BMC Genomics

Research article

A draft genome sequence and functional screen reveals the repertoire of type III secreted proteins of *Pseudomonas syringae* pathovar *tabaci* 11528

David J Studholme*[1], Selena Gimenez Ibanez[1], Daniel MacLean[1], Jeffery L Dangl[2], Jeff H Chang[3,4] and John P Rathjen[1]

Address: [1]The Sainsbury Laboratory, Norwich, NR4 7UH, UK, [2]Department of Biology, CB# 3280, Coker Hall, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA, [3]Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331, USA and [4]Center for Genome Research and Biocomputing, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331, USA

Email: David J Studholme* - david.studholme@tsl.ac.uk; Selena Gimenez Ibanez - selena.gimenez-ibanez@tsl.ac.uk; Daniel MacLean - dan.maclean@tsl.ac.uk; Jeffery L Dangl - dangl@email.unc.edu; Jeff H Chang - changj@cgrb.oregonstate.edu; John P Rathjen - john.rathjen@tsl.ac.uk

* Corresponding author

Abstract

Background: *Pseudomonas syringae* is a widespread bacterial pathogen that causes disease on a broad range of economically important plant species. Pathogenicity of *P. syringae* strains is dependent on the type III secretion system, which secretes a suite of up to about thirty virulence 'effector' proteins into the host cytoplasm where they subvert the eukaryotic cell physiology and disrupt host defences. *P. syringae* pathovar *tabaci* naturally causes disease on wild tobacco, the model member of the Solanaceae, a family that includes many crop species as well as on soybean.

Results: We used the 'next-generation' Illumina sequencing platform and the Velvet short-read assembly program to generate a 145X deep 6,077,921 nucleotide draft genome sequence for *P. syringae* pathovar *tabaci* strain 11528. From our draft assembly, we predicted 5,300 potential genes encoding proteins of at least 100 amino acids long, of which 303 (5.72%) had no significant sequence similarity to those encoded by the three previously fully sequenced *P. syringae* genomes. Of the core set of Hrp Outer Proteins that are conserved in three previously fully sequenced *P. syringae* strains, most were also conserved in strain 11528, including AvrE1, HopAH2, HopAJ2, HopAK1, HopAN1, HopI, HopJ1, HopX1, HrpK1 and HrpW1. However, the *hrpZ1* gene is partially deleted and *hopAF1* is completely absent in 11528. The draft genome of strain 11528 also encodes close homologues of HopO1, HopT1, HopAH1, HopR1, HopV1, HopAG1, HopAS1, HopAE1, HopAR1, HopF1, and HopW1 and a degenerate HopM1'. Using a functional screen, we confirmed that *hopO1, hopT1, hopAH1, hopM1', hopAE1, hopAR1*, and *hopAI1'* are part of the virulence-associated HrpL regulon, though the *hopAI1'* and *hopM1'* sequences were degenerate with premature stop codons. We also discovered two additional HrpL-regulated effector candidates and an HrpL-regulated distant homologue of *avrPto1*.

Conclusion: The draft genome sequence facilitates the continued development of *P. syringae* pathovar *tabaci* on wild tobacco as an attractive model system for studying bacterial disease on plants. The catalogue of effectors sheds further light on the evolution of pathogenicity and host-specificity as well as providing a set of molecular tools for the study of plant defence mechanisms. We also discovered several large genomic regions in *Pta* 11528 that do not share detectable nucleotide sequence similarity with previously sequenced *Pseudomonas* genomes. These regions may include horizontally acquired islands that possibly contribute to pathogenicity or epiphytic fitness of *Pta* 11528.

Background

Pseudomonas syringae is a widespread bacterial pathogen that causes disease on a broad range of economically important plant species. The species *P. syringae* is subdivided into about 50 pathovars, each exhibiting characteristic disease symptoms and distinct host-specificities. *P. syringae* pathovar *tabaci* (*Pta*) causes wild-fire disease in soybean and tobacco plants [1,2], characterised by chlorotic halos surrounding necrotic spots on the leaves of infected plants. Formation of halos is dependent on the beta-lactam tabtoxin, which causes ammonia accumulation in the host cell by inhibition of glutamine synthetase [3]. However, whether tabtoxin is an essential component of the disease process is unclear [4,5].

Pathogenicity of *P. syringae* strains is dependent on the type III secretion system (T3SS). The T3SS secretes a suite of virulence 'effector' proteins into the host cytoplasm where they subvert the eukaryotic cell physiology and disrupt host defences [6-14]. Mutants lacking the T3SS do not secrete effectors, and as a consequence do not infect plants or induce disease symptoms. Thus, understanding effector action is central to understanding bacterial pathogenesis. A single *P. syringae* strain typically encodes about 30 different effectors [14]. However, different *P. syringae* strains have different complements of effector genes. The emerging view is that of a core of common effectors encoded by most strains, augmented by a variable set. Individual effectors appear to act redundantly with each other and are individually dispensable with a small or no loss to pathogen virulence [10]. Effectors are also thought to play an important role in determining host range. This is most clearly true when infections are restricted by host defences. Some plants have evolved specific mechanisms to recognise certain effectors; such recognition induces strong host defences which curtail infection. For example, expression of the T3SS effector HopQ1-1 from *P. syringae* pathovar *tomato* (*Pto*) DC3000 was sufficient to render *Pta* 11528 avirulent on *Nicotiana benthamiana* [15]. The opposite situation, in which acquisition of a novel effector gene confers the ability to infect new host plants, has not been demonstrated and remains speculative. However, heterologous expression of the effector gene *avrPtoB* conferred a plasmid-cured strain of *P. syringae* pathovar *phaseolicola* (*Pph*) with increased virulence [16]. We hope that further identification and characterisation of effector repertoires of particular strains will shine new light on their roles in determining host range. Finally, bacterial virulence is also likely to be influenced by other non-T3SS-dependent virulence factors such as toxins which are often co-regulated with the T3SS [17].

Complete genome sequences are available for strains representing three *P. syringae* pathovars: *Pto*, pathovar *phaseolicola* (*Pph*) and pathovar *syringae* (*Psy*) [18-20].

Comparisons of these have led to the identification of core effector gene sets and to explain some of the differences in host-specificity between pathovars. However, these three sequenced strains are representatives of three distinct phylogroups within the species *P. syringae*, and as such are phylogenetically quite distant [21,22]. According to DNA-DNA hybridisation studies and ribotyping [21], *P. syringae* can be divided into 9 discrete genomospecies. Representative strains of *Psy*, *Pph* and *Pto* fell into genomospecies one, two and three respectively [21]. Recently, a strain of pathovar *oryzae* (genomospecies four) was sequenced [23]. A draft genome sequence was also published for *Pto* T1 [24], a strain closely related to *Pto* DC3000 but restricted to tomato hosts, whereas *Pto* DC3000 is able to cause disease on *Arabidopsis*. In the current study, we explore genetic differences at an intermediate phylogenetic resolution; that is, we compared the genome sequences of *Pta* 11528 to that of *P. phaseolicola* (*Pph*) 1448A, which resides within the same phylogroup but possesses a distinct host range and causes different disease symptoms.

Pto DC3000 was the first plant-pathogenic pseudomonad to have its genome sequenced, helping to establish the *Arabidopsis-Pto* system as the primary model for plant-microbe interactions. However, *Arabidopsis* is not a natural host of *Pto*, and it is important to develop alternative systems given the genetic variability of *P. syringae* strains, particularly in regard to effectors. We work on the interaction between *Pta* and the wild tobacco plant *N. benthamiana*, which offers certain advantages over *Arabidopsis*. Firstly, *N. benthamiana* is an important model for the Solanaceae, which includes many important crop species. The *Pta-N. benthamiana* interaction is a natural pathosystem. Lastly, *N. benthamiana* is an important model plant that is more amenable to biochemistry-based approaches and facile manipulation of gene expression such as virus-induced gene silencing (VIGS). Thus *N. benthamiana* provides experimental options for understanding plant-bacterial interactions. Strains of *Pta* can cause disease on *N. benthamiana*, but relatively few genetic sequence data are available for this pathovar.

In this study we generated a draft complete genome sequence of *Pta* 11528 and used a functional screen for HrpL-dependent genes to infer its repertoire of T3SS effectors and associated Hrp Outer Proteins (Hops), which differs significantly from that of its closest relative whose complete genome has previously been published (*Pph* 1448A). *Pta* 11528 does not encode functional homologues of HopAF1 or HrpZ1. This was surprising since HopAF1 was conserved in the three previously sequenced pathovars [18-20]. HrpZ1 is conserved in most strains of *P. syringae* that have been investigated, albeit with differences in amino acid sequence [25]. However, Pta strain

6605 and several other isolates from Japan, were previously shown to carry a major deletion leading to truncated HrpZ protein product [26]. *Pta* 11528 encodes several novel potential T3SS effectors for which no close orthologues have been reported. We also discovered several large genomic regions in *Pta* 11528 that do not share detectable nucleotide sequence similarity with previously sequenced *Pseudomonas* genomes. These regions may be horizontally acquired islands that possibly contribute to pathogenicity or epiphytic fitness of *Pta* 11528.

Results and discussion

Sequencing and assembly of the Pta 11528 genome

The Illumina sequencing platform provides a cost-effective and rapid means to generate nucleotide sequence data [27-29]. Although this method generates very short sequence reads, several recent studies have demonstrated that it is possible to assemble these short reads into good quality draft genome sequences [30-41].

We generated 12,096,631 pairs of 36-nucleotide reads for a total of 870,957,432 nucleotides. This represents approximately 145X depth of coverage assuming a genome size of six megabases. We used Velvet 0.7.18 [41] to assemble the reads *de novo*. Our resulting assembly had 71 supercontigs of mean length 85,604 nucleotides, an N_{50} number of eight, and N_{50} length of 317,167 nucleotides; that is, the eight longest supercontigs were all at least 317,167 nucleotides long and together covered more than 50% of the predicted genome size of six megabases. The largest supercontig was 606,547 nucleotides long. The total length of the 71 assembled supercontigs was 6,077,921 nucleotides. The G+C content of the assembly was 57.96%, similar to that of the previously sequenced *P. syringae* genomes (Table 1). The sequence data from this project have been deposited at DDBJ/EMBL/GenBank under the accession ACHU00000000. The version described in this paper is the first version, ACHU01000000. The data can also be accessed from the authors' website http://tinyurl.com/Pta11528-data and as Additional files submitted with this manuscript. In addition, an interactive genome browser is available from the authors' website http://tinyurl.com/Pta11528-browser.

We aligned the 71 *Pta* supercontigs against published complete *Pseudomonas* genome sequences using MUMMER [42]. The *Pta* 11528 genome was most similar to that of *Pph* 1448A, with 97.02% nucleotide sequence identity over the alignable portions. The next most similar genome was that of *Pto* DC3000, with less than 90% identity (Table 1). This pattern of sequence similarity is consistent with phylogenetic studies that placed strains of *Pta* in the

Table 1: Comparison of *Pta* 11528 genome properties with those of previously sequenced *P. syringae* genomes [18-20,83-85], [86-93].

