Molecular Microbiology (2008) 68(4), 871–889 ■
doi:10.1111/j.1365-2958.2008.06209.x
First published online 8 April 2008

Characterization of the PvdS-regulated promoter motif in *Pseudomonas syringae* pv. *tomato* DC3000 reveals regulon members and insights regarding PvdS function in other pseudomonads

Bryan Swingle,[1,2]* Deepti Thete,[2] Monica Moll,[1] Christopher R. Myers,[3] David J. Schneider[1,2] and Samuel Cartinhour[1,2]

[1]*United States Department of Agriculture-Agricultural Research Service, Ithaca, NY 14853, USA.*
[2]*Department of Plant Pathology and Plant-Microbe Biology, and* [3]*Computational Biology Service Unit, Center for Advanced Computing, Cornell University, Ithaca, NY 14853, USA.*

Summary

Bacteria that survive under variable conditions possess an assortment of genetic regulators to meet these challenges. The group IV or extracytoplasmic function (ECF) sigma factors regulate gene expression in response to specific environmental signals by altering the promoter specificity of RNA polymerase. We have undertaken a study of PvdS, a group IV sigma factor encoded by *Pseudomonas syringae* pv. *tomato* DC3000 (DC3000), a plant pathogen that is likely to encounter variations in nutrient availability as well as plant host defences. The gene encoding PvdS was previously identified by sequence similarity to the *Pseudomonas aeruginosa* orthologue, which directs transcription of genes encoding the biosynthesis of pyoverdine, a siderophore involved in iron acquisition, and is responsible for the characteristic fluorescence of the pseudomonads. We identified 15 promoters regulated by PvdS in DC3000 and characterized the promoter motif using computational analysis. Mutagenesis of conserved nucleotides within the motif interfered with promoter function and the degree of the effect was different depending on which region of the motif was mutated. Hidden Markov models constructed from alignments of sequence motifs extracted from DC3000 and PAO1 were used to query genomes of DC3000 and other fluorescent pseudomonads for similar motifs. We conclude that the role of PvdS as a regulator of pyoverdine synthesis is conserved among the fluorescent pseudomonads, but the promoters recognized by PvdS orthologues may differ subtly from species to species.

Accepted 7 March, 2008. *For correspondence. E-mail Bryan.Swingle@ars.usda.gov; Tel. (+1) 607 255 7893; Fax (+1) 607 255 4471.

Introduction

Bacterial survival depends on co-ordinated and systematic gene regulation in response to changing conditions. The nature and variety of the changes are reflected in the genome, with species that are free living in the environment tending to encode proportionally more transcriptional regulators (Stover *et al.*, 2000). Promoter selection by alternative forms of RNA polymerase holoenzyme containing different sigma factor subunits is a primary mechanism for regulating transcription of collections of genes (Gruber and Gross, 2003). The sigma subunit confers the sequence-specific DNA-binding determinants to RNA polymerase allowing it to recognize and correctly bind to promoter sequences (Murakami *et al.*, 2002).

Most bacteria contain multiple members of the σ^{70} family of sigma factors. The primary σ^{70}-type sigma factor is responsible for directing expression of housekeeping functions during exponential growth. The most diverse group of σ^{70}-type sigma factors is the group IV or extracytoplasmic function (ECF) sigma factors, which are generally involved in regulating transcription in response to environmental signals (Helmann, 2002; Potvin *et al.*, 2008). The model plant pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (DC3000) (Buell *et al.*, 2003) is predicted to encode 10 group IV sigma factors (Oguiza *et al.*, 2005). Based on sequence similarity, five appear to be dedicated to regulating the expression of genes whose products are involved with iron acquisition and homeostasis.

Pseudomonas syringae and other fluorescent pseudomonads are characterized by their ability to secrete pyoverdine in low-iron conditions (Meyer, 2000).

Pyoverdines are chemically diverse, yellow-green fluorescent, high-affinity siderophores capable of chelating Fe(III) as a mechanism for scavenging iron from the environment. Pyoverdines are the primary iron acquisition system in pseudomonads and give these organisms a competitive advantage in the environment and in their association with plant or animal hosts (Poole and McKay, 2003; Sarkar *et al*., 2006). In addition to their role in iron acquisition, pyoverdines also function as extracellular signalling molecules important for inducing the production of virulence factors in *Pseudomonas aeruginosa* (Lamont *et al*., 2002). Presumably, the low iron concentration in the host signals the requirement for pyoverdine synthesis as well as other factors required to mount a successful infection.

The genes coding for the biosynthesis of pyoverdine are organized in genomic clusters in all sequenced pseudomonad genomes that have been analysed (Buell *et al*., 2003; Ravel and Cornelis, 2003; Feil *et al*., 2005; Joardar *et al*., 2005; Paulsen *et al*., 2005). In *P. aeruginosa* the expression of these genes is regulated by the group IV sigma factor PvdS (Leoni *et al*., 2000; Wilson and Lamont, 2000). The gene encoding PvdS in *P. aeruginosa* is regulated in response to iron concentration in the growth media via the ferric uptake regulator, Fur (Cunliffe *et al*., 1995; Leoni *et al*., 1996). It has been hypothesized that pyoverdine synthesis is regulated by similar mechanisms in other pseudomonads (Cunliffe *et al*., 1995; Rombel *et al*., 1995; Leoni *et al*., 2000; Ravel and Cornelis, 2003). Lamont and co-workers identified a conserved TAAAT nucleotide motif known as the IS-box upstream of PvdS-regulated genes. This motif has been proposed to function as part of the –35 element of PvdS-regulated promoters (Rombel *et al*., 1995; Wilson *et al*., 2001). A conserved CGT nucleotide triplet positioned 16 bp downstream of the IS-box has also been noted in some PvdS-regulated promoters (Ochsner *et al*., 2002; Visca *et al*., 2002). The location, spacing and function of these motifs are consistent with the hypothesis that they are part of the –35 and –10 elements of PvdS-regulated promoters in *P. aeruginosa*.

The orthologue of PvdS (PSPTO_2133) was identified in DC3000 by sequence analysis (Ravel and Cornelis, 2003; Oguiza *et al*., 2005) but the function of the gene has not been demonstrated. To characterize the DC3000 PvdS regulon, PvdS-responsive promoters were obtained by screening a DC3000 genomic DNA promoter trap library and further analysed using transcriptional reporter fusions, quantitative real-time (qRT)-PCR and transcriptional start site mapping. The PvdS-box motif was identified using Gibbs sampling and was analysed by scanning mutagenesis to identify conserved nucleotides whose presence is required for full promoter function. The results indicate that the putative –10 and –35 elements are both necessary for promoter function and that the two domains differ in their sensitivity to single-nucleotide substitutions. Three hidden Markov models (HMM) were constructed using either (i) the experimentally verified PvdS-regulated promoter regions, (ii) a set of 13 IS-box promoters from *P. aeruginosa* PAO1 (PAO1) based on data summarized in Ravel and Cornelis (2003), or (iii) the combination of sequences from both (i) and (ii) to yield a combined DC3000/PAO1-based model. These models were used to scan the genomes of 10 fluorescent pseudomonads, each encoding PvdS orthologues sharing greater than 80% identity with the DC3000 PvdS and invariant with respect to the residues important for promoter binding in the PAO1 orthologue (Wilson and Lamont, 2006). As expected, many of the genes linked with the motif were pyoverdine-related but in every species some genes were detected that were not obviously involved with pyoverdine or iron metabolism. The results of the HMM scans varied from species to species, suggesting the possibility that the PvdS-box may not be strictly conserved within the genus.

Results

Genome-wide screen identifies PvdS-regulated promoters

A conventional promoter trap library screen was used to isolate PvdS-regulated promoters encoded in the DC3000 genome. The library was constructed using size-selected, partially digested DC3000 genomic DNA that was subcloned in a vector (pBS29) upstream of a promoterless *′lacZα* gene. In this construct, the *′lacZα* gene is expressed only when the genomic insert encodes an active promoter. The screening was carried out in *Escherichia coli* where previous studies have shown that PvdS-regulated promoters from *P. aeruginosa* are active only in the presence of PvdS expressed *in trans* (Cunliffe *et al*., 1995). This helped focus the screen to identify promoters that were directly regulated by PvdS and to minimize complications due to secondary regulatory effects or induction from other iron-regulated sigma factors that could occur in the native DC3000 background.

Screening for PvdS-regulated promoters was performed in two steps. First, *E. coli* cells transformed with the promoter trap library and a DC3000 PvdS expression vector were screened for a lac+ phenotype on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) indicator media, which identified clones containing active promoters. Second, the lac+ clones were evaluated to identify isolates where reporter expression was dependent on the presence of the PvdS expression construct. From 116 000 initial transformants, 2992 were isolated in step one of the screen. Of these, 97 clones were confirmed to exhibit PvdS-dependent expression in step two, and the inserts

were sequenced and mapped onto the DC3000 genome. The 25 unique sequences from this set that contained intergenic regions (see Table S1) were used to construct preliminary hidden Markov models (data not shown) for scanning the genome to identify candidate promoters and to guide subsequent experimentation as described below.

Multiple methods validate candidate promoters

The candidate promoters suggested by the process above were chosen for functional validation using reporter fusions, qRT-PCR and transcription start site mapping. This strategy reflects our goal of both associating function with particular upstream sequence fragments and confirming PvdS-regulated expression of downstream genes.

The *lux* reporter fusion assay was used to associate promoter activity with small DNA fragments and to isolate promoter activity from potentially confounding effects such as readthrough from transcripts initiated upstream in the native genomic context. Each region was cloned as a transcriptional fusion with a *′luxCDABE* operon. Promoter activity of these constructs was assessed in *E. coli* cells transformed with either a PvdS expression vector or the empty vector control. Ten regions showed strong PvdS-dependent expression, as displayed in Fig. 1A. Expression varied from 23- to 676-fold, compared with a twofold change in the promoterless reporter control, confirming promoter activity in these fragments.

A qRT-PCR assay was used to assess PvdS-dependent transcription of genes downstream of candidate promoters in their native context in DC3000 (see Fig. 1B). Relative expression for 18 genomic loci was assessed by comparing transcript levels in DC3000 cells transformed with the PvdS expression vector versus DC3000 cells transformed with the empty vector. This confirmed PvdS-dependent differential expression of all 18 genes, 15 of which are directly downstream of candidate promoters, two of which (PSPTO_2136/*daT* and PSPTO_2151) are potentially within operons preceded by candidate promoters, and one of which, PSPTO_3556 (*glcE*), is in an unconventional orientation relative to the candidate promoter (see Fig. 1B and below). Overall, differential expression in this experiment ranged from 5- to 329-fold. The results from both transcriptional fusion assays and qRT-PCR assays were consistent with these regions containing *cis*-acting elements that function as PvdS-regulated promoters. However, in one case (PSPTO_2175, *leuB*), PvdS promoter activity was observed using the promoter fusion but could not be detected using qRT-PCR at the native location (data not shown). It is possible that expression at this locus in DC3000 requires other regulators [e.g. Lrp as in *E. coli* (Newman and Lin, 1995)].

A 5′ RACE strategy was used to map the location of transcription start sites for several PvdS-regulated genes.

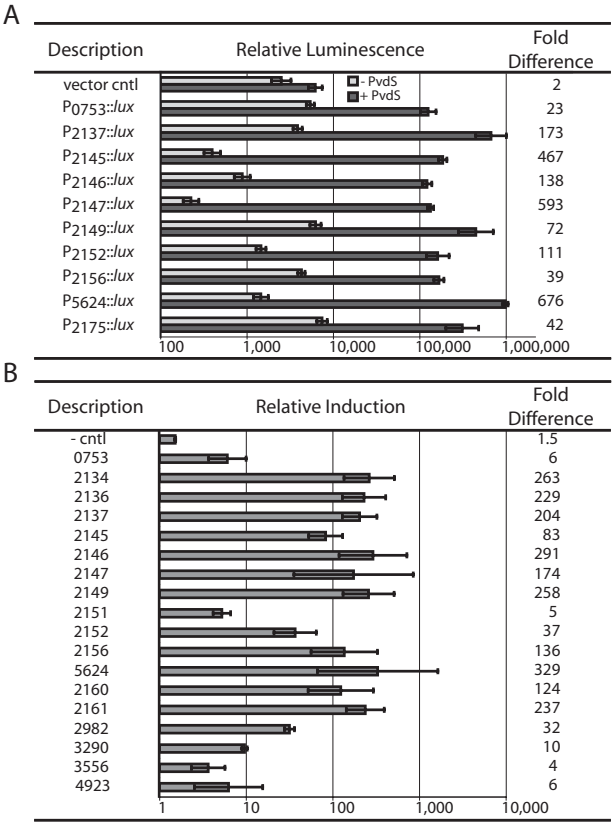

Fig. 1. PvdS-dependent promoter activity. Genomic regions assessed for PvdS-dependent promoter activity were identified by promoter trapping or by scanning with preliminary bioinformatics models constructed using sequences identified in the promoter trapping experiments. Promoter and gene numbers refer to PSPTO numbers in the annotated DC3000 genome sequence NC_004578 (Buell *et al*., 2003) here and throughout the report, except where noted. Error bars show standard deviation.
A. Relative luminescence expressed from *E. coli* cells (six biological replicas) transformed with the promoter::*lux* fusion construct and either the empty expression vector control (light bars) or the PvdS-expressing construct (dark bars). The vector control (vector cntl) was the promoter trap vector without a promoter insert.
B. qRT-PCR analysis showing PvdS-dependent differential expression in DC3000 (three biological replicas). Relative induction is the ratio of normalized transcript level in DC3000 cells transformed with the PvdS expression vector versus DC3000 cells transformed with the empty vector control. Transcript levels for each gene were normalized to *gap1*. The *gyrA* gene, which is not predicted to be PvdS-regulated, was used as a negative control (-cntl). All genes were directly downstream from the PvdS-box except for 2136 and 2151, which are potentially part of multicistronic operons preceded by PvdS-box motifs, and 3556, for which the PvdS-box is positioned within the indicated gene in the antisense orientation. For 2175, differential expression was not detected (data not shown).

