DOI 10.1099/mic.0.048645-0

Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered

Matthieu Barret,[1]† Frank Egan,[1]† Emilie Fargier,[1] John P. Morrissey[2] and Fergal O'Gara[1]

[1]BIOMERIT Research Centre, Department of Microbiology, University College Cork, Cork, Ireland

[2]Department of Microbiology, University College Cork, Cork, Ireland

Correspondence
Fergal O'Gara
f.ogara@ucc.ie

Received 28 January 2011
Revised 29 March 2011
Accepted 1 April 2011

Bacteria encode multiple protein secretion systems that are crucial for interaction with the environment and with hosts. In recent years, attention has focused on type VI secretion systems (T6SSs), which are specialized transporters widely encoded in Proteobacteria. The myriad of processes associated with these secretion systems could be explained by subclasses of T6SS, each involved in specialized functions. To assess diversity and predict function associated with different T6SSs, comparative genomic analysis of 34 *Pseudomonas* genomes was performed. This identified 70 T6SSs, with at least one locus in every strain, except for *Pseudomonas stutzeri* A1501. By comparing 11 core genes of the T6SS, it was possible to identify five main *Pseudomonas* phylogenetic clusters, with strains typically carrying T6SSs from more than one clade. In addition, most strains encode additional *vgrG* and *hcp* genes, which encode extracellular structural components of the secretion apparatus. Using a combination of phylogenetic and meta-analysis of transcriptome datasets it was possible to associate specific subsets of VgrG and Hcp proteins with each *Pseudomonas* T6SS clade. Moreover, a closer examination of the genomic context of *vgrG* genes in multiple strains highlights a number of additional genes associated with these regions. It is proposed that these genes may play a role in secretion or alternatively could be new T6S effectors.

INTRODUCTION

Gram-negative bacteria rely on several secretion systems to influence their environment by translocating protein and DNA into host cells and the extracellular milieu. These secretion systems can range from simple transporters to multi-component complexes. Type III (T3SS), type IV (T4SS) and, more recently, type VI (T6SS) secretion systems have received considerable attention because they are specialized in mediating the delivery of effectors directly to the host cytoplasm via a needle-like apparatus (Alvarez-Martinez & Christie, 2009; Cornelis, 2006; Filloux *et al.*, 2008).

The T6S machinery is the product of approximately 15 conserved genes which are generally found together inside a genomic locus (Cascales, 2008). The mechanism of T6S has yet to be fully elucidated but a putative model of T6SS assembly has been proposed (Bönemann *et al.*, 2010). Briefly, stacked, tubular haemolysin-coregulated protein (Hcp) hexamers form a 4 nm wide conduit for the passage of effectors from the cytoplasm to the environment or into another host cell (Mougous *et al.*, 2006). Hcp proteins interact with valine-glycine repeat (Vgr) proteins, which could pierce the outer bacterial membranes with the help of the TssE protein, which may disrupt the peptidoglycan layer (Bönemann *et al.*, 2010). Outgrowth of the Hcp tubules is energized by the ATPase ClpV, which produces a conformational change by disassembling the IglA–IglB complex which surrounds the Hcp rings (Bönemann *et al.*, 2009; Bröms *et al.*, 2009). The proteins IcmF and DotU act as associated inner-membrane-spanning structural proteins that anchor the secretion system in the cell membrane (Ma *et al.*, 2009; Zheng & Leung, 2007) whereas the lipoprotein TssJ extends into the periplasm from the outer membrane and interacts with IcmF (Aschtgen *et al.*, 2008). Although part of the T6S apparatus, Hcp and VgrG protein are also secreted by bacteria with a functional T6SS. Some VgrG and Hcp proteins, called evolved VgrG (Pukatzki *et al.*, 2009) or Hcp (Blondel *et al.*, 2009), have a C-terminal domain extension and therefore could also act as effectors (Ma & Mekalanos, 2010; Pukatzki *et al.*, 2007; Suarez *et al.*, 2010). Interestingly, numerous *hcp* and *vgrG*

†These authors contributed equally to this work.

Abbreviations: COG, cluster of orthologous groups of proteins; HGT, horizontal gene transfer; T3SS, T4SS and T6SS, type III, type IV and type VI type secretion system.

Supplementary material is available with the online version of this paper.

048645 © 2011 SGM *Printed in Great Britain*

paralogues are scattered around the bacterial chromosome. This raises the questions of how Hcps and VgrGs evolve, from where are they acquired, and whether all of them are secreted.

T6SSs are widespread in Proteobacteria, particularly among gamma-Proteobacteria (Shrivastava & Mande, 2008), and are more frequent than T3SSs and T4SSs in marine isolates (Persson *et al*., 2009). Like T3SSs and T4SSs, several findings suggest that T6SSs have been acquired through horizontal gene transfer (HGT). Indeed, the T6SS gene loci are frequently found inside genomic islands gained by HGT. Moreover, some T6S apparatus proteins such as Hcp and VgrG exhibit structural homology to phage tail-associated proteins, which suggests a common ancestral origin (Leiman *et al*., 2009; Mougous *et al*., 2006; Pell *et al*., 2009). Hence, it is believed that these T6SS genomic islands have been spread among bacteria by bacteriophages. Interestingly, some T6SS proteins are still able to interact with early phage protein, as demonstrated for Fha2 (PA1665) of *Pseudomonas aeruginosa* PAO1 (Roucourt *et al*., 2009). A recent phylogenetic analysis performed on T6SS core components across a range of bacterial species has shown that the T6SS loci can be divided into five clusters (Boyer *et al*., 2009). Although these phylogenetic clusters have probably evolved to adapt to various environments, it is difficult to find any correlation between clusters and ecological niches (Schwarz *et al*., 2010b).

The presence of multiple T6SS clusters in individual bacterial strains suggests that these secretion systems perform different roles for the bacterial cell (Bingle *et al*., 2008). Several phenotypes such as increased or attenuated virulence against human cells (Parsons & Heffron, 2005; Pukatzki *et al*., 2007; Robinson *et al*., 2009; Suarez *et al*., 2010), animals (Burtnick *et al*., 2010; Potvin *et al*., 2003), plants (Lesic *et al*., 2009; Liu *et al*., 2008; Wu *et al*., 2008), fish (Wang *et al*., 2009) and bacteria (Hood *et al*., 2010; MacIntyre *et al*., 2010; Schwarz *et al*., 2010a, b) have been associated with the T6SS. More general physiological roles, such as biofilm formation (Aschtgen *et al*., 2008; Southey-Pillig *et al*., 2005) and quorum-sensing regulation (Weber *et al*., 2009) have also been linked to T6SSs. Different T6SSs in a given strain may secrete different sets of effectors, or, as with the T3SS (Cornelis, 2006), the myriad of processes associated with T6SSs could also be explained by secretion of specific subsets of effectors by one T6SS. Thus far, these hypotheses cannot be tested as only a few T6S effectors have been identified (Hood *et al*., 2010; Zheng & Leung, 2007).

In this paper, we focus on T6SS distribution in the genus *Pseudomonas*. Pseudomonads have a remarkable ecological and metabolic diversity, and are of interest as agents of plant disease (*P. syringae*), plant growth promotion (*P. fluorescens*) or bioremediation (*P. putida*). Moreover, *P. aeruginosa* has emerged as one model organism for T6SS studies (Filloux *et al*., 2008). *P. aeruginosa* possess three different loci named HSI-I to III which perform different functions (Mougous *et al*., 2006; Lesic *et al*., 2009). Whereas HSI-I may be involved in interbacterial interactions through secretion of Tse2 (Hood *et al*., 2010), HSI-II and III could be linked to virulence towards animals and plants (Lesic *et al*., 2009). Reflecting the importance of this genus, many *Pseudomonas* genome-sequencing projects are currently being undertaken or have been recently completed. As previous *in silico* studies included relatively few *Pseudomonas* genomes, we set out to identify T6SS loci in all 34 *Pseudomonas* species sequenced to date and to establish their evolutionary relationship. We also assessed possible association of VgrG and Hcp paralogues with each *Pseudomonas* T6SS by combining phylogeny and meta-analysis of transcriptome datasets. Finally we found a subset of genes co-regulated with every T6SS of *P. aeruginosa*, which may include new T6S effectors.

METHODS

***Pseudomonas* genomic data acquisition.** Information about the current status of *Pseudomonas* genome sequencing projects was obtained from the Genomes Online Database (GOLD) (http://www.genomesonline.org/, updated on April 7, 2010). Among these sequencing projects, 33 having sequence data publicly available were selected for genome identification of T6SS loci. A preliminary draft of the strain *P. fluorescens* F113 (R. Rivilla, D. Dowling & F. O'Gara, unpublished) was also included in this analysis. The genomes analysed covered eight different species. Genome accession numbers and information about the *Pseudomonas* genome sequencing projects utilized in this work are detailed in Table 1.

***In silico* identification of T6SS loci.** Nucleotide and amino acid sequences of ORFs representing T6SS components were obtained from public sequence databases. The ORFs of *P. aeruginosa* PAO1 (Stover *et al*., 2000), and 'outgroup species' representing each branch in T6SS phylogenetic trees previously described (Blondel *et al*., 2009; Boyer *et al*., 2009) were used as baits in sequential BLASTN, BLASTX and BLASTP searches to identify homologues in the 34 *Pseudomonas* genomic sequences (e-value $<10^{-5}$). A T6SS locus was defined as a gene cluster encoding at least five core components. A systematic analysis of gene content and gene architecture of the identified T6SS gene clusters was performed in 20 finished or permanent draft genomes. The remaining genomes correspond to unfinished projects containing more than one contig; therefore an exhaustive ORF-by-ORF analysis of T6SS genetic architecture was not possible. In such cases, we only determined the presence/absence of T6SS core components in unassembled contigs.

