Short communication

***In silico* analysis reveals multiple putative type VI secretion systems and effector proteins in *Pseudomonas syringae* pathovars**

PANAGIOTIS F. SARRIS^{1,2,*}, NICHOLAS SKANDALIS^{1,2,3}, MICHAEL KOKKINIDIS^{1,2} AND NICKOLAS J. PANOPOULOS^{1,2,4,*}

¹Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Hellas, PO Box 1527, Heraklion 71110, Crete, Greece

²Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece

³Laboratory of Bacteriology, Department of Phytopathology, Benaki Phytopathological Institute, St. Delta 8, 14561 Kifissia, Attica, Greece

⁴Department of Environmental Science, Policy and Management, 137 Mulford Hall, University of California, Berkeley, CA 94720, USA

SUMMARY

Type VI secretion systems (T6SS) of Gram-negative bacteria form injectisomes that have the potential to translocate effector proteins into eukaryotic host cells. *In silico* analysis of the genomes in six *Pseudomonas syringae* pathovars revealed that *P. syringae* pv. *tomato* DC3000, pv. *tabaci* ATCC 11528, pv. *tomato* T1 and pv. *oryzae* 1-6 each carry two putative T6SS gene clusters (HSI-I and HSI-II; HSI: Hcp secretion island), whereas pv. *phaseolicola* 1448A and pv. *syringae* B728 each carry one. The pv. *tomato* DC3000 HSI-I and pv. *tomato* T1 HSI-II possess a highly similar organization and nucleotide sequence, whereas the pv. *tomato* DC3000, pv. *oryzae* 1-6 and pv. *tabaci* 11528 HSI-II are more divergent. Putative effector orthologues vary in number among the strains examined. The Clp-ATPases and IcmF orthologues form distinct phylogenetic groups: the proteins from pv. *tomato* DC3000, pv. *tomato* T1, pv. *oryzae* and pv. *tabaci* 11528 from HSI-II group together with most orthologues from other fluorescent pseudomonads, whereas those from pv. *phaseolicola*, pv. *syringae*, pv. *tabaci*, pv. *tomato* T1 and pv. *oryzae* from HSI-I group closer to the *Ralstonia solanacearum* and *Xanthomonas* orthologues. Our analysis suggests multiple independent acquisitions and possible gene attrition/loss of putative T6SS genes by members of *P. syringae*.

Protein secretion/translocation systems play crucial roles in the interactions of pathogenic, symbiotic and commensal bacteria with eukaryotic host organisms. Such systems include the

so-called type I (T1SS), type II (T2SS), type III (T3SS), type IV (T4SS), type V (T5SS) and type VI (T6SS) systems (Economou *et al.*, 2006). T2SS and T5SS secretion systems transport proteins to the extracellular milieu via a two-step process. In general, T1SS, T3SS and T4SS secretion systems involve complex multi-protein channels (injectisomes) dedicated to the transport of selected sets of passenger proteins (effectors) from the bacterial cytoplasm, either to the extracellular environment or directly into eukaryotic cells (interkingdom protein transfer), probably bypassing both the bacterial periplasm and the prokaryote–eukaryote intercellular space. The secretion pathway T6SS remains unclear in this respect. T4SSs are unique among these systems in their ability to transport genetic molecules (DNA), either between bacteria or across the prokaryote–eukaryote phylogenetic divide, although at least some also transport proteins (Christie and Cascales, 2005).

T6SS is a recently characterized system that is composed of 15–20 proteins whose functions are not well understood (Bingle *et al.*, 2008; Cascales, 2008; Filloux *et al.*, 2008; Shrivastava and Mande, 2008). This system initially drew attention as a conserved family of pathogenicity islands or as an atypical T4SS gene locus in Gram-negative bacteria, before being identified as encoding a protein secretory machinery required for virulence in *Vibrio cholerae* (Pukatzki *et al.*, 2006) and *Pseudomonas aeruginosa* (Paer) (Mougous *et al.*, 2006). T6SS landmarks include an AAA+ Clp-like ATPase (ClpB/ClpV), a regulatory FHA (fork-head associated) domain protein (Fha), IcmF/IcmH(DotU)-like proteins that are homologous to T4SS membrane components (Sexton *et al.*, 2004) and the secreted Hcp-like (haemolysin-coregulated) and VgrG-like (valine–glycine repeat) proteins, proposed effectors that are also essential components of the secretion machinery (reviewed in Pukatzki *et al.*, 2009). The ClpB/V homologues