RefSeq accession number	Description	G+C content (%)	Length (nucleotides)	Nucleic acid sequence identity to *P. syringae* pv *tabaci* 11528 draft assembly (%)
n. a.	*P. syringae* pv. *tabaci* 11528 draft genome assembly	57.96	6,077,921	100
NC_005773	*P. syringae* pv. *phaseolicola* 1448A, chromosome	58.01	5,928,787	97.02
NC_007274	*P. syringae* pv. *phaseolicola* 1448A large plasmid	55.14	73,661	91.59
NC_004632	*P. syringae* pv. *tomato* str. DC3000 plasmid pDC3000B	56.16	67,473	90.77
NC_007005	*P. syringae* pv. *syringae* B728a, chromosome	59.23	6,093,698	89.42
NC_004633	*P. syringae* pv. *tomato* str. DC3000 plasmid pDC3000A	58.39	6,397,126	89.36
NC_007275	*P. syringae* pv. *phaseolicola* 1448A small plasmid	54.10	131,950	89.09
NC_004578	*P. syringae* pv. *tomato* str. DC3000	58.39	6,397,126	87.65

Percentage identities were calculated over the alignable portion of the genomes using MUMMER [42].

same phylogroup as *Pph* and revealed a relatively distant relationship to *Pto* [21,22].

Comparison of the protein complement of Pta 11528 versus Pph 1448A and other pseudomonads

Using the FgenesB annotation pipeline http://www.soft berry.com, we identified 6,057 potential protein-coding genes, of which 5,300 were predicted to encode proteins of at least 100 amino acids long. Of 5,300 predicted *Pta* 11528 proteins, 575 (10.8%) had no detectable homology with *Pph* 1448A proteins (based on our criterion of an E-value less than 1e-10 using BLASTP). Of these 575 sequences, 303 had no detectable homologues in *Psy* B728a nor *Pto* DC3000. These 303 *Pta*-specific sequences had a median length of 198 amino acids whereas the median length of the 5,300 sequences was 216 amino acids. Automated gene prediction is not infallible and inevitably a subset of the predictions will be incorrect. The reliability of gene predictions is poorer for short sequences than for longer ones. This slight enrichment for very short sequences among the *Pta*-specific gene predictions might be explained by the inclusion of some open reading frames that are not functional genes among those 303. However, many of the predicted proteins showed significant similarity to other proteins in the NCBI NR databases (See Additional file 1: Table S1), confirming that these are likely to be genuine conserved genes.

Conservation of the T3SS apparatus and T3SS-dependent effectors

The Hop Database (HopDB, http://www.pseudomonas-syringae.org) provides a catalogue of confirmed and predicted *hop* genes [43]. Figure 1 lists the *hop* genes in HopDB for the three previously fully sequenced *P. syringae* genomes. A 'core' set of *hop* genes are conserved in all three previously sequenced pathovars: *avrE1*, *hopAF1*, *hopAH2*, *hopAJ2*, *hopAK1*, *hopAN1*, *hopI1*, *hopJ1*, *hopX1*, *hrpK1*, *hrpW1* and *hrpZ1*. In addition to this core set, each genome contains additional *hop* genes that are found in only a subset of the sequenced strains. The *Pta* 11528 homologues of *hop* genes are listed in Table 2. Figure 1 also indicates those *hop* genes for which a close homologue was found to be encoded in *Pta* 11528.

In sequenced strains of *P. syringae*, the gene cluster encoding the T3SS apparatus is flanked by collections of effector genes termed the exchangeable effector locus (EEL) and the conserved effector locus (CEL). Together, these three genetic components comprise the *Hrp* pathogenicity island [44]. A core set of *hop* genes is located in the *Hrp* pathogenicity island [44], which is highly conserved between *Pta* 11528 and *Pph* 1448A (Figure 2), except that in *Pta* 11528 there is a deletion in *hrpZ1* and an insertion in the *hrpV-hrcU* intergenic region. The core *hop* genes *avrE1*, *hopAH2*, *hopAJ2*, *hopAK1*, *hopAN1*, *hopI1*, *hopJ1*,

hopX1 and *hrpK1* are conserved in *Pta* 11528 and encode intact full-length proteins. *Pta* 11528 encodes a full-length HrpW1 protein, albeit with insertions of 69 and 12 nucleotides relative to the *Pph* 1448A sequence. However, there is a large deletion in *hrpZ1* that likely renders it non-functional and *hopAF1* is completely absent.

Besides the core conserved *hop* genes, the *Pta* 11528 genome assembly contains full-length orthologues of *hopR1*, *hopAS1*, *hopAE1* and *hopV1*, which are also found in *Pph* 1448A but are absent from *Psy* B728a and/or *Pto* DC3000.

The *hrpZ1* gene encodes a harpin, which is not classified as a type III effector because it is not injected directly into host cells. Harpins are characteristically acidic, heat-stable and enriched for glycine, lack cysteine residues [8] and can induce defences in both host and non-host plants [45,46]. HrpZ1 forms pores in the host membrane [47] suggesting a role in translocation of effectors across the host membrane. It also shows sequence-specific protein binding activity [48]. HrpZ1 can induce defences in both host and non-host plants and tobacco has been extensively used as the non-host plant species [45,46]. The inactivation of *hrpZ1* in *Pta* 11528 and other strains of *Pta* [26] may be an adaptive strategy and have been an important process in the stepwise progression towards compatibility, allowing *Pta* 11528 to avoid detection by the tobacco host plant. This is reminiscent of the "black holes" and other processes that inactivate genes whose expressed products are detrimental to a pathogenic lifestyle [49,50]. One excellent example is the inactivation of *cadA* in genomes of *Shigella* species as compared to the genome of their closely related but non-pathogenic *Escherichia coli* strain [51,52].

Pta 11528 contains highly conserved homologues of *hopAB2*, *hopW*, *hopO1-1*, *hopT1-1*, *hopAG1*, *hopAH1*, *hopF1* and *hopAR1*, which are absent in *Pph* 1448A. Although absent from the *Pph* 1448A genome, *hopAR1* and *hopF1* have been identified in other strains of *Pph* [53-57]. In *Pph* 1302A, *hopAR1* is located on the pathogenicity island *PPHGI-1*, though its genomic location varies between strains [56,57]. *PPHGI-1* is absent from the *Pph* 1448A genome [57]. The *Pta* 11528 genome (supercontig 1087) possesses a region of similarity to *PTPHGI-1*, but which contains a substantial number of insertions and deletions (Additional file 2: Figure S1). The *Pta* 11528 *hopAR1* homologue (C1E_2036) is not located in the *PPHGI-1* region; it falls on supercontig 672 about two kilobases upstream of a gene encoding a protein (C1E_2039) sharing 43% amino acid identity with *Pto* DC3000 *avrPto1*. In contrast to AvrPto1 from *Pto* DC3000, the AvrPto1 homologue (C1E_2039) from *Pta* 11528 is not recognised by the plant Pto/Prf system (S. Gimenez Ibanez and J. Rathjen, manuscript in preparation).

Table 2: Homologues of known *hop* genes in *Pta* 11528. Homologues were detected by searching the *Pta* 11528 FgenesB-predicted protein sequences against HopDB http://www.pseudomonas-syringae.org using BLASTP

Effector gene	Gene in *Pta* 11528 genome (location)	Hrp-box HMM score (bioinformatic evidence)	HrpL-dependent (functional screen)	Homologue in Pph 1448A
avrE1	C1E_5333 (1087:342585..346532)	18.24	Yes	PSPPH_1268 (chromosome)
avrPto1-like	C1E_2039 (672:104030..104509)	26.02	Yes	None
hopAB2	C1E_3975 (955:85214..86053)	None	Yes	None
hopAE1	C1E_0512 (174:82348..85077)	17.91	Yes	PSPPH_4326 (chromosome)
hopAG1	C1E_2305 (679:71608..73584)	19.67	No	None
hopAH1	C1E_2306 (679:74209..74976)	None	No	None
hopAH2	C1E_3507 (891:229657..230907)	None	No	PSPPH_3036 (chromosome)
hopAI1' (degenerate)	C1E_2307 (679:75143..75466)	24.85	Yes	None
hopAJ2	C1E_0586 (174:157540..158877)	None	No	PSPPH_4398 (chromosome)
hopAK1	C1E_4764 (1053:316850..318520)	21.71	No	PSPPH_1424 (chromosome)
hopAN1	C1E_1908 (661:72932..74221)	None	No	PSPPH_0456 (chromosome)
hopAR1	C1E_2036 (672:101352..102155)	15.68	Yes	None
hopAS1	C1E_1389 (554:110458..114543)	18.77	No	PSPPH_4736 (chromosome)
hopI1	C1E_0551 (174:125987..126916)	21.66	Yes	PSPPH_4366 (chromosome)
hopM1'	C1E_5336 (1087:348226..350460)	13.65	Yes	PSPPH1266
hopO1-1	C1E_5022 (1087:78582..79433)	18.48	Yes	None
hopR1	C1E_3889 (955:5054..6763)	24.65	No	PSPPH_0171
hopT1-1	C1E_5021 (1087:77437..78576)	18.48	Yes	None
hopV1	C1E_2810 (733:27251..28225)	14.09	No	PSPPH_2351 (chromosome)

Table 2: Homologues of known *hop* genes in *Pta* 11528. Homologues were detected by searching the *Pta* 11528 FgenesB-predicted protein sequences against HopDB http://www.pseudomonas-syringae.org using BLASTP *(Continued)*

hopW1	C1E_3964 (955:74860..77184)	10.73	No	None (PSPPH is a truncated HopW1 homologue)
hopX1	C1E_5300 (1087:315085..316227)	27.96	No	PSPPH_1296 (chromosome)
hrpK1	C1E_5301 (1087:316323..318641)	27.96	No	PSPPH_1295 (chromosome)
hrpW1	C1E_5341 (1087:351970..352491)	20.57	No	PSPPH_1264 (chromosome)
hrpZ1	C1E_5325 (1087:337056..337478) C1E_5324 (1087:336767..337045)	19.74	No	PSPPH_1273 (chromosome)

The locations of *Pph* 1448A homologous genes is indicated, including an indication of whether they are located on the chromosome or on the large plasmid. Also indicated is whether each gene appeared in the functional screen for Hrp-dependent transcription.

The homologues of *hopAG1*, *hopAH1* and the degenerate *hopAI1'* are found within a region of the *Pta* 11528 genome that shares synteny with the chromosome of *Psy* B728a. This region is also conserved in *Pto* DC3000A, albeit with several deletions and insertions, suggesting that these effector genes are ancestral to the divergence of the pathovars and have been lost in *Pph* 1448A rather than having been laterally transferred laterally between *Pta* 11528 and *Psy* B728a. In *Pto* DC3000, *hopAG1* (PSPTO_0901) has been disrupted by an insertion sequence (IS) element. This is consistent with a model of lineage-specific loss of certain ancestral effectors.