The point at which transcription initiates is constrained by the structure of the RNA polymerase holoenzyme bound at the promoter, and is generally not more than 10 or 11 bp downstream of the –10 element for σ^{70}-dependent promoters (Murakami *et al*., 2002; Lewis and Adhya, 2004). The transcription start points for alternative sigma factors are

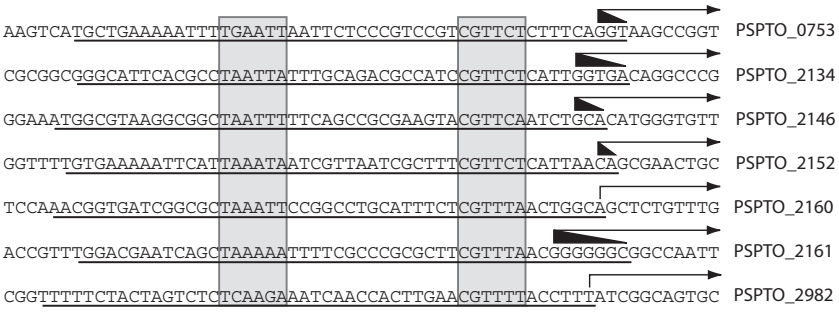

Fig. 2. 5′ RACE analysis for seven PvdS-dependent transcripts. The DC3000 genomic sequence corresponding to the region upstream of the indicated ORF is shown with the promoter elements of the PvdS-box shaded. The direction of transcription is indicated by an arrow; a triangle beneath the arrow indicates a range of possible start sites due to ambiguities in the sequence chromatogram and/or the addition of homopolymer tails during cDNA synthesis. The underlined regions show the 49 nucleotides used for Gibbs sampling.

expected to be similar to the prototypical σ^{70}-regulated promoter elements (Nonaka *et al*., 2006; Rhodius *et al*., 2006). Therefore the location of the start site was determined in order to focus our search for the promoter motif in the smallest possible region. We analysed seven transcripts (Fig. 2); five loci (PSPTO_2134, 2146, 2152, 2160 and 2161) are within the pyoverdine cluster, and two (PSPTO_0753 and 2982) are in different regions with no apparent role in pyoverdine metabolism. Transcription start sites were mapped by 5′ RACE using RNA from DC3000 cells transformed with the PvdS expression construct. For all of the genes tested, the transcription start site was determined with a maximal uncertainty of five nucleotides. The relation of the transcription start site to the promoter motif is discussed below.

Identification of the PvdS-box motif by Gibbs sampling

The combined evidence from the promoter trap library, *lux* fusions, qRT-PCR and transcription start site data was used to define a sequence input set for Gibbs sampling to identify an aligned promoter motif as described previously (Ferreira *et al*., 2006). Where applicable, sequences were trimmed as follows: (i) overlapping library isolates and *lux* fusion sequences were trimmed to the common subsequence, (ii) sequences were trimmed at the 3′ end to remove overlap with annotated coding regions, (iii) sequences were limited to extend no further than 300 nucleotides upstream of the annotated coding region, and (iv) fragments for which transcription start site data were available were trimmed to 49 nucleotides as shown in Fig. 2. The final set of 16 sequences used for Gibbs sampling is summarized in Table S2 with experimental evidence validating their PvdS-dependent regulation.

Gibbs sampling was conducted for a range of motif window sizes. One motif was consistently found in 15 of the input sequences and had a structure consistent with a promoter for an ECF sigma factor. The results of a typical sampling run with a window size of 29 nucleotides are shown in Fig. 3 along with the associated sequence logo; we find this same basic structure preserved in larger window sizes as well. Each input sequence yielded one motif except for PSPTO_2136 (*daT*), for which no motifs were found. The sequence logo suggests the presence of two conserved regions, separated by a less conserved spacer. This structure resembles promoters recognized by σ^{70}-type sigma factors (Gross *et al*., 1998), where these conserved regions are referred to as the –10 and –35 elements. Note that the IS-box TAAAT-(N16)-CGT found upstream of iron-responsive genes in *P. aeruginosa* PAO1 (Wilson *et al*., 2001; Ochsner *et al*., 2002) occupies positions 2–25 in the logo. Given the degree of sequence conservation, we adopted the following convention to represent the consensus sequence for the DC3000 PvdS-box (which differs slightly from that for the PAO1 IS-box): TAAAT[A/T]-(N15)-CGTT[C/T][T/A]. For cases where mapped transcription start sites are available, the 5′ end of the transcript was located between 3 and 9 bp downstream of the 3′ end of the conserved CGTT[C/T][T/A] domain of the motif (Fig. 2). The source of this uncertainty is due to the technical limitations of the transcript mapping method (see *Experimental procedures*).

These data are consistent with the conserved features of this motif functioning as the –10 and –35 promoter elements recognized by PvdS, although close inspection reveals that the geometry of the motif is slightly more compact than the canonical σ^{70} promoters in *E. coli* (Shultzaberger *et al*., 2007). We hypothesize that the conserved nucleotides in Fig. 3 correspond to the promoter sequence recognized by PvdS in DC3000, and henceforth refer to this motif as the PvdS-box.

Mutagenesis in the PvdS-box reduces promoter activity

Typically, σ^{70}-type promoters contain nucleotides within the –10 and –35 promoter elements that are necessary for specific recognition by the sigma factor (McClure *et al*., 1983; Gross *et al*., 1998; Murakami *et al*., 2002); therefore, some nucleotide substitutions within the conserved elements are expected to alter promoter function.

The region upstream of PSPTO_2152, a putative TonB-dependent receptor in the pyoverdine locus, was chosen as an example of a promoter containing consensus sequences in the –10 and –35 elements (Fig. 3). A total of 23 mutant derivatives were constructed from the wild-type promoter sequence using transversion mutations

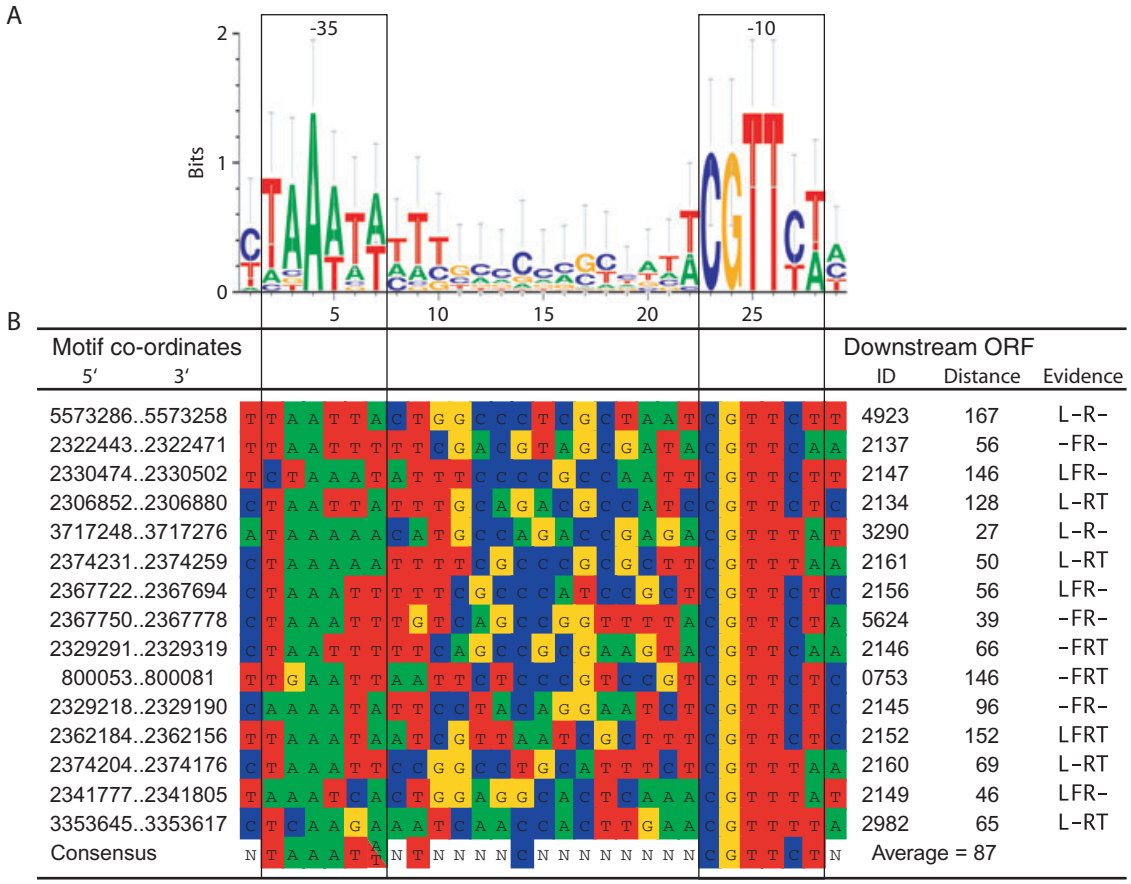

Fig. 3. PvdS-box sequence motif. This alignment was produced by Gibbs sampling analysis of sequences confirmed to contain a promoter regulated by PvdS. A summary of the propensity of each nucleotide to occupy each position is represented above the alignment as a sequence logo.
A. The –35 motif (TAAAT[A/T]) is on the left-hand side and is separated most often by a 15 bp spacer from the –10 motif which contains the invariant CGTT nucleotides.
B. Individual sequence motifs aligned and shown with relevant position information. The distance is from the invariant T at position 26 and the first nucleotide of the annotated start codon. The evidence column summarizes the experimental evidence supporting PvdS-dependent regulation: L, promoter trap library isolate; F, *lux* fusion; R, confirmed by qRT-PCR; T, transcript start site located.

(A↔C and T↔G) to test specific bases for a functional role in PvdS-dependent expression (Table 1). Twenty-one contained single-base substitutions and two contained multiple substitutions in either the –10 or –35 regions. Plasmids containing transcriptional fusions of the wild-type promoter region for PSPTO_2152 or the mutant derivatives with the *′luxCDABE* operon were introduced into *E. coli* along with the PvdS expression construct (pBS49). The effect of the mutations on reporter expression was assessed and compared with the wild-type control to infer transcriptional activity.

Several mutations severely reduced transcription, while none of the mutant promoters was significantly more active than wild type at $P \leq 0.05$. The single-nucleotide mutations with the greatest impact on transcription were all in the –10 region. Substitution of any of the four invariant nucleotides in the –10 element reduced transcription 86% or more compared with wild type (see C162A, G163T, T164G and T165G in Table 1). As expected, similar levels were observed when the entire –10 region was mutated. One position within the spacer region adjacent to the –10 element that was not identified as being conserved among the promoters was associated with modest reductions in expression (T160G), suggesting a role for flanking nucleotides. In contrast to the –10 element, single-base-pair mutations in the –35 element did not have a statistically significant effect on transcription, except in one case (T141G; 65% reduction). Multiple substitutions disrupting the –35 element eliminated promoter activity, confirming the essential role of this element.

These results indicate that both the –10 and –35 elements are required for promoter function and that an important difference exists between the two domains of this motif in terms of their sensitivity to single-nucleotide substitutions. The scanning mutagenesis of the con-

Table 1. Mutagenesis of PvdS-box motif.

Plasmid	Promoter sequence	Log RLU	Promoter activity
pBS102 (WT)	TTCATTAAATAATCGTTAATCGCTTTCGTTCTCAT	5.70 (±0.45)	1.00
pBS102 (A139C)	TTCCTTAAATAATCGTTAATCGCTTTCGTTCTCAT	5.73 (±0.02)	1.06 (0.87)
pBS102 (T140G)	TTCAGTAAATAATCGTTAATCGCTTTCGTTCTCAT	5.72 (±0.04)	1.05 (0.91)
pBS102 (T141G)	TTCATGAAATAATCGTTAATCGCTTTCGTTCTCAT	5.25 (±0.04)	0.35 (0.03)
pBS102 (A142C)	TTCATTCAATAATCGTTAATCGCTTTCGTTCTCAT	5.42 (±0.40)	0.52 (0.21)
pBS102 (A143C)	TTCATTACATAATCGTTAATCGCTTTCGTTCTCAT	5.45 (±0.38)	0.57 (0.26)
pBS102 (A144C)	TTCATTAACTAATCGTTAATCGCTTTCGTTCTCAT	5.62 (±0.40)	0.84 (0.72)
pBS102 (T145G)	TTCATTAAAGAATCGTTAATCGCTTTCGTTCTCAT	5.46 (±0.06)	0.58 (0.18)
pBS102 (A146C)	TTCATTAAATCATCGTTAATCGCTTTCGTTCTCAT	5.78 (±0.41)	1.21 (0.70)
pBS102 (A147C)	TTCATTAAATACTCGTTAATCGCTTTCGTTCTCAT	5.64 (±0.05)	0.88 (0.73)
pBS102 (T148G)	TTCATTAAATAAGCGTTAATCGCTTTCGTTCTCAT	5.66 (±0.06)	0.91 (0.81)
pBS102 (C149A)	TTCATTAAATAATAGTTAATCGCTTTCGTTCTCAT	5.57 (±0.35)	0.77 (0.52)
pBS102 (G150T)	TTCATTAAATAATCTTTAATCGCTTTCGTTCTCAT	5.78 (±0.03)	1.21 (0.62)
pBS102 (141–145)	TTCATGCCCGAATCGTTAATCGCTTTCGTTCTCAT	3.74 (±0.02)	0.01 (6E-6)
pBS102 (T160G)	TTCATTAAATAATCGTTAATCGCTGTCGTTCTCAT	5.22 (±0.04)	0.33 (0.02)
pBS102 (T161G)	TTCATTAAATAATCGTTAATCGCTTGCGTTCTCAT	5.43 (±0.49)	0.54 (0.27)
pBS102 (C162A)	TTCATTAAATAATCGTTAATCGCTTTAGTTCTCAT	4.28 (±0.07)	0.04 (4E-5)
pBS102 (G163T)	TTCATTAAATAATCGTTAATCGCTTTCTTTCTCAT	3.58 (±0.08)	0.01 (2E-6)
pBS102 (T164G)	TTCATTAAATAATCGTTAATCGCTTTCGGTCTCAT	4.08 (±0.24)	0.02 (3E-6)
pBS102 (T165G)	TTCATTAAATAATCGTTAATCGCTTTCGTGCTCAT	4.83 (±0.35)	0.14 (9E-4)
pBS102 (C166A)	TTCATTAAATAATCGTTAATCGCTTTCGTTATCAT	5.64 (±0.34)	0.88 (0.78)
pBS102 (T167G)	TTCATTAAATAATCGTTAATCGCTTTCGTTCGCAT	5.39 (±0.36)	0.49 (0.15)
pBS102 (C168A)	TTCATTAAATAATCGTTAATCGCTTTCGTTCTAAT	5.39 (±0.04)	0.49 (0.09)
pBS102 (162–165)	TTCATTAAATAATCGTTAATCGCTTTATGGCTCAT	3.80 (±0.04)	0.01 (6E-6)

The activity of promoters with single- and multiple-nucleotide substitutions within and flanking conserved positions of the PvdS-box motif was assessed *in vivo* by comparing the relative luminescence produced from wild type (WT = PSPTO_2152) and mutated PvdS-box derivatives. Log RLU shows the average ($n = 6$) of the $\log_{10}$ relative light units and standard deviation. Promoter activity is the ratio of expression from the indicated promoter relative to wild type with the *P*-value shown in parentheses. The –35 and –10 promoter elements are indicated by underscoring and the mutated nucleotides are highlighted.

served regions of the PvdS-box, along with the transcription start mapping, provides extremely strong evidence for the association of promoter activity with this motif.