*Correspondence: Email: panopoul@imbb.forth.gr; sarrispanos@science.agrool.gr

belong to a subtype of ATPases, namely the AAA+ family (Schlieker *et al.*, 2005), and appear to be essential for T6SS-dependent secretion (Mougous *et al.*, 2006). The ClpV group lacks the protein-disaggregating function of ClpB-type proteins, although still provides energy to the protein secretion process (Schlieker *et al.*, 2005). Several additional proteins have also been implicated in T6SS-mediated Hcp protein secretion in *Paer* strain PA01 (Mougous *et al.*, 2007): PpkA (a serine/threonine protein kinase) and PppA (a phosphatase) control T6SS function at the post-translational level. These proteins act on a common protein substrate, Fha1, necessary for ClpV1 localization to T6SS foci. PpkA becomes activated by autophosphorylation under certain environmental conditions and, once these are removed, it dephosphorylates Fha1 and possibly PpkA, switching off T6SS. Several other genes, as well as quorum sensing (QS), also control T6SS gene transcription in various bacteria (Filloux *et al.*, 2008; Ishikawa *et al.*, 2009; Lesic *et al.*, 2009; Liu *et al.*, 2008). In *Paer* strain PA01, one of the three T6SS systems (HSI-I; HSI: Hcp secretion island) is transcriptionally coregulated with T3SS by the *retS/ladS* genes (Filloux *et al.*, 2008). Furthermore, in the multi-host strain PA14 of *Paer*, HSI-I is negatively regulated by both the homoserine lactone transcription factor LasR and the 4-hydroxyquinoline (HAQ) transcriptional regulator MvfR. The other two T6SSs (HSI-II and HSI-III) are positively controlled by LasR and MvfR, whereas PqsE, a key component of the MvfR regulon, is required for the expression of part of HSI-III, but not HSI-II (Lesic *et al.*, 2009). The presence of genes coding for a putative σ^{54} -dependent transcriptional factor, *sfa*, in HSIs further suggests an additional level of regulation (Lesic *et al.*, 2009).

In addition to the mammalian pathogens mentioned above, T6SS clusters have been implicated in the virulence of certain plant pathogenic bacteria, such as *Agrobacterium tumefaciens* (Wu *et al.*, 2008), *Pectobacterium atrosepticum* (Liu *et al.*, 2008; Mattinen *et al.*, 2007) and *Xanthomonas oryzae* (Bingle *et al.*, 2008), as well as in the multihost pathogen of *Paer* strain PA14 (Lesic *et al.*, 2009). In addition to pathogenesis, T6SS may modulate root colonization/nodule formation by the nitrogen-fixing plant symbionts/mutualists *Mesorhizobium loti*, *Rhizobium leguminosarum* (Bingle *et al.*, 2008; Bladergroen *et al.*, 2003) and *Cupriavidus taiwanensis*, which belongs to the phylogenetically distinct β -Rhizobium group (Amadou *et al.*, 2008). Recent *in silico* analyses of prokaryotic genomes (Bingle *et al.*, 2008; Boyer *et al.*, 2009) have further revealed that T6SSs are widespread among Proteobacteria, suggesting that they may be involved in yet unknown pathogenic or symbiotic lifestyles or other types of cell–cell communication.

In a previous study, Shrivastava and Mande (2008) reported the presence of putative core T6SS genes in some members of *Pseudomonas syringae*. In this study, we present a comparative *in silico* analysis of putative T6SS components in six strains of *P. syringae*. For three of these strains [*P. syringae* pv. *tomato* DC3000

(*PstomDC*), *P. syringae* pv. *syringae* B728a (*Psyr*) and *P. syringae* pv. *phaseolicola* 1448A (*Psph*)], whole genome sequences have been extensively annotated (<http://pseudomonas-syringae.org>), whereas, for three others [*P. syringae* pv. *oryzae* 1-6 (*Psoryz*), *P. syringae* pv. *tabaci* ATCC 11528 (*Pstab*) and *P. syringae* pv. *tomato* T1 (*PstomT1*)], full or partial genomic sequences are available at the draft stage. Our analysis allowed us to identify genes encoding for putative T6SS core components and candidate effectors that had not been described previously in the aforementioned bacteria. In addition, we carried out a phylogenetic analysis of the ClpV/B ATPases and lcmF orthologues, which provides a basis for an understanding of the phylogenetic roots of T6SSs in pathogenic members of *P. syringae*.

To retrieve T6SS homologue coding loci in the phytopathogenic *P. syringae* genomes, BLASTP and reverse BLASTP analyses (Supporting data file, see Supporting Information) were initially performed in the annotated proteomes of *PstomDC*, *Psyr* and *Psph* (PPI Home page, <http://pseudomonas-syringae.org>) and in the *Pseudomonas* genome database (Stover *et al.*, 2000; Winsor *et al.*, 2009; <http://pseudomonas.com>), using standard/default BLAST parameters. Bait protein sequences coded by the T6SS gene islands (HSI-I, HSI-II and HSI-III) of *Paer* strains PA01 and PA14 were used, as they have been reported by Mougous *et al.* (2006) and Lesic *et al.* (2009). Subsequently, BLASTP searches were carried out against all KEGG bacterial genomes (>500) using, as bait, all putative T6SS proteins (HSI-Is) from *P. syringae* pv. *phaseolicola* and *P. syringae* pv. *syringae* B728a. The closest orthologues identified were from *Paer*, *P. fluorescens* and *P. putida* plus those of HSIs in other *P. syringae* strains.

Gene clusters potentially coding for putative T6SS core components also exist in the other strains, with either complete or segmental synteny to those in *Paer*. For example, lcmF-lcmH-HsiJ and ClpV-HsiH-HisG-HsiF proteins are coded by adjacent genes, in the same order and orientation, in the HSIs of all six strains, as in *Paer* (Fig. 1). Some HSIs of *P. syringae* strains also have adjacent *vgr-hcp* genes, whereas others do not, as is the case with *Paer* HSIs. These physical arrangements may indicate common regulation and possible physical or functional interactions. The key features of the HSIs in the different strains examined are described below.

