University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Papers in Plant Pathology

Plant Pathology Department

11-2006

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

Adriana O. Ferreira
Boyce Thompson Institute for Plant Research

Christopher R. Myers
Cornell University

Jeffrey S. Gordon
Boyce Thompson Institute for Plant Research,

Gregory B. Martin
Cornell University

Monica Vencato
Cornell University

Ferreira, Adriana O.; Myers, Christopher R.; Gordon, Jeffrey S.; Martin, Gregory B.; Vencato, Monica; Collmer, Alan; Wehling, Misty D.; Alfano, James R.; Moreno-Hagelsieb, Gabriel; Lamboy, Warren F.; DeClerck, Geneviene; Schneider, David J.; and Cartinhour, Samuel W., "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" (2006). *Papers in Plant Pathology*. 300.
http://digitalcommons.unl.edu/plantpathpapers/300

Authors

Adriana O. Ferreira, Christopher R. Myers, Jeffrey S. Gordon, Gregory B. Martin, Monica Vencato, Alan Collmer, Misty D. Wehling, James R. Alfano, Gabriel Moreno-Hagelsieb, Warren F. Lamboy, Geneviene DeClerck, David J. Schneider, and Samuel W. Cartinhour

MPMI Vol. 19, No. 11, 2006, pp. 1167–1179. DOI: 10.1094/MPMI-19-1167.
Molecular Plant-Microbe Interactions Nov 2006, Volume 19, Number 11

e-Xtra*

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

Adriana O. Ferreira,[1] Christopher R. Myers,[2] Jeffrey S. Gordon,[1] Gregory B. Martin,[1,3] Monica Vencato,[3] Alan Collmer,[3] Misty D. Wehling,[4] James R. Alfano,[4] Gabriel Moreno-Hagelsieb,[5] Warren F. Lamboy,[5] Genevieve DeClerck,[5] David J. Schneider,[5] and Samuel W. Cartinhour[5]

[1]Boyce Thompson Institute for Plant Research, Ithaca, NY 14853, U.S.A.; [2]Cornell Theory Center, Cornell University, Ithaca, NY 14853, U.S.A.; [3]Department of Plant Pathology, Cornell University, Ithaca, NY 14853, U.S.A.; [4]The Plant Science Initiative and the Department of Plant Pathology, University of Nebraska-Lincoln, Lincoln, 68588, U.S.A.; [5]United States Department of Agriculture–Agricultural Research Service, Ithaca, NY 14853, U.S.A.

Submitted 28 January 2006. Accepted 2 July 2006.

Pseudomonas syringae **pv.** ***tomato*** **DC3000 is a model pathogen of tomato and** ***Arabidopsis*** **that uses a hypersensitive response and pathogenicity (Hrp) type III secretion system (T3SS) to deliver virulence effector proteins into host cells. Expression of the Hrp system and many effector genes is activated by the HrpL alternative sigma factor. Here, an open reading frame-specific whole-genome microarray was constructed for DC3000 and used to comprehensively identify genes that are differentially expressed in wild-type and Δ*hrpL* strains. Among the genes whose differential regulation was statistically significant, 119 were upregulated and 76 were downregulated in the wild-type compared with the Δ*hrpL* strain. Hierarchical clustering revealed a subset of eight genes that were upregulated particularly rapidly. Gibbs sampling of regions upstream of HrpL-activated operons revealed the Hrp promoter as the only identifiable regulatory motif and supported an iterative refinement involving real-time polymerase chain reaction testing of additional HrpL-activated genes and refinements in a hidden Markov model that can be used to predict Hrp promoters in *P. syringae* strains. This iterative bioinformatic-experimental approach to a comprehensive analysis of the HrpL regulon revealed a mix of genes controlled by HrpL, including those encoding most type III effectors, twin-arginine transport (TAT) substrates, other regulatory proteins, and proteins involved in the synthesis or metabolism of phytohormones, phytotoxins, and myo-inositol. This analysis provides an extensively verified, robust method for predicting Hrp promoters in *P. syringae* genomes, and it supports subsequent identification of effectors and other factors that likely are important to the host-specific virulence of *P. syringae.***

Additional keywords: weight matrix model.