Construction and use of PvdS-box hidden Markov models

The 15 DC3000 promoter sequences that were experimentally validated (Fig. 3) and detected by Gibbs sampling were manually realigned to optimize conservation of the –10 and –35 elements, and are summarized in the sequence logo shown in Fig. 4A. HMM model A was trained on this realigned set and used to scan the genomes of 10 fully sequenced fluorescent pseudomonads, including human pathogens (*P. aeruginosa*), an insect pathogen (*P. entomophila*), saprophytes (*P. fluorescens*), bioremediation agents (*P. putida*) and plant pathogens (*P. syringae*). The results of the genome scans are summarized in Fig. 5A, where we characterize the number of 'plausible' and 'implausible' matches to model A for each genome at two different HMM score value cut-offs.

Scanning the DC3000 genome identified 27 matches to the motif (16 plausible and 11 implausible) at a HMM score cut-off of 10.0 (Table 2 and Fig. 5A). All 15 sequences in the model A training set were recovered with scores greater than or equal to 11.2. Of these, 11 were located within the pyoverdine cluster, while four mapped to other regions of the genome (Fig. 6). This HMM scan identified one additional PvdS-box that is plausibly positioned to function as a promoter, upstream of PSPTO_3172, which is predicted to encode a NorM-like member of the MATE family of efflux transporters. The 11 implausible motifs were either positioned within genes, were located or oriented in ways that are not typically associated with promoter function based on existing annotation, and generally had lower scores ($\leq$ 11.2). We cannot, however, rule out the possibility that one or more are functional. For example, the PvdS-box embedded in PSPTO_3556 (*glcE*) exhibits modest but significant differential expression (see Fig. 1B) although it is positioned to generate an antisense transcript within this gene, and in fact evidence for transcriptional activity generated from promoters in unconventional orientations has been observed in at least one other system (Eiamphungporn and Helmann, 2008). In general the scanning results in DC3000 suggest that model A provides a reasonably complete description of the PvdS regulon in DC3000.

The other pseudomonads appear to group into two classes based on the tendency for sequences similar to the DC3000 PvdS-box to be located upstream of anno-

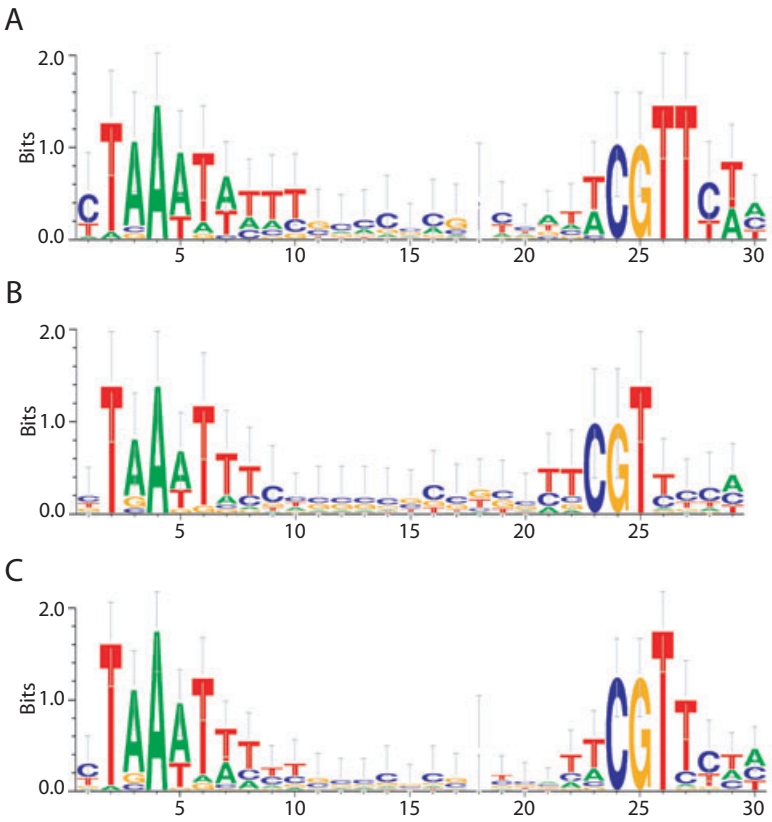

Fig. 4. Comparison of PvdS-regulated promoter motifs. Sequence logos have been scaled to reflect a G/C nucleotide bias of 61%, the approximate average G/C content in the 10 genomes analysed. Model A (A), DC3000-based; model B (B), PAO1-based; model C (C), DC3000 and PAO1 combined. Aligned sequences represented in these sequence logos and used to train HMMs (models A, B and C) are available in Clustal format in Supplementary material at http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2008.06209.x

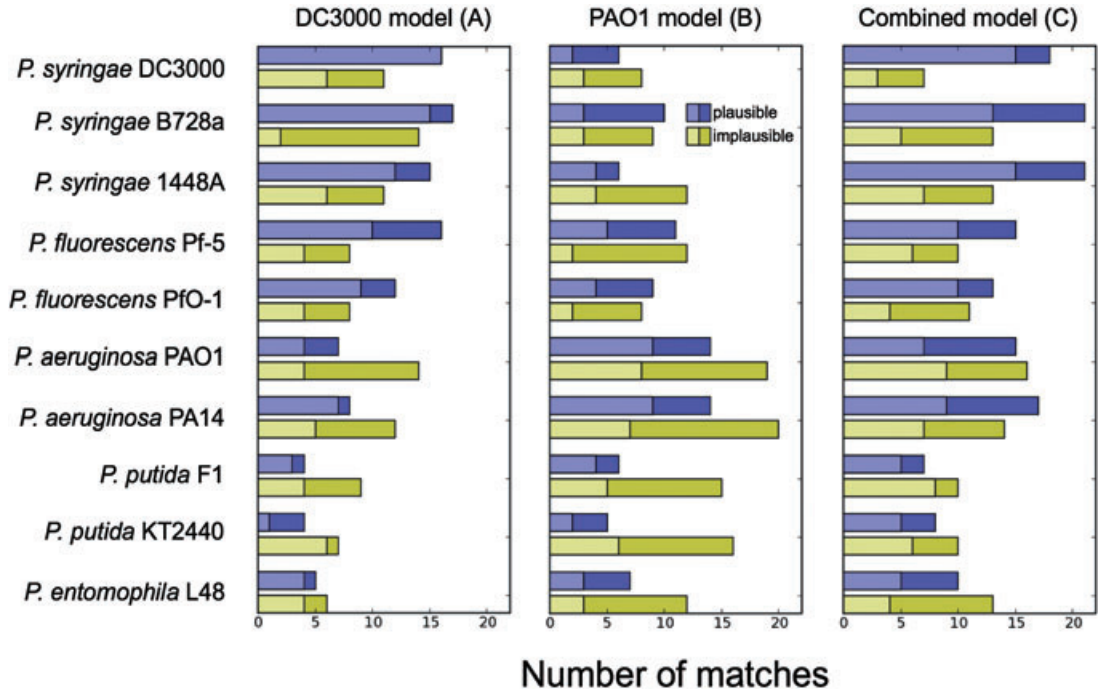

Fig. 5. PvdS-box matches in 10 pseudomonad genomes. HMMs were used to scan genomes at two different cut-off scores (10.0 and 11.0). Matches were considered plausible if they were oriented correctly with respect to a downstream gene, and if their 5′ ends were located in the interval beginning 300 nt upstream of an annotated gene start and 50 nt downstream of this start. We included 50 nt of the N-terminus to account for the possibility of misannotated start codons. Both plausible (blue) and implausible (yellow) matches are shown for each genome. The results from scans using a cut-off of 11.0 are shown using the lighter colours. The more relaxed cut-off (10.0, with darker colours) was included to demonstrate the general degradation of the model as the cut-off was reduced. (A) model A; (B) model B; (C) model C.

Table 2. PvdS-box HMM scan results in DC3000.

Co-ordinates 5′	Co-ordinates 3′	PvdS-box sequence	G	HMM score	Orientation
2306853	2306880	TAATTATTTGCAGACGCCATCCGTTCTC	•	16.9	U/S 2134
2322444	2322471	TAATTTTTCGACGTAGCGATACGTTCAA	•	16.4	U/S 2137
2374232	2374259	TAAAAATTTTCGCCCGCGCTTCGTTTAA	•	16.1	U/S 2161
5573285	5573258	TAATTACTGGCCCTCGCTAATCGTTCTT	•	16.0	U/S 4923
2367751	2367778	TAAATTTGTCAGCCGGTTTTACGTTCTA	•	15.8	U/S 5624
2329292	2329319	TAATTTTTCAGCCGCGAAGTACGTTCAA	•	15.4	U/S 2146
2330476	2330502	TAAATATTTCCCCGCCAATTCGTTCTT	•	15.4	U/S 2147
3717249	3717276	TAAAAACATGCCAGACCGAGACGTTTAT	•	15.2	U/S 3290
2367721	2367694	TAAATTTTTCGCCCATCCGCTCGTTCTC	•	15.0	U/S 2156
4013974	4014001	AAAAGAACGGCAGACGCTGTTCGTTCAA		13.2	E/A 3556
2374203	2374176	TAAATTCCGGCCTGCATTTCTCGTTTAA	•	12.4	U/S 2160
3565679	3565652	AAAATATAGCCACGCGACAGACGATCAT		12.3	U/S 3172
800054	800081	TGAATTAATTCTCCCGTCCGTCGTTCTC	•	12.1	U/S 0753
2329217	2329190	AAAATATTCCTACAGGAATCTCGTTCTC	•	12.1	U/S 2145
3200665	3200692	TAAATATTTCATACTGCACTCCGTTCTT		11.8	I/A 2850
5055286	5055259	TGAAGTTTTCCGAGCACAGTTCGTTTTA		11.6	E/S 4488
2341777	2341805	TAAATCACTGGAGGCACTCAAACGTTTAT	•	11.3	U/S 2149
2362183	2362156	TAAATAATCGTTAATCGCTTTCGTTCTC	•	11.3	U/S 2152
3353644	3353617	TCAAGAAATCAACCACTTGAACGTTTTA	•	11.2	U/S 2982
646696	646723	TCGATATTCCAGCCAGCAAGACGTTCAA		11.1	E/S 0587
1209925	1209952	TAAAAAATGCTCCACAAGATACGATCAC		11.1	I/A 1098
5758588	5758561	TAAAGAAAAGCACACGCTAATCATTCTT		11.1	E/S 5056
225903	225876	TAATTAGCGGAACCACTGTTACGTTTTT		10.8	E/S 0202
3717077	3717050	TAAAAAAAGCGCCCCGAAGGACGCTTTA		10.7	I/A 3289
3244256	3244283	AAAAAACCCGCCGAAGCGGGTCGTTTTT		10.6	I/A 2881
312912	312939	TAAGATTTTGAACGCATTACACGCTCAA		10.3	U/S 0287
6224284	6224257	TGAATTTTGCGCCGAGCGATTCGATCTG		10.3	E/A 5463

All regions of the genome with significant matches (score ≥ 10.0) to the DC3000-based PvdS-box model A are shown with the location indicated by genome co-ordinates and are listed in descending order by HMM score. Shaded entries indicated association with pyoverdine. The –10 and –35 elements are underlined. A dot in the Gibbs column (G) indicates membership in both the Gibbs input and output (see Table S2 and Fig. 3). The HMM score represents the degree of similarity between the genomic sequence and the HMM. The orientation column describes the positioning of the promoter relative to annotated genes: U/S (upstream/sense), upstream and positioned to function as a promoter in the sense orientation for the indicated gene. The remaining three categories describe matches that are unlikely to function as conventional promoters: E/S (embedded/sense), embedded in a gene and oriented to transcribe in the sense direction. I/A (intergenic/antisense), intergenic but oriented to generate an antisense transcript for the indicated gene. E/A (embedded/antisense), embedded in a gene and oriented to generate an antisense transcript.

tated reading frames (see Fig. 5A). In the first class (*P. syringae* B728a, *P. syringae* 1448A, *P. fluorescens* Pf-5 and *P. fluorescens* PfO-1), we find larger numbers of plausible matches and a higher ratio of plausible-to-implausible promoters; between 67% (12/18) and 88% (15/17) of the matches to the DC3000-based model A with scores ≥ 11.0 are located in positions consistent with their function as promoters. In the five remaining genomes that comprise the second class, significantly smaller numbers of plausible matches were identified, and lower plausible-to-implausible ratios, suggesting the possibility that the PvdS-regulated promoters of these organisms are measurably different from those found in DC3000.

To understand the differences in specificity of the DC3000-based model described above, we constructed two additional HMMs using sequences conferring PvdS-dependent expression in PAO1. Model B was constructed using 13 sequences from PAO1 containing IS-boxes as described in Ravel and Cornelis (2003) (see Fig. S1 and *Experimental procedures*). Model C is a consistently aligned combination of the DC3000-based model A and the PAO1-based model B. Sequence logos of the training sets for models A, B and C are shown in Fig. 4. Table 3 lists the plausible promoters resulting from scanning pseudomonad genomes with models A, B and C.