Two contiguous clusters coding for potential T6SS proteins were found in the *PstomDC* genome and were designated HSI-I and HSI-II. The HSI-I cluster comprised 15 genes [open reading frames (ORFs), locus numbers PSPTO_2538 to PSPTO_2554], whereas the HSI-II cluster contained 22 (PSPTO_5415 to PSPTO_5436; homology between corresponding loci in the two clusters is given in Table S1, see Supporting Information). The contiguity of each cluster is interrupted between *hsiB* and *hcp* orthologues by the presence of a putative ISPsy transposase gene (PSPTO_2541 in HSI-I and PSPTO_5428 in HSI-II) bracketed by uncharacterized DNA segments, and between *hsiJ* and *sfa*

Fig. 1 Putative type VI secretion system (T6SS) clusters from *Pseudomonas syringae* pathovars. The amino acid sequences of proteins encoded by the three T6SS gene clusters of *P. aeruginosa* (two, HSI-I and HSI-III, are included in this diagram) were used as queries for BLASTP and reverse BLAST searches against the genomes of the *P. syringae* strains. The putative T6SS clusters deduced for the two *P. syringae* pv. *tomato* strains (DC3000 and T1), *P. syringae* pv. *syringae* strain B728a, *P. syringae* pv. *phaseolicola* strain 1448a, *P. syringae* pv. *tabaci* strain ATCC 11528 and *P. syringae* pv. *oryzae* strain 1-6. Orthologues (indicated by the same colours) are based on a BLASTP *E*-value of less than e^{-10} or on the annotated or previously published data (Lindeberg *et al.*, 2008; Wu *et al.*, 2008). Genes adjacent or assigned to the T6SS clusters, but not recognized as orthologues, are indicated by light beige arrows. Arrows indicate the direction of transcription. In each cluster, the four-digit locus number is shown only for the left-most gene and the last two digits for the other genes. Published or annotated gene/protein designations are indicated below each gene. Bacterial name abbreviations are as follows: *P. aeruginosa* (*Pseudomonas aeruginosa* PA01); *Pstom* DC3000 (*Pseudomonas syringae* pv. *tomato* DC3000); *Pstom* T1 (*Pseudomonas syringae* pv. *tomato* T1); *Psyr* B728a (*Pseudomonas syringae* pv. *syringae* B728a); *Psph* 1448a (*Pseudomonas syringae* pv. *phaseolicola* 1448a); *Pstab* 11528 (*Pseudomonas syringae* pv. *tabaci* ATCC 11528); *Psoryz* 1-6 (*Pseudomonas syringae* pv. *oryzae* 1-6).

orthologues by two small putative ORFs (PSPTO_2550 and PSPTO_2551 in HSI-I and PSPTO_5421 and PSPTO_5423 in HSI-II) unrelated to T6SS components (Fig. 1). A gene coding for a VgrG homologue is terminally located in HSI-I, whereas two such genes are found at either end of HSI-II. *fha* and *pppA* orthologues (PSPTO_5422 and PSPTO_5417) are present only in HSI-II, between or adjacent to small putative ORFs unrelated to T6SS. A *pppA/stp* locus is embedded in HSI-II, but is absent from HSI-I; *pppB* orthologues appear to be absent, but a *ppkA* orthologue is found in a distant location (PSPTO_2874), next to a gene coding for the putative type III effector HopL1 (PSPTO_2872) plus two putative ABC transporters (Fig. 1). The possible significance of this linkage is not apparent and it should be noted that the *P. syringae* HopL proteins were recently invalidated as type III effectors (see List of Assigned Hop names at http://pseudomonas-syringae.org/hop_used_names.htm). Nevertheless, the linkage of T6SS and putative T3SS effector genes apparently occurs near one of the HSI-II and HSI-III termini in the multihost pathogen *Paer* strain PA14 (Lesic *et al.*, 2009).

In the *Psyr* genome, six contiguous genes (*hsiB*, *hsiC*, *hsiF*, *hsiG*, *hsiH* and *clpB/V*) coding for orthologous proteins are found in the same order and orientation to those in the *PstomDC* HSI-I counterparts, as in each HSI of PAO1 and PA14 (Fig. 1, and data not shown). Amino acid sequence identities between these and the *Paer* PAO1 orthologues did not exceed 56% (Table S3, see Supporting Information). A nearly identical cluster to *Psyr* HSI is found in *Psph*, with amino acid identity values to *Paer* PAO1 counterparts up to 60% (Table S2, see Supporting Information). Distinct features of the HSIs in both *Psyr* and *Psph* are the presence of genes unlinked to the core clusters coding for putative VgrG-like proteins, the presence of a *ppkA* orthologue in *Psyr* but not in *Psph* and the extra copies of *hcp*-like genes, especially in *Psph*, that are scattered throughout the genome, except for one copy which is found next to the *ompA*-like genes near the terminus of the HSI clusters of both bacteria. A *pppA* orthologue is also present, but at a distant location from the main clusters in both strains (Fig. 1).