In *Pto* DC3000, *hopO1-1* and *hopT1-1* are located on the large plasmid pDC3000A; homologues of these effector-encoding genes are not found in *Pph* 1448A. The *Pta* 11528 genome contains a three kilobase region of homology to pDC3000 comprising homologues of these two effector genes and a homologue of the ShcO1 chaperone-encoding gene. These three genes are situated in a large (at least 50 kilobase) region of the *Pta* 11528 genome that has only limited sequence similarity with *Pph* 1448A. Two tRNA genes (tRNA-Pro and tRNA-Lys) are located at the boundary of this region (Figure 3), which would be consistent with this comprising a mobile island.

In plasmid pMA4326B from *P. syringae* pathovar *maculicola* (*Pma*), the *hopW1* effector gene is immediately adjacent to a three-gene cassette comprising a resolvase, an integrase and *exeA*. This cassette is also found in plasmids and chromosomes of several human-pathogenic Gram-negative bacteria [58]. We found a homologue of this cassette along with a *hopW1* homologue on supercontig 955 of the *Pta* 11528 genome assembly. Stavrinides and Gutt-

man [58] proposed that the boundaries of the cassette lay upstream of the resolvase and upstream of *hopW1*. The presence of this four-gene unit in a completely different location in *Pta* 11528 is indeed consistent with the hypothesis that it represents a discrete mobile unit.

Several *hop* genes are located on the large plasmid of *Pph* 1448A. We found no homologues of these genes in *Pta* 11528, suggesting that the plasmid is not present in *Pta* 11528. Consistent with this, only a small proportion of the plasmid was alignable to our 36-nucleotide Illumina sequence reads (Figure 4). This reveals that a large component of the pathogen's effector arsenal is determined by its complement of plasmids. However, simple loss or gain of a plasmid does not explain all of the differences in effector complement since *Pta* 11528 lacks homologues of several *Pph* 1448A chromosomally-located effector-encoding *hop* genes *hopG1*, *hopAF1*, *avrB4*, *hopF3* and *hopAT1* as well as the non-effector *hopAJ1*. It also lacks homologues of the *Pph* 1448A degenerate effector gene *hopAB3'*.

The regions of the *Pph* 1448A large plasmid that are apparently conserved in *Pta* 11528 include genes encoding the conjugal transfer system, suggesting the presence of one or more plasmids in this strain. We found an open reading frame (C1E_3950, located on supercontig 955 coordinates 59126-60394) encoding a protein with about 97% sequence identity to the RepA proteins characteristically encoded on pT23A-family plasmids (*e.g.* AAW01447; reviewed in [59]), suggesting that this 236 kilobase supercontig might represent a plasmid.
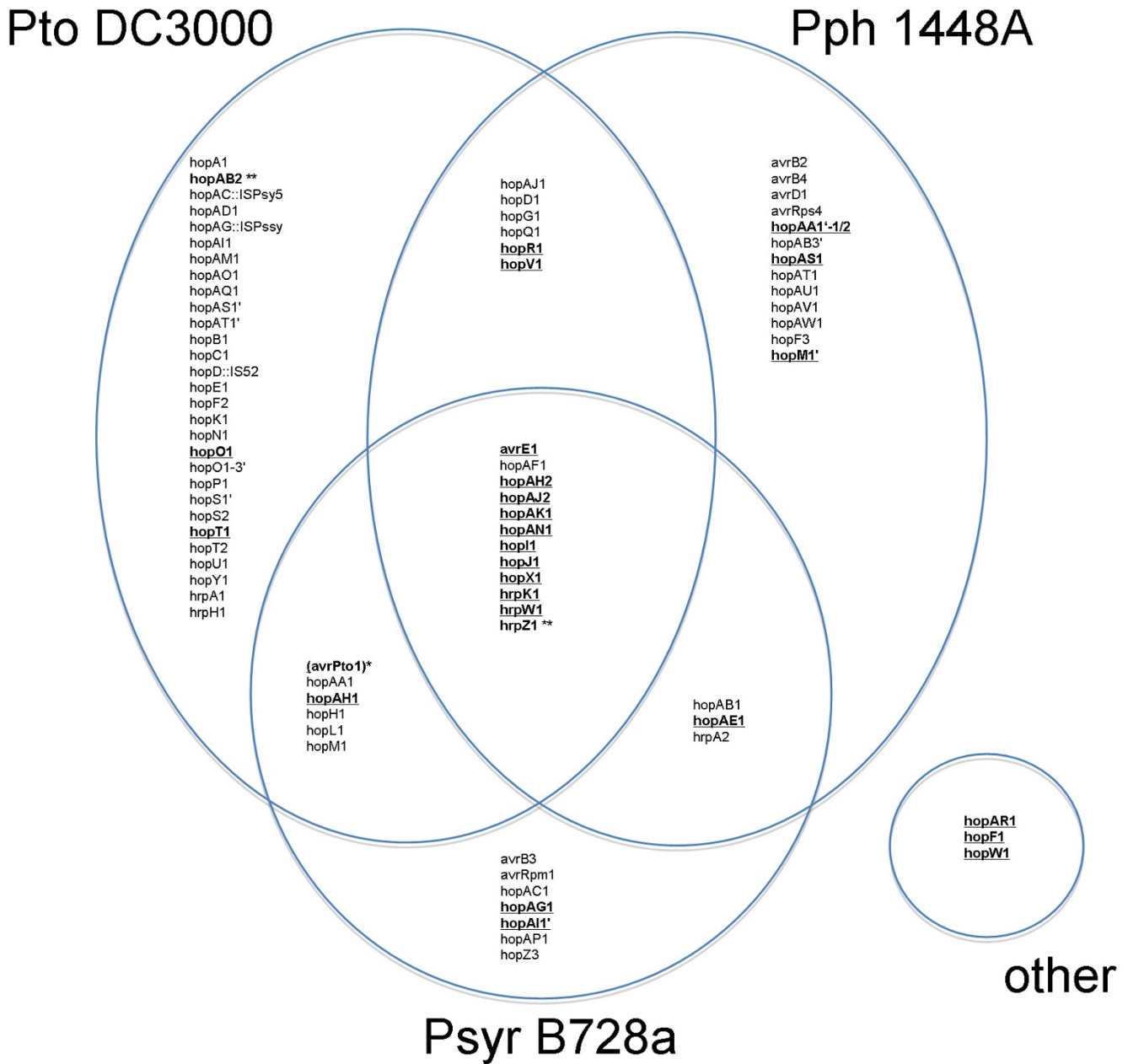
Figure 1
Comparison of the *hop* gene complements of the three previously fully sequenced *P. syringae* genomes. Those *hop* genes that are conserved in *Pta* 11528 are shown in boldface and underlined. *Pta* 11528 also contains three *hop* genes that do not have orthologues in the sequenced genomes: *hopAR1*, *hopF1* and *hopW1*. * No close homologue of *avrPto1* was found in *Pta* 11528; however, there is a gene encoding a protein that shares 43% amino acid identity with Avr *Pto*1 from *Pto* DC3000. ** In the *Pta* 11528 genome *hrpZ1* appears to be a pseudogene.

A functional screen for HrpL-regulated genes
We used a previously described functional screen [60] to complement our bioinformatics-based searches for type III effectors of *Pta* 11528. Our functional screen was based on two steps. The first step was employed to identify genes whose expression was regulated by the T3SS alternative sigma factor, HrpL. The second step was used to identify the subset of HrpL-regulated genes that encoded effectors. For *Pta* 11528, we employed only the first step to identify candidate effector genes based on induced expression by HrpL. A library was constructed from *Pta* 11528 into a broad-host range vector carrying a promoter-less GFP and
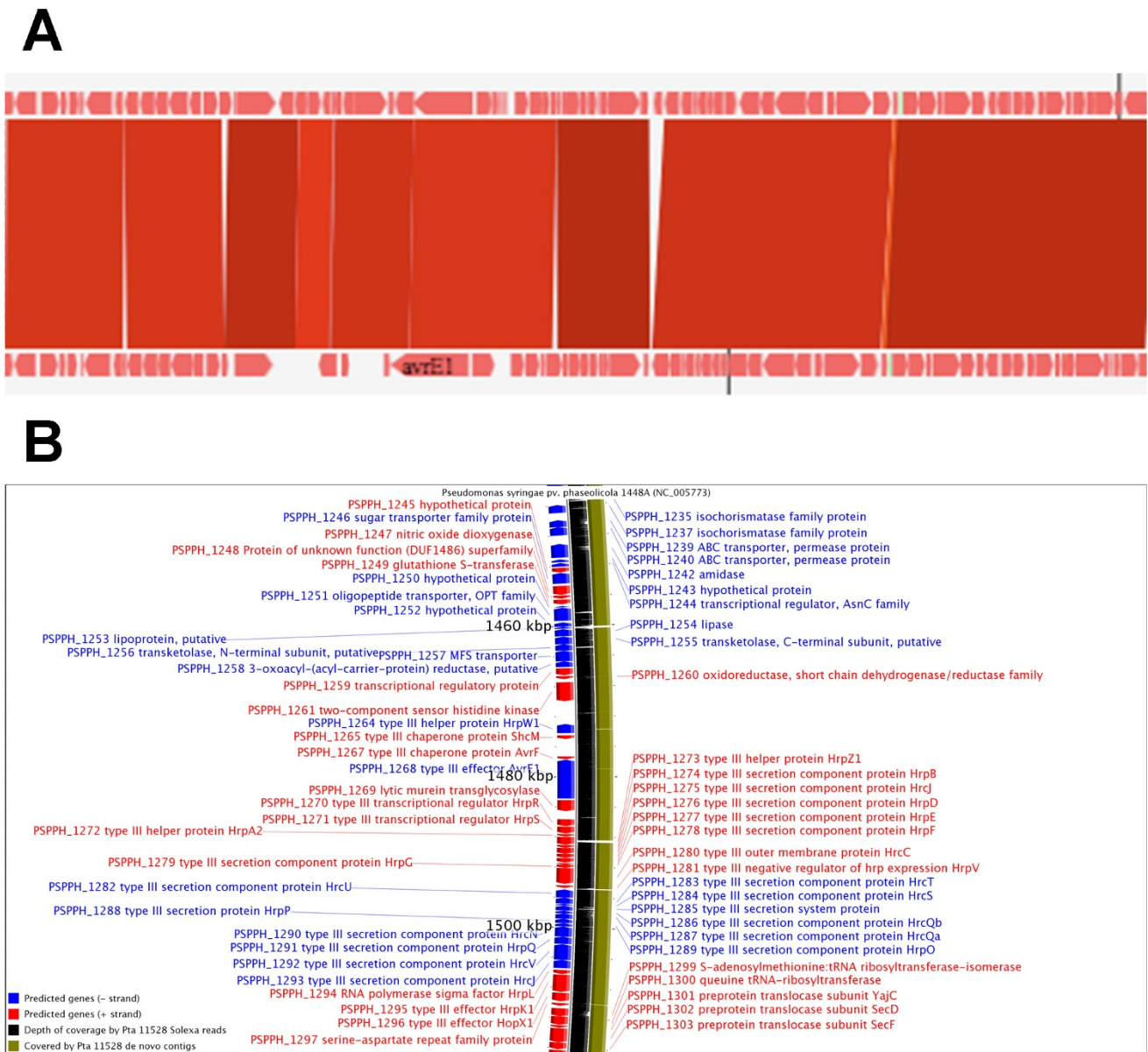
Figure 2
Conservation of the *Hrp* pathogenicity island between *Pph* 1448A and *Pta* 11528. Panel A shows an alignment of the *Pph* 1448A *Hrp* pathogenicity island (lower track) against the homologous region in *Pta* 11528 (upper track), prepared using GenomeMatcher, which indicates similarity values by colour with dark blue, green, yellow and red representing increasing degrees of similarity [78]. Panel B shows the MAQ [79] alignment of the *Pta* 11528 Illumina reads (in black) and the BLASTN [80] alignment of the *Pta* 11528 *de novo* assembly (in green) against the *Hrp* region of the *Pph* 1448A genome.