Pseudomonas syringae pv. *tomato* DC3000 is a pathogen of tomato and *Arabidopsis* that translocates virulence effector proteins into host cells via a type III secretion system (T3SS). Regulation of the T3SS is known to depend on the ECF-type sigma factor HrpL (Xiao et al. 1994). The identification of hypersensitive response and pathogenicity (Hrp) promoters and genes expressed in a HrpL-dependent manner is key to understanding *P. syringae* as a plant pathogen for several reasons. First, most T3SS effectors are associated with Hrp promoters, and identifying Hrp promoters is an efficient first step in identifying candidate effector genes (Chang et al. 2005; Fouts et al. 2002; Zwiesler-Vollick et al. 2002). Second, genes with no apparent function in the T3SS that nevertheless are activated (directly or indirectly) by HrpL are likely to have a significant role in pathogenesis. A notable example is found in the genes directing synthesis of the phytotoxin coronatine (Chang et al. 2005; Fouts et al. 2002; Peñaloza-Vázquez et al. 2000). Third, *P. syringae* strains are divided into more than 50 pathovars and multiple races based on host specificity and have remarkably diverse interactions with plants (Hirano and Upper 2000). Polymorphisms in the HrpL regulon (particularly among the effectors) are known to underlie some of this diversity (Arnold et al. 2003; Rohmer et al. 2004), and methods to efficiently identify Hrp promoters in genome sequences should improve our understanding of this phenomenon as well as provide useful diagnostic tools. Our purpose here is to use a microarray-based iterative approach to comprehensively identify all of the Hrp promoters in the model pathogen *P. syringae* pv. *tomato* DC3000 and to develop improved methods for identifying Hrp promoters in the sequences of any new *P. syringae* strains.

A. O. Ferreira and C. R. Myers contributed equally to this work.

Corresponding author: Samuel W. Cartinhour; E-mail: sc167@cornell.edu

Current address of G. Moreno-Hagelsieb: Department of Biology, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada.

*The *e*-Xtra logo stands for "electronic extra" and indicates the HTML abstract available on-line contains supplemental material not included in the print edition. There are four supplemental tables and one supplemental figure.

In previous work, we initiated our analysis of the HrpL regulon in *P. syringae* pv. *tomato* DC3000 by using a reporter transposon to enlarge the set of known, functional Hrp promoters, which enabled the training of a hidden Markov model (HMM) that was used to search a draft sequence of the DC3000 genome (Fouts et al. 2002). Six of the candidate promoters we found with strong expectation values were tested experimentally and confirmed to be activated by HrpL. Downstream of these promoters were several known effectors and candidate effectors, as well as genes encoding factors unrelated to the T3SS, such as *iaaL* and *corR*. The process of experimentally identifying HrpL-responsive genes in DC3000 was carried to near saturation using differential fluorescence induction (Chang et al. 2005), which resulted in the identification of 49 HrpL-responsive genes and corroborated our previous findings (Fouts et al. 2002). A limitation of these two studies is that both involved overexpression of HrpL, and neither provided time-course data on the expression of the HrpL regulon. Thus, it is important to note that the microarray analysis presented in this study follows changes in gene expression of wild-type and Δ*hrpL* strains for several hours following transfer to Hrp-inducing conditions.

Here, we report i) the use of an open reading frame (ORF)-specific whole genome microarray for *P. syringae* pv. *tomato* DC3000 to comprehensively identify genes that are differentially expressed in wild-type and HrpL-deficient strains; ii) Gibbs sampling for de novo identification of Hrp promoter sequences upstream of HrpL-activated genes that were identified in the microarray; iii) development of a hidden Markov model and position-specific weight matrix (PSWM) model based on a microarray-derived training set, which was used to comprehensively identify candidate Hrp promoters in the complete genome of DC3000; iv) real-time polymerase chain reaction (PCR) confirmation of candidate Hrp promoters that were missed by the microarray analysis; and v) analysis of Hrp promoters and downstream genes in the context of predicted operons, genome annotation errors, and functions of the HrpL regulon in pathogenesis.

RESULTS

Clustering of microarray expression data.

Our analysis of microarray expression levels, described in Materials and Methods, yielded a set of 119 genes that satisfied the criteria for significant differential upregulated expression (DEG-UP), and 76 genes with significant downregulated expression (DEG-DOWN) in the wild-type compared with the Δ*hrpL* strain. (One gene, PSPTO0067, exhibited both significant upregulation and downregulation at different time points.) The full set we refer to collectively as DEG-ALL, whose differential expression levels, along with their gene identifiers, are presented in Figure 1A. Hierarchical clustering of the expression data reveals a subcluster with highly correlated expression (Fig. 1A-1) which, upon further inspection, is seen to contain known T3SS components, including *hrp*/*hrc* genes and effectors. The general character of these clustering results is insensitive to specific details of the algorithm, particularly if correlation-based distance metrics are used to relate expression patterns. Clustering of the upregulated gene set DEG-UP (data not shown), using fold changes rather than their logs, yields a similar tightly correlated subcluster, including the known T3SS components. Subclusters 3 and 4 constitute the bulk of the downregulated gene set DEG-DOWN, a set that, to our knowledge, has not been described previously. Subcluster 2 consists largely of a set of genes that are upregulated at time $t = 0$ h.