Genomic islands analysis. The T6SS loci sequences of seven representative strains (see Supplementary Table S1, available with the online version of this paper) were examined for sequence composition bias such as aberrant G+C percentage or dinucleotide frequency. The dinucleotide frequency analysis calculates the genomic dissimilarity values δ^* (the average dinucleotide relative abundance difference) between T6SS loci sequences and the associated genome sequence using a web-based application, $\delta\rho$-web (van Passel *et al*., 2005). Mol% G+C and dinucleotide frequencies of each T6SS locus were assessed in 5 kb windows and compared to the overall chromosomal signature. As the *Pseudomonas* genome is a flexible genome with numerous genes acquired by HGT (Gross & Loper, 2009), loci were arbitrarily defined as being of heterologous origin when the percentage of the genomic fragments with lower genomic

Table 1. Distribution of T6SS loci in *Pseudomonas*

The strain *P. fluorescens* F113 highlighted in bold has been sequenced recently and is currently under annotation.

Strain	Status	NCBI reference sequence	Locus 1	Locus 2	Locus 3	Locus 4A	Locus 4B	Locus 5	Total	VgrG	Hcp
P. aeruginosa PAO1	Complete	NC_002516	1	0	1	1	0	0	3	10	5
P. aeruginosa UCBPP-PA14	Complete	NC_008463	1	0	1	1	0	0	3	11	6
P. aeruginosa PA7	Complete	NC_009656	1	0	1	1	0	0	3	7	3
P. aeruginosa LESB58	Complete	NC_011770	1	0	1	1	0	0	3	9	5
P. aeruginosa 2192	Draft	NZ_AAKW00000000	1	0	1	1	0	0	3	7	5
P. aeruginosa C3719	Draft	NZ_AAKV000000000	1	0	1	1	0	0	3	9	5
P. aeruginosa PACS2	Draft	NZ_AAQW00000000	1	0	1	1	0	0	3	9	4
P. aeruginosa PAb1	Draft	NZ_ABKZ000000000	1	0	1	1	0	0	3	6*	6*
P. entomophila L48	Complete	NC_008027	1	0	0	0	0	0	1	10	8
P. fluorescens Pf-5	Complete	NC_004129	0	0	1	0	0	0	1	3	1
P. fluorescens PfO-1	Complete	NC_007492	1	0	0	1	0	0	2	10	4
P. fluorescens SBW25	Complete	NC_012660	1	0	1	0	0	0	2	3	3
***P. fluorescens* F113**	**Draft**		**1**	**0**	**1**	**1**	0	**0**	**3**	**8**	**3**
P. mendocina ymp	Complete	NC_009439	1	0	1	0	0	0	2	4	2
P. putida KT2440	Complete	NC_002947	2	0	0	0	1	0	3	4	5
P. putida GB-1	Complete	NC_010322	1	0	0	0	1	0	2	4	5
P. putida W619	Complete	NC_010501	0	1	0	0	1	0	2	3	3
P. putida F1	Complete	NC_009512	1	0	0	0	1	0	2	3	2
P. stutzeri A1501	Complete	NC_009434	0	0	0	0	0	0	0	0	0
P. savastanoi pv. *savastanoi* NCPPB 3335	Draft	NZ_ADMI01000000	1	0	0	0	1	0	2*	3*	6*
P. syringae pv. *aesculi* 2250	Draft	NZ_ACXT00000000	0	0	0	0	1	0	1*	4*	3*
P. syringae pv. *aesculi* NCPPB3681	Draft	NZ_ACXS00000000	1	0	0	0	1	0	2*	1*	5*
P. syringae pv. *oryzae* 1_6	Draft	NZ_ABZR00000000	1	0	0	0	1	0	2*	4*	7*
P. syringae pv. *phaseolicola* 1448A	Complete	NC_005773	0	0	0	0	1	0	1	3	7
P. syringae pv. *syringae* B728a	Complete	NC_007005	0	0	0	0	1	0	1	5	4
P. syringae pv. *syringae* FF5	Draft	NZ_ACXZ00000000	1	0	0	0	1	0	2*	2*	8*
P. syringae pv. *syringae* 642	Draft	NZ_ACXZ00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tabaci* 11528	Draft	NZ_ACHU00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* DC3000	Complete	NC_004578	2	0	0	0	0	0	2	6	3
P. syringae pv. *tomato* T1	Draft	NZ_ABSM00000000	1	0	0	0	1	0	2	4	7
P. syringae pv. *tomato* K40	Draft	NZ_ADFY00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* Max13	Draft	NZ_ADFZ00000000	1	0	0	0	1	0	2*	4*	6*
P. syringae pv. *tomato* NCPPB 1108	Draft	NZ_ADGA00000000	1	0	0	0	1	0	2*	4*	5*
Pseudomonas sp. UK_4	Draft	NZ_ACOQ00000000	1	0	0	0	0	0	1*	0*	0*

*Only presence or absence of T6SS loci could be performed in unfinished genomes. Therefore, we cannot rule out the possibility of the presence of additional T6SS clusters in these draft genomes.

dissimilarity was above 80 % and when mol% G+C was above or below 90 % of the genomic fragments. Presence of insertion elements, flanking direct repeats, and proximity of tRNA was also assessed in the vicinity of T6SS loci.

Phylogenetic analyses. In the case of T6SS loci, the prevalence of each COG (cluster of orthlogous groups of proteins) defined by Boyer *et al.* (2009) was analysed for every locus. COGs with frequencies higher than 90 % were considered as *Pseudomonas* T6SS core components. Phylogenetic analyses were performed on the amino acids sequences from each of the 11 selected 'core' COGs (COG0542, COG3455, COG3515, COG3516, COG3517, COG3518, COG3519, COG3520, COG3521, COG3522 and COG3523). Maximum-likelihood trees with 1000 bootstrap replicates were built with PhyML (Guindon & Gascuel, 2003) using the WAG amino acid substitution model of evolution (Whelan & Goldman, 2001) and four categories of substitution rates. To test the homogeneity between trees, the split distances were determined by TOPD/FMTS (Puigbò *et al.*, 2007) then an average-linkage method of clustering was applied on these distances. In order to compute the supertree, for each of the individual trees a matrix representation with parsimony (MRP) was built using Mesquite V.2.72 then concatenated into a supermatrix. The supertree was built with the Pars program from PHYLIP (Felsenstein, 2005) with 1000 bootstrap replicates.

Besides this phylogenetic supertree construction, individual phylogenetic trees from 180 VgrG and 163 Hcp amino acid sequences were also generated by maximum-likelihood by the method outlined above.

Comparative analysis of transcriptomes. All *P. aeruginosa* transcriptome datasets publicly available were retrieved from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) when available, or directly from publications. T6SS inducing or repressing conditions were defined for each locus when at least 50 % of the genes within that locus were differentially regulated. Based on these criteria, 19 transcriptome datasets were considered (Supplementary Table S2). Genes were clustered according to their expression profiles using the MultiExperiment Viewer Software v4.5.1 (http://www.tm4.org/mev/). Genes were grouped with either Pearson correlation or Euclidean distance as distance for *K*-means clustering. The optimal number of expression clusters was chosen after a figure-of-merit algorithm. For robustness, only genes present in expression clusters obtained with Euclidean distance and Pearson correlation were defined as T6SS co-regulated genes.

RESULTS

Identification of T6SS gene clusters in *Pseudomonas*

A list of genes encoded in T6SS loci of *P. aeruginosa* PAO1 was manually established. PAO1 gene sequences and the corresponding protein sequences were used as baits in sequential TBLASTN, BLASTX and BLASTP searches to identify homologues in 34 *Pseudomonas* chromosome and plasmid sequences available (Table 1). The genomes analysed covered eight different species of pseudomonads. To maximize the power of the screen, representatives of each of the core components of T6SS (Boyer *et al.*, 2009) from different micro-organisms used in a previous study (Blondel *et al.*, 2009) were also used as baits. A T6SS locus was defined when at least five genes predicted to encode proteins showed significant similarities (e-value $<10^{-5}$) to T6SS bait genes (Boyer *et al.*, 2009).

The analysis revealed that every *Pseudomonas* sequenced to date, except *P. stutzeri* A1501, possesses at least one putative T6SS locus (Table 1). The analysis also highlights that 27 *Pseudomonas* strains possess multiple T6SS gene clusters. Indeed, all *P. aeruginosa* species sequenced to date encode three complete T6SS loci, whereas other species encode one to three loci per genome.

Phylogenetic relationship between *Pseudomonas* T6SSs

In order to study the evolutionary relationship among all T6SSs of *Pseudomonas*, we performed phylogenetic analyses on 11 COGs which occur within more than 90 % of *Pseudomonas* T6SS loci (Supplementary Fig. S1). These 11 COGs, as well as Hcp and VgrG, were already identified as core components (Boyer *et al.*, 2009). However, we decided to perform independent phylogenetic analyses of Hcp and VgrG proteins, as they are often encoded outside T6SS loci and may be subject to higher evolutionary pressures due to potential interaction with host-cell membranes. A tree was built on protein sequences of each of the 11 core components. According to the split distances determined by TOPD/FMTS, the 11 trees were congruent. A supertree was then built based on the amino acid sequences from each of the 11 selected 'core' COGs.