To identify candidate T6SS clusters in the draft versions of the genome sequences of the other *P. syringae* strains, BLAST analyses were performed in GENBANK (National Center for Biotechnol-

ogy Information, NCBI) using, as bait, the predicted protein sequences coded in the HSIs of *PstomDC* and *Psph*. Two putative T6SS clusters were found in *PstomT1*, *Psoryz* and *Pstab* (Fig. 1; Tables S4a,b and S5a,b, see Supporting Information). In the *PstomT1* genome, a cluster of 11 highly homologous genes (99–100%) is found in near-perfect synteny to the HSI-I counterparts of *PstomDC* (Table S4a). When *PstomDC* HSI-II-predicted gene sequences were used as bait, the only orthologue found in *PstomT1* corresponded to PSPTO_5430 (coding for a protein of unknown function), with 45% identity to the bait sequence. Nucleotide identities between *Psoryz* and *Pstab* genes and the *PstomDC* HSI-I bait sequences varied from 54% to 96% and from 50% to 76%, respectively, except for orthologues of PSPTO_5422 (coding for an Fha domain protein), which has a disrupted orthologue with 57% homology in *Psoryz* HSI-II (Psrp01_18588) and seems to be absent from *Pstab* (HSI-II) (Tables S4a and S5b). Finally, a gene coding for a PppA orthologue is also distant from the putative HSIs in *PstomT1* (PSP-TOT1_4968) and *Psoryz* (see Table S4), as in *Psph* and *Psyr* (Fig. 1). A putative *ppkA* orthologue is present at some distance (70 ORFs) away from the putative *PstomT1* HSI and in the same gene context (i.e. near genes coding a putative HopL1 and ABC transporter) as its orthologues in *PstomDC* and *Psyr*.

Similar searches in the *PstomT1*, *Psoryz* and *Pstab* genomes revealed a second putative T6SS cluster (HSI-II) which has an identical organization and high nucleotide sequence homology (93–100%; Tables S4b and S5a) to the *Psph* T6SS.

Based on the studies of Mougous *et al.* (2006) and Lesic *et al.* (2009), we further sought to identify candidate regulatory gene orthologues to *retS/ladS*, *mvfR*, *lasR*, *pqsE* and *sfa* in the *P. syringae* strains by BLAST searches using the aforementioned *Paer* proteins as bait. Genes coding for putative RetS homologues have been annotated in the PPI Home site (<http://pseudomonas-syringae.org>) for *PstomDC* (PSPTO_4868), *Psyr* (Psr_4088) and *Psph* (Psph_4451), and candidate homologues were identified in our BLAST searches for *PstomT1* (annotated as PSPTOT1_0312 in NCBI), *Pstab* and *Psoryz* (data not shown). LadS homologues also exist in *PstomDC* (PSPTO_4796), *Psyr* (Psr_4339), *Psph* (PSPPH_4381), *PstomT1*, *Pstab* and *Psoryz* (data not shown). No candidate MvfR, LasR or PqsE homologues

Fig. 2 Evolutionary relationships of 47 ClpB/V orthologue proteins from various bacterial species. The evolutionary history was inferred using the neighbour-joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 6.63970056 is shown. The percentage of replicate trees in which the associated ClpB/V proteins clustered together in the bootstrap test (1000 replicates) is shown next to each branch (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965), and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 782 positions in the final dataset. Phylogenetic analyses were conducted by means of MEGA4 software (Tamura *et al.*, 2007). Bacterial name abbreviations are as follows: *PstomDC* (*Pseudomonas syringae* pv. *tomato* DC3000); *PstomT1* (*Pseudomonas syringae* pv. *tomato* T1); *Psyr* (*Pseudomonas syringae* pv. *syringae* B728a); *Psph* (*Pseudomonas syringae* pv. *phaseolicola* 1448a); *Psoryz* (*Pseudomonas syringae* pv. *oryzae* 1-6); *Pstab* (*Pseudomonas syringae* pv. *tabaci* 11528). Abbreviated genus names: E, *Escherichia*; P, *Pseudomonas*; R, *Ralstonia*; X, *Xanthomonas*.

were identified in any of the six *P. syringae* strains. Finally, genes for Sfa homologues (σ^{54} -dependent transcriptional activator) were found in both HSI of *PstomDC*, in the HSI-II cluster of *Pstab* and the HSI clusters of *PstomT1* and *Psoryz*, but were seemingly absent from those of *Psyr*, *Psph* and from the *Pstab* HSI-I cluster.

Phylogenetic trees were constructed (see Supporting data file) for the Clp-ATPases and IcmF-like proteins of various bacterial species of different phylogenetic proximity to *P. syringae* pathovars. The Clp protein tree shows two deep branches (designated cluster I and cluster II in Fig. 2). Cluster I carries all but one *P. syringae* orthologue and is divided into two broad groups. One (called the 'Pst' group) includes the two *PstomDC* ClpB/V orthologues (PSPTO_2548 and PSPTO_5425), together with the *PstomT1*, *Psoryz* and *Pstab* ClpB/V proteins PSPTOT1_2217, Psyrpo1_18603 and PsyrptA_32530, as well as most Clp orthologues from other fluorescent pseudomonads [*Pseudomonas entomophila*, *Pseudomonas fluorescens* and *Pseudomonas mendocina* plus the ClpV2 protein (PA1662) of the HSI-II cluster of *Paer*]. In the other group ('Psph' group), the *Psph* ClpB/V (PSPPH_0129), *Pstab* ClpB/V (PsyrptA_10435), *Psyr* ClpB/V (Psyr_4958), *PstomT1* ClpB (PSPTOT1_4113) and *Psoryz* ClpB/V (Psyrpo1_25205) proteins cluster very close to each other, having as nearest neighbours ClpB/V orthologues from phylogenetically distant pathogens (*Xanthomonas oryzae* pv. *oryzicola*, *Xanthomonas oryzae* pv. *oryzae* and *Ralstonia solanacearum*). The putative ClpB/V protein of *Psph*, coded by the solitary locus PSPH_0742, appears in cluster II of the Clp tree, close to the *P. putida* (Pput_0665) and *P. fluorescens* (PFL_5304) orthologues. Finally, *Paer* ClpV1 (HSI-I), ClpV2 (HSI-II) and ClpV3 (HSI-III) are located in separate clades of the tree.