Using the PAO1-based model B to scan the 10 pseudomonad genomes revealed that this model identified plausible promoters in the two *P. aeruginosa* strains, and also identified other plausible promoters associated with pyoverdine in several species not found using the DC3000-based model A. However, Fig. 5 reveals that there are systematic differences in the performance of models A and B, suggesting that neither model is transferable across the genus. This observation provided our motivation for constructing the combined model C. Figure 5C indicates that model C generally performs better across the genus than either A or B separately, identifying greater numbers of plausible promoters while

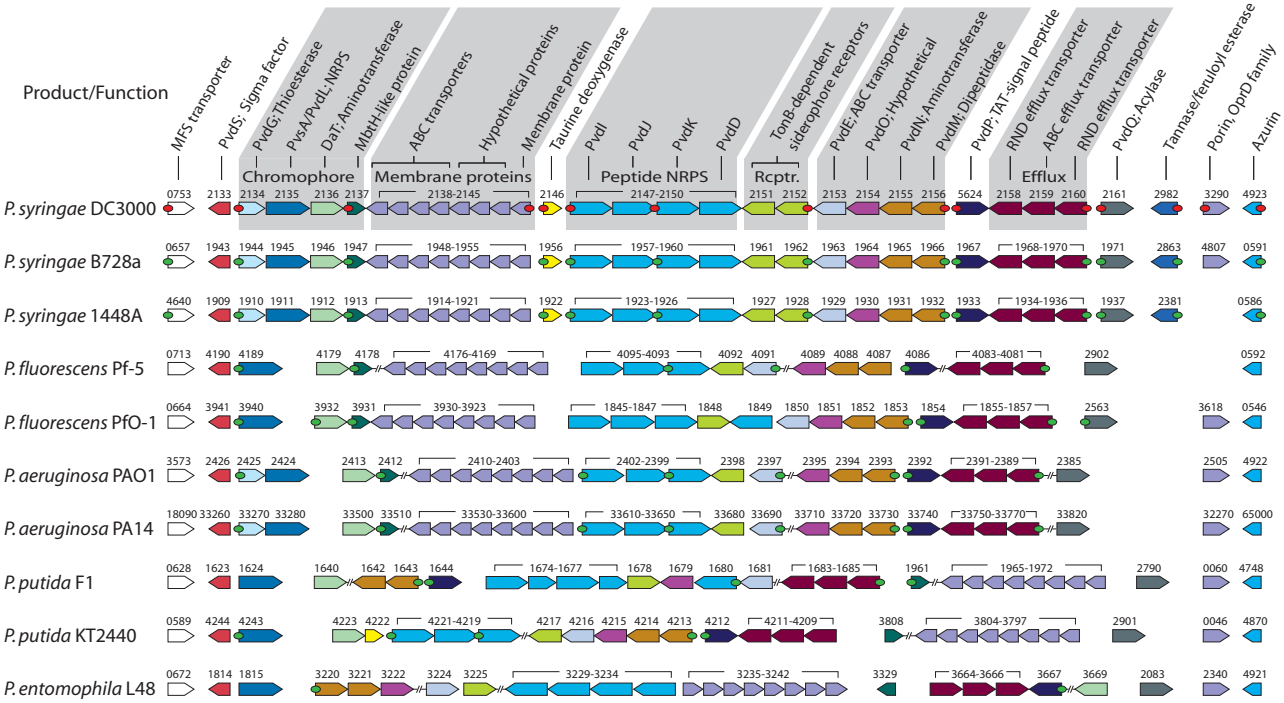

Fig. 6. The DC3000 PvdS regulon with predicted orthologues in nine pseudomonad genomes. Genes in DC3000 downstream of experimentally validated PvdS-boxes (red dots) and predicted operon members (shaded) are shown in their genomic context. The DC3000 pyoverdine cluster is located in a single locus that spans 70.5 kb (from 2133 to 2161). In the other pseudomonads, matches (scores ≥ 10.0) to any of the three models detailed in Table 3 are shown with green dots. Numbers above each gene refer to annotation in the respective genomes.

retaining an acceptable ratio of plausible-to-implausible matches. In addition, model C identifies PvdS-boxes upstream of genes related to pyoverdine metabolism that were missed by the single-species models A and B. Finally, other genes are identified that appear to vary from organism to organism (see Table 3), and the union of results from scans A, B and C provides novel hypotheses regarding species-specific PvdS regulation across the fluorescent pseudomonads.

Discussion

Genome-wide approaches are well suited for investigating the role and scope of individual sigma factors, both because of the essential role in transcription of these regulators and because of their potential for co-ordinating global organismal responses. We used a combination of computational and molecular genetic approaches to identify the DC3000 promoter motif recognized by PvdS, a group IV/ECF sigma factor, members of which regulate transcription in response to specific environmental signals. This work is an important step towards understanding the regulation of pyoverdine metabolism, which makes vital contributions to pseudomonad fitness in environmental and host interactions (Meyer *et al.*, 1996; Poole and McKay, 2003). The results confirm the role of PvdS in the expression of pyoverdine-related genes and suggest that PvdS integrates iron status with other functions beyond pyoverdine metabolism.

These experiments establish two important aspects pertaining to the details of regulation by PvdS. First, we report a *de novo* identification of a PvdS-regulated promoter motif using sequences that exhibit PvdS-dependent activity in DC3000. The observed sequence conservation in the confirmed promoters, along with mutagenesis data, suggests important and distinct roles for the –10 and –35 elements in the interactions of PvdS with cognate promoters. Second, we have developed computational models that can be used to predict regulation by PvdS. We have applied these models both to DC3000, confirming that our inventory of PvdS-dependent promoters in DC3000 is nearly complete, and to other fully sequenced pseudomonads to generate predictions about PvdS-dependent regulation in those organisms. The DC3000 PvdS-box refines and extends the IS-box previously described in *P. aeruginosa* (Rombel *et al.*, 1995; Wilson *et al.*, 2001; Ochsner *et al.*, 2002; Visca *et al.*, 2002) and supports the prediction that pyoverdine genes are regulated by similar mechanisms among the fluorescent pseudomonads (Rombel *et al.*, 1995; Ravel and Cornelis, 2003).

Table 3. Genes downstream of HMM matches in DC3000 and nine other sequenced pseudomonad genomes.

Co-ordinates 5′	3′	Sequence	HMM A	HMM B	HMM C	ORF	Product
P. syringae pv. *tomato* DC3000		NC_004578					
2306853	2306880	TAATTATTTGCAGACGCCATCCGTTCTC	16.9		14.6	PSPTO_2134	PVD chromophore thioesterase
2322444	2322471	TAATTTTTCGACGTAGCGATACGTTCAA	16.4		15.3	PSPTO_2137	MbtH-like protein
2374232	2374259	TAAAAATTTTCGCCCGCGCTTCGTTTAA	16.1		15.5	PSPTO_2161	PvdQ; acylase
5573285	5573258	TAATTACTGGCCCTCGCTAATCGTTCTT	16.0		14.1	PSPTO_4923	Azurin
2367751	2367778	TAAATTTGTCAGCCGGTTTTACGTTCTA	15.8	12.5	16.4	PSPTO_5624	PvdP: TAT signal
2329292	2329319	TAATTTTTCAGCCGCGAAGTACGTTCAA	15.4	10.6	15.4	PSPTO_2146	Taurine deoxygenase
2330476	2330502	TAAATATTTCCCCGCCAATTCGTTCTT	15.4		14.2	PSPTO_2147	PVD peptide NRPS
3717249	3717276	TAAAAACATGCCAGACCGAGACGTTTAT	15.2		11.3	PSPTO_3290	Porin; OprD family
2367721	2367694	TAAATTTTTCGCCCATCCGCTCGTTCTC	15.0		15.3	PSPTO_2156	Dipeptidase
2374203	2374176	TAAATTCCGGCCTGCATTTCTCGTTTAA	12.4	11.8	14.2	PSPTO_2160	RND efflux transporter, porin
3565679	3565652	AAAATATAGCCACGCGACAGACGATCAT	12.3			PSPTO_3172	Efflux transporter
800054	800081	TGAATTAATTCTCCCGTCCGTCGTTCTC	12.1		10.7	PSPTO_0753	Drug resistance transporter
2329217	2329190	AAAATATTCCTACAGGAATCTCGTTCTC	12.1			PSPTO_2145	Membrane protein
2362183	2362156	TAAATAATCGTTAATCGCTTTCGTTCTC	11.3		11.1	PSPTO_2152	TonB-dependent receptor
2341777	2341805	TAAATCACTGGAGGCACTCAAACGTTTAT	11.3			PSPTO_2149	PVD peptide NRPS
3353644	3353617	TCAAGAAATCAACCACTTGAACGTTTTA	11.2			PSPTO_2982	Tannase/feruloyl esterase
4725717	4725744	TGAATTCGACTGTCGGTGATTCGTTTCA		10.8		PSPTO_4196	Glucose dehydrogenase
4378372	4378345	TGAGTTTCAGGCGGCCAGTTACGTTCAT		10.7	10.8	PSPTO_3864	PsyI; *N*-acylhomoserine lactone synthase
4131225	4131252	TAAAGCCTACCGCGCCGCCATCGTTCAC		10.0	11.3	PSPTO_3663	Ganine deaminase
4357542	4357569	TAAATGCCTGCGCCAGAACATCGTTTCA			12.3	PSPTO_3849	Rhs element Vgr protein
1925512	1925539	AAAAATTCAGGCCGAGTTGATCGTTCCC			11.1	PSPTO_1754	UDP-glucose 4-epimerase
693151	693178	TAAAGTTCTGCTTGGCCCGCACGTTTCC			11.0	PSPTO_0628	50S ribosomal protein L23
2394871	2394898	TAAATTTTTTGCCGTATTCAGCGTTATA			10.5	PSPTO_2175	LeuB; 3-isopropylmalate dehydrogenase
P. syringae pv. *syringae* B728a		NC_007005					
2216773	2216800	TAATTATTTGCAGACGCCATCCGTTCTC	16.9		14.5	Psyr_1944	PVD chromophore thioesterase
2232356	2232383	TAATTTTTCGACGTAGTGATACGTTCAA	16.4		15.6	Psyr_1947	MbtH-like protein
2239795	2239822	TAATTTTTCAGCGACGCAGTACGTTCAA	14.8		14.4	Psyr_1956	Taurine deoxygenase
3435473	3435446	TGAATAATCGCCAGGCTTAAACGTTTTA	14.8		13.6	Psyr_2863	Tannase/feruloyl esterase
2285734	2285707	TAAATTCCTGCCGGCACGCCTCGTTTAA	14.5	14.0	17.0	Psyr_1970	RND efflux transporter, porin
2279024	2278997	TAAATTTATCGGCGATCCGCTCGTTCTC	13.9	10.6	15.2	Psyr_1966	Dipeptidase
2241308	2241334	TAAATATTTCCTCGCCAATTCGTTCTT	13.9		12.4	Psyr_1957	PVD peptide NRPS
679878	679905	TAATTACCGGGTCCCGTGAATCGTTCTT	13.8	10.4	14.4	Psyr_0591	Azurin
2279054	2279081	TAAATTTGTCTGCCGGGTTTACGTTCTA	13.7	11.6	15.1	Psyr_1967	PvdP: TAT signal
744682	744709	TGAATTAATTCCCCCGTTCGTCGTTCTC	13.0		11.6	Psyr_0657	Drug resistance transporter
2252732	2252760	TAAATCACTGGAGGCACTCAAACGTTTAT	11.4			Psyr_1959	PVD peptide NRPS
849265	849238	GAAATCCTTCAAGACCCAAGTCGTTCAA	11.3			Psyr_0746	IS66 Orf2 like
3187125	3187098	TAAATACTGGGGCCAGGTGCTCTTTCTA	11.2		11.2	Psyr_2626	Hypothetical protein
2605313	2605340	TAAAGCTTTGGTAGCGACAGACGTTAAC	11.1			Psyr_2241	Chemotaxis sensory transducer
5703584	5703611	TAAAAAAAGCGCCCCGAAGGACGCTTAA	11.1			Psyr_4808	Hypothetical protein
1648899	1648926	TAAATTCAGAGCCGTGGCGCTCGTTTTT	10.3	11.8	13.3	Psyr_1462	GCN5-related *N*-acetyltransferase
1417924	1417951	TCAATAACTGGCTGGATGCGTCGTTCAC	10.1		10.9	Psyr_1261	GuaB
5017688	5017661	TGAATTTGCCCGACGGCTGCGCGTTTCA		10.7	10.0	Psyr_4211	ABC transporter
1085896	1085923	TGAAGTTCTGGCGGGCGGGGTCGATGCA		10.5		Psyr_0950	HemK; methyl transferase
4713662	4713689	TAAAAAACAGGCGCCCGGATGCGTCTGA		10.1		Psyr_3966	Membrane transport
2086497	2086470	TAAAGCCTACCGCGCCGCCATCGTTCAC		10.0	11.3	Psyr_1812	Guanine deaminase
1012883	1012856	TGAAGTCCTGCCTGCGCCTTGCGTTTCA		10.0	10.5	Psyr_0891	ChrR; transcription activator
5401838	5401811	TAAAGTTCTGCTTGGCCCGCACGTTTCC			10.9	Psyr_4546	50S ribosomal protein L23
2285763	2285790	TAAACATCAGCGTCCGTGTTCCGTTTAA			10.3	Psyr_1971	PvdQ; acylase
2489035	2489008	TAAATTGTCGGCCCGACGGATCGTTTCG			10.3	Psyr_2129	MFS transporter
4025673	4025700	TGATTGTCGAAGGCCGCGGAACGTTCAA			10.2	Psyr_3367	Iron permease FTR1
2273489	2273462	TAAATAAACGTTATTCGCTTTCGTTCTC			10.1	Psyr_1962	TonB-dependent receptor
P. syringae pv. *phaseolicola* 1448A		NC_005773					
2215579	2215606	TAATTATTTGCAGACGCCATCCGTTCTC	17.0		14.6	PSPPH_1910	PVD chromophore thioesterase
2231162	2231189	TAATTTTTCGACGTAGCGATACGTTCAA	16.4		15.3	PSPPH_1913	MbtH-like protein
2239548	2239574	TAAATATTTCCCCGCCAATTCGTTCTT	15.4		14.3	PSPPH_1923	PVD peptide NRPS
2276764	2276737	TAAATTTTTCGCCCATCCGCTCGTTCTC	15.1		15.4	PSPPH_1932	Dipeptidase
2762672	2762699	TGAATAATTGCCTCGCTTAAACGTTTTA	14.2		12.6	PSPPH_2381	Tannase/feruloyl esterase
2238138	2238165	TAATTTTCTCGTGGCGAAGTACGTTCAA	14.0	10.7	14.6	PSPPH_1922	Taurine deoxygenase
2276794	2276821	TAAATTTGTCTGCCGGGTTTACGTTCTA	13.7	11.6	15.1	PSPPH_1933	PvdP; TAT signal
693174	693201	TAATTACCTGCTCCCGTGAATCGTTATT	12.7		12.9	PSPPH_0586	Azurin
1689357	1689330	TAAATTCATCCACACTGGATACGATCAA	11.7		11.0	PSPPH_1448	Hypothetical protein
5279635	5279608	TGAATTAATTACCCTGTCCTTCGTTCTC	11.6		12.0	PSPPH_4640	Drug resistance transporter
2283545	2283518	TAAATTCCCGTCCTCATCCCTCGTTTAA	11.3	12.0	14.0	PSPPH_1936	RND efflux transporter, porin
2250912	2250940	TAAATCACTGGAGGCACTCAAACGTTTAT	11.2			PSPPH_1925	PVD peptide NRPS
2300785	2300812	TAAATTTTTCGCCGCTGCGAGCGTTATA	10.3	13.0	13.9	PSPPH_1954	LeuB; 3-isopropylmalate dehydrogenase
1850987	1851014	GAAATTCCAGCGGGCGCAAGTCGTTCAT	10.2		10.6	PSPPH_1597	ISPsy18, transposase
1543459	1543486	TCAATAACTGGCTGGATGCGTCGTTCAC	10.0		10.8	PSPPH_1333	GuaB
4485522	4485549	TGAATTCGACCGTCGGTGATTCGTTTCA		12.5	11.1	PSPPH_3927	Glucose dehydrogenase
1191967	1191994	TGAAGTTCTGGCGGGCGGGGTCGATGCA		10.1		PSPPH_0998	HemK; methyl transferase
2283574	2283600	TAAATTTTGCGTCTGCGCTTCGTTTAA			11.3	PSPPH_1937	PvdQ; acylase
5225800	5225773	TAAAGTTCTGCTTGGCCCGCACGTTTCC			11.0	PSPPH_4590	50S ribosomal protein L23
593310	593337	TGATTACCCGTGCCCGCCAGACGTTCCA			10.9	PSPPH_0509	GlnE
2201685	2201658	TCAATGTCTGCGGGGGTATCTCGTTTTT			10.4	PSPPH_1897	Integrase
2271222	2271195	TAAATAAACGTTATTCGCTTTCGTTCTC			10.2	PSPPH_1928	TonB-dependent receptor
184783	184810	TAAATCCCCGGCATACGGGACCGTTCAT			10.0	PSPPH_0157	Hypothetical protein
P. fluorescens Pf-5		NC_004129					
4876034	4876007	TAATTATTTCAAGACGTCATCCGTTCTC	15.1		13.2	PFL_4189	PvdL; PVD chromophore NRPS
4589567	4589540	TAAATTCCTCACGGGCTGTTACGTTTTA	13.8	13.5	16.2	PFL_3974	Hypothetical protein
4738601	4738573	TAAATCTTTGGAGGCCCTCAAACGTTTAT	13.6		11.4	PFL_4093	PVD peptide NRPS
1032891	1032918	TAAATTGAGCATAGAGGAAAACGTTTTC	12.4		10.7	PFL_0889	Auxin Efflux Carrier