In the resulting phylogeny, the *Pseudomonas* T6SSs are grouped in five major phylogenetic clusters (1, 2, 3, 4A and 4B), whereas a sixth cluster (5) comprises selected out-group species (Fig. 1). Previous independent analyses have already reported the presence of three *Pseudomonas* T6SS subtrees, termed cluster 1 (or HSI-II), cluster 3 (or HSI-I) and cluster 4 (or HSI-III) (Bingle *et al.*, 2008; Boyer *et al.*, 2009). Whereas a similar T6S phylogeny is observed for loci in clusters 1 and 3, the previously described cluster 4 is not supported in our analysis and is split into two distinct subtrees. To be consistent with previous nomenclature used by Boyer *et al.* (2009), these two distinct subtrees are termed 4A and 4B. Cluster 4A is encoded in *P. aeruginosa* and some *P. fluorescens* strains, whereas cluster 4B is found in *P. putida* and *P. syringae.* The present work also identified a novel *Pseudomonas* T6SS subtree related to cluster 2. This subtree is only found in *P. putida* W619 and in one *Pseudomonas*-related strain, called uncultured proteobacterium QS1 (Williamson *et al.*, 2005). Although T6SSs of *Pseudomonas* are divided into five main clades, cluster 1 could also be subdivided into two subclusters. Subcluster 1.2 is specific to *P. putida* strains, whereas cluster 1.1 is also present in other *Pseudomonas* species. In summary, we can divide pseudomonad T6SS loci into six groups: 1.1 (or HSI-II), 1.2, 2, 3 (or HSI-I), 4A (or HSI-III) and 4B.

Interestingly, only two strains, namely *P. syringae* pv. *tomato* DC3000 and *P. putida* KT2440, possess two copies

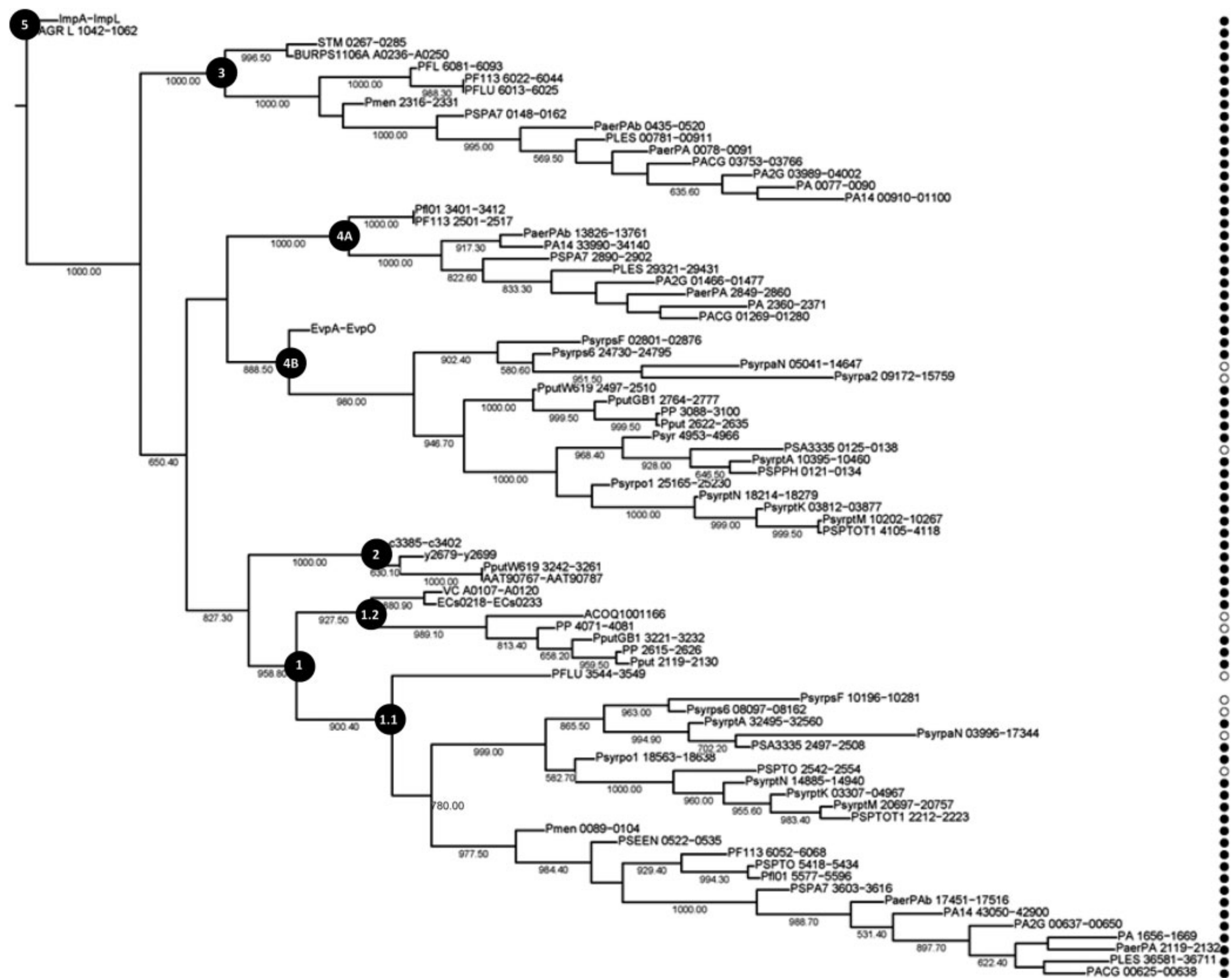

Fig. 1. Phylogenetic distribution of T6SS clusters in *Pseudomonas* species. Maximum-likelihood trees with 1000 bootstrap replicates were built with PhyML for each 'core' protein. In order to compute the supertree, for each of the individual trees a matrix representation with parsimony (MRP) was built using Mesquite V.2.72 and then concatenated into a supermatrix. The supertree was built with the Pars program from PHYLIP with 1000 bootstrap replicates. T6SS cluster nomenclature (Boyer *et al.*, 2009) is used to show the major phylogenetic clusters (1, 2, 3 and 5). However, cluster 4 was not supported as a single clade in this work and was therefore divided into 4A and 4B. The phylogenetic cluster 1 could be subdivided into two subclusters represented in the nodes labelled 1.1 and 1.2. Black circles to the right of the figure indicate a complete T6SS locus, whereas white circles represent T6SS loci in which at least one core component is missing or mutated.

of the same T6SS cluster. However, a closer examination of these loci reveals that two T6SS core genes in *P. putida* KT2440 (PP4083 and PP4075) have frameshift mutations while one core gene from *P. syringae* pv. *tomato* DC3000 (PSPTO_2542) is disrupted by a transposase, which suggest that these loci are non-functional. Some other *Pseudomonas* T6SS loci are also incomplete, with multiple gene deletions. The most striking example comes from *P. fluorescens* SBW25 cluster 1.1, where five T6SSs core components are missing. This suggests that not every T6SS gene cluster should be expected to encode a functioning secretion system.

Variations among *Pseudomonas* T6SS clusters

As stated by Bingle *et al.* (2008), it is clear that multiple T6SS clusters have not arisen by duplications within a given lineage but rather by independent HGT events. Hence, an analysis of genomic dissimilarities between *Pseudomonas* T6SS loci and the associated genomes was performed on seven representative strains (Supplementary Table S1) using the program $\delta\rho$-web (van Passel *et al.*, 2005). The dinucleotide frequencies and mol% G+C of each *Pseudomonas* T6SS locus were overall not very different when compared to the genome in which they

presently reside. Therefore, these T6SS loci are probably ancient acquired sequences that have been subject to amelioration.

Based on *Pseudomonas* T6SS cluster organization (Fig. 2), it is believed that T6SSs assemble from at least 11 structural proteins and two structural/effector proteins (namely Hcp and VgrG) called 'core components' (Boyer *et al.*, 2009). However, most T6SS loci encode additional proteins whose roles are unknown. These 'non-core' proteins could form an accessory complex, which could possibly be required for efficient secretion or in a process specific to the T6SS cluster (Aschtgen *et al.*, 2010). Multiple accessory elements have been found in every *Pseudomonas* T6SS cluster (Supplementary Table S3). The majority of these non-core proteins are shared in phylogenetically related T6SS loci present in other bacterial genera. For example, the accessory protein SciZ of the *E. coli* T6SS *sci-1* locus, which is only required for efficient secretion of this locus (Aschtgen *et al.*, 2010), is also present in the phylogenetically related loci in *P. putida* W619 and in QS1, suggesting a specific role of this protein for cluster 2. However, whereas non-core proteins involved in the assembly of the putative secretion apparatus are well conserved, some specific regulatory components of T6SSs (Bernard *et al.*, 2010) are only encoded in pseudomonads. For instance, the post-translational regulatory element TagR, which promotes efficient phosphorylation of the kinase PpkA leading to activation of HSI-I (Hsu *et al.*, 2009; Mougous *et al.*, 2007), is only present in cluster 3 of *P. aeruginosa* and *P. fluorescens*. Taken together, these data suggest that each T6SS cluster encodes specialized secretion machinery, well conserved through evolution, but that individual T6SSs can be integrated into specific bacterial regulatory circuits.