Gibbs sampling upstream of differentially expressed genes.

From the set of differentially expressed genes, Gibbs sampling was carried out to search for conserved regulatory motifs. Gibbs sampling of the set of candidate regulatory regions, targeted toward finding promoter-sized objects (e.g., size 30 to 40 bases), consistently identified and clustered a motif consistent with the consensus sequence previously associated with the Hrp promoter, 5′-GGAAC-$N_{16\text{-}17}$-CCACNNA-3′ (Fouts et al. 2002; Innes et al. 1993; Shen and Keen 1993; Xiao and Hutcheson 1994). This observation is true for sampling done on both the upstream regions of the full DEG-ALL set and the upregulated DEG-UP set. For sampling done on the DEG-UP set, with motif size 33, the average cluster occupancies of all sequences clustered in the Hrp promoter motif are plotted in Figure 2. Sequences with separations of length 16 and 17 between the –35 and –10 boxes are clustered together by the Gibbs sampler. A few stray sequences occasionally are included in the Hrp promoter cluster during the sampling process, but Figure 2 demonstrates that there is a clear separation between a set of sequences that is strongly clustered and a few outliers that are not. For further analysis, we have kept only those sequences whose average occupancy was greater than 50%. Because of the differing separation lengths within the population of putative Hrp promoters, it was useful for some purposes to align the set of sequences clustered by the Gibbs sampler for further analysis. Alignment of the motifs with separation of length 17 to the set of length 16 using ClustalX (Thompson et al. 1997) yielded a combined model with gaps (for construction of HMMs).

Although Gibbs sampling located the Hrp promoter motif, sampling of the upstream sequence regions associated with the DEG-ALL, DEG-UP, and DEG-DOWN sets and various subsets did not consistently identify any other strongly conserved sequence motifs. Furthermore, models of other regulatory motifs that we have constructed (e.g., Fur, RpoN, RpoS, and RpoD promoters) do not reveal putative binding sites in the DC3000 genome that are correlated with the sets of differentially expressed genes identified by the microarray.

Construction of initial Gibbs-derived motif models.

The set of motifs clustered by the Gibbs sampler (with occupancy greater than 50%) was used to construct HMM and PSWM models of the Hrp promoter motif. (These are denoted as our “level 1” training set.) The entire DC3000 genome then was scanned using these models to identify putative Hrp promoters whose targets might not have been identified by the microarray. Several such genes were identified, with upstream promoters scoring greater than 10 by the HMM. Some of those genes were not represented on our microarray due to technical problems, some did not meet our criteria for statistically significant differential expression but are in polycistronic operons with genes that were differentially expressed, and some simply did not meet our cutoff criteria.

Real-time PCR confirmation of additional Hrp promoters.

Our hypothesis was that genes with strong upstream Hrp promoters should exhibit significant upregulation in the microarray. That was largely true; however, for a handful of genes that did not fit that scenario, real time-PCR was used as a more sensitive but less high-throughput method of determining HrpL-dependent expression. Ten genes with strong Hrp promoters (HMM level 1 score >10) but no significant differential expression in the microarray were analyzed by real time-PCR. Of those 10, PSPTO0474 and PSPTO4776 had exhibited differential expression under HrpL overexpression (Fouts et al. 2002), and PSPTO1394 was not on our microarray. One unannotated region downstream of a strong putative promoter, upstream of PSPTO0869, also was tested for real time-PCR. Two genes (PSPTO4732 and PSPTOB0005) that were in the DEG-UP set but are rather far downstream from a putative Hrp promoter also were tested, as was one gene (PSPTO2691) that was also in

DEG-UP but does not have any identifiable upstream promoter. Two putative transcriptional regulators (PSPTO3576 and PSPTO1645) that were not in DEG-UP also were tested; PSPTO3576 has a rather weak promoter (HMM level 1 score = 7.1), whereas PSPTO1645 has a stronger promoter (HMM level 1 = 12.4) which is embedded in an upstream gene roughly 700 bp from the transcription start of PSPTO1645. Finally, two housekeeping genes were included as negative controls. For these 18 cases, real time-PCR assays were done using both the condition investigated here (*hrpL* deletion mutant in HrpMM) and that investigated in earlier work (*hrpL* overexpression in AB media).