Table 3. *cont.*

Co-ordinates 5′	Co-ordinates 3′	Sequence	HMM A	HMM B	HMM C	ORF	Product
4715783	4715756	TAAATTTCCCTGCAAATCCAACGTTCTA	12.3		13.2	PFL_4086	PvdP; TAT signal
4708544	4708571	TAAATTCCCCCCCCGGCATCTCGTTTAC	12.1		13.0	PFL_4081	RND efflux transporter
5186603	5186576	AAAAAAACGCCCTGACCGAGTCGTTTTA	11.3			PFL_4475	DNA-binding response regulator
2947943	2947916	GCAATATTCGCAAATGTAAAACGTTCTC	11.2			PFL_2665	PfeR: transcriptional activator
4851182	4851155	TAATTTTTTGCCCCGCCGTTTCGTTCCT	11.1	11.0	14.3	PFL_4178	MbtH-like protein
5895395	5895368	TAAATTAATGGACCGGCGTCTCGTTACT	11.0	10.1	13.5	PFL_5130	Hypothetical protein
4485923	4485896	TAAAAAGGTTGACGCCGCGGACGTTTAT	10.9		10.4	PFL_3886	ClpS; protease
4720593	4720620	TAAATACTGGCGTCGGCGCTTCGTTTGT	10.6		12.3	PFL_4091	PvdE; PVD ABC export
4894277	4894250	AAAATAAATTCCGACGACAGACTTTTTC	10.4			PFL_4207	Lipid kinase
5934877	5934904	TAAATTTTTTAACAGCCGGTATGTTAAA	10.2			PFL_5174	ABC transporter
5168623	5168651	AAAAGAACTGCCGTCGCCCATTCGTTTTT	10.1			PFL_4460	Ais protein
5857039	5857067	TAAATAAATTTTTAAGCAGGAACGTTCAT	10.0			PFL_5096	Aminotransferase
4877230	4877203	TAAATTTCCGGCGGCGGCCCACGTCACT		14.4	12.5	PFL_4191	PvdY; acetylase
6754482	6754455	TAAATTGCCGGTGGCCTTGTCCGTACCC		11.8		PFL_5941	*N*-acetylglutamate synthase
4720560	4720533	TAAATTGCCCTGCCTGTGGTTCGTTCTT		11.7	12.5	PFL_4090	PvdF; transformylase
6723328	6723355	TAAATTCAAGGAGTGACCATGCGTTTCC		10.4	10.9	PFL_5913	ABC transporter
4300935	4300908	TCAAGATCCCCGCGTCTGCTGCGTCGCA		10.4		PFL_3705	Dyp-type peroxidase family protein
573556	573583	TAAATGCGCGGCCCGCCTGATCGTCTCC		10.2		PFL_0496	Proline permease
3849065	3849038	TCAATAGGCAAGCGGGAGCTTCGTCGAA		10.2		PFL_3330	Sensory box protein
3486863	3486890	TAATTAGCCCGCCGCCGCTTTCGTCTGT		10.1		PFL_3024	SCO1/SenC family protein
6360907	6360880	TAAAGTTCTGCTTGGCCCGCACGTTTCC			10.7	PFL_5580	50S ribosomal protein L23
4924976	4925003	TAATTTTGCGCTGCGCTGTAACGTTTCC			10.1	PFL_4236	Activation/secretion; TPS family
P. fluorescens PfO-1		NC_007492					
4445728	4445701	TAAATTTTTCGCCGCGTCGAACGTTCTT	18.8	13.0	18.7	Pfl_3931	MbtH-like protein
2705993	2705966	TAAATAACTTAAAGAAATAATCGTTCTA	15.3		12.9	Pfl_2344	Lipoprotein
4467547	4467520	TAATTATTTCAAGACGTCATCCGTTCTC	15.2		13.3	Pfl_3940	PvdL; PVD chromophore NRPS
2110662	2110689	TAAATTTGTCGGGGCAGCGTACGTTCTA	14.7	15.7	17.6	Pfl_1854	PvdP: TAT signal
2117054	2117027	TAAATTCCCCACCCCGTCTCTCGTTTAA	14.0	12.9	16.1	Pfl_1857	RND efflux transporter
3780368	3780395	TAAATAAATGTCGACGCGATGCGTTATC	13.7		12.2	Pfl_3297	Hypothetical protein
6036707	6036680	TAAATATTCGAACCCCGGCGACGTTTGC	12.3		13.1	Pfl_5373	Thymidylate synthase
4957069	4957096	AAAACAATTGAACAAGCTAATCGTATAA	11.3			Pfl_4395	Cold-shock DNA-binding protein
2110632	2110605	TAAATTTCGTTGCCAATCACTCGTTCTT	11.1		11.5	Pfl_1853	Dipeptidase
2935846	2935873	TAAATTTTTTTTCCACGGCTTCGTTTTC	10.8		11.6	Pfl_2563	PvdQ, acylase
3198120	3198093	TAAATAAACTACAAAGCGATCCGGTCAT	10.7			Pfl_2772	TonB-dependent receptor
4525835	4525862	AAAATTTTTCATCTCGTTTTACATTCAA	10.5			Pfl_3997	OsmC-like protein
4469455	4469428	TAAATTTCCCGCCCCGGCTCACGTCCCA		14.5	14.8	Pfl_3942	PvdY; acetylase
5536599	5536626	TGAATTCCCGGCGCTCGCGCTCGTCAAT		10.6		Pfl_4913	Metallopeptidase
4447264	4447237	TAAATCCTCGACACGCTGGTGCGTTTCC		10.4	10.8	Pfl_3932	DaT; aminotransferase
1065742	1065768	TAAATCTCCGGGGCATGCCTCGTTTTA		10.3		Pfl_0915	Dehydrogenase
509936	509909	TGATTTTCCTGGCGGCTCGACCGTTCTT		10.1		Pfl_0447	Methyl-accepting chemotaxis protein
518914	518941	TAAATGCGCGGCCGTGCGATTCGTCTGT		10.1		Pfl_0453	Proline permease
5703984	5703957	TAAAGTTCTGCTTGGCCCGCACGTTTCC			10.8	Pfl_5077	50S ribosomal protein L23
4752445	4752472	TCAAGTAATGCGCAGGTGGCTCGTTCCT			10.3	Pfl_4211	Histidine kinase
P. aeruginosa PAO1		NC_002516					
2648478	2648451	TAAATTTGCCGACGGAAGGAACGTTCTA	16.5	14.7	17.5	PA2392	PvdP; TAT signal
1857013	1857040	TAAAGAAGTGCAAGTCTCGTTCGTTTTC	13.6	10.6	13.9	PA1713	ExsA; transcriptional regulator
4593899	4593872	GAAAAATTTGGAGCAGAGATTGGTTCGA	13.0			PA4109	AmpR; transcriptional regulator
4433361	4433388	GAAATATTGGAGCACCCTGCACGATCTC	11.1			PA3953	Isochorismatase hydrolase
2591870	2591843	TAGAAAAATCAAAGAGTTGGAGGTTTTC	10.7			PA2345	Transmembrane protein
5486200	5486226	AAAATATTTCAGTGAACGATCGTTCAC	10.2			PA4890	Transcriptional regulator
5778096	5778069	CAAATTTAGTGGGAACATTGTCGTTCAA	10.0			PA5130	Rhodanese-like domain protein
2653190	2653163	TAAATTGCAGGCGATGCCGTTCGTTGCA		15.4	13.8	PA2396	PvdF; transformylase
2648528	2648555	TAAATTTTCCCGCTCCGGCCTCGTCCCA		14.5	13.1	PA2393	PvdM; dipeptidase
2687257	2687230	TAATTTTCACGATGTGTCGTCCGTTTCA		13.2	10.6	PA2402	PvdI; peptide NRPS
1871089	1871116	TGAATGTCTCGACAGGAGGTGCGTCTCA		12.8	10.4	PA1728	Hypothetical protein
2665166	2665139	TGATTTCCTAAGAGGCGGTAGCGTGCAA		12.7		PA2399	PvdD; peptide NRPS
2687227	2687200	TGAATGACCGAGCCCCGCGCTCGTTGCC		12.4	10.6	PA2402	PvdI; peptide NRPS
2653223	2653250	TAAATACCGGGCATCCTGCTTCGTCTGT		12.3	11.0	PA2397	PvdE; ABC transporter
2723276	2723249	TAAATTCGCGGCGGGATGCGACGTTACT		11.5	11.6	PA2427	PvdY; acetylase
5843906	5843879	TAAGTAGCTGCTGTGGGGCTGCGTCGGA		10.9		PA5190	Nitroreductase
2694844	2694817	TAATTTTCCCGCCGGGCTTTTCGTTATC		10.6	12.0	PA2412	MbtH-like protein
2642039	2642066	TAAATTTAGCCGCCCTGGCCTCGTATAT		10.6	10.8	PA2389	RND efflux transporter
4253623	4253596	GAATGTTCAGCCGGCGCGCAGCGTCCTA		10.4		PA3794	Transcriptional regulator
1243321	1243348	TGAATTTTGAAAGTGGGAGTTCGTTAAT			10.9	PA1150	Pyocin S2
2721687	2721660	TAATTATTTGCCGTTGTTATCCGTTCCC			10.7	PA2425	PvdG; thioesterase
4921090	4921063	TGAACAGTTGGGCGCGTCTTTCGTTCCT			10.3	PA4390	Hypothetical protein
2665216	2665188	TAAATCCCTGGAGGCGCTCAAACGTCTAT			10.0	PA2399	PvdD; peptide NRPS
P. aeruginosa PA14		NC_008463					
2995315	2995342	TAAATTTGCCGACGGAAGGAACGTTCTA	16.3	14.6	17.4	PA14_33740	PvdP: TAT signal
2479149	2479176	TAAATTATTGCCAAACATGCTCGTTCGA	14.6		14.0	PA14_28800	Hypothetical protein
5774350	5774323	TAAAAAACTTCAGGAATAAATCGCTCAA	13.7			PA14_64790	MFS transporter
3777786	3777759	TAAAGAAGTGCAAGTCTCGTTCGTTGTC	11.8	10.9	13.2	PA14_42390	ExsA; transcriptional regulator
933954	933981	GAAAAATTTGGAGTAGAGATTGGTTCGA	11.7			PA14_10800	AmpR; transcriptional regulator
4362208	4362181	TAAATAGGTTGTAGAAAGGATCGTTTTT	11.2		10.2	PA14_49100	Glutathione S-transferase
1091645	1091618	GAAATATTGGAGCACCCTGCACGATCTC	11.0			PA14_12730	Isochorismate hydrolase
2026752	2026779	TAAAGACAGGGTGTCAGAATACGTTTCA	10.4		11.5	PA14_23360	O-antigen chain length regulator
2990603	2990630	TAAATTGCAGGCGATGCCGTTCGTTGCA		15.3	13.8	PA14_33700	PvdF; transformylase
2995265	2995238	TAAATTTTCCCGCTCCGGCCTCGTCCCA		14.5	13.1	PA14_33730	PvdM; dipeptidase
2956536	2956563	TAATTTTCACGATGTGTCGTCCGTTTCA		13.2	10.6	PA14_33610	PvdI; peptide NRPS
3763716	3763689	TGAATGTCTCGACAGGAGGTGCGTCTCA		12.7	10.3	PA14_42200	Hypothetical protein
2978627	2978654	TGATTTCCTAAGAGGCGGTAGCGTGCAA		12.6		PA14_33650	PvdD; peptide NRPS
2956566	2956593	TGAATGACCGAGCCCCGCGCTCGTTGCC		12.4	10.6	PA14_33610	PvdI; peptide NRPS