mobilized into *Pto* lacking its endogenous *hrpL* but conditionally complemented with an arabinose-inducible *hrpL*. We used a fluorescence activated cell sorter (FACS) to select clones that carried HrpL-inducible promoters based on expression of GFP after growth in arabinose. Clones were sequenced and sequences were assembled. Clones representative of assembled supercontigs were verified again for HrpL regulation using FACS. Among the genes

whose expression was confirmed to be HrpL-dependent were those encoding effectors *hopAE1*, *hopI1*, *hopAR1*, the *avrPto1*-like gene, *hopF1*, *hopT1-1*, *hopO1-1*, *avrE1*, *hopX1*, and the degenerate *hopM1'* and *hopAI1'* as well as known T3SS-associated genes *hrpH* (ORF1 of the CEL; [61]) and *hrpW1*. Interestingly, the screen also confirmed HrpL-dependent regulation of genes encoding a major facilita-
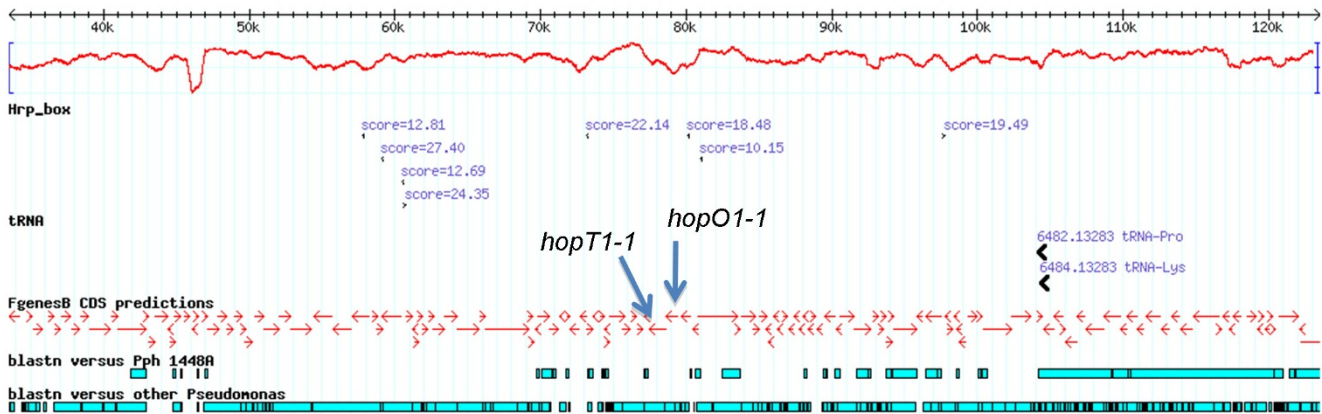
Figure 3
A 90-kilobase region of the *Pta* 11528 genome containing homologues of *hopT1-1* and *hopO1-1*. The G+C content is indicated by the plot near the top of the figure.

tor superfamily (MFS) permease and a putative peptidase (Table 3).

Other differences in predicted proteomes of P. syringae strains

Host range and pathogenicity are likely to be further influenced by genes other than those associated with type III secretion. Virulence determinants in *P. syringae* include toxins as well as epiphytic fitness; that is, the ability to acquire nutrients and survive on the leaf surface [14]. Epiphytic fitness depends on quorum-sensing [62], chemotaxis [63], osmo-protection, extracellular polysaccharides, glycosylation of extracellular structures [64] iron uptake [65] and the ability to form biofilms. Cell-wall-degrading hydrolytic enzymes play a role in virulence in at least some plant-pathogenic pseudomonads [66]). Secretion systems (including type I, type II, type IV, type V, type VI and twin arginine transporter) may also contribute to both virulence and epiphytic fitness [67], whilst multidrug efflux pumps may confer resistance to plant-derived antimicrobials [68].

To identify differences between *Pta* 11528 and the previously sequenced *Pph* 1448A, *Psy* B728a and *Pto* DC3000 with respect to their repertoires of virulence factors, we performed BLASTP searches between the predicted proteomes. We found no significant differences in the repertoires of secretion systems between the proteomes. However, we found that *Pta* 11528 lacks homologues of several *Pph* 1448A polysaccharide modifying enzymes (glycosyl transferase PSPPH_0951, polysaccharide lyase PSPPH_1510, glycosyl transferase PSPPH_3642). Conversely, *Pta* 11528 encodes two glycosyl transferases (C1E_0355 and C1E_0361) and a thermostable glycosylase (C1E_4802) that do not have homologues in any of the three fully sequenced *P. syringae* genomes. This may

imply differences in the extracellular polysaccharide profiles. In contrast to *Pph* 1448A, *Pta* 11528 lacks homologues of RhsA insecticidal toxins (PSPPH_4042 and PSPPH_4043). However, a tabtoxin biosynthesis gene cluster is found in the *Pta* 11528 genome and shows a high degree of conservation with the previously sequenced *Pta* BR2 tabtoxin biosynthesis cluster [69].

Pta 11528 encodes several enzymes that do not have homologues in any of the three fully sequenced *P. syringae* genomes (Table 4), including a predicted gluconolactonase (C1E_2553), a predicted dienelactone hydrolase (C1E_2589), a predicted nitroreductase (C1E_6026), and a sulphotransferase (C1E_6026). C1E_0903 shares 71.4% amino acid sequence identity with a predicted epoxide hydrolase (YP_745600.1) from *Granulibacter bethesdensis* CGDNIH1 [70] and has a significant match to the epoxide hydrolase N-terminal domain in the Pfam database (PF06441) [71,72]. Epoxide hydrolases are found in *P. aeruginosa* and *P. fluorescens* PfO-1, but not in any other pseudomonads. It is possible that this gene product has a function in detoxification of host-derived secondary metabolites.

Pta protein C1E_6026 has a significant match to the sulphotransferase domain (Pfam:PF00685). Examples of this protein domain have not been found in other pseudomonads except for *P. fluorescens* PfO-1. Sulphotransferase proteins include flavonyl 3-sulphotransferase, aryl sulphotransferase, alcohol sulphotransferase, estrogen sulphotransferase and phenol-sulphating phenol sulphotransferase. These enzymes are responsible for the transfer of sulphate groups to specific compounds. The sulphotransferase gene (C1E_6026, 82% amino acid identity to *P. fluorescens* Pfl01_0157) overlaps a two kilobase *Pta* 11528-specific genomic island that also encodes a phage
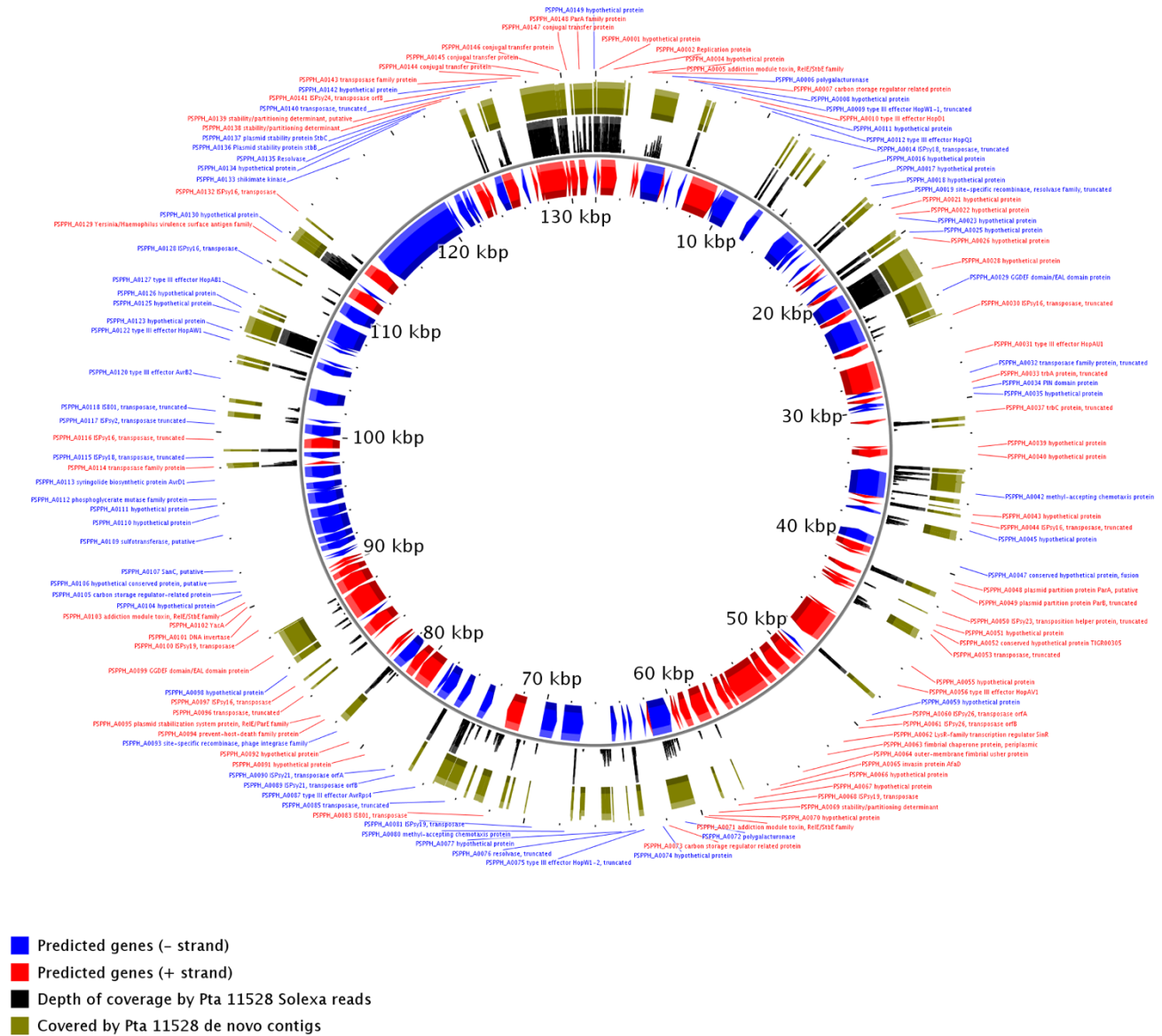
Figure 4
Limited conservation between the *Pta* 11528 genome sequence and the sequence of the *Pph* 1448A large plasmid. The MAQ [79] alignment of the *Pta* 11528 Illumina reads is shown in black. The thickness of the black track is proportional to the depth of coverage by Illumina reads. The BLASTN [80] alignment of the *Pta* 11528 *de novo* assembly against the plasmid sequence is shown in a green track, with the thickness of this single green track being proportional to sequence identity.