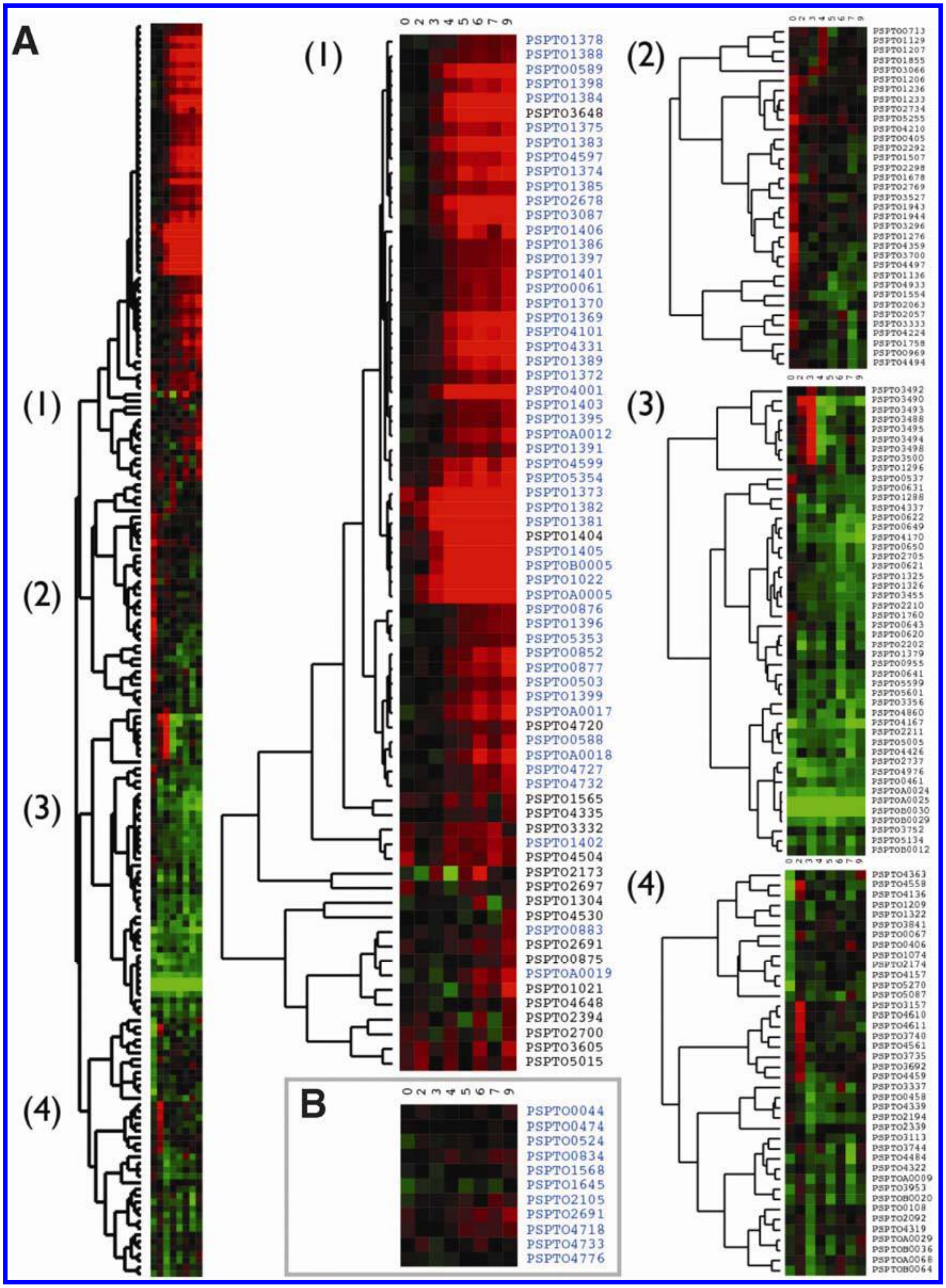


Fig. 1. Differentially expressed genes as detected by **A,** microarray or **B,** real-time polymerase chain reaction (RT-PCR). The microarray expression levels are hierarchically clustered, and broken out into four clusters (1 to 4) as shown. Gene names presented in blue font are those that either were found by the Gibbs sampler to contain an upstream hypersensitive response and pathogenicity (Hrp) promoter or subsequently inferred to be part of an operon with an upstream Hrp promoter (PSPTOA0018, PSPTOA0019, and PSPTOB0005). Subcluster 1 is the most tightly clustered (as indicated by the short depth of the clustering tree to the left), and all Hrp promoters associated with upregulated genes in the microarray are found in that subcluster.