Phylogenetic relationship between Hcps/VgrGs and T6SS clusters

The VgrG and Hcp proteins are believed to interact and form part of the extracellular secretion apparatus. A very interesting feature of the genetics of T6SSs is the finding that, in addition to VgrG and Hcp genes inside T6SS loci, all the *Pseudomonas* strains examined possess numerous VgrG and Hcp paralogues that are encoded elsewhere in the genome (Table 1). Whereas VgrG and Hcp proteins located inside a T6SS locus are likely to be associated with their respective secretion systems, the role of 'orphan' Hcp and VgrG proteins is less clear. It is not known whether

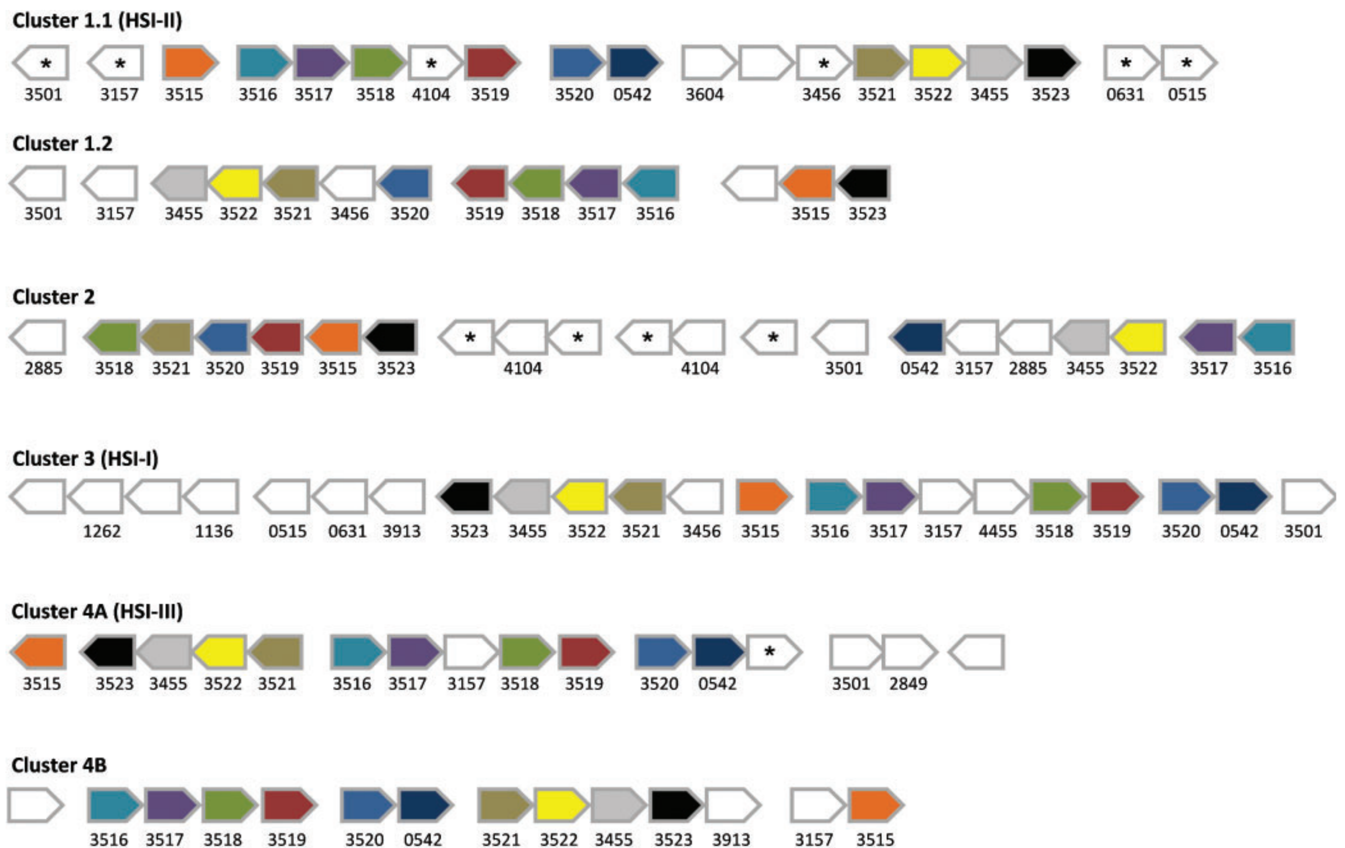

Fig. 2. Genomic organization of the *Pseudomonas* T6SS clusters. Genes are represented as blocked arrows showing the direction of their transcription. Numbers represent the COG number. A unique colour is assigned to T6SS core component genes used in the phylogeny. Asterisks indicate that genes are not always conserved among all strains.

they are recruited by specific T6SSs or whether they have other functions, nor how they have arisen or evolved.

In order to ascertain whether 'orphan' VgrG and Hcp proteins can be linked to specific secretion systems, the evolutionary relationships among VgrG/Hcp paralogues were investigated. Phylogenetic analyses of 180 VgrGs and 163 Hcps from all *Pseudomonas* genomes as well as 'outgroup species' representing each branch in T6SS phylogenetic trees were performed. Hcps of *Pseudomonas* are divided into six clusters (1.1, 1.2, 2, 3, 4A, 4B) that correspond to each of the phylogenetic T6SS clusters (Fig. 3),

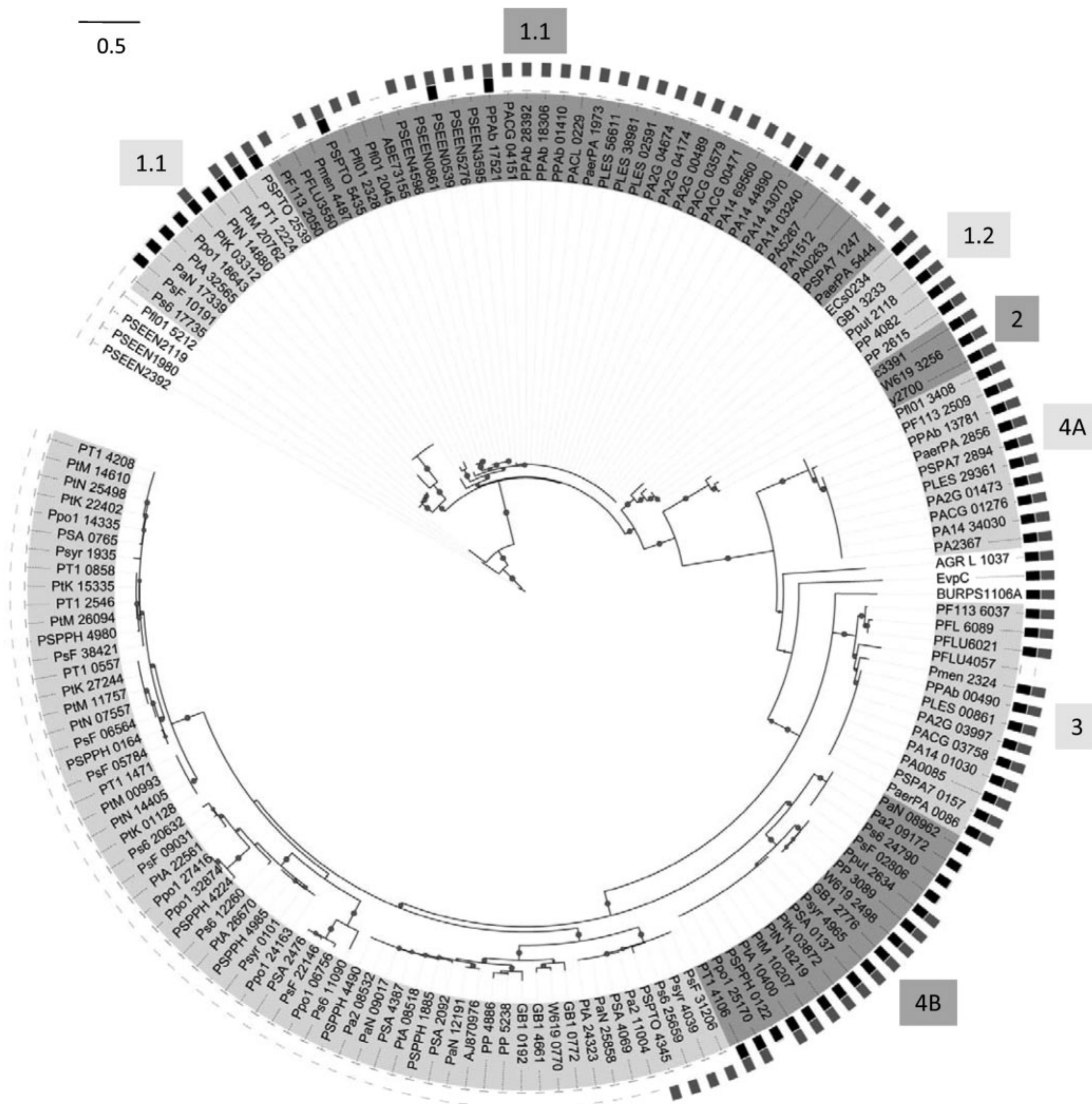

Fig. 3. Phylogenetic distribution of Hcp proteins within *Pseudomonas* species. A distance tree (maximum-likelihood) was calculated from 163 Hcp protein sequences of *Pseudomonas* spp. Black circles indicate branches with bootstrap support values >75 % (1000 replicates). Black rectangles indicate that Hcp proteins are encoded inside a T6SS locus. Dark grey rectangles indicate that *hcp* is linked to *vgrG*. The scale bar indicates genetic distance.

which suggests that every gene in each Hcp clade may be specifically recruited by the corresponding T6SS cluster. Such correlation between VgrGs and T6SS phylogenetic clusters is also clear for clusters 2, 3, 4A and 4B (Fig. 4). However, the picture is more complex for cluster 1. A subset of orphan VgrGs which group with cluster 1 are present in strains where this cluster is absent (e.g. PFL2812 from *P. fluorescens* Pf-5). The link between PFL2812 and phylogenetic cluster 1.1 could be explained by loss of cluster 1.1 from *P. fluorescens* Pf-5 or by an acquisition of this *vgrG* through HGT.

