pathogenic taxa. However, in these specific instances, direction was addressed inferentially.

**RESULTS**

The hrcC DNA alignment consisted of 29 taxa and 855 characters, of which 458 were informative, whereas the amino acid alignment consisted of 31 taxa, including the outgroups, and 291 characters of which 150 were informative. The final ML search of the nucleotide data resulted in one tree (Fig. 3a), and the (most parsimonious) MP analysis of the amino acid data yielded 16 equally most parsimonious trees [272 steps; consistency index (CI)=0.87; rescaled consistency (RC)=0.81], from which the strict consensus was derived (Fig. 3b). The hrcV DNA alignment consisted of 28 taxa and 508 characters, of which 384 were informative, while the amino acid alignment consisted of 29 taxa and 170 characters, of which 133 were informative. The final ML search of the nucleotide data resulted in one tree (Fig. 4a), and the MP analysis of the amino acid data yielded two equally most parsimonious trees (175 steps; CI=0.93; RC=0.97), from which a strict consensus was derived (Fig. 4b). The hrcJ alignment consisted of 52 taxa and 454 characters after truncating the beginning and end of the dataset in order to eliminate gaps resulting from the absence of data due to varying sequence lengths. The GARLI tree with the best log-likelihood score for hrcJ is presented in Fig. 5(a). The hrcR alignment consisted of 28 taxa and 1536 characters, of which only the first 910 were used in the phylogenetic analysis because the remaining 630 base pairs could only be meaningfully aligned for the enterobacterial plant pathogens. After the 800th base pair approximately, the sequences of Pseudomonas, Xanthomonas, Ralstonia, Erwinia, Salmonella and Yersinia began diverging from the sequences of Erwinia, Bremeria, Pectobacterium, Dickeya and Pantoea, and by the 910th nucleotide position, they could no longer be aligned. The GARLI tree with the best score for hrcR is presented in Fig. 5(b).
absence of data due to varying sequence lengths. The GARLI tree with the best log-likelihood score for gapA/gyrA is presented in Fig. 2(c).

Erwinia, Brenneria and Pectobacterium are non-monophyletic on the hrcC and hrcR phylogenies, a conclusion strongly supported by clades C and D (Figs 3a and 4a). Both clades contain multiple Erwinia, Brenneria and Pectobacterium taxa with bootstrap support values of 90 and 98% for clades C and D, respectively. In addition, strains of Erwinia mallotivora 1 and 2, Pectobacterium carotovorum subsp. carotovorum 1 and 2, and Pectobacterium betavasculorum 1 and 2 are broken up, each being more closely related to other Erwinia and Pectobacterium species than to each other. We resequenced hrcC and hrcR for Erwinia tracheiphila 1 and Brenneria salicis 1 to confirm that the original sequences were not artefacts, since in the hrcC phylogeny, the sequences are indistinguishable and on a long branch (Fig. 3a). Even though the hrcC and hrcR reconstructions include only enterobacterial plant pathogen representatives, we suggest a monophyletic origin of the virulence determinants comprising their T3SS because of the extensive divergence between the sequences of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoaea from sequences of the enterobacterial animal pathogens and plant pathogenic c and b proteobacteria, at both the nucleotide and the amino acid levels.

The hrcJ and hrcV trees (Fig. 5) are generally congruent (Table 4), indicating a substantial degree of evolutionary and genetic linkage between these loci. The relationships are largely unaffected by the number of representative sequences, which differs between the two phylogenies.

### Table 3. Bacterial strains and accession numbers for gene sequences used in the housekeeping gene phylogenetic analyses