The results of the real time-PCR experiments are shown in Table 1. HrpL-dependent expression was confirmed or reconfirmed for 14 genes or ORFs with strong Hrp promoters (PSPTO0044, PSPTO0474, PSPTO0524, PSPTO0834, PSPTO1394, PSPTO1568, PSPTO1645, PSPTO2105, PSPTO4718, PSPTO4732, PSPTO4733, PSPTO4776, PSPTOB0005, and the unannotated region upstream of PSPTO0869). PSPTO2691 was reconfirmed to be HrpL dependent despite the lack of a Hrp promoter. Only one of the two putative regulators (PSPTO1645) with weak or dubious Hrp promoters exhibited differential expression based upon real time-PCR. The two housekeeping genes used as negative controls were shown to not be regulated by HrpL in this experiment. Generally, little significant difference in the log of the expression fold change was observed between the two conditions (HrpL deletion versus HrpL overexpression) for the genes tested here.

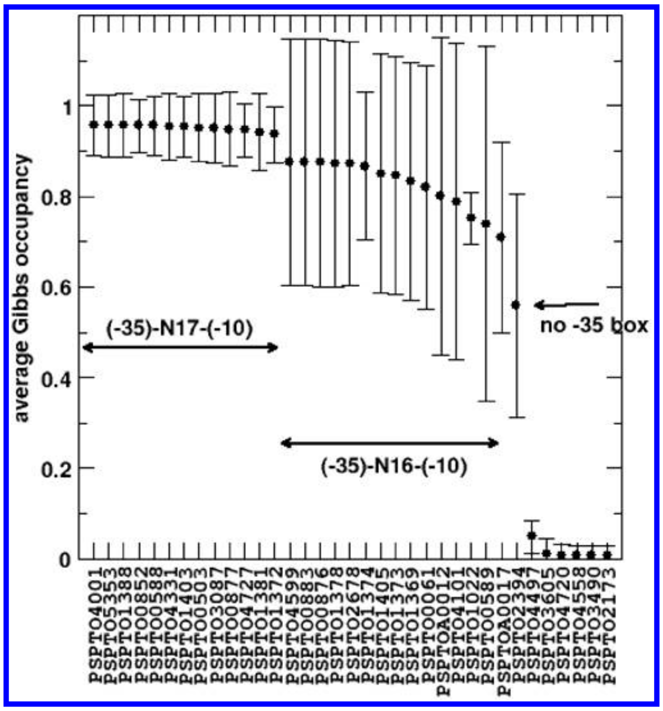


Fig. 2. Average occupancy of upstream sequence motifs within the dominant hypersensitive response and pathogenicity (Hrp)-promoter-related Gibbs cluster (averaged over 10 runs). The set of 13 genes highly clustered with occupancy near 1.0 represent most of the Hrp promoters with spacer length 17. The next set of 15 genes with slightly lower occupancies are primarily those with spacer length 16. Error bars represent standard deviations from the set of 10 occupancy scores (even though no occupancy scores can exceed 1.0 in a given run); if a given gene fails to be included in the cluster in a given run, it is assigned an occupancy score of 0 for that run.

Correlation of Hrp promoter motifs and microarray clustering.

With the identification of Hrp promoters from the combination of bioinformatics and further bioinformatically directed experimentation, we can revisit the clustering of microarray expression data in Figure 1 to examine correlations between clustering and sequence motifs. The genes annotated in Fig. 1A with blue font are those with upstream Hrp promoters. All upstream Hrp promoters are associated with the tightly correlated subcluster in Fig. 1A-1 and, in fact, most of that subcluster involves direct regulation via an Hrp promoter. Furthermore, those genes without Hrp promoters lie in the less tightly clustered subset at the bottom of subcluster 1. Also shown in blue font in Fig. 1B are those genes with Hrp promoters that were identified to be HrpL dependent via real time-PCR. (Their microarray expression levels also are shown, but not clustered; the fact that those expression levels are mostly black reflects the fact that those genes did not exhibit significant differential expression in the microarray.)

Construction of refined motif models from Gibbs sampling and real time-PCR.