In the IcmF tree (Fig. 3), two distinct subgroups were also formed (also designated 'Pstom' and 'Psph' groups). The *PstomDC* IcmF paralogues (coded by PSPTO_2554 and PSPTO_5418), together with the *PstomT1* (PSPTOT1_2212), *Psoryz* (Psyrpo1_18563) and *Pstab* (PsyrptA_32495) proteins, clustered together with the IcmF proteins from *P. entomophila*, *P. fluorescens*, *P. mendocina* and *Paer* IcmF2 of the HSI-II locus (PA1669) in the *Pstom* group. The IcmF orthologues of *Psph* (PSPPH_0125), *Psyr* (Psyr_4962), *Pstab* (PsyrptA_10410), *PstomT1* (PSPTOT1_4109) and *Psoryz* (Psyrpo1_25185) clustered very close to each other, with the *R. solanacearum* and *X. oryzae* pv. *oryzicola* IcmF homologues (RSp0763 and Xoryz_6365) being their nearest neighbours. Interestingly, the *Paer* IcmF3 protein (PA2361) appeared as an outgroup in our analysis, whereas the *Paer* IcmF1 homologue (PA0077) clustered with the *X. campestris* pv. *citri*, *X. axonopodis* pv. *citri* and *X. campestris* pv. *vesicatoria* IcmF proteins (XAC4119, XAC4119, XCV2138; Fig. 3).

Identical tree topologies were obtained by phylogenetic tree construction for three additional T6SS core proteins (ImpB, ImpC and DotU) (Figs S2–S4, see Supporting Information).

Our computational genomic screening approach has revealed the presence of contiguous genomic clusters coding for putative T6SS core components, plus single or multiple unlinked loci coding for putative T6SS effector proteins, in all six *P. syringae* strains examined. Duplicate T6SS clusters occur in four of the strains examined (*PstomDC*, *PstomT1*, *Pstab* and *Psoryz*), representative of pathovars with distinct host specialization. The other two strains (*Psph* and *Psyr*) harbour single T6SS clusters. Several of the strains carry multiple copies of genes coding for Hcp-Vgr- and Hsi-like proteins. We have not assigned specific gene/protein designations in the *P. syringae* HSI at this stage of our analysis to avoid confusion in ongoing genome annotations. In addition to structural core proteins and putative effectors that serve as landmarks of T6SS clusters, all six strains examined carry genes potentially coding for the regulatory proteins RetS and LadS, but not MvfR or PsqE homologues, and these are not linked to the HSI, as is also the case for other bacteria. Genes for Sfa orthologues were found in both HSI of *PstomDC* and in HSI-II of *Pstab*, *PstomT1* and *Psoryz*, but were seemingly absent from the HSI of *Psyr* and *Psph* and HSI-I of *Pstab*, *PstomT1* and *Psoryz*, where they were apparently replaced by homologues annotated as VirB (VasD) in the KEGG database. Genes for putative PpkA orthologues were found in three strains (*PstomDC*, *Psyr* and *PstomT1*) that also carry a gene for a T3SS effector HopL1-like protein, but appear to be absent from the other strains. With regard to *pppA*, the PppA/Stp protein embedded in HSI-II of *PstomDC* matches its PA01 HSI-I orthologue along its entire length, whereas the predicted proteins in all other strains/HSIs are longer (marked by a star in Fig. 1) and their N-terminal portion matches the PP2Cc superfamily of serine/threonine protein kinases.

Partial or full sets of T6SS orthologues have been described in several phytopathogenic and symbiotic bacteria, including *P. syringae*, *Erwinia carotovora*, *Pe. atrosepticum*, *R. solanacearum* and *Burkholderia cepacia* (Bingle *et al.*, 2008; Boyer *et al.*, 2009; Lindeberg *et al.*, 2008; Liu *et al.*, 2008; Shrivastava and Mande, 2008; Wu *et al.*, 2008), and have been implicated in the virulence of *A. tumefaciens* (Wu *et al.*, 2008), *Pe. atrosepticum* (Liu *et al.*, 2008; Mattinen *et al.*, 2007), *Xanthomonas* spp. (Bingle *et al.*, 2008) and in the multihost pathogen of *Paer* strain PA14 (Lesic *et al.*, 2009). Whether the putative HSI in *P. syringae* pathovars are involved in pathogenesis or are functional protein secretion/translocation channels, and the environmental conditions under which they might be expressed, remain to be determined. Some observations raise questions as to whether or not these systems are functional in some of the strains studied. For example, the orthologue of the PSPTO_5418 gene in *Pstab* HSI-II, coding for an IcmF-related protein, has a point mutation which creates an in-frame stop codon at position 741 (codon 247), and some ORFs in *Psoryz* and *Pstab* HSI-II seem to be disrupted (Fig. 1). Further, a *ppkA* orthologue appeared to be absent from *Psph* and, as

Fig. 3 Evolutionary relationships of 36 lcmF orthologue proteins of various bacterial species. Tree construction and other details are given in the legend to Fig. 2.

already noted, the predicted PppA proteins in most strains examined (except for *PstomDC*) are coded outside the HSI, and their N-terminal portion matches well the PP2Cc superfamily of serine/threonine protein kinases in BLASTP analysis. Finally, an *impA* gene in *PstomDC* HSI-I seems to be nonfunctional because of a transposon insertion (PSPTO_2541, ISPsy transposase) in the corresponding genomic region.