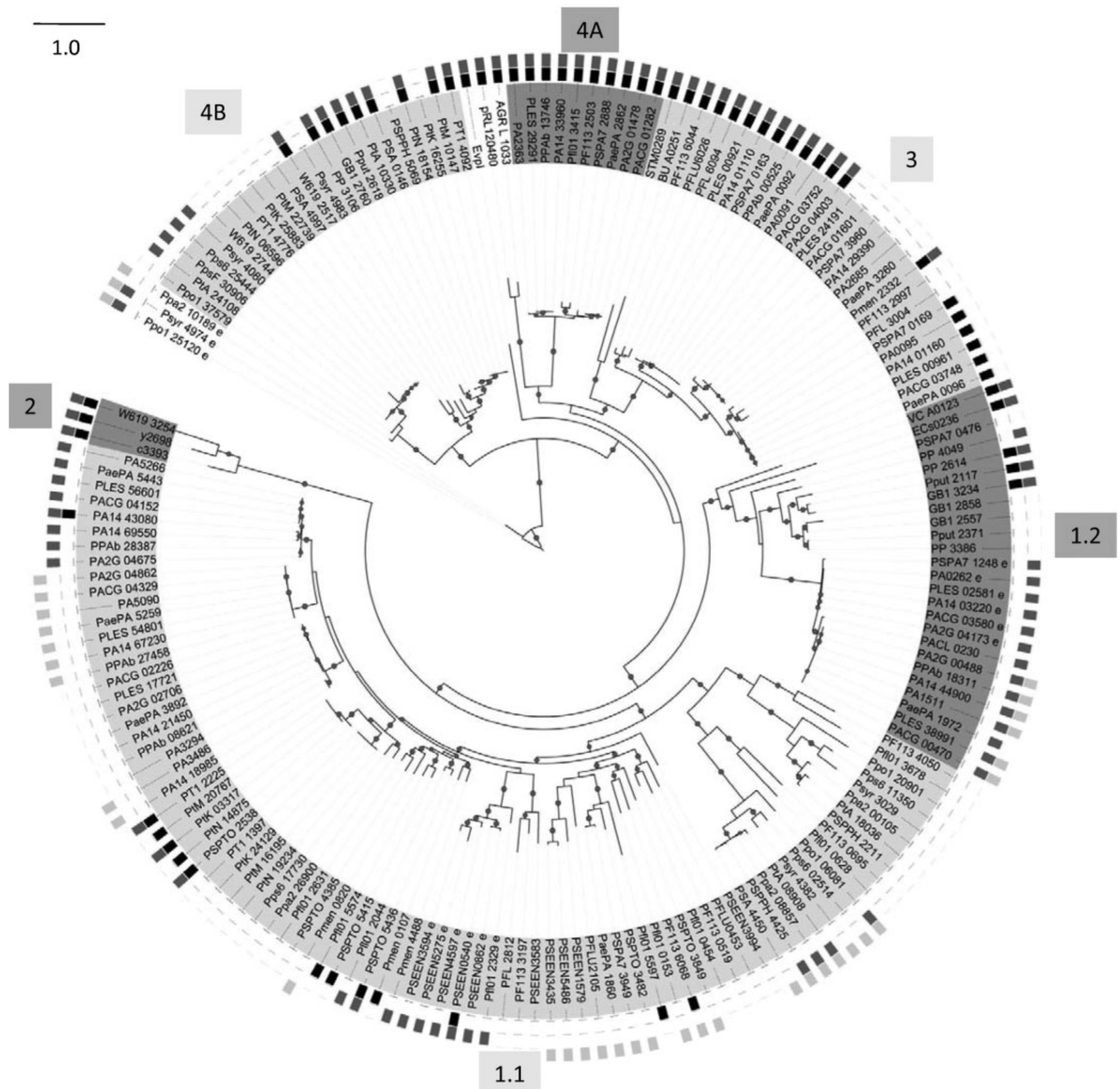

Fig. 4. Phylogenetic distribution of VgrG proteins within *Pseudomonas* species. A distance tree (maximum-likelihood) was calculated from 180 VgrG protein sequences of *Pseudomonas* spp. Black circles indicate branches with bootstrap support values >75 % (1000 replicates). Black rectangles indicate VgrG proteins encoded inside a T6SS locus. Dark grey rectangles indicate that *vgrG* is linked to *hcp*, whereas light grey rectangles indicate that a gene encoding a lipase is found in the vicinity of *vgrG*. Evolved VgrG proteins are labelled with the letter e. The scale bar indicates genetic distance.

Orphan VgrGs and Hcps are frequently encoded in the same genomic location, consistent with the view that they interact as part of the T6SS. In some cases, however, genes encoding these proteins were found singly, not in proximity to any other T6S-related gene. A closer examination of the phylogenetic trees shows that Hcps without a partner VgrG often grouped together (Fig. 3), whereas no such pattern is seen for VgrGs (Fig. 4). Therefore, it is possible that these particular Hcp proteins are no longer part of a T6SS and are linked to other biological functions, like the HilE protein of *Salmonella enterica* (Blondel *et al.*, 2009).

The VgrG phylogeny also highlights three distinct groups of VgrG proteins containing C-terminal extensions (Fig. 4). Because these extensions could possibly function as effector domains, these VgrGs have been named evolved VgrGs (Pukatzki *et al.*, 2009). The first evolved VgrG group contains PA0262 orthologues, which are only found in *P. aeruginosa* genomes. The second group is restricted to one VgrG of *P. fluorescens* Pf0-1 (Pfl01_2329) and five VgrGs of *P. entomophila* L48. Finally, the third group is shared by three proteins (Psyr_4974, Psyrpol_ 25120 and Psyrpa2_ 10189) each belonging to one *P. syringae* strain. A motif search was systematically performed with INTERPRO scan to predict a specific function for VrgG C-terminal extensions but none of the domains shows homology with domains associated with biological functions.

Genomic organization of orphan Hcp and VgrG

Some *vgrG* and *hcp* genes of *P. aeruginosa* PAO1 have been reported to be located inside genomic islands (Ernst *et al.*, 2003; Wilderman *et al.*, 2001). To see if this trend was applicable to all orphan *vgrG* and *hcp* genes of *P. aeruginosa* PAO1, mol% G+C and dinucleotide frequencies of each genomic region associated with *vgrG* or *hcp* were assessed in 5 kb windows and compared to the overall chromosomal signature (Table 2). Five out of the ten *vgrG* regions analysed showed atypical sequence composition when compared to the genome in which they presently reside. Therefore, acquisition of these five '*vgrG* islands' by HGT seems to be relatively recent.

We reasoned that these '*vgrG* islands' could have been acquired along with other genes (hereinafter referred to as cargo genes). The VgrG phylogenetic analysis and the conserved genome organization of *P. aeruginosa* were then used to determine whether orphan *vgrG*s were always linked to cargo genes. Remarkably, each orphan *vgrG* was associated with a set of specific cargo genes in all strains of *P. aeruginosa* carrying that *vrgG* gene (Fig. 5). The function(s) of each cargo gene was predicted using BLASTP and INTERPRO scan. Although most cargo genes encode hypothetical proteins or putative lipoproteins, genes encoding predicted lipases/esterases are also frequently linked to *vgrG* in these islands. These proteins are identified by the presence of a phospholipase D domain (e.g. PldA or PA5089) or a PGAP1 domain (e.g. PA1510). To determine whether this association between VgrG genes and lipase-encoding genes extended beyond *P. aeruginosa*, the presence of gene encoding lipases in the vicinity of *vgrG* was also assessed in the other *Pseudomonas* species. A total of 34 genes encoding lipases or esterases were encoded in the vicinity of 118 orphan VgrGs, indicating that the association is found in multiple *Pseudomonas* species (Fig. 4). The lipases/esterases can be separated into six distinct clades. Three clades are *P. aeruginosa* specific and contain orthologues of PA1510, PA3487 (or PldA) and PA5089. A fourth clade is linked to three *P. syringae* strains and contains orthologues of Psyr_4970, a predicted phospholipase D. Finally, clades five and six represent proteins of the lipase/esterase superfamily and predicted triglyceride lipases, respectively. Whereas clade five (e.g. PSPPH_4425) is only present in some *P. syringae* strains, clade six (e.g. PSPA7_ 3949) is found in some *P. aeruginosa*, *P. entomophila*, *P. syringae*, *P. mendocina* and *P. fluorescens* strains. The phylogenetic separation of lipases/esterases may indicate that functionally distinct proteins are associated with specific VgrG proteins in particular species of *Pseudomonas*.