Table 3: *Pta* 11528 genes confirmed by the functional screen to be under the transcriptional control of HrpL

Gene	Gene in *Pta* 11528 gnome	Hrp-box HMM score (bioinformatic evidence)	Orthologue in Pph 1448A
avrE1	C1E_5333 (1087:342585..346532)	18.24	PSPPH_1268
avrF	C1E_5335 (1087:347759..348148)	13.65	PSPPH_1267
avrPto1	C1E_2039 (672:104030..104509)	26.02	None
hopAB2	C1E_3975 (955:85214..86053)	None	None
hopAE1	C1E_0512 (174:82348..85077)	17.91	PSPPH4326 (chromosome)
hopAI1'	C1E_2307 (679:75143..75466)	24.85	None
hopAR1	C1E_2036 (672:101352..102155)	15.68	None
hopF1	C1E_5009 (1087:72050..72664)	22.14	None
hopI1	C1E_0551 (174:125987..126916)	21.66	PSPPH_4366
hopM1'	(C1E_5336) 1087:348226..350460	13.65	PSPPH_1266
hopO1-1	C1E_5022 (1087:78582..79433)	18.48	None
hopT1-1	C1E_5021 (1087:77437..78576)	18.48	None
hopW1	C1E_3964 (955:74860..77184)	10.73	None
hopX1	C1E_5300 (1087:315085..316227)	27.96	PSPPH_1296
hrpH	C1E_5332 (1087:341340..342365)	17.46	PSPPH_1269
hrpW1	C1E_5339 (1087:351255..351707) C1E_5341 (1087:351970..352491)	20.57	PSPPH_1264
Major facilitator superfamily permease	C1E_4990 1087:59133..60425	12.69	None
Putative M20 peptidase	C1E_1425 (554:155221..156516)	15.05	None
schF	C1E_5010 (1087:72735..73127)	22.14	None
schO1	C1E_5023 1087:79682..80245	10.50	None

tail collar-protein encoding gene (C1E_5461, 61% amino acid identity to *P. fluorescens* Pfl01_0155) and an acetyltransferase (C1E_5459, 76% amino acid identity to *P. fluorescens* Pfl01_0148). We speculate that this region has been horizontally acquired in the *Pta* 11528 lineage *via* a bacteriophage.

An 80 kilobase region of *Pta* 11528 supercontig 684 contains two open reading frames (ORFs) (C1E_2584 and C1E_2585) whose respective predicted protein products

show 48 and 55% amino acid identity to the C- and N-termini of a *P. putida* methyl-accepting chemotaxis protein (MCP) (PP_2643) and little similarity to any *P. syringae* protein. Since the N- and C-termini are divided into separate reading frames, this probably represents a degenerate pseudogene. Immediately downstream of these ORFs is a gene (C1E_2583) that specifies a MCP showing greatest sequence identity (70%) to PP_2643 from *P. putida*, whilst sharing only 65% identity to its closest homologue in *P. syringae* (PSPPH_4743). This region also encodes

Table 4: Proteins encoded by the draft *Pta* 11528 genome that have no detectable homologues on three previously fully sequenced *P. syringae* genomes.

Genomic coordinates	Locus tag	Length (amino acids)	Predicted function (FgenesB automated annotation)
122:73423..74418	C1E_0355	331	Glycosyltransferase involved in cell wall biogenesis
122:82519..83649	C1E_0361	376	Glycosyltransferase
195:60030..60371	C1E_0654	113	Integrase
195:63783..67199	C1E_0659	1138	ATP-binding protein
195:67192..68034	C1E_0660	280	Phosphoadenosine phosphosulfate reductase
195:68040..68927	C1E_0661	295	Serine/threonine protein kinase
256:43490..44116	C1E_0901	208	TetR family ranscriptional regulator. 49% amino acid sequence identity to *G. bethesdensis* GbCGDNIH1_1777 [70].
256:45146..46300	C1E_0903	384	Hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
419:2876..4135	C1E_1014	419	Biotin carboxylase
419:12230..12925	C1E_1023	231	Tabtoxin biosynthesis enzyme, TblA
419:14223..15053	C1E_1025	276	Tetrahydrodipicolinate N-succinyltransferase
554:156634..157137	C1E_1426	167	Histone acetyltransferase HPA2
554:289047..290207	C1E_1572	386	Integrase
554:300540..302342	C1E_1580	600	P-loop ATPase
661:62161..62487	C1E_1896	108	Amine oxidase, flavin-containing
672:22725..23156	C1E_1956	143	Rhs family protein
672:25358..25786	C1E_1961	142	RHS protein
672:118239..118676	C1E_2056	145	Xenobiotic response element family of transcriptional regulator. 37% amino acid sequence identity to *Xylella fastidiosa* PD0954 [81].
672:276912..277349	C1E_2209	145	Histone acetyltransferase HPA2
679:1273..1647	C1E_2227	124	Similar to Mucin-1 precursor (MUC-1)
679:22846..23553	C1E_2251	235	Site-specific recombinases, DNA invertase Pin homologs
679:27901..29442	C1E_2260	513	Phage integrase
679:50793..51164	C1E_2286	123	LacI family transcriptional regulator. 57% amino acid sequence identity to *Rhizobium leguminosarum* plasmid-encoded pRL1201 [82].
679:95310..95744	C1E_2329	144	Tfp pilus assembly protein, major pilin PilA. 42% amino acid sequence identity to P. aeruginosa UniProt:P17838 [74].
684:105988..106806	C1E_2502	272	Histone acetyltransferase HPA2
684:114136..118092	C1E_2506	1318	NTPase (NACHT family)

Table 4: Proteins encoded by the draft *Pta* 11528 genome that have no detectable homologues on three previously fully sequenced *P. syringae* genomes. *(Continued)*

684:137829..138611	C1E_2527	260	Permeases of the major facilitator superfamily
684:150262..150585	C1E_2541	107	Short-chain dehydrogenase/reductase SDR
684:151896..152792	C1E_2545	298	Nucleoside-diphosphate-sugar
684:160330..161241	C1E_2553	303	Gluconolactonase
684:166181..167533	C1E_2556	450	ASPIC/UnbV domain-containing protein
684:174586..174987	C1E_2563	133	Xenobiotic response element family of transcriptional regulator. 38% amino acid sequence identity to *P. aeruginosa* PACL_0260 [83].
684:178452..178925	C1E_2570	157	Cro/CI family transcriptional regulator. 36% amino acid sequence identity to *Pto* DC3000 PSPTO_2855 [18].
684:192825..193367	C1E_2584	180	Methyl-accepting chemotaxis sensory transducer (C terminus)
684:193364..193720	C1E_2585	118	Methyl-accepting chemotaxis sensory transducer (N terminus)
684:197314..198498	C1E_2589	394	Dienelactone hydrolase
891:113887..115026	C1E_3396	379	Pectate lyase
891:121799..122809	C1E_3401	336	Type II secretory pathway, component PulK
955:27327..28100	C1E_3914	257	DNA-binding HTH domain-containing
955:37638..38900	C1E_3925	420	Outer membrane efflux protein
955:67953..68687	C1E_3957	244	Plasmid stability protein
1053:122043..123497	C1E_4568	484	Phage integrase family protein
1053:127685..128557	C1E_4572	290	Superfamily I DNA or RNA helicase
1053:130636..131022	C1E_4576	128	ATP-dependent DNA helicase, UvrD/Rep family
1053:363558..364328	C1E_4802	256	Thermostable 8-oxoguanine DNA glycosylase
1053:364986..365591	C1E_4804	201	PP-loop superfamily ATPase
1053:365588..366790	C1E_4805	400	Sugar kinase, ribokinase
1053:371237..371890	C1E_4809	217	Restriction endonuclease
1053:384339..385847	C1E_4823	502	ATPase
1053:387166..387564	C1E_4827	132	ATP-dependent DNA helicase, UvrD/Rep family
1053:403085..403480	C1E_4845	131	ATP-dependent DNA helicase, UvrD/Rep family
1053:405835..406488	C1E_4849	217	Restriction endonuclease
1087:462184..462693	C1E_5459	169	Histone acetyltransferase HPA2

Table 4: Proteins encoded by the draft *Pta* 11528 genome that have no detectable homologues on three previously fully sequenced *P. syringae* genomes. *(Continued)*

1087:463060..463650	C1E_5461	196	Phage collar protein
1087:464015..464851	C1E_5462	278	Sulfotransferase
1087:466231..466944	C1E_5464	237	S-layer domain protein
1087:466941..471797	C1E_5465	1618	Pyrrolo-quinoline quinone
1102:92687..93796	C1E_5711	369	Major facilitator superfamily (MFS) permease
1102:97423..98163	C1E_5715	246	IclR-like transcriptional regulator. 62% amino acid sequence identity to *Acinetobacter baumanii* ACICU_01897 [84].
1160:302149..302877	C1E_6026	242	Nitroreductase
1160:303074..303616	C1E_6027	180	TetR family transcriptional regulator. 56% amino acid sequence identity to *Ralstonia solanacearum* RSc0820 [85].

Protein-coding genes were predicted and automatically annotated using the FgenesB pipeline http://www.softberry.com. Only those proteins are shown for which a predicted function could be proposed.

another MCP (C1E_2587) that shares only 50% amino acid identity with any previously sequenced *P. syringae* homologue. It remains to be tested whether these MCPs play a role in pathogenesis and/or epiphytic fitness.