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Strain number</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gapA</td>
<td>gyrA</td>
</tr>
<tr>
<td><strong>Erwinia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Er. amylovora</em></td>
<td>(Ea)178</td>
<td>EU340558</td>
</tr>
<tr>
<td><em>Er. rhapontici</em></td>
<td>1, ATCC 29283*</td>
<td>EU340561, EU340562</td>
</tr>
<tr>
<td><em>Er. mallotivora</em></td>
<td>(Em)8645, (Em)8646</td>
<td>EU340559, EU340560</td>
</tr>
<tr>
<td><em>Er. psidii</em></td>
<td>8423, (Ep)8429</td>
<td>EU340565, EU340566</td>
</tr>
<tr>
<td><strong>Brenneria</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Br. rubrifaciens</em></td>
<td>(Br)5951, 29291*</td>
<td>EU340569, EU340570</td>
</tr>
<tr>
<td><em>Br. nigrifluens</em></td>
<td>(En)1391, 4789</td>
<td>EU340571, EU340572</td>
</tr>
<tr>
<td><em>Br. alni</em></td>
<td>pfli20*</td>
<td>EU340574</td>
</tr>
<tr>
<td><em>Br. quercina</em></td>
<td>(Bq)1846, 1895</td>
<td>EU340575, EU340676</td>
</tr>
<tr>
<td><strong>Pectobacterium</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pe. carotovorum</em></td>
<td>SCR1482</td>
<td>EU340589</td>
</tr>
<tr>
<td>subsp. carotovorum</td>
<td>Eca4, 15713*</td>
<td>EU340579, EU340580</td>
</tr>
<tr>
<td>subsp. brasiliensis</td>
<td>1692, 1695</td>
<td>EU340583, EU340584</td>
</tr>
<tr>
<td>subsp. odoriferum</td>
<td>SCR11043, WPP158</td>
<td>EU340581, EU340582</td>
</tr>
<tr>
<td><em>Pe. betavasculorum</em></td>
<td>ECB1, ECB2</td>
<td>EU340587, EU340588</td>
</tr>
<tr>
<td><em>Pe. atrosepticum</em></td>
<td>SCR11043, WPP158</td>
<td>EU340585, EU340586</td>
</tr>
<tr>
<td><em>Pe. wasabiae</em></td>
<td>WPP163, SCR1488</td>
<td>EU34057, EU340577</td>
</tr>
<tr>
<td><em>Pe. cyripedii</em></td>
<td>29267*, 29269</td>
<td>EU340579, EU340580</td>
</tr>
<tr>
<td><strong>Dickeya</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Di. dadantii</em></td>
<td>11663*, EC16</td>
<td>EU340590, EU340591</td>
</tr>
<tr>
<td><strong>Pantoaea</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pa. agglomerans</em></td>
<td>33243*</td>
<td>EU340592</td>
</tr>
<tr>
<td><em>Pa. stewartii</em></td>
<td>8199, 29227</td>
<td>EU340593, EU340594</td>
</tr>
<tr>
<td><em>Pa. ananatis</em></td>
<td>11530, 19321</td>
<td>EU340595</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Shigella flexneri</em></td>
<td>5 strain 8401</td>
<td>NC_008258</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>UT189</td>
<td>NC_007946</td>
</tr>
<tr>
<td><em>Salmonella Typhimurium</em></td>
<td>LT2</td>
<td>NC_003197</td>
</tr>
<tr>
<td><em>Salmonella Typhi</em></td>
<td>Ty2</td>
<td>NC_004631</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>ATCC 13883</td>
<td>M66869</td>
</tr>
<tr>
<td><em>Serratia marcescens</em></td>
<td>DB11</td>
<td>M63373</td>
</tr>
<tr>
<td><strong>Outgroup taxon</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>86-028NP</td>
<td>NC_007146</td>
</tr>
</tbody>
</table>

*Indicates the type culture for each of the species represented as originally obtained from ATCC.
because hrcJ sequences were more conserved, allowing us to include a greater number of animal pathogen and \( \gamma \) and \( \beta \) proteobacterial plant pathogen representatives. Both trees place Xanthomonas and Ralstonia as sisters to one another and to the other plant pathogens, and they also group Escherichia and Salmonella together, and as a sister group to Yersinia. Moreover, the trees differentiate the plant and animal pathogens into two well-defined clades. Furthermore, both phylogenies group Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea as closest relatives, with Pseudomonas syringae as their sister group. Additionally, the enterobacterial plant pathogens form two well-supported clades, designated A and B (Fig. 5a). Clade A consists of Pectobacterium and Brenneria species, along with two strains of Erwinia rhapontici, while clade B is dominated by Erwinia and Pantoaea. However, in the hrcV tree (Fig. 5b), Erwinia lupinicola 1 is placed outside both clade A and clade B (although this position is not well supported), while P. carotovorum subsp. carotovorum 1 is found in clade B. Taken together, these results suggest that the hrcJ and hrcV virulence genes of the enterobacterial plant pathogens and Pseudomonas share a recent common ancestor.