Our phylogenetic analysis for five core proteins of the T6SS secretion machinery (ClpB/V, IcmF, DotU, ImpB and ImpC) revealed a *Pstom* group, which includes one subgroup with closer phylogenetic relationship among the Clp proteins of *PstomDC*, *PstomT1*, *Psoryz* and *Pstab* from HSI-II, as well as another subgroup formed by the ClpB/V2 protein of *PstomDC*, the *Paer* ClpV2 protein and by homologues from other nonphytopathogenic fluorescent *Pseudomonas*. This group is distinct from ClpB/V of other *P. syringae* pathovars and from *Paer* ClpV2 and ClpV3. The Clp proteins of *Psph*, *Psyr*, *Pstab*, *Psoryz* and *PstomT1* from HSI-I grouped closer to those of *R. solanacearum*, *Xanthomonas* spp. and the ClpV1 and ClpV3 proteins from *Paer*, whereas the second ClpB/V orthologue of *Psph*, coded by the solitary PSPPH_0742 gene, was placed in cluster II separately from all other *P. syringae* Clp orthologues. An analogous situation occurs with IcmF and IcmF homologues from these strains, with the IcmF3 protein of the *Paer* HSI-III locus appearing as an outgroup in the bootstrap analysis. The incongruent phylogenies of these proteins with those of the 16S rRNA gene and other conserved genes, such as *rpoB*, for members of the *Pseudomonas* genus (Tayeb *et al.*, 2005), coupled with the absence of a *PstomDC* HSI-II orthologue cluster in *PstomT1* and the absence of a *PstomT1* HSI-I orthologue cluster in *PstomDC*, suggest that core T6SS genes in the various strains probably originated from multiple independent acquisitions by horizontal gene transfer from different sources.

The involvement of T6SS in pathogenicity has been established for several bacterial pathogens, including *Vibrio cholerae*, *Paer* and *Burkholderia pseudomallei* in mammalian and other eukaryote host models (Mougous *et al.*, 2006; Pukatzki *et al.*, 2006, 2007; Schell *et al.*, 2007). T6SS clusters have also been implicated in the virulence of certain plant pathogenic bacteria, such as *A. tumefaciens* (Wu *et al.*, 2008), *Pe. atrosepticum* (Liu *et al.*, 2008; Mattinen *et al.*, 2007), *Xanthomonas* spp. (Bingle *et al.*, 2008) and *Paer* strain PA14 (Lesic *et al.*, 2009). However, there is no information so far on whether T6SS in members of *P. syringae* is functional and what its biological role might be. To obtain preliminary evidence for transcriptional expression, we carried out reverse transcriptase-polymerase chain reaction (RT-PCR) in total RNA extracts from cultures grown in rich (Luria-Bertani) and minimal (M9) media, after exhaustive treatment with RNase-free DNase I (Roche Applied Science, Indianapolis, IN, USA) to detect putative RNA transcripts of *icmF* orthologues in *PstomDC* (PSPTO_5418 in HSI-II) and *Psph* (PSPPH_0125).

Such transcripts were indeed found under both growth conditions tested (Fig. S1, see Supporting Information), indicating that the gene is probably expressed in both strains.

Multiple copies of T6SS in a single bacterial strain appear to be a frequent phenomenon. Recent studies (Boyer *et al.*, 2009; Filloux *et al.*, 2008) have indicated that multiple copies of apparently complete and/or degenerate T6SS loci occur in about one-third of over 500 of the proteobacterial genomes examined, that they generally display different phylogenetic origins and are not a result of recent duplication events, suggesting sustained and constrained mechanisms that favour this trend. In the case of pathogenic bacteria, such mechanisms may involve host/vector biology, adaptations via regulatory mechanisms and specialization of effectors to host cellular targets. Our study provides a basis for focused investigations on this newly discovered and poorly understood secretion system in this group of plant pathogens.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Greek Ministry of Education (EPEAEK-Protein Biotechnology and Plant Molecular Biology and Biotechnology graduate programs) and by Grants 05NON-EU-170, PEP-DEL4 and PENED875-03ED, implemented within the framework of the 'Reinforcement Programme of Human Research Manpower' (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from EU-European Social Fund from the General Secretariat for Research and Technology of Greece). We thank Dr Laurence Rahme for sharing unpublished data with us.

REFERENCES

- Amadou, C., Pascal, G., Mangenot, S., Glew, M., Bontemps, C., Capela, D., Carrère, S., Cruveiller, S., Dossat, C., Lajus, A., Marchetti, M., Poinot, V., Rouy, Z., Servin, B., Saad, M., Schenowitz, C., Barbe, V., Batut, J., Médigue, C. and Masson-Boivin, C. (2008) Genome sequence of the rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res.* **18**, 1472–1483.
- Bingle, L.E.H., Bailey, C.E. and Pallen, M.J. (2008) Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* **11**, 3–8.
- Bladergroen, M.R., Badelt, K. and Spink, H.P. (2003) Infection blocking genes of a symbiotic *Rhizobium leguminosarum* strain that are involved in temperature-dependent protein secretion. *Mol. Plant-Microbe Interact.* **16**, 53–64.
- Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. and Attree, I. (2009) Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics*, **10**, 1–14.
- Cascales, E. (2008) The type VI secretion toolkit. *EMBO Rep.* **9**, 735–741.
- Christie, P.J. and Cascales, E. (2005) Structural and dynamic properties of bacterial type IV secretion systems. *Mol. Membr. Biol.* **22**, 51–61.