The Gibbs-derived motif training set (level 1, described above) was augmented and realigned with Hrp promoter sequences upstream of genes confirmed to be HrpL dependent via real time-PCR, resulting in a refined training set ("level 2") and associated HMM and PSWM motif models which were

Table 1. Results of real-time polymerase chain reaction analysis

PSPTO no.[a]	ΔLogC: WT vs. Δ*hrpL*[b]	Fold change: WT vs. Δ*hrpL*	ΔLogC: Δ*hrp*/*hrc*+*hrpL* vs. Δ*hrp*/*hrc*	Fold change: Δ*hrp*/*hrc*+hrpL vs. Δ*hrp*/*hrc*	Protein
0044	0.85 ± 0.55	7.10	0.38 ± 0.22	2.41	HopK1
0474	0.45 ± 0.27	2.80	0.65 ± 0.64	4.51	Hypothetical protein
0524	0.83 ± 0.42	6.71	0.86 ± .045	7.22	Peptidase, M20/M25/M40
0834	1.89 ± 0.53	76.80	1.82 ± 0.33	65.76	Alcohol dehydrogenase
5617*	0.52 ± 0.37	3.34	0.22 ± 0.27	1.64	Unannotated
1394	1.95 ± 0.51	88.70	N/A	N/A	HrcS
1568	1.71 ± 0.47	51.10	1.13 ± 0.32	13.38	HopAF1
1645	0.65 ± 0.56	4.50	0.19 ± 0.15	1.54	Transcr. regulator, MarR family
2105	1.51 ± 0.41	32.02	1.10 ± 0.39	12.61	ApbE-family protein
2691	0.80 ± 0.34	6.28	0.13 ± 0.15	1.35	Membrane prot., TerC family
3576	–0.05 ± 0.27	0.90	–0.03 ± 0.15	0.93	TvrR (Preiter et al. 2005)
3648	1.41 ± 0.49	25.59	0.57 ± 0.37	3.70	Acid phosphatase
4718	1.82 ± 0.43	66.46	1.07 ± 0.28	11.80	HopAA1-2
4732	1.64 ± 0.48	43.23	0.71 ± 0.30	5.10	HopQ1-2
4733	1.01 ± 0.39	10.22	0.26 ± 0.14	1.83	Hypothetical protein
4776	2.51 ± 0.65	323.10	1.52 ± 0.62	33.49	HopI1
B0005	1.39 ± 0.45	24.73	0.59 ± 0.30	3.92	Phosphoesterase
1287	0.07 ± 0.29	1.18	–0.08 ± 0.12	0.82	Gap-1 (housekeeping)
1745	–0.03 ± 0.23	0.93	0.10 ± 0.15	1.26	GyrA (housekeeping)

[a] PSPTO number indicates gene tested. An asterisk indicates a new gene call based on this work.

[b] ΔLogC is the difference of the logarithm (base 10) of relative mRNA concentrations between the two identified strains or conditions of DC3000. The values reported represent the mean ± the standard deviation of three biological replicates for each sample; normalization to levels for *gap-1* and *gyrA* as described in Materials and Methods. Fold change represents $10^{\Delta \mathrm{LogC}}$. Differential expression for Δ*hrp*/*hrc*+*hrpL* versus Δ*hrp*/*hrc* was not available (N/A) for PSPTO1394 because it is deleted in that set of strains. Abbreviations in the protein designations: transcr. = transcriptional, biosynth. = biosynthesis, prot. = protein.

used to rescan the DC3000 genome for putative Hrp promoters. The refinement of the motif model associated with the incremental sequences added in level 2 did not produce HMM and PSWM scores that differed dramatically from their level 1 values. Nor did either the level 1 or level 2 scores differ substantially from those associated with our level 0 training set; namely, that reported by Fouts and associates (2002) with corrections as described below. Receiver-operator characteristic (ROC) scores indicate (data not shown) that all three sets of models (levels 0, 1, and 2) have roughly equal discriminating power (area under the ROC curve). All are highly discriminating, and our new level 2 models do not represent a significant refinement over the previous level 0 models. A sequence logo (Crooks et al. 2004) for this refined motif is shown in Figure 3, juxtaposed with that derived from the level 0 sequence training set.

The results of our genomewide scans for Hrp promoter motifs are summarized in Figure 4, which plot maximal expression fold change versus HMM level 2 score, for each operon for which an HMM score above a threshold of 6 is registered in the upstream promoter region (Fig. 4A); and distance from the start of a putative Hrp promoter to the annotated transcription start of the downstream target operon (Fig. 4B). Distances less than zero (i.e., below the dashed line) represent Hrp promoters upstream of the start site; negative distances near 0 (e.g., within 500 bases of the start site) most plausibly represent functional promoters. (As noted below, however, in some cases, where strong Hrp promoters occur far upstream of the nominal transcription start, we suspect that small unannotated genes lie more closely downstream of the promoter.) The lack of strong expression shown in Figure 4A for HMM scores <10 and the prevalence of presumably nonfunctional promoter start sites in Figure 4B for HMM score <10 suggest that this represents a reasonable cutoff for functional Hrp promoters. The HMM was constructed from training data for genes exhibiting significant differential expression; therefore, the fact that highly scoring Hrp promoters in Figure 4A correspond to upregulated genes is not surprising. Because the HMM sequence model reflects sequence similarity rather than actual binding affinity, however, we should not necessarily expect a strong correlation between score and expression level within the upregulated set (nor did we observe one).