Table 3. *cont.*

Co-ordinates 5′	Co-ordinates 3′	Sequence	HMM A	HMM B	HMM C	ORF	Product
2990570	2990543	TAAATACCGGGCATCCTGCTTCGTCTGT		12.2	11.0	PA14_33690	PvdE; ABC transporter
2909516	2909543	TAAATTCGCGGCGGGATGCGACGTTACT		11.4	11.5	PA14_33250	PvdY; acetylase
2948949	2948976	TAATTTTCCCGCCGGGCTTTTCGTTATC		10.6	12.0	PA14_33510	MbtH-like protein
3001754	3001727	TAAATTTAGCCGCCCTGGCCTCGTATAT		10.5	10.8	PA14_33770	RND efflux transporter
1272858	1272885	GAATGTTCAGCCGGCGCGCAGCGTCCTA		10.4		PA14_15000	Transcriptional regulator
1313040	1313067	TAAATTGCGGCGTGCTGCACGCGACACA		10.4		PA14_15520	TrbJ; conjugation protein
2911105	2911132	TAATTATTTGCCGTTGTTATCCGTTCCC			10.8	PA14_33270	PvdG; thioesterase
4405751	4405724	TGAATTTTGAAAGTGGGAGTTCGTTAAT			10.8	PA14_49520	Pyocin S-type
5084645	5084618	TGAACAGTTGGGCGCGTCTTTCGTTCCT			10.3	PA14_57060	Hypothetical protein
P. putida F1		NC_009512					
1929264	1929237	TAAATTTTGCGGGCGGGTTCTCGTTCAA	13.6	13.8	16.1	Pput_1685	RND efflux transporter
1848958	1848985	TAAATTTCAGTGCAGGTACAACGTTCTA	12.6	11.1	14.7	Pput_1644	PvdP; TAT signal
1921760	1921733	TAAATTTGCCGGGCGGCGGCTCGTTTAC	12.1	13.8	15.3	Pput_1680	PVD peptide NRPS
2480320	2480347	TAAAACGATGGCAGTGCTTATCGTTCAT	10.4			Pput_2186	RND efflux transporter
2241240	2241213	TAAATTCCCCCTCGGCTGGCTCGTTCCA		15.1	14.9	Pput_1973	PvdA; L-ornithine hydroxylase
821479	821506	CAAATGCCTGGCGCTGGGTTTCGTCGCA		10.3		Pput_0734	MFS transporter
1848938	1848911	TAAATTTCCTCCGTGTCGACTCGTTCAC		10.0	11.9	Pput_1643	Dipeptidase
2230663	2230690	TAATTTTTACAACGCTTGCTACGTCTAC			10.4	Pput_1961	MbtH-like protein
1924616	1924589	TAAACATCGGCAGGTGCGGTTCGTCTTT			10.1	Pput_1682	PvdY; acetylase
P. putida KT2440		NC_002947					
4759862	4759835	TAAATTTTCTTGCCAGCTCAACGTTCTA	14.0		12.8	PP_4212	PvdP; TAT signal
4793920	4793893	TAATTCGCCCTCCCTCCGCTCGTTCTA	10.7		12.2	PP_4221	PVD peptide NRPS
4073526	4073499	TAAAACGATGGCAGTGCTTATCGTTCAT	10.3			PP_3585	RND efflux transporter
2182167	2182140	TAATTTTTGGGTGCAGGAAGTCGTTCCC	10.0		11.0	PP_1935	Transcriptional regulator
4325423	4325450	TAAATTCCCCCTCGGCTGGCTCGTTCCA		15.1	14.9	PP_3796	PvdA; L-ornithine hydroxylase
4759900	4759927	TAAATTTCTGCCGCGTGCACTCGTTCCC		14.5	14.8	PP_4213	PvdM; dipeptidase
4335850	4335823	TAATTTTTACAGCGCTTGCTACGTCTAC		10.2	10.7	PP_3809	Hypothetical protein
814215	814242	CAAATGCCTGGCGCTGGGTTTCGTCGCA		10.2		PP_0701	MFS transporter
1911046	1911073	TAAATTGCAGAGTTTGGACATCGTACAA		10.2		PP_1711	Transcriptional regulator
4779487	4779459	TAAATCCTTGGCCGCCTTGAAACGTCTTA			10.5	PP_4219	PVD peptide NRPS
4831302	4831275	TAATTTCTTCCCCTGCCCATCCGTTCCC			10.5	PP_4243	PvdL; chromophore NRPS
P. entomophila L48		NC_008027					
3446461	3446488	TAAATTTGCCCCGCCGCCATTCGTCCAC	12.2	12.1	14.6	PSEEN3220	Dipeptidase
2589825	2589852	TAATGATTCGCATCTGTGAATCGTTCTC	12.0		11.2	PSEEN2482	TonB-dependent receptor
3942523	3942496	TGAAATTTTCCCGCCCCGGAACGTTTTT	11.6		11.1	PSEEN3658	Electron transfer flavoprotein
971446	971419	GAAAAAATTAAGCTGCTGAATCGTTTAA	11.5		10.4	PSEEN0931	FruR; transcription repressor
5597091	5597064	TAAATATTCCAAAAGGCTATAACTTTAA	10.3			PSEEN5280	TonB-like protein
1890002	1890029	TAAATTCCACGGGGGTGGCCACGTATTC		12.2		PSEEN1813	PvdY; acetylase
3953874	3953847	TAAATTTTCAGCCCCAGCCTACGTCCTA		12.0	14.0	PSEEN3667	PvdP; TAT signal
2284551	2284578	TAAAGTTGTTGCCCGGTGATTCGTGTCT		10.5	10.3	PSEEN2176	Activation/secretion protein, TPS family
3173660	3173633	TAAAGCCCTGGGGCCTGGGTGCGTGCGC		10.5		PSEEN3007	MFS Transporter
982336	982309	AAAATATCAGCCCGGCTGTTACGTCTCT		10.4	10.9	PSEEN0940	Porin; OprD family
615140	615167	AAATTGCCCGGGCCTCTCGTTCGTCGCA		10.0		PSEEN0603	AlgI-like protein
3450106	3450133	TAAATTGTCCCACCCGCTCCTCGTCCTA			11.8	PSEEN3223	PvdA; L-ornithine hydroxylase
1632514	1632541	TAAAGAACTGAAACCCGGTACCGTTCTG			10.0	PSEEN1562	Elongation factor P
2463786	2463759	TAAATTTTCCTCCCCTCGCCTCGTCCAC			10.0	PSEEN2373	Transcription regulator

Models were constructed from training sets based on DC3000 (model A); PAO1 (model B); DC3000 and PAO1 combined (model C). Models were calibrated for each organism and used to identify matches with HMM scores greater than or equal to 10.0. Matches were evaluated to identify those that were plausibly positioned to function as promoters (see Fig. 5). Matches are listed in descending order based on HMM score. Pyoverdine-related genes are shaded in grey. Full results of HMM scans for all genomes considered are available in standard GFF format at http://www.pseudomonas-syringae.org.

Functional characterization of the PvdS-box motif

The PvdS-box sequence logo and alignment data in Fig. 3 show the positions of conserved nucleotides and suggest features of the model that are most likely to be responsible for identifying PvdS-regulated genes. Similarly, the degree of conservation in the motif also provides an indication of which nucleotides are likely to be most important for PvdS-holoenzyme promoter-binding interactions.

The most prominent features of the DC3000 PvdS-box are the invariant CGTT nucleotides in the –10 region and the A/T-rich –35 domain. The observation of the strict requirement for the CGTT nucleotides in the –10 region suggests that this is possibly the primary discriminating feature for PvdS binding. This hypothesis is supported by the mutagenesis experiments, which showed that changes to any of these positions severely reduced transcription (see Table 1). In contrast, A/T nucleotides were found at most locations in the –35 region. The lack of a strict nucleotide requirement at most positions in the –35 element suggests that the PvdS interactions here are more tolerant of base substitutions. This hypothesis is consistent with the mutagenesis experiments where only modest effects on transcription were observed with point mutations in the region (maximum approximately three-fold; Table 1). Mutational analysis of the DNA-binding determinants within the *P. aeruginosa* PvdS protein (Wilson and Lamont, 2006) mirrored our analysis of the promoter DNA. Using alanine substitutions they showed that disruption of any amino acids important for contacts

with the –10 element eliminated normal DNA binding and promoter activity; furthermore, mutations that disrupt interactions with the –35 element had smaller effects. To emphasize the validity of this comparison the DC3000 and *P. aeruginosa* PvdS proteins are 84% identical overall and all amino acids that Wilson and Lamont analysed for effects on DNA binding and transcription are conserved.

Collectively, these data suggest a revised concept of the relative contribution of each motif in PvdS promoter selection. Studies of PvdS regulation in *P. aeruginosa* have focused on the IS-box (TAAAT/–35 element) as the hallmark of PvdS-regulated genes (Rombel *et al*., 1995; Wilson *et al*., 2001; Ochsner *et al*., 2002). Our results show that both the –10 and –35 elements are important for PvdS-dependent promoter activity, but it is the –10 element that is likely to contribute the primary sequence-specific contacts for promoter binding.

Localization of the primary specificity determinants in the –10 element may be a general property of the group IV sigma factors (Lonetto *et al*., 1994). Other sigma factors tolerate degeneration in the –35 region, including FecI, which is an *E. coli* group IV sigma factor that also regulates genes involved with iron uptake (Enz *et al*., 2003). A mechanistic explanation for these observations is provided by the recent structural determination of the –35 DNA-binding domain from the *E. coli* group IV sigma factor σ^E in complex with its cognate promoter element (Lane and Darst, 2006). This showed that the structure conferred by the AT tract in the –35 element is more important for binding than base-specific contacts.

The DC3000 PvdS regulon

Characterizing the promoter motif that confers PvdS regulation was essential to our strategy for identifying PvdS-regulated genes. In general, we found that the experimentally confirmed PvdS-boxes were positioned less than 100 bp from the annotated start codon of the closest downstream genes (Fig. 3), consistent with other σ^{70}-type promoters (Shultzaberger *et al*., 2007). In five cases the region between the promoter and the start codon was greater than 100 bp. These genes may be preceded by a long 5′ untranslated sequence that could be involved with other types of regulation; alternatively, the translation start codon may be annotated incorrectly or the 'untranslated' region may contain unannotated genes. Scans identified one additional PvdS-box motif in DC3000 (upstream of PSPTO_3172) beyond those already experimentally confirmed that could plausibly function as a promoter (Table 2). This implies that the promoter trap screen identified nearly all (13/15; Table S1) of the PvdS-dependent promoters in DC3000.

The genes that are downstream of the experimentally confirmed PvdS-box motifs can be classified into two general categories. As expected, one category contains genes that are involved in pyoverdine metabolism and accounts for 75% (11/15) of the PvdS-dependent promoters that we identified. These are positioned within a 70.5 kb region spanning from PSPTO_2133 to PSPTO_2161 known as the pyoverdine cluster (shaded entries in Tables 2 and 3). PSPTO_2136 and PSPTO_2151 are not directly downstream of a PvdS-box but are PvdS-regulated (Fig. 1B). Based on the experimental and computational evidence presented here, along with the proposed operon structures as shown in Fig. 6, we predict that the entire pyoverdine cluster is under the direct control of PvdS except for *pvdS* itself.

The second category contains four genes that do not have an obvious connection to pyoverdine metabolism and are dispersed in the genome. PSPTO_0753 encodes an MFS permease of unknown specificity with similarity to the Bcr/CflA subfamily. PSPTO_3290 also encodes a putative membrane protein with similarity to OprD-type porins, which are involved with uptake of amino acids and related compounds (Tamber *et al*., 2006). It is possible that this gene is induced by PvdS to transport nutrients for pyoverdine synthesis. PSPTO_2982 is predicted to encode a protein with similarity to feruloyl esterases and tannases. The PSPTO_2982 protein may function to detoxify antimicrobial feruloyl-containing compounds, which are secreted by a wide range of plants in response to bacterial infection (Tamber *et al*., 2006; Zacares *et al*., 2007) or may degrade tannins as a mechanism to obtain iron from the host plant (Karamanoli and Lindow, 2006). The fourth gene, PSPTO_4923, codes for azurin, a well-studied periplasmic protein in *P. aeruginosa* believed to have a role in electron transport and/or redox stress (Vijgenboom *et al*., 1997). Azurin may be induced to help protect the cell from iron-catalysed oxidative stress or reactive oxygen bursts produced as part of the plant defence response to bacterial infection (Delledonne *et al*., 2001; Torres *et al*., 2005).

Comparative analysis with other pseudomonads

Comparative and functional genomic studies (Rombel *et al*., 1995; Ravel and Cornelis, 2003; Oguiza *et al*., 2005) suggest that the regulation of pyoverdine synthetic genes is conserved among the fluorescent pseudomonads, in spite of considerable inter-species (and in some cases inter-strain) divergence within the coding sequences of some genes involved in pyoverdine synthesis (Smith *et al*., 2005). As noted in *Results*, none of the HMMs (models A, DC3000-based; B, PAO1-based; and C, combined) adequately identify PvdS-boxes upstream of pyoverdine-related genes in all species considered. However, the union of the three sets of predictions provides an effective guide for identifying plausible PvdS-regulated genes

across the fluorescent pseudomonads (Table 3 and Fig. 6).