- Economou, A., Christie, P.J., Fernandez, R.C., Palmer, T., Plano, G.V. and Pugsley, A.P. (2006) Secretion by numbers: protein traffic in prokaryotes. *Mol. Microbiol.* **62**, 308–319.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.
- Filloux, A., Hachani, A. and Bleves, S. (2008) The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology*, **154**, 1570–1583.
- Ishikawa, T., Rompikuntal, P.K., Lindmark, B., Milton, D.L. and Wai, S.N. (2009) Quorum sensing regulation of the two *hcp* alleles in *Vibrio cholerae* O1 strains. *PLoS ONE*, **4**, e6734.
- Lesic, B., Starkey, M., He, J., Hazan, R. and Rahme, G.L. (2009) Quorum sensing differentially regulates *Pseudomonas aeruginosa* type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. *Microbiology*, **155**, 2845–2855 (DOI 10.1099/mic.0.029082-0).
- Lindeberg, M., Myers, C.R., Collmer, A. and Schneider, D.J. (2008) Roadmap to new virulence determinants in *Pseudomonas syringae*: insights from comparative genomics and genome organization. *Mol. Plant-Microbe Interact.* **21**, 685–700.
- Liu, H., Coulthurst, S.J., Pritchard, L., Hedley, P.E., Ravensdale, M., Humphris, S., Burr, T., Takle, G., Brurberg, M.B., Birch, P.R., Salmond, G.P. and Toth, I.K. (2008) Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. *PLoS Pathog.* **4**, e1000093.
- Mattinen, L., Nissinen, R., Riipi, T., Kalkkinen, N. and Pirhonen, M. (2007) Host-extract induced changes in the secretome of the plant pathogenic bacterium *Pectobacterium atrosepticum*. *Proteomics*, **7**, 3527–3537.
- Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordonez, C.L., Lory, S., Walz, T., Joachimiak, A. and Mekalanos, J.J. (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*, **312**, 1526–1530.
- Mougous, J.D., Gifford, C.A., Ramsdell, T.L. and Mekalanos, J.J. (2007) Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nat. Cell Biol.* **9**, 797–803.
- Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg J.F. and Mekalanos, J.J. (2006) Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. USA*, **103**, 1528–1533.
- Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D. and Mekalanos, J.J. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it crosslinks actin. *Proc. Natl. Acad. Sci. USA*, **104**, 15508–15513.
- Pukatzki, S., McAuley, S.B. and Miyata, S.T. (2009) The type VI secretion system: translocation of effectors and effector-domains. *Curr. Opin. Microbiol.* **12**, 11–17.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Schell, M.A., Ulrich, R.L., Ribot, W.J., Brueggemann, E.E., Hines, H.B., Chen, D., Lipscomb, L., Kim, H.S., Mrázek, J., Nierman, W.C. and Deshazer, D. (2007) Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol. Microbiol.* **64**, 1466–1485.
- Schlieker, C., Zentgraf, H., Dersch, P. and Mogk, A. (2005) ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria. *Biol. Chem.* **386**, 1115–1127.
- Sexton, J.A., Miller, J.L., Yoneda, A., Kehl-Fie, T.E. and Vogel, J.P. (2004) *Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex. *Infect. Immun.* **72**, 5983–5992.
- Shrivastava, S. and Mande, S.S. (2008) Identification and functional characterization of gene components of Type VI secretion system in bacterial genomes. *PLoS ONE*, **3**, 1–11.
- Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E., Lory, S. and Olson, M.V. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, **406**, 959–964.
- Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.
- Tayeb, A.L., Ageron, E., Grimont, F. and Grimont, P.A.D. (2005) Molecular phylogeny of the genus *Pseudomonas* based on *rpoB* sequences and application for the identification of isolates. *Res. Microbiol.* **156**, 763–773.
- Winsor, G.L., Van Rossum, T., Lo, R., Khaira, B., Whiteside, M.D., Hancock, R.E. and Brinkman, F.S. (2009) *Pseudomonas* genome database: facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Res.* **37** (Database issue), D483–D488.
- Wu, H.Y., Chung, P.C., Shih, H.W., Wen, S.R. and Lai, E.M. (2008) Secretome analysis uncovers an Hcp-family protein secreted via a type VI secretion system in *Agrobacterium tumefaciens*. *J. Bacteriol.* **190**, 2841–2850.
- Zuckerandl, E. and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In: *Evolving Genes and Proteins* (Bryson, V. and Vogel, H.J., eds), pp. 97–166. New York: Academic Press.