Co-regulation of orphan VgrG/Hcp and T6SS clusters

Previous studies have demonstrated that there is often co-regulation of the genes encoding the structural components of secretion systems and the genes encoding substrates of those secretion systems (Alvarez-Martinez & Christie, 2009; Cornelis, 2006). To ascertain whether such links could be established for T6SS and VgrG/Hcp proteins, we carried out a meta-analysis of selected *P. aeruginosa* transcriptome datasets for each HSI locus. Transcriptome datasets were analysed when at least 50 % of the structural T6SS genes were differentially expressed (Supplementary Table S2). Expression profiles of all *P. aeruginosa* genes were generated using either Pearson correlation or Euclidean distance as distance for *K*-means clustering. A list of gene co-regulated with the T6SS structural genes was obtained for each locus (Supplementary Table S4).

A total of 36, 90 and 96 genes were co-regulated with the phylogenetic clusters 1, 3 and 4a, respectively (Supplementary Table S4). We specifically examined whether orphan *vgrG* and *hcp* genes showed patterns of co-regulation. Two genes encoding VgrG1 and VgrG4 (PA2685) along with two genes encoding Tse proteins (PA1844 and PA3484) were co-regulated with phylogenetic cluster 3 (or HSI-I); PA2373 was co-regulated with phylogenetic cluster 4A (or HSI-III) and PA3294 and PA3486 with phylogenetic cluster 1 (HSI-II). The other five orphan *vgrG* genes, PA0095, PA0262, PA1511, PA5090 and PA5266, were unassigned to any particular cluster. Among the five genes encoding Hcp proteins, PA0085 was linked with phylogenetic cluster 3, PA0263 with cluster 1.1 (or HSI-II) and PA2367 with cluster 4A, whereas PA1512 and PA5267 were unassigned. Taken together these results show that some orphan VgrGs or Hcps are co-regulated with a T6SS phylogenetic cluster (Table 3). The lack of co-regulation between T6SS loci and some orphan VgrGs/Hcps

Table 2. Genomic dissimilarities between *P. aeruginosa* PAO1 VgrG regions and the associated genome

Dinucleotide frequency and mol% G+C of each VgrG genomic region were calculated using the software $\delta\rho$-web (van Passel *et al.*, 2005). Rows in bold indicate that the regions differ significantly in mol% G+C and dinucleotide frequency compared to the PAO1 genome signature. References indicate that these regions have been already reported to be located inside genomic islands.

VgrG island	Chromosome position	Locus	Mol% G+C	Genomic fragments with lower mol% G+C (%)	1000×δ* (genomic dissimilarity)	Genomic fragments with lower δ* (%)	Mobile elements	Reference
PA0091 (*vgrG1*), cluster 3 (HSI-I)	112506–117524	PA0091–0095	66	33	59	81	None	
PA0095	117524–122541	PA0095–0101	67	50	17	1	None	Ernst *et al.* (2003)
	122541–127541	PA0101–0105	67	50	36	32	None	
PA0262	286154–291154	PA0254–0260	63	7	37	33	Putative integrase	Ernst *et al.* (2003)
	291154–296154	PA0260–0262	65	24	34	26	None	
	296154–301154	**PA0262–0265**	**62**	**5**	**95**	**98**	**tRNA-Arg**	
PA1511	1638379–1643379	PA1508–1511	60	3	58	79	None	
	1643379–1648379	PA1511–1515	63	10	49	62	None	
PA2373 (*vgrG3*) cluster 4A (HSI-III)	2621552–2626757	PA2371–2374	64	12	57	79	None	Ernst *et al.* (2003)
	2626757–2631962	PA2374–2378	68	75	46	58	None	
PA2685 (*vgrG4*)	**3030752–3035752**	**PA2683–2685**	**70**	**92**	**60**	**82**	**RHS element**	
	3035752–3040752	PA2685–2688	67	45	36	31	None	
PA3294	**3681584–3686584**	**PA3290–3294**	**58**	**2**	**104**	**99**	None	
	3686584–3691584	PA3294–3297	67	50	42	47	None	
PA3486	3895233–3900233	PA3482–3486	67	44	39	38	None	Wilderman *et al.* (2001)
	3900233–3905233	PA3486–3487	66	44	42	47	None	
PA5090	**5724473–5729473**	**PA5085–5090**	**58**	**2**	**87**	**97**	**Tetratricopeptide-like helical**	Ernst *et al.* (2003)
	5729473–5734473	PA5090–5093	69	79	23	6	none	
PA5266	**5925602–5930602**	**PA5263–5266**	**61**	**4**	**65**	**87**	None	Ernst *et al.* (2003)
	5930602–5935602	PA5266–5270	66	27	36	31	None	

could possibly be explained by (i) the conditions used for transcriptome experiments; (ii) the high degree of identity between some *hcp* and *vgrG* paralogues, which could mask the expression pattern of specific genes during array hybridization; or (iii) the fact that these orphan genes are not part of a T6SS.

Some genes linked with *vgrG* islands (e.g. PA0097, PA2684 and PA3487 or *pldA*) were also co-regulated with some T6SS phylogenetic clusters, which could possibly suggest that these genes are also related to T6SS. Interestingly, *pldA*, associated with the orphan *vgrG* PA3486, is co-regulated with cluster 1.1. This might suggest that the phospholipase D PldA could possibly be secreted by this T6SS.

DISCUSSION

This study demonstrates the prevalence of type VI secretion in *Pseudomonas* spp. We have shown that every *Pseudomonas* genome sequenced (except *P. stutzeri* A1501) possesses at least one putative T6SS. Among these strains, five different T6SS clusters, having a distinct evolutionary origin, have been found. Whereas phylogenetic clusters 1, 3 and 4 have already been reported for the genus *Pseudomonas* (Boyer *et al.*, 2009), the present work identified a novel *Pseudomonas* T6SS locus related to cluster 2 and a clear separation of cluster 4 into clusters 4A and 4B.

The different T6S clusters have been acquired a long time ago, and have probably evolved into specialized secretion machinery with different gene organization and gene regulation patterns. This observation could possibly indicate that these clusters play different functional roles, providing a competitive advantage in certain niches. Although a systematic prediction of the function of each T6SS cluster will require more studies on individual strains, some correlation between function and phylogenetic

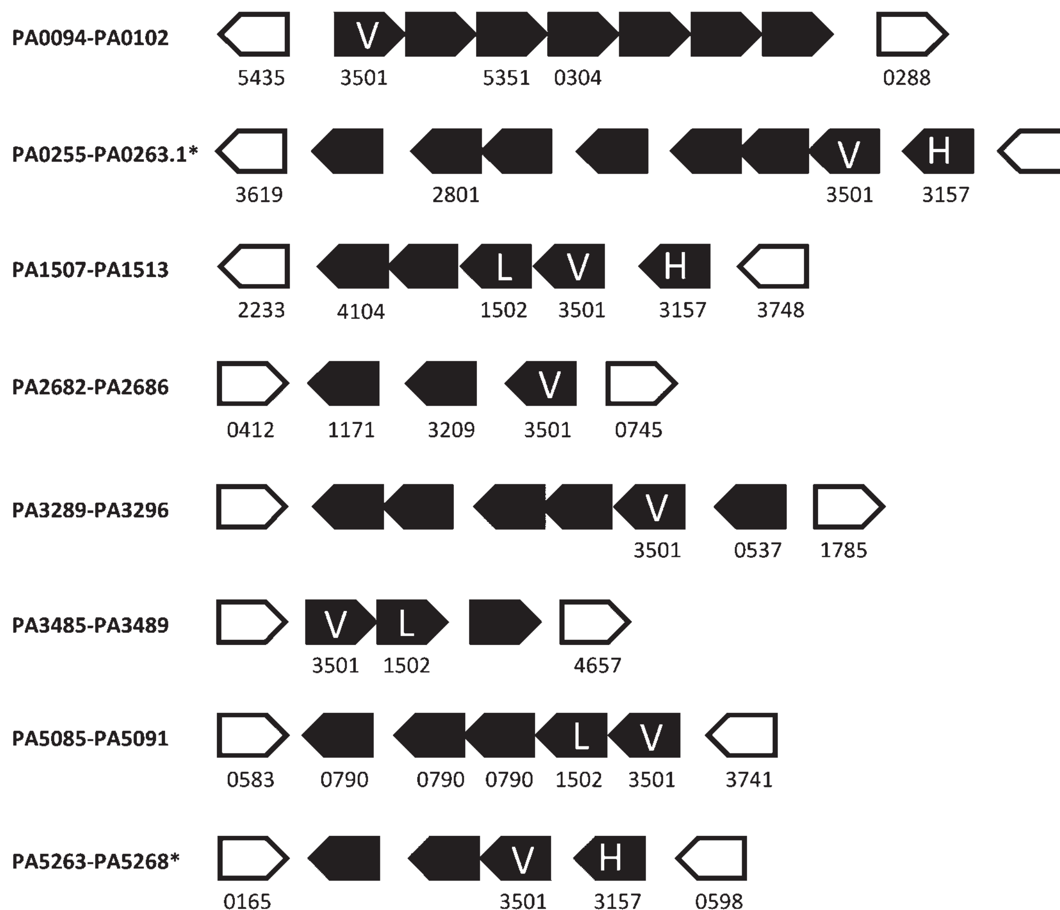

Fig. 5. Genomic organization of the *P. aeruginosa vgrG* islands. Genes are represented as blocked arrows showing the direction of their transcription. Numbers represent COG number. Black arrows indicate genes associated with *vgrG*, while white arrows indicate flanking conserved genes. Letters V, H and L represent *vgrG*, *hcp* and genes encoding lipases, respectively. Asterisks indicate that gene content among different strains could vary.