Furthermore, the hrcJ and hrcV trees are congruent to one another with respect to the resulting relationships between the enterobacterial plant pathogens (Table 4). The trees are also consistent in the placement of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoaea with the other plant pathogens and not with the enteric animal pathogens, a result in agreement with and expanding on a previous study of hrcJ sequences (Brown et al., 1998). It is notable that the hrcC and hrcR trees are incongruent to one another and to both the hrcJ and hrcV phylogenies (Table 4), which could be the result of multiple interspecific HGT events among the enterobacterial plant pathogens after the acquisition of the T3SS into this lineage.

On the other hand, all of the housekeeping gene trees (Fig. 2) support multiple independent origins of the plant pathogenic genera within the Enterobacteriaceae, and all loci are consistent in their taxonomic classification of the four genera. Additionally, in all three housekeeping gene
phylogenies, it is evident that housekeeping gene evolution – taken here to depict the vertical evolution of the strains analysed – contrasts with virulence gene evolution and does not support a single origin of a plant pathogenic ancestor followed by speciation into the numerous plant pathogenic genera and species.

Finally, ILD testing comparing the virulence genes with one another and with several housekeeping loci confirmed the hypothesis of incongruence ($P=0.001$, Table 4). Strikingly, each pairwise comparison of the virulence genes with the housekeeping loci resulted in a statistically significant $P$ value ($P<0.05$, Table 4), leading us to reject the null hypothesis. This result indicates that the virulence determinants examined in this study have all been subject to some level of lateral transfer between the enterobacterial plant pathogens. The only gene dataset comparison for which we could not reject the null hypothesis (i.e. for which the dataset partitions were found to be congruent based on the ILD test) was between the hrcJ and hrcV partitions. These results support our hypothesis that both ancestral and recent HGT events have been driving the evolution of virulence in the enterobacterial plant pathogens.

**DISCUSSION**

The nine proteins involved in forming the needle complex of the T3SS are conserved across multiple animal and plant pathogens and are believed to share a common ancestor with the flagellar export mechanism used by many microorganisms (Gophna et al., 2003). However, it is evident from the results of this study that vertical acquisition (Fig. 2) has been interjected by numerous HGT events, which undeniably have been a driving force in shaping the evolutionary history of the enterobacterial plant pathogen virulence genes (Figs 3, 4 and 5).

Despite the sampling differences between the hrcJ and hrcV datasets, analyses of the conserved gene segments resulted in congruent phylogenies for the two genes (Fig. 5, Table 4), which are in conflict with the phylogenies generated from hrcC and hrcR gene sequences (Figs 3 and 4, Table 4). Unlike the hrcJ and hrcV datasets, the partial hrcC and hrcR

---

**Fig. 4.** (a) ML tree comprising hrcR gene sequences from 28 enterobacterial plant pathogen species representatives. The amplified products encompass the 210th nucleotide to the 630th nucleotide based on the *Pectobacterium atrosepticum* genome (NC_004547). ML solutions were generated in GARLI and visualized in PAUP*. ML bootstrap nodal support was generated in PAUP*, and subsequent values are reported for each node in parentheses. Bayesian nodal support values are presented for all nodes without parentheses. (b) Strict consensus tree of the two equally most parsimonious trees resulting from the hrcR amino acid dataset. Species abbreviations for the taxa used are as follows: *Erwinia*, Er; *Brenneria*, Br; *Pectobacterium*, Pe; *Dickeya*, Di; *Pantoea*, Pa; *Escherichia*, Es; *Yersinia*, Ye. Multiple strains are denoted as 1 or 2 in the order that they have been listed in Table 1. For legibility purposes we have not italicized species names.
gene sequences we studied were highly variable, much like the C-terminal region of hrcV, and only sequences from Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea could be meaningfully aligned. In addition, because the taxonomic placement of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea in the family Enterobacteriaceae has been confirmed by analysis of multiple housekeeping genes (Fig. 2; Brown et al., 2000; Hauben et al., 1998; Young & Park, 2007), it is striking that the hrcJ and hrcV trees (Fig. 5) reveal that the enterobacterial plant pathogens likely acquired their hrcJ and hrcV genes during a single HGT event from the same ancestor as Pseudomonas syringae prior to their speciation.