SUPPORTING INFORMATION

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

Fig. S1 Detection of putative RNA transcripts for the *icmF* loci (A) of *Pseudomonas syringae* pv. *tomato* DC3000 (PSPTO_5418, HSI-II, lanes 1, 2) and *P. syringae* pv. *phaseolicola* 1448A (PSPH_0125, lanes 3, 4) and rRNA (B) by reverse transcriptase-polymerase chain reaction (RT-PCR) carried out in DNAase-treated total RNA extracts from 24-h cultures grown in Luria-Bertani (LB) (lanes 1, 3) and M9-glucose (lanes 2, 4) media. Negative control PCR (C) was performed on total RNA from *P. syringae* pv. *tomato* DC3000 cultivated in M9 medium (lane 1), *P. syringae* pv. *tomato* DC3000 cultivated in LB medium (lane 2), *P. syringae* pv. *phaseolicola* 1448a cultivated in M9 medium (lane 3) and *P. syringae* pv. *phaseolicola* 1448a cultivated in M9 medium (lane 4), without reverse transcriptase assay. Primers were designed to amplify 642- and 771-nucleotide regions of PSPTO_5418 and PSPH_0125 (A), and

a 540-nucleotide region of rRNA (B). Primer sequences were as follows: PSPTO_5418 upper, 5'-GCACCCAGCACTGCGACTGGTACTTC-3'; lower, 5'-CCAGATTGGCGCCGATGCCAGCG-3'; PSPPH_0125 upper, 5'-GCACCCAGCACTGCGACTGGTACTTC-3'; lower, 5'-GGGTCAGGTCACGGTCAGGAAAATG-3'; rRNA upper, 5'-CGGGTACTTGTTACCTGGTGGC-3'; lower, 5'-CTTGCCAGTTTTGGATGCAGTTC-3'. Molecular weight markers were phage λ DNA digested with *HindIII/EcoRI* (A) and *PstI* (B and C). The experiment was repeated with different cultures and with similar results.

Fig. S2 Evolutionary relationships of 36 DotU/ImpK homologue proteins from various bacterial species. The evolutionary history was inferred using the neighbour-joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 10.80240760 is shown. The percentages of replicate trees in which the associated DotU/ImpK clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). There were a total of 205 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura *et al.*, 2007).

Fig. S3 Evolutionary relationships of 25 ImpB homologue proteins from various bacterial species. Tree construction and other details are given in the legend to Fig. S2.

Fig. S4 Evolutionary relationships of 23 ImpG homologue proteins from various bacterial species. Tree construction and other details are given in the legend to Fig. S2.

Table S1 Homologies between hypothetical proteins of the putative type VI secretion system (T6SS) clusters (HSI-I and HSI-II) of *Pseudomonas syringae* pv. *tomato* DC3000. Gene locus numbers are as annotated at PPI Home (<http://pseudomonas-syringae.org>).

Table S2 (a) Putative type VI secretion system (T6SS) proteins in *Pseudomonas syringae* pv. *syringae* B728a (*Psyr*). Locus numbers (*Psyr*) and protein names are as annotated at <http://genome.jgi-psf.org/psesy/psesy.home.html>. Protein identities with PAO1 homologues (as annotated at <http://www.pseudomonas.com>; Stover *et al.*, 2000; Winsor *et al.*, 2009) are included. (b) Reverse BLAST results for putative type VI secretion system (T6SS) proteins

of *Pseudomonas syringae* pv. *syringae* B728a (*Psyr*). Locus numbers (*Psyr*) and protein names are as annotated at <http://genome.jgi-psf.org/psesy/psesy.home.html>. Protein identities with PAO1 homologues (as annotated at <http://www.pseudomonas.com>; Stover *et al.*, 2000; Winsor *et al.*, 2009) are included.

Table S3 (a) Putative type VI secretion system (T6SS) genes in *Pseudomonas syringae* pv. *phaseolicola* 1448A (*Psph*). Locus numbers (*Psph*) and protein names are as annotated at <http://pseudomonas-syringae.org>. Percentage protein identities with PAO1 homologues (PAO), as annotated at <http://www.pseudomonas.com> (Stover *et al.*, 2000; Winsor *et al.*, 2009), are included. (b) Reverse BLAST results for the putative type VI secretion system (T6SS) genes of *Pseudomonas syringae* pv. *phaseolicola* 1448A (*Psph*). Locus numbers (*Psph*) and protein names are as annotated at <http://pseudomonas-syringae.org>. Percentage protein identities with PAO1 homologues (PAO), as annotated at <http://www.pseudomonas.com> (Stover *et al.*, 2000; Winsor *et al.*, 2009), are included.

Table S4 (a) Putative type VI secretion system (T6SS) genes in *Pseudomonas syringae* pv. *tomato* T1 (*PstomT1*) and *P. syringae* pv. *oryzae* 1-6 (*Psoryz*) and protein identities with *P. syringae* pv. *tomato* DC3000 (*PstomDC*) and/or *P. syringae* pv. *phaseolicola* 1448A (*Psph*) homologues. Gene locus numbers for *PstomDC* and *Psph* are as annotated at PPI Home (<http://pseudomonas-syringae.org>). (b) Putative type VI secretion system (T6SS) genes in *Pseudomonas syringae* pv. *tomato* T1 (*PstomT1*) and *P. syringae* pv. *oryzae* 1-6 (*Psoryz*) and protein identities with *P. syringae* pv. *phaseolicola* 1448A (*Psph*) homologues. Gene locus numbers for *Psph* are as annotated at PPI Home (<http://pseudomonas-syringae.org>).

Table S5 (a, b) Putative type VI secretion system (T6SS) genes of *Pseudomonas syringae* pv. *tabaci* ATCC 11528 (*Pstab*) identified by BLASTN analysis of the indicated contigs deposited at <http://www.ncbi.nlm.nih.gov/genomes/geblast.cgi?gi=6282> and homologies to *P. syringae* pv. *phaseolicola* 1448A (*Psph*) HSI and *P. syringae* pv. *tomato* DC3000 (*PstomDC*) HSI-II genes used as bait (locus numbers as annotated at PPI Home, <http://pseudomonas-syringae.org>).

Supporting data file Materials and methods.

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