Transcriptional regulators are not normally considered to be virulence factors. However, expression of virulence factors may be coordinated by and dependent on regulators. Moreover, heterologous expression of the RscS regulator was recently shown to be sufficient to transform a fish symbiont into a squid symbiont [73]. *Pta* 11528 encodes several predicted transcriptional regulators that are not found in *Pto* DC3000, *Psy* B728a and *Pph* 1448A. These include two predicted TetR-like proteins (C1E_0901 and C1E_6027), two predicted xenobiotic response element proteins (C1E_2056 and C1E_2563), a LacI-like protein (C1E_2286), a Cro/CI family protein (C1E_2570) and an IclR family protein (C1E_5715).

Pta 11528 encodes a novel pilin (C1E_2329) not found in previously sequenced *P. syringae* strains but sharing significant sequence similarity with a type IV pilin from *P. aeruginosa* [74]. Pilin is the major protein component of the type IV pili, which have functions in forming micro-colonies and biofilms, host-cell adhesion, signalling, phage-attachment, DNA uptake and surface motility, and have been implicated as virulence factors in animal-pathogenic bacteria [75]. The precise function of the C1E_2329 pilin is unknown but it may be involved in epiphytic fitness or plant-pathogenesis or could even be involved in an interaction with an insect vector.

Pta-specific genomic islands
We identified 102 genomic regions of at least one kilobase in length which gave no BLASTN matches against previously sequenced *Pseudomonas* genomes (Additional file 3: Table S2). Ten of the *Pta* 11528-specific regions are longer than 10 kilobases, the longest being 37.7, 21.8, 18.7, 17.9 and 16.6 kilobases. The 16.6 kilobase region corresponds to the tabtoxin biosynthesis gene cluster [69]. These regions will be good candidates for further study of the genetic basis for association of *Pta* with the tobacco host. For example, several of the islands encode MFS transporters and other efflux proteins that might be involved in protection from plant-derived antimicrobials (Additional file 3: Table S2).

Conclusion
We have generated a draft complete genome sequence for the *Pta* 11528 a pathogen that naturally causes disease in wild tobacco, an important model system for studying plant disease and immunity. From this sequence, combined with a functional screen, we were able to deduce the pathogen's repertoire of T3SS-associated Hop proteins. This has revealed some important differences between *Pta* and other pathovars with respect to the arsenal of T3SS effectors at their disposal for use against the host plant. We also revealed more than a hundred *Pta*-specific genomic regions that are not conserved in any other sequenced *P. syringae*, providing many potential leads for the further study of the *Pta*-tobacco disease system.

Methods

Sequence data

The previously published sequences of *P. syringae* pathovar *phaseolicola* 1448A [20], *P. syringae* pathovar *syringae* B728a [19], *P. syringae* pathovar *tomato* DC3000 [18] were downloaded from the NCBI FTP site ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pseudomonas_syringae_pv_B728a. The NCBI non-redundant (NR) Proteins database was downloaded from the NCBI FTP site ftp://ftp.ncbi.nih.gov/blast/db/ on 10th December 2008.

De novo *sequence assembly and annotation*

Solexa sequence data were assembled using Velvet 0.7.18 [41]. We used Softberry's FgenesB pipeline http://www.softberry.com to predict genes encoding rRNAs, tDNAs and proteins. Annotation of protein-coding genes by FgenesB was based on the NCBI NR Proteins database.

Prediction of HrpL-binding sites (Hrp boxes)

We built a profile hidden Markov model (HMM) based on a multiple sequence alignment of 26 known Hrp boxes from *Pto* DC3000 using *hmmb* from the HMMER 1.8.5 package http://hmmer.janelia.org. DNA sequence was scanned against this profile-HMM using *hmmls* from HMMER 1.8.5 with a bit-score cut-off of 12.0.

Functional screen for candidate type III effectors

Library preparation and the Flow cytometric-based screen for HrpL-induced genes of *Pta* 11528 were done according to [60].

Visualisation of data

We generated graphical views of genome alignments using CGView [76]. To visualise the annotation draft genome assembly of *Pta*11528, we used the 'gbrowse' Generic Genome Browser [77].

Library preparation for Illumina sequencing

DNA was prepared from bacteria grown in L-medium using the Puregene Genomic DNA Purification Kit (Gentra Systems, Inc., Minneapolis, USA) according to manufacturer's instructions. A library for Illumina Paired-End sequencing was prepared from 5 mg DNA using a Paired-End DNA Sample Prep Kit (Pe-102-1001, Illumina, Inc., Cambridge, UK). DNA was fragmented by nebulisation for 6 min at a pressure of 32 psi. For end-repair and phosphorylation, sheared DNA was purified using QIAquick Nucleotide Removal Kit (Quiagen, Crawley, UK). The end repaired DNA was A-tailed and ada *Pto*rs were ligated according to manufacturer's instructions.

Size fractionation and purification of ligation products was performed using a 5% polyacrylamide gel run in TBE at 180V for 120 min. Gel slices were cut containing DNA in the 500 to 10 bp range. DNA was than extracted using 0.3 M sodium acetate and 2 mM EDTA [pH 8.0] followed by ethanol precipitation. Using 18 PCR cycles with primer PE1.0 and PE2.0 supplied by Illumina, 5' ada *Pto*r extension and enrichment of the library was performed. The library was finally purified using a QIAquick PCR Purification Kit and adjusted to a concentration of 10 nM in 0.1% Tween. The stock was kept at -20°C until used.

Sequencing

The flow cell was prepared according to manufacturer's instructions using a Paired-End Cluster Generation Kit (Pe-103-1001) and a Cluster Station. Sequencing reactions were performed on a 1G Genome Analyzer equipped with a Paired-End Module (Illumina, Inc., Cambridge, UK). 5 pM of the library were used to achieve ~20,000 to 25,000 clusters per tile. Capillary sequencing of *avrE*, *HrpW1* and other individual genes was done on an ABI 3730. PCR products were directly sequenced after treatment with ExoI and SAP. Primer sequences are available upon request from JHC.

Verification of Illumina sequence data

Three of the core *hop* genes in *Pta* 11528 appeared to be degenerate, based on the *de novo* assembly of short Illumina sequence reads. The *avrE1* gene appeared to have a 20-nucleotide deletion, *hrpZ1* a 325-nucleotide deletion, whilst *hrpW1* appeared to have three insertions of 22, 6 and 12 nucleotides. Currently, the reliability of *de novo* sequence assembly from short Illumina reads has not been fully characterised. In particular, repetitive and low-complexity sequence might generate artefacts in assembled supercontigs. Therefore, we checked these putative insertions and deletions by aligning the Illumina sequence reads against the relevant regions of both the *Pph* 1448A reference genome sequence and our *Pta* 11528 assembly. As an additional control, we also performed Velvet assemblies on previously published Illumina short-read data from *Psy* B728a [35]. We found that the B728a *avreE1*, *hrpZ1*, *hrpW1* and *hopAF1* were assembled intact [Additional file 4: Figure S2], indicating that there is nothing inherently 'un-assemble-able' about these gene sequences. Sequence alignment is much more robust than *de novo* assembly and is not subject to assembly artefacts. The alignments supported the presence of a large deletion in *hrpZ1*. However, the alignments were not consistent with the assembly for *avrE1* and *hrpW1*. Therefore, we amplified the *Pta* 11528 *avrE1* and *hrpW1* genes by PCR and verified their sequences by capillary sequencing [Additional file 5: Table S3]. This confirmed that the apparent deletion in *avrE1* was an artefact of the *de novo* assembly and that the *avrE1* sequence encodes a full-length protein product. Furthermore, transient expression of *avrE1* in *N. benthamiana* induces cell death (S. Gimenez Ibanez and J. Rathjen, unpublished). Capillary sequenc-

ing also confirmed that the *de novo* assembly of *hrpW1* was incorrect and that *Pta* 11528 encodes a full-length HrpW1 protein, albeit with repetitive sequence insertions of 69 and 12 nucleotides relative to the *Pph* 1448A sequence.

The absence of *hopAF1* from Pta 11528 is supported not only by the *de novo* assembly, but also by the absence of aligned (unassembled) reads. As an additional control for the degeneracy of *hopAF1* and *hrpZ1*, we performed the same bioinformatics and sequencing protocols to *Psy* B728a [35] and recovered *hopAF1* and *hrpZ1* intact in the *de novo* assembly assembly (Additional file 4: Figure S1).

Sequence data
In addition to the data available from Genbank accession <u>ACHU00000000</u>, the Velvet assembly and predicted protein sequences are provided in FastA format in Additional files 6 and Additional file 7.

Bioinformatics tools
We used GenomeMatcher [78] for generating and visualising whole-genome alignments. For aligning short Illumina sequence reads against a reference genome, we used MAQ [79] and for other sequence alignments and searches we used BLAST [80]. We used previously published complete genomes as reference sequences for comparative analyses [81-85].

List of abbreviations
CEL: conserved effector locus; EEL: exchangeable effector locus; HMM: hidden Markov model; HopDB: Hop database; MCP: methyl-accepting chemotaxis protein; PCR: polymerase chain reaction; *Pma*: *Pseudomonas syringae* pathovar *maculicola*; *Pph*: *Pseudomonas syringae* pathovar *phaseolicola*; PPHGI-1: *Pph* genomic island 1; *Pta*: *Pseudomonas syringae* pathovar *tabaci*; *Psy*: *Pseudomonas syringae* pathovar *syringae*; *Pto*: *Pseudomonas syringae* pathovar *tomato*; VIGS: virus-induced gene silencing; IS: insertion sequence.

Authors' contributions
DJS and DM performed the sequence assembly and all subsequent bioinformatics analyses. SGI prepared the DNA libraries and performed phenotypic characterisation of *Pta* 11528. JHC performed the functional screen for HrpL-dependent genes and analysed the resulting data. JR, DJS and JHC conceived of the study, participated in its design. DJS, JR and JHC wrote the manuscript. All authors read, approved and made contributions to the manuscript.