Several Hrp promoter hits plotted in Figure 4A have high scores but are upstream of genes whose expression changes fell below our cutoff. A significant fraction of those composed the set retested via real time-PCR, as described above. Five additional genes showed differential expression with a fold change greater than 1.25, but missed one or more of the threshold criteria (PSPTO0371, PSPTO1377, PSPTO4589, PSPTO4718, and PSPTO4722). Three genes (PSPTO1392, PSPTO1393, and PSPTO1400) are members of the large HrpJ operon. All the remaining members of that operon exhibited significant differential expression and are included in the DEG-ALL set, and we infer that the missing three also should exhibit differential expression. In addition, several genes that were not on the microarray (PSPTO1387, PSPTO1390, PSPTO1394, PSPTO1408, PSPTO2679, and PSPTO4588) are not shown in Figure 4A. HrpL-dependent expression of PSPTO1394 was confirmed via real time-PCR. Based on their strong promoters, however, we would expect the rest of that missing set to show significant differential expression.

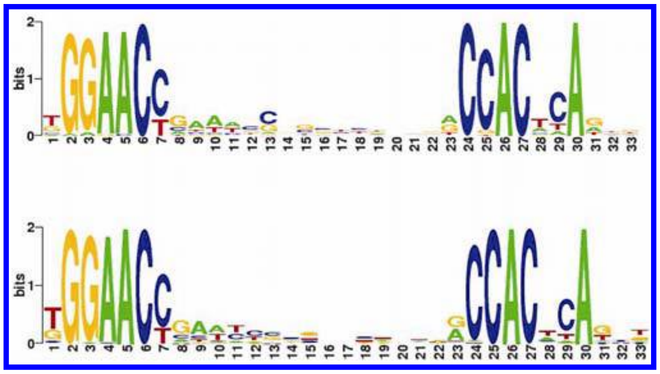


Fig. 3. Sequence logos of hypersensitive response and pathogenicity promoter motifs derived from Gibbs sampling of candidate regulatory regions identified by differential expression in microarray experiment (top) and by Fouts and associates (2002) based on literature searching and mini-Tn5-gus mutagenesis (bottom).