In particular, the union of the three sets of predictions identifies PvdS-boxes upstream of many pyoverdine-related genes in all genomes, consistent with the idea that the PvdS 'core' regulon is associated with pyoverdine metabolism (Table 3). The genomic context of the *pvdS* gene provides independent support for this hypothesis. Comparative analyses of the pyoverdine genes showed that in all cases the *pvdS* orthologue is linked with the gene encoding the non-ribosomal peptide synthetase (NRPS) responsible for producing the pyoverdine chromophore (Fig. 6). These results suggest that the essential features of PvdS regulation are broadly conserved among the fluorescent pseudomonads, but that the details can differ substantially even among closely related species. The HMM scans also identified high-scoring PvdS-boxes upstream of genes that are not in the pyoverdine cluster and which may have roles unrelated to siderophore production (the 'extended' regulon). These genes have diverse functions and vary from species to species, presumably to meet the specific needs of these organisms in the environments they encounter (see Table 3). For example, azurin is predicted to exhibit PvdS-dependent expression in the three *P. syringae* strains but not in the other pseudomonads. Also, we predict a PvdS-dependent promoter upstream of *exsA* in both *P. aeruginosa* strains, suggesting a possible role in ExoS production (Hovey and Frank, 1995).

Although the general function of PvdS is conserved, as is the general structure of the PvdS-box promoter motif, the comparison of sequence logos reveals subtle differences that give rise to variation among sequences identified by the respective models. As noted, gaps were inserted into the original set of sequences aligned by Gibbs sampling (Fig. 3A) to form the DC3000-based model A (and the combined model C); this was done in order to maximize conservation of the −35 and −10 promoter elements and permit for alternative spacer lengths when searching (Fig. 3A versus Fig. 4A). Further differences arise from the variation in sequences in the training sets for each model. These subtle differences lead to distinct but overlapping sets of predictions within each target genome.

What might explain these differences? One possibility is that the PvdS orthologues are interchangeable and that differences are simply due to random bias arising from the limited size of the training sets. Alternatively, they could reflect intrinsic structural and/or functional differences among PvdS orthologues leading to changes in sequence specificity. Finally, PvdS orthologues might be interchangeable, but individual promoter strengths are tuned to accommodate different physiological demands. Currently, we do not have sufficient data to distinguish among these possibilities. However, we suggest that the union of the three models provides a method for identifying the essential features of PvdS-regulated promoters and is therefore generally useful for formulating specific hypothesis across the fluorescent pseudomonads.

Experimental procedures

General methods, bacterial strains and plasmids

All general molecular biology protocols were performed as described in Sambrook and Russel (2001) unless noted otherwise. Table 4 contains a list of strains and plasmids used in this study.

Promoter trap vector constructions. The promoter trap vectors were constructed from a derivative of pUCP26 (West *et al.*, 1994). To adapt pUCP26 for use as a promoter trap vector, the *lacZ*α gene and promoter were deleted and a cassette containing stop codons in all reading frames followed by a ribosome binding site (RBS) sequence was inserted. This cassette was flanked by the recognition sequences for *Eco*RI, *Kpn*I and *Bam*HI upstream and *Hind*III downstream of the stop codons and RBS. Next, the *′lacZ*α gene was added as a *Hind*III fragment from pYES2/CtlacZ (Invitrogen, Carlsbad, CA). The kanamycin resistance encoding omega cassette (ΩKmΩ) from pHP45 ΩKm (Prentki and Krisch, 1984) was added to prevent transcription from upstream regions from affecting the expression of the reporter to produce pBS29, which was used as the base vector for construction of the *′lacZ*α promoter trap library.

The *lux* transcriptional fusion assays used a promoter trap vector that was derived from pBS29 by replacement of the *′lacZ*α gene with a *Hind*III fragment containing the *′luxCDABE* operon of *Photohabdus luminescens* (Winson *et al.*, 1998). The product of this ligation was pBS44, which was used as the empty vector control in the *lux* fusion assays. This vector was modified for efficient cloning by addition of the Gateway cassette (Invitrogen, Carlsbad, CA) to produce pBS58 and pBS59, which were used for LR-based Gateway cloning of specific regions to test for PvdS-dependent regulation. Mutant derivatives of the promoter region for PSPTO_2152 were constructed by Geneart AG, Regensburg, Germany.

PvdS expression vector constructions. Two pBBR1-based expression vectors were constructed for *pvdS* expression. First, pDT28 was constructed from pJN105 by introduction of the Gateway cassette downstream of the arabinose-inducible (pBAD) promoter (pBS1). The *sacB* gene was then added to facilitate sucrose-selective elimination of the vector, yielding pBS1-*sacB*. Second, pBS49 was derived from pBS1 by replacement of the arabinose-inducible P_{BAD} promoter with the constitutive *nptII* promoter to yield pBS46. An empty vector control (pBS60) was also constructed from pBS46 by deletion of the Gateway cassette. The *pvdS* gene was cloned into the two expression vectors (pBS1-*sacB* and pBS46) using Gateway cloning technology as per the manufacturer's specifications.

Library construction and screening

The promoter trap library was constructed using size-selected (0.75–1 kb), *Sau*3AI partially digested DC3000

Table 4. Description of strains and plasmids used in this study.

Strain/plasmid	Relevant characteristic	Reference
P. syringae pv. *tomato* DC3000	Wild-type strain	Buell *et al.* (2003)
DH5α	*E. coli* [F′ ϕ80*lacZ ΔM15 Δ*(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17* (rK-mK+)*phoA supE44* lambda- thi-1]	Hanahan (1983)
pJN105	pBBR1 derivative with the P_{BAD} promoter upstream of a multiple cloning site	Newman and Fuqua (1999)
pBS1-*sacB*	pJN105 with Gateway cloning cassette (*Spe*I *Nhe*I)	This work
pDT28	pBS1(*sacB*):: PSPTO2133 (*pvdS*)	This work
pBS60	pJN105 derivative with the P_{NPT2} promoter cloned upstream of cloning site; used as the empty vector control	This work
pBS49	pBS60 derivative for expression of PvdS P_{NPT2}::*pvdS* expression construct	This work
pUCP26	Cloning shuttle vector capable of replicating in *E. coli* and *Pseudomonas* species	West *et al.* (1994)
pSB417	Contains the *′luxCDABE* operon from *Photohabdus luminescens*	Winson *et al.* (1998)
pBS29	pUCP26 derivative used as library cloning vector; ΩkmΩ::MCS::3x stop::RBS::*′lacZα*	This work
pBS44	pBS29 derivative with promoterless *′luxCDABE* operon	This work
pBS105	pBS44 with DC3000 genomic co-ordinates 799976..800149 upstream of the *luxCDABE* operon (P_{PSPTO_0753}::*lux*)	This work
pBS130	pBS44 with DC3000 genomic co-ordinates 2322344..2322586 upstream of the *luxCDABE* operon (P_{PSPTO_2137}::*lux*)	This work
pBS54	pBS44 with DC3000 genomic co-ordinates 2329464..2329079 upstream of the *luxCDABE* operon (P_{PSPTO_2145}::*lux*)	This work
pBS55	pBS44 with DC3000 genomic co-ordinates 2329079..2329464 upstream of the *luxCDABE* operon (P_{PSPTO_2146}::*lux*)	This work
pBS51	pBS44 with DC3000 genomic co-ordinates 2330367..2330735 upstream of the *′luxCDABE* operon (P_{PSPTO_2147}::*lux*)	This work
pBS131	pBS44 with DC3000 genomic co-ordinates 2341677..2341919 upstream of the *luxCDABE* operon (P_{PSPTO_2149}::*lux*)	This work
pBS102	pBS44 with DC3000 genomic co-ordinates 2361993..2362319 upstream of the *luxCDABE* operon (P_{PSPTO_2152}::*lux*)	This work
pBS57	pBS44 with DC3000 genomic co-ordinates 2367905..2367497 upstream of the *luxCDABE* operon (P_{PSPTO_2156}::*lux*)	This work
pBS56	pBS44 with DC3000 genomic co-ordinates 2367497..2367905 upstream of the *luxCDABE* operon (P_{PSPTO_5624}::*lux*)	This work
pBS134	pBS44 with DC3000 genomic co-ordinates 2394774..2395015 upstream of the *luxCDABE* operon (P_{PSPTO_2175}::*lux*)	This work

genomic DNA that was subcloned into the *Bam*HI site of the pBS29 promoter trap vector to create a transcriptional fusion of the genomic DNA with the promoterless *′lacZα* gene. The sequences from a random sample of 89 library isolates showed that 91% of the constructs contained an insert with an average size of 615 ± 341 bp.

Screening for clones containing PvdS-dependent promoters was carried out in two steps. First, *E. coli* DH5α cells were transformed with the PvdS expression vector (pDT28) and the *′lacZα* promoter trap library. Transformants were selected for growth on LB agar supplemented with 50 μg ml^{-1} kanamycin, 10 μg ml^{-1} gentamicin and 20 μg ml^{-1} X-gal. Transformants were screened for expression of the *lacZα* reporter gene, which was indicated by the ability of the clone to metabolize the X-gal substrate as assessed by the production of blue colour, and lac+ clones were arrayed in 96-well format. In the second step, clones containing PvdS-dependent promoters were identified. This involved elimination of the PvdS expression vector by selecting for growth on 5% sucrose. Loss of the vector was confirmed by testing sensitivity to gentamicin. Clones were then patched to assess promoter activity in the absence of the PvdS expression vector and compared with the parent clone retaining the PvdS expression vector. This procedure yielded 25 unique non-overlapping intergenic regions that were used to produce preliminary HMMs. The majority of these library isolates contained *bona fide* PvdS-box motifs, while the remaining 12/25 are potentially false positives in that we have not been able to explain the basis of the observed PvdS-dependent promoter activity. However, it should be noted that the *′lacZ*-based promoter-trapping protocol that was used is extremely sensitive, and we may be able to detect low levels of transcription from cryptic PvdS-boxes.

Promoter fusion assay

Discrete regions of the DC3000 genome and the PSPTO_2152 substitution mutant variants were tested for PvdS-dependent promoter activity using the *′luxCDABE* promoter trap vectors. The DC3000 genome co-ordinates of the regions tested are shown in Table 4. To assess the region of interest, fusion vectors were used to transform *E. coli* DH5α cells containing either the PvdS expression vector (pBS49) or the isogenic empty vector control (pBS60). Reporter expression was assayed in mid-logarithmic growth phase using a

GENios Pro 96-well plate reader (Tecan Group) to measure the production of light (100 ms integration). The relative luminescence in Fig. 1A and Table 1 is the base 10 logarithm of the ratio of the observed luminescence/optical density at 620 nm (OD_{620}). The average and standard deviations were calculated from six biological replicates. The fold difference in Fig. 1A is the ratio of the average relative luminescence +PvdS expression versus –PvdS expression expressed as the geometric mean. The Student's *t*-test was used to calculate *P*-values in Table 1.

qRT-PCR

PvdS-dependent differential expression was assessed at 19 genomic loci by comparing transcript levels in DC3000 cells transformed with either the PvdS expression vector (pBS49) or the isogenic empty vector control (pBS60). These strains were grown under conditions where the endogenous *pvdS* allele is repressed so as to facilitate detection of PvdS-dependent differential regulation. Accordingly, total RNA was prepared from cells grown in SOC media supplemented with 10 μg ml^{-1} gentamicin. Cells were harvested at an OD_{600} of 0.65 and RNA was prepared as described in Vencato *et al*. (2006) and used for both qRT-PCR and 5′ RACE (see below). Primer sets for qRT-PCR analysis were designed using Beacon Designer::TaqMan probe design software. The sequence of the primers and their DC3000 co-ordinates are shown in Table S3.

Real-time PCR was performed with the My IQ5 Sequence Detection System (Bio-Rad) and iTaq SYBR Green Supermix (Bio-Rad) following manufacturer's protocols and as described previously (Vencato *et al*., 2006). Ct values were calculated using the My IQ5 software and analysed using the relative standard curve method (separate tubes) described in ABI User Bulletin #2. In each strain, the Ct values of each gene tested were normalized to the Ct values of the housekeeping gene *gap-1*, and then the relative induction was calculated as the ratio of the normalized expression levels in RNA prepared from cells transformed with either the PvdS expression vector (pBS49) or the isogenic empty vector control (pBS60). A second housekeeping gene *gyrA* was also tested and is presented as a negative control. Averages and standard deviations were calculated from three independent biological replicates. The fold difference in Fig. 1B is the geometric mean of the relative induction.

5′ RACE

The transcription start points were determined using the Invitrogen 5′ RACE System for Rapid Amplification of cDNA Ends Version 2.0 (Catalogue No. 18374-058). RNA isolated from cells expressing PvdS (see qRT-PCR, above) was used as the template for 5′ RACE analysis. The primers used in this analysis are shown in Table S4. The 5′ RACE reactions were performed as recommended by the manufacturer. 5′ RACE reactions were analysed by agarose gel electrophoresis to assess purity and product size. Single bands were observed for six of seven reactions; for PSTPO_2146, there was a second minor product that was not investigated further. Products were gel purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) and sequenced using a nested gene-specific primer to locate the 5′ end of the transcript. The sequencing results were interpreted by pair-wise alignment of the 5′ RACE product sequence with the DC3000 genomic sequence. A range of possible transcription start positions were designated if the chromosomal sequence matched the nucleotide used in the tailing reaction or if the identity of a nucleotide could not be determined.