Assessing the evolutionary history of hrcC and hrcR gene sequences has, however, been a more difficult task. First, there are relatively few hrcC and hrcR homologues currently in the nucleotide databases (GenBank and EMBL), and obtaining an appropriately broad sample from throughout the family would have been outside the scope of this study. Second, based on available sequences, it is evident that homologous gene sequences from the enterobacterial animal pathogens as well as other plant pathogenic c and b proteobacteria are so divergent from those of Erwinia, Brenneria, Pectobacterium, Dickeya, Pantoea, Pseudomonas, Xanthomonas, Ralstonia, Escherichia, Salmonella, Yersinia that they cannot be meaningfully compared in a phylogenetic context. This extensive divergence is interpreted here as important, albeit indirect, evidence that the T3SS determinants of the enterobacterial plant pathogens together form a monophyletic group. Virulence sequences from within the enterobacterial phytopathogens are far

Table 4. Pair-wise ILD values among eight virulence loci in 12 representative enterobacterial plant pathogens

<table>
<thead>
<tr>
<th>Locus</th>
<th>P value (1000 partitions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hrcV</td>
</tr>
<tr>
<td>hrcJ</td>
<td>0.174</td>
</tr>
<tr>
<td>hrcV</td>
<td>0.001</td>
</tr>
<tr>
<td>hrcC</td>
<td>0.001</td>
</tr>
<tr>
<td>hrcR</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Fig. 5. (a) ML tree resulting from the phylogenetic analysis of the hrcJ sequence dataset including enterobacterial plant and animal pathogenic representatives, and non-enterobacterial plant pathogenic \( \gamma \) and \( \beta \) proteobacterial taxa. The amplified products encompass the 93rd nucleotide to the 414th nucleotide based on the Pectobacterium atrosepticum genome (NC_004547). (b) ML tree resulting from the phylogenetic analysis of the hrcV sequence dataset. The amplified products encompass the 124th nucleotide to the 1359th nucleotide of the Pectobacterium atrosepticum genome (NC_004547), with only the first 910 nucleotides (124–1034) included in the analyses. The resultant maximum trees for these datasets were produced in GARLI and viewed in PAUP*. ML bootstrap nodal support was generated in PAUP*, and subsequent values are reported for each node in parentheses. Bayesian nodal support values are presented for all nodes without parentheses. Species abbreviations for the taxa used are as follows: Erwinia, Er; Brenneria, Br; Pectobacterium, Pe; Dickeya, Di; Pantoea, Pa; Pseudomonas, Ps; Xanthomonas, Xa; Ralstonia, Ra; Escherichia, Es; Salmonella, Sa; Yersinia, Ye. Multiple strains are denoted as 1 or 2 in the order that they have been listed in Table 1. For legibility purposes we have not italicized species names.
more similar to one another than they are to any available sequences from animal and other plant pathogenic taxa.

A conclusion of monophyly, however, assumes a singular evolutionary origin for the enterobacterial plant pathogens that is not supported by the currently accepted taxonomy of the group (Fig. 2). The 16S rDNA, atpD/cara/recA and gapA/gyrA phylogenies suggest that the enterobacterial plant pathogenic genera of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea arose independently multiple times within the family. Therefore, if we recognize that the hypothesis of polyphyly accurately reflects the evolutionary history of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea then there must be an alternative explanation for the relationships resulting from the virulence gene phylogenies. For example, if plant pathogenicity evolved independently multiple times and was associated with a reduced rate of virulence gene evolution in the resulting lineages, then we would expect the plant pathogens to exhibit relatively high sequence similarity in spite of their polyphyletic origins. Conversely, if an animal pathogenic lineage has arisen from within a lineage of plant pathogens, and the host switch was associated with an increased rate of sequence evolution, then eventually the plant pathogens will display greater sequence similarity to one another than to the animal pathogen in spite of their paraphyly relative to the derived animal pathogen. However, both of these scenarios would require rate changes that affect hrcC and hrcR simultaneously and in the same way, while at the same time hrcJ and hrcV remain unaffected. These hypotheses would be more plausible if hrcC and hrcR encoded extracellular proteins, as it has been suggested that extracellular protein components undergo bursts of rapid polypeptide evolution to facilitate the mechanical adaptations necessary for the T3SS to interface with different eukaryotic cell types (Dale & Moran, 2006). However, since neither gene product is known to be expressed extracellularly, these explanations seem unlikely.