clusters could be found. For example, the phylogenetic cluster 1 has been found to be involved in virulence of *Aeromonas hydrophila*, *Vibrio cholerae* and *Burkholderia thailandensis* towards eukaryotic cells (Pukatzki *et al.*, 2007; Schwarz *et al.*, 2010b; Suarez *et al.*, 2010). In *P. aeruginosa* this cluster, called HSI-II, is also involved in bacterial virulence towards plants and animals (Lesic *et al.*, 2009). Remarkably, cluster 1 is the only T6SS encoded in the genome of the entomopathogenic bacterium *P. entomophila* L48, which, assuming that these genes prove to be functional, also suggests a role of this system in bacterial virulence. Another interesting example is the presence of a unique cluster 2 in two strains isolated from poplar, namely *P. putida* W619 and uncultured proteobacterium QS1 (Taghavi *et al.*, 2005; Williamson *et al.*, 2005). Whereas the lifestyle of strain QS1 is not known, *P. putida* W619 is an endophytic strain. Interestingly, the phylogenetic cluster 2 is the most frequent T6SS group in an endophytic microbiome from rice (GOLD stamp Gm00046, data not shown). Therefore cluster 2 in *P. putida* W619 could have an endophytic-specific function.

Phylogeny could be useful for predicting function associated with different T6SSs. However, the specificity of such functions could also be linked to the activity of specific effectors. For example, the cluster 3 of *P. aeruginosa* has been shown to target a toxin to other bacteria, which suggests a role for this system in bacterial–bacterial interaction (Hood *et al.*, 2010). Whether this is the case with each *Pseudomonas* cluster 3 is unclear, as the bacterial toxin is uniquely encoded in *P. aeruginosa* genomes. Unfortunately, no other T6SS effectors have yet been identified within the genus *Pseudomonas*, which hampers further understanding of these secretion systems.

The fact that multiple VgrGs and Hcps could be associated with one specific secretion apparatus could also possibly explain the myriad of phenotypes associated with T6SS. Indeed, one may propose that several subsets of effectors, including different Hcp and VgrG proteins, are associated with particular T6SSs under certain conditions. According to the phylogenetic and expression clusters of VgrGs and Hcps (Table 3), it is believed that three VgrGs (VgrG1,

Table 3. Relation between *P. aeruginosa* T6SS locus and VgrGs/Hcps

Relationship between orphan VgrGs/Hcps and associated *P. aeruginosa* T6SS locus are summarized in this table. PC, relationship based on VgrG or Hcp phylogenetic clusters (Figs 3 and 4); EC, relationship based on expression cluster (Supplementary Table S4). Y indicates that the corresponding protein is associated with cluster 1, 3 or 4A.

VgrG/Hcp	Cluster 1 (HSI-II)		Cluster 3 (HSI-I)		Cluster 4A (HSI-III)		Biological evidence
	PC	EC	PC	EC	PC	EC	
PA0085 (Hcp1)			Y				Mougous *et al.* (2006)
PA0091 (VgrG1)			Y	Y			Hood *et al.* (2010)
PA0095 (VgrG)			Y				None
PA2685 (VgrG4)			Y	Y			Hood *et al.* (2010)
PA0262 (VgrG)	Y						None
PA0263 (Hcp)	Y	Y					None
PA1511 (VgrG)	Y						None
PA1512 (Hcp)	Y						None
PA2367 (Hcp3)					Y	Y	None
PA2373 (VgrG3)					Y	Y	None
PA3294 (VgrG)		Y					None
PA3486 (VgrG)		Y					None
PA5090 (VgrG)							None
PA5266 (VgrG)	Y						None
PA5267 (Hcp)	Y						None

PA0095 and VgrG4) and one Hcp (Hcp1) are associated with locus 3 (or HSI-I) whereas VgrG3 and Hcp3 are part of locus 4A (or HSI-III) of *P. aeruginosa*. Among these proteins, VgrG1, VgrG4 and Hcp1 have already been isolated from the supernatant of *P. aeruginosa* with an 'on-state' locus 3 (Hood *et al.*, 2010; Mougous *et al.*, 2006). The other orphan Hcps and VgrGs of *P. aeruginosa* seem to be related to locus 1.1 (HSI-I). However, we cannot rule out the possibility that some orphan Hcps and VgrGs have evolved non-T6SS specialized functions. Indeed, Hcp phylogeny clearly highlights one clade of orphan Hcp unlinked to any *vgrG*. Whether these Hcps are involved in other biological functions, such as the transcriptional regulator HilE in *Salmonella enterica* (Blondel *et al.*, 2009), remains to be demonstrated.

A novel observation from this study was the identification of several conserved gene arrangements consisting of two to eight genes containing *vgrG* and often *hcp* (Fig. 5). These '*vgrG* islands' are frequently observed in other bacterial species (unpublished observation). Thus far, only one *vgrG* island has been associated with a biological function (Gibbs *et al.*, 2008). Genes encoded within these *vgrG* islands could be part of the secretion machinery or secreted substrates. Among these genes, some encode proteins with potential lipase or esterase activity. For example PA1510 possesses a PFAM domain of a PGAP1-like protein. PGPA1 is an endoplasmic reticulum (ER) membrane protein which is involved in inositol deacylation of glycosylphosphatidylinositol (GPI)-anchored proteins (Tanaka *et al.*, 2004). This process is important for efficient ER-to-Golgi transport of GPI-anchored proteins (Tanaka *et al.*, 2004). Thus the presence of the PGPA1 domain could imply a role for PA1510 in host protein trafficking. Two other *P. aeruginosa*-specific lipases, PA3487 (or PldA) and PA5089, encoded within VgrG islands possess phospholipase D (PLD) domains, and PLD activity has been demonstrated for PldA (Wilderman *et al.*, 2001). Bacterial PLDs are often associated with virulence as they specifically hydrolyse phosphatidylcholine, although phosphatidylinositol-specific PLDs have also been identified (Liscovitch *et al.*, 2000). PldA is involved in virulence in a chronic pulmonary infection model (Wilderman *et al.*, 2001). However, during its characterization, this enzyme, which lacks the type 2 signal sequence, was not detected in culture supernatants under the conditions used (Wilderman *et al.*, 2001). Interestingly, the fact that PA1510 and PA5089 also lacked a canonical hydrophobic signal peptide suggested that these three proteins could possibly be secreted by the T6SS. However, the extracellular signals that may trigger their secretion seem to be different, as only *pldA* was found to be co-regulated with HSI-II of *P. aeruginosa* PAO1.

This work demonstrates that phylogeny and transcriptome analyses could be powerful tools to predict proteins associated with secretion systems. While we are aware that some of the candidates identified in our meta-analysis of transcriptome data are likely to be unrelated to T6SS function, we believe that some of them could be new T6S effectors. Some of these predicted effectors linked with HSI-II are currently under study in our laboratory.

ACKNOWLEDGEMENTS

This research was supported in parts by grants awarded by the Science Foundation of Ireland (07/IN.1/B948, 08/RFP/GEN1295, 08/RFP/

GEN1319, 09/RFP/BMT2350); the Department of Agriculture, Fisheries and Food (RSF grants 06-321 and 06-377; FIRM grants 06RDC459 06RDC506 and 08RDC629); the European Commission (MTKD-CT2006-042062, Marie Curie TOK:TRAMWAYS); ERCSET (05/EDIV/FP107/INTERPAM), the Marine Institute Beaufort award (C&CRA 2007/082) and the HRB (RP/2006/271, RP/2007/290, HRA/2009/146). The authors thank members of the BIOMERIT Research Centre for their support and scientific input.

REFERENCES

Alvarez-Martinez, C. E. & Christie, P. J. (2009). Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, 775–808.

Aschtgen, M. S., Bernard, C. S., De Bentzmann, S., Lloubès, R. & Cascales, E. (2008). SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *J Bacteriol* **190**, 7523–7531.

Aschtgen, M. S., Gavioli, M., Dessen, A., Lloubès, R. & Cascales, E. (2010). The SciZ protein anchors the enteroaggregative *Escherichia coli* Type VI secretion system to the cell wall. *Mol Microbiol* **75**, 886–899.

Bernard, C. S., Brunet, Y. R., Gueguen, E. & Cascales, E. (2010). Nooks and crannies in type VI secretion regulation. *J Bacteriol* **192**, 3850–3860.

Bingle, L. E. H., Bailey, C. M. & Pallen, M. J. (2008). Type VI secretion: a beginner's guide. *Curr Opin Microbiol* **11**, 3–8.

