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

Mathieu Barret,1,† Frank Egan,1,† Emilie Fargier,1 John P. Morrissey2 and Fergal O’Gara1

1BIOMERIT Research Centre, Department of Microbiology, University College Cork, Cork, Ireland
2Department of Microbiology, University College Cork, Cork, Ireland

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 T6S assembly has been proposed (Bönnemann 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 (Bönnemann 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 (Bönnemann 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 T6S. 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

---

**Correspondence**
Fergal O’Gara
f.ogara@ucc.ie

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

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.

†These authors contributed equally to this work.

Supplementary material is available with the online version of this paper.
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 PA01 (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 PA01 (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 <10⁻¹⁷). 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 loci 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 δ² (the average dinucleotide relative abundance difference) between T6SS loci sequences and the associated genome sequence using a web-based application, δ²p-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.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Status</th>
<th>NCBI reference sequence</th>
<th>Locus 1</th>
<th>Locus 2</th>
<th>Locus 3</th>
<th>Locus 4A</th>
<th>Locus 4B</th>
<th>Locus 5</th>
<th>Total</th>
<th>VgrG</th>
<th>Hcp</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. aeruginosa</em> PAO1</td>
<td>Complete</td>
<td>NC_002516</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> UCBPP-PA14</td>
<td>Complete</td>
<td>NC_008463</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PA7</td>
<td>Complete</td>
<td>NC_009656</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> LESB58</td>
<td>Complete</td>
<td>NC_011770</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> 2192</td>
<td>Draft</td>
<td>NZ_AAKV000000000000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> C3719</td>
<td>Draft</td>
<td>NZ_AAKV000000000000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PACS2</td>
<td>Draft</td>
<td>NZ_AAQW000000000000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PAb1</td>
<td>Draft</td>
<td>NZ_ABKZ000000000000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>6*</td>
</tr>
<tr>
<td><em>P. entomophila</em> L48</td>
<td>Complete</td>
<td>NC_008027</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td><em>P. fluorescens</em> Pf-5</td>
<td>Complete</td>
<td>NC_004129</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><em>P. fluorescens</em> PFO-1</td>
<td>Complete</td>
<td>NC_007492</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td><em>P. fluorescens</em> SBW25</td>
<td>Complete</td>
<td>NC_012660</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>P. fluorescens</em> F113</td>
<td>Draft</td>
<td>NC_009439</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td><em>P. mendocina</em> ymp</td>
<td>Complete</td>
<td>NC_009439</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><em>P. putida</em> KT2440</td>
<td>Complete</td>
<td>NC_002947</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><em>P. putida</em> GB-1</td>
<td>Complete</td>
<td>NC_010322</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><em>P. putida</em> W619</td>
<td>Complete</td>
<td>NC_010501</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>P. putida</em> F1</td>
<td>Complete</td>
<td>NC_009512</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><em>P. stutzeri</em> A1501</td>
<td>Complete</td>
<td>NC_009434</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>P. savastanoi</em> pv. savastanoi NCPPB 3335</td>
<td>Draft</td>
<td>NC_012660</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. aesculi 2250</td>
<td>Draft</td>
<td>NZ_ACUO000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. aesculi NCPPB3681</td>
<td>Draft</td>
<td>NZ_ACUO000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. oryzae 1_6</td>
<td>Draft</td>
<td>NZ_AB2R000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. phaseolicola 1448A</td>
<td>Complete</td>
<td>NC_005773</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. syringae B728a</td>
<td>Complete</td>
<td>NC_007005</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. syringae FF5</td>
<td>Draft</td>
<td>NZ_ACUZ000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>8*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. syringae 642</td>
<td>Draft</td>
<td>NZ_ACUZ000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tabaci 11528</td>
<td>Draft</td>
<td>NZ_ACHU000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato DC3000</td>
<td>Complete</td>
<td>NC_004578</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato T1</td>
<td>Draft</td>
<td>NZ_ABSM000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato K40</td>
<td>Draft</td>
<td>NZ_ADFY000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato Max13</td>
<td>Draft</td>
<td>NZ_ADFZ000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato NCPPB 1108</td>
<td>Draft</td>
<td>NZ_ADGA000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>5*</td>
</tr>
<tr>
<td><em>P. syringae</em> pv. tomato NCPPB 1108</td>
<td>Draft</td>
<td>NZ_ACOQ000000000000</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0*</td>
</tr>
</tbody>
</table>

*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 orthologous groups of proteins) defined by Boyer et al. (2009) was analysed for each 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, COG5315, COG5316, COG3517, COG5318, COG5319, COG5320, COG5321, COG5322 and COG5323). 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 (Puigbo 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 V2.72 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 core-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 subclusters, 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 T6 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 subclusters. To be consistent with previous nomenclature used by Boyer et al. (2009), these two distinct subclusters 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...
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 6p-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

---

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 PHYLP 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.
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/effecter 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 \textit{P. fluorescens} Pf-5). The link between PFL2812 and phylogenetic cluster 1.1 could be explained by loss of cluster 1.1 from \textit{P. fluorescens} Pf-5 or by an acquisition of this \textit{vgrG} through HGT.

\textbf{Fig. 4.} Phylogenetic distribution of VgrG proteins within \textit{Pseudomonas} species. A distance tree (maximum-likelihood) was calculated from 180 VgrG protein sequences of \textit{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 \textit{vgrG} is linked to \textit{hcp}, whereas light grey rectangles indicate that a gene encoding a lipase is found in the vicinity of \textit{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 HiE 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* PFO-1 (Pfl01_2329) and five VgrGs of *P. entomophila* L48. Finally, the third group is shared by three proteins (Psyrl_4974, Psyrp1_25120 and Psyrp2_10189) each belonging to one *P. syringae* strain. A motif search was systematically performed with INTERPRO scan to predict a specific function for VgrG 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 vgrGs 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 vgrG 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 genes 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-II); PA3486 and PA3294 and PA3486 with phylogenetic cluster 1 (HSI-II). The other five orphan vgrG genes, PA0095, PA0262, PA1511, PA0590 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 (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
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 T6SS 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
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,
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/...
REFERENCES


Editted by: W. Bitter

http://mic.sgmjournals.org