Alternatively, repeated exchanges of common T3SS alleles among the enteric plant pathogens could provide a similar phylogenetic outcome – that is one of monophyly – among the T3SS genes used in this study. Clearly, the repeated horizontal transfer of a few preferred alleles would result in the homogenization of population structure and subsequently result in the evolutionary divergence of the phytopathogens from closely related animal pathogenic neighbours that lie just beyond the barrier of these extensive allelic exchanges. The effects of continued interspecific HGT are clearly seen in the differing phylogenetic placement of the bacterial strains from within species among the phylogenies (Figs 3, 4 and 5). There are cases where strains of a species group together on one tree, but are separated on the other tree, suggesting that one of the two strains has acquired a phylogenetically distinct gene copy through HGT. For example, the two E. malolitivora sequences are nearly identical, and group together, on the hrcJ and hrcV trees (Fig. 5), but they are widely separated from one another on the hrcC and hrcR trees (Figs 3 and 4). The two P. carotovorum subsp. carotovorum strains, on the other hand, group closely together on the hrcJ tree, but are widely separated on the hrcC (Fig. 3a), hrcR (Fig. 4a) and hrcV (Fig. 5b) trees. Conflicts involving interspecific relationships are evident as well. For example, the grouping of E. tracheiphila 1 with B. salicis 1 at the base of the hrcC tree (Fig. 3a) is in conflict with both of their positions on the hrcR tree (Fig. 4a) (neither is included on the hrcJ or hrcV tree). A likely explanation for the conflicting placements of these bacterial strains and species is the acquisition of different gene copies. Since the strains were isolated from different individuals of the same host species, and transfer events are believed to be influenced more by physical proximity of strains than by phylogenetic relatedness (Matte-Taillez et al., 2002; Toth et al., 2006), it is possible that the bacterial strains acquired their copies of the genes from neighbouring bacteria on the same plant.

Finally, even though the lack of an appropriate outgroup has encumbered our ability to detect the direction of the noted HGT events, we can infer that the enterobacterial plant pathogens probably acquired their T3SS prior to their radiative speciation, from the same ancestor as *Pseudomonas syringae*, a conclusion that is strengthened by the observed % GC content, which is uniform between these four virulence genes and housekeeping genes such as 16S rDNA, gapA and gyrA. Thus, it is ironic and somewhat unexpected that the virulence genes may have provided an additional and important clue to buttress the hypothesis of a monophyletic plant pathogenic lineage in the family *Enterobacteriaceae*, as originally proposed by Brown et al. (2000). Moreover, the differences in the phylogenetic positions of the enterobacterial plant pathogenic species among the hrcJ, hrcV, hrcC and hrcR trees are possibly the result of continued HGT events followed by genetic recombination during and possibly after the speciation of Erwinia, Brenneria, Pectobacterium, Dickeya and Pantoea, as clearly seen in the phylogenetic placement of the enterobacterial plant pathogenic strains. Additionally, the ineffectiveness of the flagellar gene homologues as outgroups corroborates the hypothesis proposed by Gophna et al. (2003) stating that the T3SS and the flagellar export mechanism share a common ancestor but have evolved independently of one another since. Taken together, these results have provided important clues about the natural history of acquisition of virulence among a highly significant group of bacterial plant pathogens, and although the precept that HGT has forged virulence evolution is certainly not a novel one (Brown et al., 1998; Cohan, 1996; Comas et al., 2006; Dale & Moran, 2006; Gophna et al., 2003), we know of no other analysis encompassing the enterobacterial plant pathogens which has demonstrated that the diversification of the virulence operons is likely to be ancestral to the radiation of the plant pathogens themselves. The monophyly of the T3SS genes of the enterobacterial phytopathogens is evidenced by the extraordinary observed diversification at the amino acid
level (in hrcC, hrcR and hrcV), which was so advanced that the inclusion of relevant enteric homologues was largely an impossible task. This diversification among the plant and animal pathogens within the Enterobacteriaceae has likely persisted for epochs among these unique lineages, and continues to be maintained by ongoing HGT events.

ACKNOWLEDGEMENTS

Thanks to A. Charkowsky of the University of Wisconsin, Madison, for very generously providing the Pectobacterium strains used in this study, and to Diane McCarthy, Michael Jorgensen and Keith Lampel for valuable comments on an earlier version of the manuscript. Also, a special thanks to Peter Naum and Sergio Mojica for helping us create the illustration in Fig. 1. The research was supported by National Science Foundation (NSF) grant number DEB 0426194 to R. J. M.-G., and a University of Illinois at Chicago (UIC) Provosts Award to M. N.

REFERENCES


Edited by: I. K. Toth