Additional material

Additional file 1
Table S1. Proteins encoded in the **Pta** *11528 draft with no detectable homologue in previously sequenced* P. syringae *genomes (***Pto DC3000**, **Psy B728a** *and* **Pph 1448A***). Proteins implicated in mobile genetic elements are shaded in cyan. Other proteins for which a function could be predicted by homology are shaded in yellow.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S1.html]

Additional file 2
Figure S1. Alignment of the **Pph** *1302A PPHGI-1 pathogenicity island against the* **Pta** *11528 genome assembly. The* Pta *11528 genome sequence is in the upper track, aligned against the* Pph *1302A PPHGI-1 pathogenicity island sequence. (Genbank:* <u>AJ870974</u>*).*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S2.pdf]

Additional file 3
Table S2. Regions of the **Pta** *11528 genome with no nucleotide sequence similarity to the genomes of* **Pto DC3000**, **Pss B728a** *and* **Pph 1448A**.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S3.html]

Additional file 4
Figure S2. The **avrE1**, **hrpZ1**, **hrpW1** *and* **hopAF1** *genes are recovered intact in a* de novo *sequence assembly of Illumina short sequence reads from* **Psy B728a**. *We assembled a 40 × deep dataset (reference 35) of paired 36-nucleotide reads from* Psy *B728a genomic DNA using Velvet 0.7.18, using the same protocol as for the* Pta *11528 data. Panel A shows the MAQ alignment of the B728a Illumina reads (in black) and the* blastn *alignment of the B728a* de novo *assembly (in green) against the* avrE1 *gene in the B728a genome. Panel B shows the alignments against* hrpZ1. *Panel C shows the alignments against* hrpW1. *Panel D shows the alignments against* hopAF1.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S4.pdf]

Additional file 5
Table S3. Verification of predicted genes by capillary sequencing. We verified a selection of genes predicted from the Illumina-based Pta11528 genome sequence assembly by capillary sequencing of cloned PCR products. Sequence reads were trimmed to remove poor quality nucleotide calls and the trimmed sequences were aligned against predicted proteins using TBLASTN.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S5.html]

Additional file 6
The **Pta** *11528 draft genome assembly, in FastA format, generated using Velvet*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-395-S6.fna]

References

1. Gasson MJ: **Indicator Technique for Antimetabolic Toxin Production by Phytopathogenic Species of *Pseudomonas*.** *Appl Environ Microbiol* 1980, **39**:25-29.
2. Ribeiro R de LD, Hagedorn DJ, Durbin RD, Uchytil TF: **Characterization of the Bacterium Inciting Bean Wildfire in Brazil.** *Phytopathology* 1979, **69**:208-212.
3. Thomas MD, Langston-Unkefer PJ, Uchytil TF, Durbin RD: **Inhibition of Glutamine Synthetase from Pea by Tabtoxinine-beta-lactam.** *Plant Physiol* 1983, **71**:912-915.
4. Durbin RD: **Bacterial phytotoxins: Mechanisms of action.** *Cell Mol Life Sci* 1991, **47**:776-783.
5. Turner JG, Taha RR: **Contribution of tabtoxin to the pathogenicity of *Pseudomonas syringae* pv. *tabaci*.** *Physiol plant pathol* 1984, **25**:55-69.
6. Chang JH, Goel AK, Grant SR, Dangl JL: **Wake of the flood: ascribing functions to the wave of type III effector proteins of phytopathogenic bacteria.** *Curr Opin Microbiol* 2004, **7**:11-18.
7. Collmer A, Lindeberg M, Petnicki-Ocwieja T, Schneider DJ, Alfano JR: **Genomic mining type III secretion system effectors in *Pseudomonas syringae* yields new picks for all TTSS prospectors.** *Trends Microbiol* 2002, **10**:462-469.
8. Cornelis GR, Van Gijsegem F: **Assembly and function of type III secretory systems.** *Annu Rev Microbiol* 2000, **54**:735-774.
9. Cunnac S, Lindeberg M, Collmer A: ***Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions.** *Curr Opin Microbiol* 2009, **12**:53-60.
10. Grant SR, Fisher EJ, Chang JH, Mole BM, Dangl JL: **Subterfuge and manipulation: type III effector proteins of phytopathogenic bacteria.** *Annu Rev Microbiol* 2006, **60**:425-449.
11. Greenberg JT, Vinatzer BA: **Identifying type III effectors of plant pathogens and analyzing their interaction with plant cells.** *Curr Opin Microbiol* 2003, **6**:20-28.
12. Jin Q, Thilmony R, Zwiesler-Vollick J, He SY: **Type III protein secretion in *Pseudomonas syringae*.** *Microbes Infect* 2003, **5**:301-310.
13. Lindeberg M, Cartinhour S, Myers CR, Schechter LM, Schneider DJ, Collmer A: **Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains.** *Mol Plant Microbe Interact* 2006, **19**:1151-1158.
14. Lindeberg M, Myers CR, Collmer A, Schneider DJ: **Roadmap to new virulence determinants in *Pseudomonas syringae*: insights from comparative genomics and genome organization.** *Mol Plant Microbe Interact* 2008, **21**:685-700.
15. Wei CF, Kvitko BH, Shimizu R, Crabill E, Alfano JR, Lin NC, Martin GB, Huang HC, Collmer A: **A *Pseudomonas syringae* pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant *Nicotiana benthamiana*.** *Plant J* 2007, **51**:32-46.
16. de Torres M, Mansfield JW, Grabov N, Brown IR, Ammouneh H, Tsiamis G, Forsyth A, Robatzek S, Grant M, Boch J: ***Pseudomonas syringae* effector AvrPtoB suppresses basal defence in Arabidopsis.** *Plant J* 2006, **47**:368-382.
17. Peñaloza-Vázquez A, Preston GM, Collmer A, Bender CL: **Regulatory interactions between the Hrp type III protein secretion system and coronatine biosynthesis in *Pseudomonas syringae* pv. *tomato* DC3000.** *Microbiology* 2000, **146**:2447-2456.
18. Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF, Madupu R, Daugherty S, Brinkac L, Beanan MJ, Haft DH, Nelson WC, Davidsen T, Zafar N, Zhou L, Liu J, Yuan Q, Khouri H, Fedorova N, Tran B, Russell D, Berry K, Utterback T, Van Aken SE, Feldblyum TV, D'Ascenzo M, Deng WL, Ramos AR, Alfano JR, Cartinhour S, Chatterjee AK, Delaney TP, Lazarowitz SG, Martin GB, Schneider DJ, Tang X, Bender CL, White O, Fraser CM, Collmer A: **The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000.** *Proc Natl Acad Sci USA* 2003, **100**:10181-10186.
19. Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A, Lykidis A, Trong S, Nolan M, Goltsman E, Thiel J, Malfatti S, Loper JE, Lapidus A, Detter JC, Land M, Richardson PM, Kyrpides NC, Ivanova N, Lindow SE: **Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000.** *Proc Natl Acad Sci USA* 2005, **102**:11064-11069.
20. Joardar V, Lindeberg M, Jackson RW, Selengut J, Dodson R, Brinkac LM, Daugherty SC, Deboy R, Durkin AS, Giglio MG, Madupu R, Nelson WC, Rosovitz MJ, Sullivan S, Crabtree J, Creasy T, Davidsen T, Haft DH, Zafar N, Zhou L, Halpin R, Holley T, Khouri H, Feldblyum T, White O, Fraser CM, Chatterjee AK, Cartinhour S, Schneider DJ, Mansfield J, Collmer A, Buell CR: **Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition.** *J Bacteriol* 2005, **187**:6488-6498.
21. Gardan L, Shafik H, Belouin S, Broch R, Grimont F, Grimont PA: **DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae sp. nov.* and *Pseudomonas cannabina sp. nov.* (ex Sutic and Dowson 1959).** *Int J Syst Bacteriol* 1999, **49**:469-478.
22. Sarkar SF, Guttman DS: **Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen.** *Appl Environ Microbiol* 2004, **70**:1999-2012.
23. Reinhardt JA, Baltrus DA, Nishimura MT, Jeck WR, Jones CD, Dangl JL: **De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*.** *Genome Res* 2009, **19**:294-305.
24. Almeida NF, Yan S, Lindeberg M, Studholme DJ, Schneider DJ, Condon B, Liu H, Viana CJ, Warren A, Evans C, Kemen E, Maclean D, Angot A, Martin GB, Jones JD, Collmer A, Setubal JC, Vinatzer BA: **A draft genome sequence of *Pseudomonas syringae* pv. *tomato* T1 reveals a type III effector repertoire significantly divergent from that of *Pseudomonas syringae* pv. *tomato* DC3000.** *Mol Plant Microbe Interact* 2009, **22**:52-62.
25. Inoue Y, Takikawa Y: **The *hrpZ* and *hrpA* genes are variable, and useful for grouping *Pseudomonas syringae* bacteria.** *J Gen Plant Pathol* 2006, **72**:26-33.
26. Taguchi F, Tanaka R, Kinoshita S, Ichinose Y, Imura Y, Andi S, Toyoda K, Tomonori Shiraishi T, Yamada T: **Harpin$_{psta}$ from *Pseudomonas syringae* pv. *tabaci* Is Defective and Deficient in Its Expression and HR-inducing Activity.** *Pseudomonas syringae* 2001, **67**:116-123.
27. Holt RA, Jones SJ: **The new paradigm of flow cell sequencing.** *Genome Res* 2008, **18**:839-846.
28. Mardis ER: **Next-generation DNA sequencing methods.** *Annu Rev Genomics* 2008, **9**:387-402.
29. MacLean D, Jones JDG, Studholme DJ: **Application of 'next-generation' sequencing technologies to microbial genetics.** *Nature Rev Microbiol* 2009, **7**:287-296.
30. Aury JM, Cruaud C, Barbe V, Rogier O, Mangenot S, Samson G, Poulain J, Anthouard V, Scarpelli C, Artiguenave F, Wincker P: **High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies.** *BMC Genomics* 2008, **9**:603.
31. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB: **ALLPATHS: de novo assembly of**

whole-genome shotgun microreads. *Genome Res* 2008, **18**:810-820.

32. Chaisson M, Pevzner P, Tang H: **Fragment assembly with short reads.** *Bioinformatics* 2004, **20**:2067-2074.

33. Chaisson MJ, Pevzner PA: **Short read fragment assembly of bacterial genomes.** *Genome Res* 2008, **18**:324-330.

34. Dohm JC, Lottaz C, Borodina T, Himmelbauer H: **SHARCGS, a fast and highly accurate short-read assembly algorithm for *de novo* genomic sequencing.** *Genome Res* 2007, **17**:1697-706.

35. Farrer RA, Kemen E, Jones JDG, Studholme DJ: ***De novo* assembly of the *Pseudomonas syringae* pv. *syringae* B728a genome using Illumina/Solexa short sequence reads.** *FEMS Microbiol Lett* 2009, **291**:103-111.

36. Hernandez D, Franvßois P, Farinelli L, Ostervs M, Schrenzel J: **De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer.** *Genome Res* 2008, **18**:802-809.

37. Jeck WR, Reinhardt JA, Baltrus DA, Hickenbotham MT, Magrini V, Mardis ER, Dangl JL, Jones CD: **Extending assembly of short DNA sequences to handle error.** *Bioinformatics* 2007, **23**:2942-2944.

38. Salzberg SL, Sommer DD, Puiu D, Lee VT: **Gene-boosted assembly of a novel bacterial genome from very short reads.** *PLoS Comput Biol* 2008, **4**:e1000186.

39. Sundquist A, Ronaghi M, Tang H, Pevzner P, Batzoglou S: **Whole-genome sequencing and assembly with high-throughput, short-read technologies.** *PLoS ONE* 2007, **2**:e484.

40. Warren RL, Sutton GG, Jones SJ, Holt R: **Assembling millions of short DNA sequences using SSAKE.** *Bioinformatics* 2007, **23**:500-501.

41. Zerbino DR, Birney E: **Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**:821-829.

42. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**:R12.