Gibbs sampling

The methods used here are minor variations of those reported by us previously (Ferreira *et al*., 2006). The input data set consisting of unaligned sequences associated with PvdS-dependent transcription was sampled using PhyloGibbs (Siddharthan *et al*., 2005; Siddharthan, 2007) to identify candidate motifs associated with promoters and transcription factor binding sites. Sampling was performed for non-symmetric motifs from 16 to 52 nucleotides in length using the second-order nucleotide statistics for the complete DC3000 genome as the background model. Motifs with significant information content were identified only in the range of 27–32 nucleotides, and all appeared to be variations of the same basic structure. The 29-nucleotide motif containing the largest number of representatives from the input training set was selected as the most complete summary of the available experimental data. The sequence logo for this model constructed by weblogo (version 3.0b13, http://code.google.com/p/weblogo/) using the CG content of the DC3000 genome as the background model is shown in Fig. 3. Additional sampling runs to identify motifs with internal reverse complementation symmetry were also performed for the same range of sizes, but none of the resulting motifs had sufficient information content to warrant further investigation.

Construction and use of hidden Markov models

As in previous studies on HrpL-dependent promoters (Fouts *et al*., 2002; Ferreira *et al*., 2006), the HMMer package (http://hmmer.janelia.org) was used to construct and calibrate hidden Markov models (HMMs), and to scan the genomes of the fully sequenced pseudomonads to identify putative PvdS-dependent promoters. Functions of downstream genes were assigned by evaluation of existing genome annotation and the Protein Clusters and BLink web pages at NCBI. The sequences identified by PhyloGibbs were manually realigned to maximize the conservation of the –35 element by introducing gaps and used to train model A. Monomer nucleotide statistics were used to calibrate the model for each target genome. An *E*-value cut-off of 10.0 was selected to balance the number of predicted promoters with plausible and implausible orientation with respect to downstream genes. Analysis of the sensitivity of the models to variation in this cut-off – in a manner reminiscent of that used for receiver–operator characteristic (ROC) plots – is shown in Fig. 5. Additional computational experiments based on generation of artificial sequences using Markov models of order 2–5 constructed using the seq++ package (Miele *et al*., 2005) and trained on the DC3000 genome suggest that this cut-off corresponds to

a false positive rate of three to six hits per genome equivalent for model A. This false positive rate is consistent with the low number of hits (data not shown) in the genomes of the non-fluorescent pseudomonad strains, *Pseudomonas stutzeri* A1501 and *P. mendocina* ymp (Holt *et al.*, 1994).

Inference of PAO1 IS-box model

The inference of the IS-box motif in *P. aeruginosa* PAO1 from the description in Ravel and Cornelis (2003) proceeded as follows. Figure 1 in that source indicated motifs representing the consensus IS-box (TAAAT-N16-CGT) upstream of the following 13 genes: PA2386, PA2387, PA2389, PA2392, PA2393, PA2396, PA2397, PA2399, PA2402, PA2403, PA2412, PA2413 and PA2425. The fuzznuc utility from the EMBOSS application suite (Rice *et al.*, 2000) was used to scan the sequenced PAO1 genome for the stated IS-box consensus sequence plus four non-specific flanking nucleotides on either side: N4-TAAAT-N16-CGT-N4. Matches upstream of the designated genes were identified. As not all IS-boxes matched the consensus sequence exactly, fuzznuc was again used to search for mismatched sequences, with a total of one, two and three mismatches until all 13 PAO1 genes were accounted for. In cases where multiple matches occurred upstream of a gene with equal score (i.e. number of mismatches), the match closest to the annotated gene start (but not overlapping with the annotated coding region) was chosen. As all mismatches were in the –35 element (TAAAT), we were able to align the 13 sequences by their conserved –10 elements. This alignment was used to train model B. The extra non-specific flanking sequence was then trimmed to arrive at a PAO1 IS-box logo of the same size and in alignment with our experimentally derived DC3000 PvdS-box (prior to realignment).

Construction of a combined DC3000-PAO1 PvdS-regulated promoter model

The training set for model C was obtained by realigning the training set for the PAO1-based model B to the training set for the DC3000-based model A. The effect of this realignment was to introduce a single gap in the spacer region of the PAO1-derived sequences relative to model B.

Acknowledgements

We thank Melanie Filiatrault and Philip Bronstein for critical reading of the manuscript, Alan Collmer for strains and plasmids, and Genevieve DeClerck and Paul Stodghill for bioinformatics support. C.R. Myers acknowledges support from USDA-ARS Project 1907-21000-027-03. We thank the reviewers for helpful suggestions leading to the construction of the combined model.

References

Buell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T., Gwinn, M.L., *et al.* (2003) The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proc Natl Acad Sci USA* **100:** 10181–10186.

Cunliffe, H.E., Merriman, T.R., and Lamont, I.L. (1995) Cloning and characterization of *pvdS*, a gene required for pyoverdine synthesis in *Pseudomonas aeruginosa*: PvdS is probably an alternative sigma factor. *J Bacteriol* **177:** 2744–2750.

Delledonne, M., Zeier, J., Marocco, A., and Lamb, C. (2001) Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response. *Proc Natl Acad Sci USA* **98:** 13454–13459.

Eiamphungporn, W., and Helmann, J.D. (2008) The *Bacillus subtilis* sigma(M) regulon and its contribution to cell envelope stress responses. *Mol Microbiol* **67:** 830–848.

Enz, S., Mahren, S., Menzel, C., and Braun, V. (2003) Analysis of the ferric citrate transport gene promoter of *Escherichia coli*. *J Bacteriol* **185:** 2387–2391.

Feil, H., Feil, W.S., Chain, P., Larimer, F., DiBartolo, G., Copeland, A., *et al.* (2005) Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a Pv *tomato* DC3000. *Proc Natl Acad Sci USA* **102:** 11064–11069.

Ferreira, A.O., Myers, C.R., Gordon, J.S., Martin, G.B., Vencato, M., Collmer, A., *et al.* (2006) Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae* pv. *tomato* DC3000, allows *de novo* reconstruction of the Hrp *cis* element, and identifies novel coregulated genes. *Mol Plant Microbe Interact* **19:** 1167–1179.

Fouts, D.E., Abramovitch, R.B., Alfano, J.R., Baldo, A.M., Buell, C.R., Cartinhour, S., *et al.* (2002) Genomewide identification of *Pseudomonas syringae* pv. *tomato* DC3000 promoters controlled by the HrpL alternative sigma factor. *Proc Natl Acad Sci USA* **99:** 2275–2280.

Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63:** 141–155.

Gruber, T.M., and Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57:** 441–466.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557–580.

Helmann, J.D. (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* **46:** 47–110.

Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T., and Williams, S.T. (eds) (1994) *Bergey's Manual of Determinative Bacteriology.* Baltimore, MD: Williams & Wilikins.

Hovey, A.K., and Frank, D.W. (1995) Analyses of the DNA-binding and transcriptional activation properties of ExsA, the transcriptional activator of the *Pseudomonas aeruginosa* exoenzyme S regulon. *J Bacteriol* **177:** 4427–4436.

Joardar, V., Lindeberg, M., Jackson, R.W., Selengut, J., Dodson, R., Brinkac, L.M., *et al.* (2005) Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J Bacteriol* **187:** 6488–6498.

Karamanoli, K., and Lindow, S.E. (2006) Disruption of *N*-acyl

homoserine lactone-mediated cell signaling and iron acquisition in epiphytic bacteria by leaf surface compounds. *Appl Environ Microbiol* **72:** 7678–7686.

Lamont, I.L., Beare, P.A., Ochsner, U., Vasil, A.I., and Vasil, M.L. (2002) Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **99:** 7072–7077.

Lane, W.J., and Darst, S.A. (2006) The structural basis for promoter –35 element recognition by the group IV sigma factors. *PLoS Biol* **4:** e269.

Leoni, L., Ciervo, A., Orsi, N., and Visca, P. (1996) Iron-regulated transcription of the *pvdA* gene in *Pseudomonas aeruginosa*: effect of Fur and PvdS on promoter activity. *J Bacteriol* **178:** 2299–2313.

Leoni, L., Orsi, N., de Lorenzo, V., and Visca, P. (2000) Functional analysis of PvdS, an iron starvation sigma factor of *Pseudomonas aeruginosa*. *J Bacteriol* **182:** 1481–1491.

Lewis, D.E., and Adhya, S. (2004) Axiom of determining transcription start points by RNA polymerase in *Escherichia coli*. *Mol Microbiol* **54:** 692–701.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994) Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci USA* **91:** 7573–7577.

McClure, W.R., Hawley, D.K., Youderian, P., and Susskind, M.M. (1983) DNA determinants of promoter selectivity in *Escherichia coli*. *Cold Spring Harb Symp Quant Biol* **47** (Part 1)**:** 477–481.

Meyer, J.M. (2000) Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. *Arch Microbiol* **174:** 135–142.

Meyer, J.M., Neely, A., Stintzi, A., Georges, C., and Holder, I.A. (1996) Pyoverdin is essential for virulence of *Pseudomonas aeruginosa*. *Infect Immun* **64:** 518–523.

Miele, V., Bourguignon, P.Y., Robelin, D., Nuel, G., and Richard, H. (2005) seq++: analyzing biological sequences with a range of Markov-related models. *Bioinformatics* **21:** 2783–2784.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme–DNA complex. *Science* **296:** 1285–1290.

Newman, E.B., and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. *Annu Rev Microbiol* **49:** 747–775.

Newman, J.R., and Fuqua, C. (1999) Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli araBAD* promoter and the *araC* regulator. *Gene* **227:** 197–203.

Nonaka, G., Blankschien, M., Herman, C., Gross, C.A., and Rhodius, V.A. (2006) Regulon and promoter analysis of the *E-coli* heat-shock factor, sigma(32), reveals a multifaceted cellular response to heat stress. *Genes Dev* **20:** 1776–1789.

Ochsner, U.A., Wilderman, P.J., Vasil, A.I., and Vasil, M.L. (2002) GeneChip expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: identification of novel pyoverdine biosynthesis genes. *Mol Microbiol* **45:** 1277–1287.

Oguiza, J.A., Kiil, K., and Ussery, D.W. (2005) Extracytoplasmic function sigma factors in *Pseudomonas syringae*. *Trends Microbiol* **13:** 565–568.

Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S., Mavrodi, D.V., *et al.* (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* **23:** 873–878.

Poole, K., and McKay, G.A. (2003) Iron acquisition and its control in *Pseudomonas aeruginosa*: many roads lead to Rome. *Front Biosci* **8:** d661–686.

Potvin, E., Sanschagrin, F., and Levesque, R.C. (2008) Sigma factors in *Pseudomonas aeruginosa*. *FEMS Microbiol Rev* **32:** 38–55.

Prentki, P., and Krisch, H.M. (1984) *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **29:** 303–313.

Ravel, J., and Cornelis, P. (2003) Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol* **11:** 195–200.

Rhodius, V.A., Suh, W.C., Nonaka, G., West, J., and Gross, C.A. (2006) Conserved and variable functions of the sigma(E) stress response in related genomes. *PLoS Biol* **4:** 43–59.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16:** 276–277.

Rombel, I.T., McMorran, B.J., and Lamont, I.L. (1995) Identification of a DNA sequence motif required for expression of iron-regulated genes in pseudomonads. *Mol Gen Genet* **246:** 519–528.

Sambrook, J., and Russel, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sarkar, S.F., Gordon, J.S., Martin, G.B., and Guttman, D.S. (2006) Comparative genomics of host-specific virulence in *Pseudomonas syringae*. *Genetics* **174:** 1041–1056.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007) Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* **35:** 771–788.

Siddharthan, R. (2007) Parsing regulatory DNA: general tasks, techniques, and the PhyloGibbs approach. *J Biosci* **32:** 863–870.

Siddharthan, R., Siggia, E.D., and van Nimwegen, E. (2005) PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput Biol* **1:** e67.

Smith, E.E., Sims, E.H., Spencer, D.H., Kaul, R., and Olson, M.V. (2005) Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J Bacteriol* **187:** 2138–2147.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406:** 959–964.

Tamber, S., Ochs, M.M., and Hancock, R.E. (2006) Role of the novel OprD family of porins in nutrient uptake in *Pseudomonas aeruginosa*. *J Bacteriol* **188:** 45–54.

Torres, M.A., Jones, J.D., and Dangl, J.L. (2005) Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana*. *Nat Genet* **37:** 1130–1134.

Vencato, M., Tian, F., Alfano, J.R., Buell, C.R., Cartinhour, S.,

DeClerck, G.A., *et al.* (2006) Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A. *Mol Plant Microbe Interact* **19:** 1193–1206.

Vijgenboom, E., Busch, J.E., and Canters, G.W. (1997) *In vivo* studies disprove an obligatory role of azurin in denitrification in *Pseudomonas aeruginosa* and show that *azu* expression is under control of *rpoS* and ANR. *Microbiology* **143** (Part 9)**:** 2853–2863.

Visca, P., Leoni, L., Wilson, M.J., and Lamont, I.L. (2002) Iron transport and regulation, cell signalling and genomics: lessons from *Escherichia coli* and *Pseudomonas*. *Mol Microbiol* **45:** 1177–1190.

West, S.E., Schweizer, H.P., Dall, C., Sample, A.K., and Runyen-Janecky, L.J. (1994) Construction of improved *Escherichia–Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Gene* **148:** 81–86.

Wilson, M.J., and Lamont, I.L. (2000) Characterization of an ECF sigma factor protein from *Pseudomonas aeruginosa*. *Biochem Biophys Res Commun* **273:** 578–583.

Wilson, M.J., and Lamont, I.L. (2006) Mutational analysis of an extracytoplasmic-function sigma factor to investigate its interactions with RNA polymerase and DNA. *J Bacteriol* **188:** 1935–1942.

Wilson, M.J., McMorran, B.J., and Lamont, I.L. (2001) Analysis of promoters recognized by PvdS, an extracytoplasmic-function sigma factor protein from *Pseudomonas aeruginosa*. *J Bacteriol* **183:** 2151–2155.

Winson, M.K., Swift, S., Hill, P.J., Sims, C.M., Griesmayr, G., Bycroft, B.W., *et al.* (1998) Engineering the *luxCDABE* genes from *Photorhabdus luminescens* to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn*5* constructs. *FEMS Microbiol Lett* **163:** 193.

Zacares, L., Lopez-Gresa, M.P., Fayos, J., Primo, J., Belles, J.M., and Conejero, V. (2007) Induction of p-coumaroyldopamine and feruloyldopamine, two novel metabolites, in tomato by the bacterial pathogen *Pseudomonas syringae*. *Mol Plant Microbe Interact* **20:** 1439–1448.

Supplementary material

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2008.06209.x
(This link will take you to the article abstract).

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