Blondel, C. J., Jiménez, J. C., Contreras, I. & Santiviago, C. A. (2009). Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics* **10**, 354.

Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. (2009). Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J* **28**, 315–325.

Bönemann, G., Pietrosiuk, A. & Mogk, A. (2010). Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**, 815–821.

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & 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**, 104.

Bröms, J. E., Lavander, M. & Sjöstedt, A. (2009). A conserved alpha-helix essential for a type VI secretion-like system of *Francisella tularensis*. *J Bacteriol* **191**, 2431–2446.

Burtnick, M. N., DeShazer, D., Nair, V., Gherardini, F. C. & Brett, P. J. (2010). *Burkholderia mallei* cluster 1 type VI secretion mutants exhibit growth and actin polymerization defects in RAW 264.7 murine macrophages. *Infect Immun* **78**, 88–99.

Cascales, E. (2008). The type VI secretion toolkit. *EMBO Rep* **9**, 735–741.

Cornelis, G. R. (2006). The type III secretion injectisome. *Nat Rev Microbiol* **4**, 811–825.

Ernst, R. K., D'Argenio, D. A., Ichikawa, J. K., Bangera, M. G., Selgrade, S., Burns, J. L., Hiatt, P., McCoy, K., Brittnacher, M. & other authors (2003). Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environ Microbiol* **5**, 1341–1349.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philos Trans R Soc Lond B Biol Sci* **360**, 1427–1434.

Filloux, A., Hachani, A. & Bleves, S. (2008). The bacterial type VI secretion machine: yet another player for protein transport across membranes. *Microbiology* **154**, 1570–1583.

Gibbs, K. A., Urbanowski, M. L. & Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science* **321**, 256–259.

Gross, H. & Loper, J. E. (2009). Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat Prod Rep* **26**, 1408–1446.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Hood, R. D., Singh, P., Hsu, F. S., Güvener, T., Carl, M. A., Trinidad, R. R. S., Silverman, J. M., Ohlson, B. B., Hicks, K. G. & Plemel, R. L. (2010). A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**, 25–37.

Hsu, F. S., Schwarz, S. & Mougous, J. D. (2009). TagR promotes PpkA-catalysed type VI secretion activation in *Pseudomonas aeruginosa*. *Mol Microbiol* **72**, 1111–1125.

Leiman, P. G., Basler, M., Ramagopal, U. A., Bonanno, J. B., Sauder, J. M., Pukatzki, S., Burley, S. K., Almo, S. C. & Mekalanos, J. J. (2009). Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**, 4154–4159.

Lesic, B., Starkey, M., He, J., Hazan, R. & Rahme, L. G. (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.

Liscovitch, M., Czarny, M., Fiucci, G. & Tang, X. Q. (2000). Phospholipase D: molecular and cell biology of a novel gene family. *Biochem J* **345**, 401–415.

Liu, H., Coulthurst, S. J., Pritchard, L., Hedley, P. E., Ravensdale, M., Humphris, S., Burr, T., Takle, G., Brurberg, M. B. & other authors (2008). Quorum sensing coordinates brute force and stealth modes of infection in the plant pathogen *Pectobacterium atrosepticum*. *PLoS Pathog* **4**, e1000093.

Ma, A. T. & Mekalanos, J. J. (2010). In vivo actin cross-linking induced by *Vibrio cholerae* type VI secretion system is associated with intestinal inflammation. *Proc Natl Acad Sci U S A* **107**, 4365–4370.

Ma, L. S., Lin, J. S. & Lai, E. M. (2009). An IcmF family protein, ImpLM, is an integral inner membrane protein interacting with ImpKL, and its Walker A motif is required for type VI secretion system-mediated Hcp secretion in *Agrobacterium tumefaciens*. *J Bacteriol* **191**, 4316–4329.

MacIntyre, D. L., Miyata, S. T., Kitaoka, M. & Pukatzki, S. (2010). The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci U S A* **107**, 19520–19524.

Mougous, J. D., Cuff, M. E., Raunser, S., Shen, A., Zhou, M., Gifford, C. A., Goodman, A. L., Joachimiak, G., Ordoñez, C. L. & other authors (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. & Mekalanos, J. J. (2007). Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nat Cell Biol* **9**, 797–803.

Parsons, D. A. & Heffron, F. (2005). *sciS*, an *icmF* homolog in *Salmonella enterica* serovar Typhimurium, limits intracellular replication and decreases virulence. *Infect Immun* **73**, 4338–4345.

Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. (2009). The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci U S A* **106**, 4160–4165.

Persson, O. P., Pinhassi, J., Riemann, L., Marklund, B. I., Rhen, M., Normark, S., González, J. M. & Hagström, A. (2009). High abundance of virulence gene homologues in marine bacteria. *Environ Microbiol* **11**, 1348–1357.

Potvin, E., Lehoux, D. E., Kukavica-Ibrulj, I., Richard, K. L., Sanschagrin, F., Lau, G. W. & Levesque, R. C. (2003). In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environ Microbiol* **5**, 1294–1308.

Puigbò, P., Garcia-Vallvé, S. & McInerney, J. O. (2007). TOPD/FMTS: a new software to compare phylogenetic trees. *Bioinformatics* **23**, 1556–1558.

Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. (2007). Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* **104**, 15508–15513.

Pukatzki, S., McAuley, S. B. & Miyata, S. T. (2009). The type VI secretion system: translocation of effectors and effector-domains. *Curr Opin Microbiol* **12**, 11–17.

Robinson, J. B., Telepnev, M. V., Zudina, I. V., Bouyer, D., Montenieri, J. A., Bearden, S. W., Gage, K. L., Agar, S. L., Foltz, S. M. & Chauhan, S. (2009). Evaluation of a *Yersinia pestis* mutant impaired in a thermoregulated type VI-like secretion system in flea, macrophage and murine models. *Microb Pathog* **47**, 243–251.

Roucourt, B., Lecoutere, E., Chibeu, A., Hertveldt, K., Volckaert, G. & Lavigne, R. (2009). A procedure for systematic identification of bacteriophage-host interactions of *P. aeruginosa* phages. *Virology* **387**, 50–58.

Schwarz, S., Hood, R. D. & Mougous, J. D. (2010a). What is type VI secretion doing in all those bugs? *Trends Microbiol* **18**, 531–537.

Schwarz, S., West, T. E., Boyer, F., Chiang, W. C., Carl, M. A., Hood, R. D., Rohmer, L., Tolker-Nielsen, T., Skerrett, S. J. & Mougous, J. D. (2010b). *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog* **6**, e1001068.

Shrivastava, S. & Mande, S. S. (2008). Identification and functional characterization of gene components of type VI secretion system in bacterial genomes. *PLoS ONE* **3**, e2955.

Southey-Pillig, C. J., Davies, D. G. & Sauer, K. (2005). Characterization of temporal protein production in *Pseudomonas aeruginosa* biofilms. *J Bacteriol* **187**, 8114–8126.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S. L., Hufnagle, W. O., Kowalik, D. J. & other authors (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964.

Suarez, G., Sierra, J. C., Erova, T. E., Sha, J., Horneman, A. J. & Chopra, A. K. (2010). A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. *J Bacteriol* **192**, 155–168.

Taghavi, S., Barac, T., Greenberg, B., Borremans, B., Vangronsveld, J. & van der Lelie, D. (2005). Horizontal gene transfer to endogenous endophytic bacteria from poplar improves phytoremediation of toluene. *Appl Environ Microbiol* **71**, 8500–8505.

Tanaka, S., Maeda, Y., Tashima, Y. & Kinoshita, T. (2004). Inositol deacylation of glycosylphosphatidylinositol-anchored proteins is mediated by mammalian PGAP1 and yeast Bst1p. *J Biol Chem* **279**, 14256–14263.

van Passel, M. W. J., Luyf, A. C. M., van Kampen, A. H. C., Bart, A. & van der Ende, A. (2005). Deltarho-web, an online tool to assess composition similarity of individual nucleic acid sequences. *Bioinformatics* **21**, 3053–3055.

Wang, X., Wang, Q., Xiao, J., Liu, Q., Wu, H., Xu, L. & Zhang, Y. (2009). *Edwardsiella tarda* T6SS component *evpP* is regulated by *esrB* and iron, and plays essential roles in the invasion of fish. *Fish Shellfish Immunol* **27**, 469–477.

Weber, B., Hasic, M., Chen, C., Wai, S. N. & Milton, D. L. (2009). Type VI secretion modulates quorum sensing and stress response in *Vibrio anguillarum*. *Environ Microbiol* **11**, 3018–3028.

Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691–699.

Wilderman, P. J., Vasil, A. I., Johnson, Z. & Vasil, M. L. (2001). Genetic and biochemical analyses of a eukaryotic-like phospholipase D of *Pseudomonas aeruginosa* suggest horizontal acquisition and a role for persistence in a chronic pulmonary infection model. *Mol Microbiol* **39**, 291–304.

Williamson, L. L., Borlee, B. R., Schloss, P. D., Guan, C. H., Allen, H. K. & Handelsman, J. (2005). Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol* **71**, 6335–6344.

Wu, H. Y., Chung, P. C., Shih, H. W., Wen, S. R. & 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.

Zheng, J. & Leung, K. Y. (2007). Dissection of a type VI secretion system in *Edwardsiella tarda*. *Mol Microbiol* **66**, 1192–1206.

Edited by: W. Bitter