43. Lindeberg M, Stavrinides J, Chang JH, Alfano JR, Collmer A, Dangl JL, Greenberg JT, Mansfield JW, Guttman DS: **Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*.** *Mol Plant Microbe Interact* 2005, **18**:275-282.

44. Alfano JR, Charkowski AO, Deng WL, Badel JL, Petnicki-Ocwieja T, van Dijk K, Collmer A: **The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants.** *Proc Natl Acad Sci USA* 2000, **97**:4856-4861.

45. Tampakaki AP, Panopoulos NJ: **Elicitation of hypersensitive cell death by extracellularly targeted HrpZPsph produced in planta.** *Mol Plant Microbe Interact* 2000, **13**:1366-74.

46. Preston G, Huang HC, He SY, Collmer A: **The HrpZ proteins of *Pseudomonas syringae* pvs. *syringae*, *glycinea*, and *tomato* are encoded by an operon containing *Yersinia ysc* homologs and elicit the hypersensitive response in tomato but not soybean.** *Mol Plant Microbe Interact* 1995, **8**:717-732.

47. Lee J, Klessig DF, Nürnberger T: **A harpin binding site in tobacco plasma membranes mediates activation of the pathogenesis-related gene HIN1 independent of extracellular calcium but dependent on mitogen-activated protein kinase activity.** *Plant Cell* 2001, **13**:1079-1093.

48. Li CM, Haapalainen M, Lee J, Nürnberger T, Romantschuk M, Taira S: **Harpin of *Pseudomonas syringae* pv. *phaseolicola* harbors a protein binding site.** *Mol Plant Microbe Interact* 2005, **18**:60-66.

49. Maurelli AT: **Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens.** *FEMS Microbiol Lett* 2007, **267**:1-8.

50. Maurelli AT, Fernández RE, Bloch CA, Rode CK, Fasano A: **"Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*.** *Proc Natl Acad Sci USA* 1998, **95**:3943-3948.

51. Casalino M, Latella MC, Prosseda G, Ceccarini P, Grimont F, Colonna B: **Molecular evolution of the lysine decarboxylase-defective phenotype in *Shigella sonnei*.** *Int J Med Microbiol* 2005, **294**:503-512.

52. Day WA Jr, Fernández RE, Maurelli AT: **Pathoadaptive mutations that enhance virulence: genetic organization of the *cadA* regions of *Shigella* spp.** *Infect Immun* 2001, **69**:7471-7480.

53. Tsiamis G, Mansfield JW, Hockenhull R, Jackson RW, Sesma A, Athanassopoulos E, Bennett MA, Stevens C, Vivian A, Taylor JD, Murillo J: **Cultivar-specific avirulence and virulence functions assigned to *avrPphF* in *Pseudomonas syringae* pv. *phaseolicola*, the cause of bean halo-blight disease.** *EMBO J* 2000, **19**:3204-3214.

54. Jackson RW, Athanassopoulos E, Tsiamis G, Mansfield JW, Sesma A, Arnold DL, Gibbon MJ, Murillo J, Taylor JD, Vivian A: **Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar *phaseolicola*.** *Proc Natl Acad Sci USA* 1999, **96**:10875-10880.

55. Jackson RW, Mansfield JW, Arnold DL, Sesma A, Paynter CD, Murillo J, Taylor JD, Vivian A: **Excision from tRNA genes of a large chromosomal region, carrying avr *Pph* B, associated with race change in the bean pathogen, *Pseudomonas syringae* pv. *phaseolicola*.** *Mol Microbiol* 2000, **38**:186-197.

56. Jenner C, Hitchin E, Mansfield J, Walters K, Betteridge P, Teverson D, Taylor J: **Gene-for-gene interactions between *Pseudomonas syringae* pv. *phaseolicola* and *Phaseolus*.** *Mol Plant Microbe Interact* 1991, **4**:553-562.

57. Pitman AR, Jackson RW, Mansfield JW, Kaitell V, Thwaites R, Arnold DL: **Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants.** *Curr Biol* 2005, **15**:2230-2235.

58. Stavrinides J, Guttman DS: **Nucleotide sequence and evolution of the five-plasmid complement of the phytopathogen *Pseudomonas syringae* pv. *maculicola* ES4326.** *J Bacteriol* 2004, **186**:5101-5115.

59. Sundin GW: **Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts.** *Annu Rev Phytopathol* 2007, **45**:129-151.

60. Chang JH, Urbach JM, Law TF, Arnold LW, Hu A, Gombar S, Grant SR, Ausubel FM, Dangl JL: **A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*.** *Proc Natl Acad Sci USA* 2005, **102**:2549-2554.

61. Oh HS, Kvitko BH, Morello JE, Collmer A: ***Pseudomonas syringae* lytic transglycosylases coregulated with the type III secretion system contribute to the translocation of effector proteins into plant cells.** *J Bacteriol* 2007, **189**:8277-8289.

62. Von Bodman SB, Bauer WD, Coplin DL: **Quorum sensing in plant-pathogenic bacteria.** *Annu Rev Phytopathol* 2003, **41**:455-482.

63. Brencic A, Winans SC: **Detection of and response to signals involved in host-microbe interactions by plant-associated bacteria.** *Microbiol Mol Biol Rev* 2005, **69**:155-294.

64. Taguchi F, Shibata S, Suzuki T, Ogawa Y, Aizawa S, Takeuchi K, Ichinose Y: **Effects of glycosylation on swimming ability and flagellar polymorphic transformation in *Pseudomonas syringae* pv. *tabaci* 6605.** *J Bacteriol* 2008, **190**:764-768.

65. Cha JY, Lee JS, Oh JI, Choi JW, Baik HS: **Functional analysis of the role of Fur in the virulence of *Pseudomonas syringae* pv. *tabaci* 11528: Fur controls expression of genes involved in quorum-sensing.** *Biochem Biophys Res Commun* 2008, **366**:281-287.

66. Jakob K, Kniskern JM, Bergelson J: **The role of pectate lyase and the jasmonic acid defense response in *Pseudomonas viridiflava* virulence.** *Mol Plant Microbe Interact* 2007, **20**:146-158.

67. Preston GM, Studholme DJ, Caldelari I: **Profiling the secretomes of plant pathogenic Proteobacteria.** *FEMS Microbiol Rev* 2005, **29**:331-360.

68. Barabote RD, Johnson OL, Zetina E, San Francisco SK, Fralick JA, San Francisco MJ: ***Erwinia chrysanthemi tolC* is involved in resistance to antimicrobial plant chemicals and is essential for phytopathogenesis.** *J Bacteriol* 2003, **185**:5772-5778.

69. Kinscherf TG, Willis DK: **The biosynthetic gene cluster for the beta-lactam antibiotic tabtoxin in *Pseudomonas syringae*.** *J Antibiot* 2005, **58**:817-821.

70. Greenberg DE, Porcella SF, Zelazny AM, Virtaneva K, Sturdevant DE, Kupko JJ 3rd, Barbian KD, Babar A, Dorward DW, Holland SM: **Genome sequence analysis of the emerging human pathogenic acetic acid bacterium *Granulibacter bethesdensis*.** *J Bacteriol* 2007, **189**:8727-8736.

71. Arand M, Hemmer H, Dürk H, Baratti J, Archelas A, Furstoss R, Oesch F: **Cloning and molecular characterization of a soluble epoxide hydrolase from *Aspergillus niger* that is related to**

mammalian microsomal epoxide hydrolase. *Biochem J* 1999, **344:**273-280.

72. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2008, **36:**D281-8.

73. Mandel MJ, Wollenberg MS, Stabb EV, Visick KL, Ruby EG: **A single regulatory gene is sufficient to alter bacterial host range.** *Nature* 2009, **12:**215-8.

74. Pasloske BL, Sastry PA, Finlay BB, Paranchych W: **Two unusual pilin sequences from different isolates of *Pseudomonas aeruginosa.*** *J Bacteriol* 1988, **170:**3738-3741.

75. Pizarro-Cerdá J, Cossart P: **Bacterial adhesion and entry into host cells.** *Cell* 2006, **124:**715-727.

76. Stothard P, Wishart DS: **Circular genome visualization and exploration using CGView.** *Bioinformatics* 2005, **21:**537-539.

77. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: **The generic genome browser: a building block for a model organism system database.** *Genome Res* 2002, **12:**1599-1610.

78. Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M: **GenomeMatcher: a graphical user interface for DNA sequence comparison.** *BMC Bioinformatics* 2008, **9:**376.

79. Li H, Ruan J, Durbin R: **Mapping short DNA sequencing reads and calling variants using mapping quality scores.** *Genome Res* 2008, **18:**1851-1858.

80. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

81. Camargo LE, da Silva AC, Moon DH, Takita MA, Lemos EG, Machado MA, Ferro MI, da Silva FR, Goldman MH, Goldman GH, Lemos MV, El-Dorry H, Tsai SM, Carrer H, Carraro DM, de Oliveira RC, Nunes LR, Siqueira WJ, Coutinho LL, Kimura ET, Ferro ES, Harakava R, Kuramae EE, Marino CL, Giglioti E, Abreu IL, Alves LM, do Amaral AM, Baia GS, Blanco SR, Brito MS, Cannavan FS, Celestino AV, da Cunha AF, Fenille RC, Ferro JA, Formighieri EF, Kishi LT, Leoni SG, Oliveira AR, Rosa VE Jr, Sassaki FT, Sena JA, de Souza AA, Truffi D, Tsukumo F, Yanai GM, Zaros LG, Civerolo EL, Simpson AJ, Almeida NF Jr, Setubal JC, Kitajima JP: **Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa.*** *J Bacteriol* 2003, **185:**1018-26.

82. Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J: **The genome of *Rhizobium leguminosarum* has recognizable core and accessory components.** *Genome Biol* 2006, **7:**R34.

83. Hayden HS, Gillett W, Saenphimmachak C, Lim R, Zhou Y, Jacobs MA, Chang J, Rohmer L, D'Argenio DA, Palmieri A, Levy R, Haugen E, Wong GK, Brittnacher MJ, Burns JL, Miller SI, Olson MV, Kaul R: **Large-insert genome analysis technology detects structural variation in *Pseudomonas aeruginosa* clinical strains from cystic fibrosis patients.** *Genomics* 2008, **91:**530-537.

84. Iacono M, Villa L, Fortini D, Bordoni R, Imperi F, Bonnal RJ, Sicheritz-Ponten T, De Bellis G, Visca P, Cassone A, Carattoli A: **Whole-genome pyrosequencing of an epidemic multidrug-resistant *Acinetobacter baumannii* strain belonging to the European clone II group.** *Antimicrob Agents Chemother* 2008, **52:**2616-2625.

85. Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choisne N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguier P, Thébault P, Whalen M, Wincker P, Levy M, Weissenbach J, Boucher CA: **Genome sequence of the plant pathogen *Ralstonia solanacearum.*** *Nature* 2002, **415:**497-502.