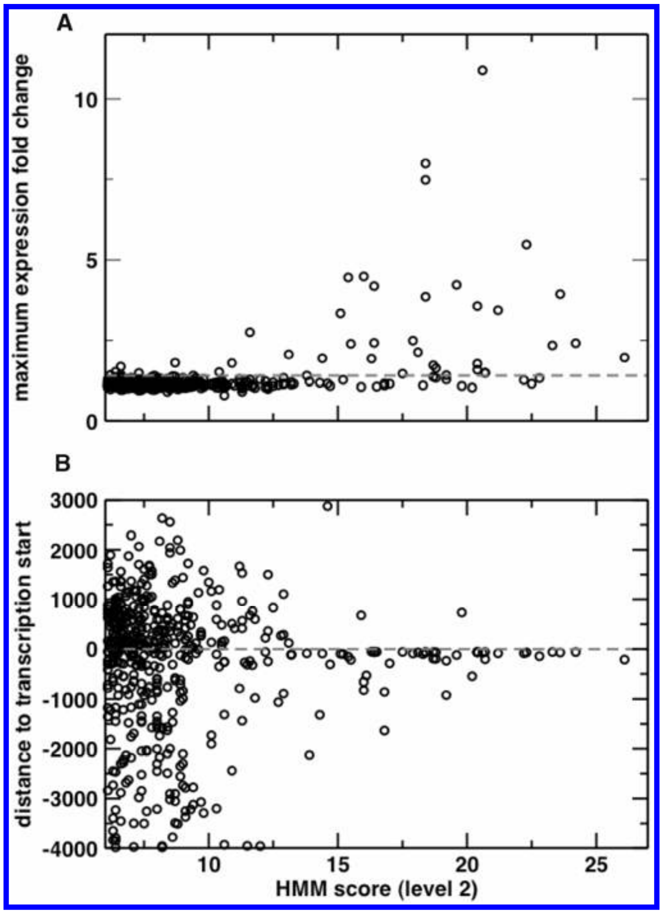


Fig. 4. Relationship between hidden Markov model (HMM) level 2 score and maximum HrpL-dependent expression change and distance of promoter from transcription start. **A,** Maximal expression fold change of target operon versus hypersensitive response and pathogenicity (Hrp) promoter HMM level 2 score (circles), along with the expression fold change cutoff used in our microarray analysis (dashed line). Some target operons have multiple putative Hrp promoter hits; in these cases, we keep only the Hrp promoter with the highest score. **B,** Distance from start of Hrp promoter to annotated start of transcription versus Hrp promoter HMM score. Distances less than zero (i.e., below the dashed line) represent Hrp promoters upstream of the start site; negative distances near 0 (e.g., within 500 bases of the start site) most plausibly represent functional promoters. The lack of strong expression in A for HMM score <10 and the prevalence of presumably nonfunctional promoter start sites in (b) for HMM score <10 suggest that this represents a reasonable cutoff for functional Hrp promoters. Hrp promoters in A that have a high score but exhibited differential expression below the threshold indicated either were among those subsequently tested for differential expression via quantitative real-time polymerase chain reaction, were not on the microarray, or belonged within putative operons with other genes that did exhibit differential expression.

Our weight matrix (WM) models form a complementary perspective on candidate Hrp promoters. Generally, the HMM and WM scores are well correlated with each other, although candidate promoter regions identified by the two models can differ at the margins (i.e., near an HMM cutoff score of 10.0 and a WM cutoff score of 4.0). We do not combine these separate scores into an aggregate score; rather, each model serves as a check on the other. We have found some candidate promoters for which there is experimental evidence of HrpL dependence, but whose HMM scores fall below our cutoff of 10 (both in DC3000, as reported here, and in *P. syringae* pv. *phaseolicola* 1448A, as reported by Vencato and associates [2006]). In all these cases, we find evidence of moderately strong coexisting WM scores of each variant (i.e., 16- and 17-bp spacers). It is possible that cooperative binding among the two promoter variants helps to stabilize binding of the RNA polymerase holoenzyme to the DNA, resulting in higher affinity binding than might be suggested by the bioinformatic models themselves.

Corrections to genome annotation from bioinformatics and experimental evidence.

Our experimental results and analysis of Hrp promoters suggest several revisions to the original genome annotation presented by Buell and associates (2003), which we have compiled in an updated submission to GenBank to accompany this article. In several cases, we have observed strong Hrp promoters somewhat far upstream of genes (e.g., more than 200 bp) for which there is experimental evidence of HrpL-dependent expression. One such region, upstream of PSPTO0869, was tested and verified for HrpL-dependent transcription via real time-PCR (Table 1). Subsequent analysis reveals this region, which we have annotated with the new identifier PSPTO5617, to be a pseudogene with similarity to annotated genes in other pseudomonads. Another new gene call, upstream of PSPTOB0005 and now designated as PSPTOB0078, corresponds to a region downstream of a putative promoter in an unannotated region. A translated peptide fragment matching this region was detected by mass spectroscopy (data not shown). Furthermore, comparative analysis of long (nominally) untranslated regions (UTRs) downstream of candidate Hrp promoters, as well as unannotated regions in the vicinity of candidate promoters, reveals strong sequence similarity with orthologous regions in other sequenced pseudomonads. Given this evidence, we have decided to make calls for nine new genes, designated as PSPTO5616 to PSPTO5622, PSPTOA0072, and PSPTOB0078. A fragmentary pseudogene in the operon containing PSPTOA0017-A0019 (Guo et al. 2005) is consistent with our bioinformatic analysis indicating the existence of a single Hrp promoter upstream of the entire set of genes. Finally, our updated genome annotation file explicitly contains Hrp promoter features that were identified bioinformatically and which have experimental evidence of HrpL-dependent transcription downstream. It should be noted that the corrected annotations typically occur in regions that have an anomalously low GC content relative to the rest of the genome or involve pseudogenes that are interrupted by premature stop codons or transposon insertions. These pseudogenes were detected by DNA- and protein-based sequence comparisons with apparently intact genes in other *Pseudomonas* spp.

Summary of genes regulated by Hrp promoters and methods used to identify them.

Genes preceded by Hrp promoters for which we have reported either direct experimental evidence for HrpL-dependent transcription via microarray or real time-PCR (corresponding to the genes highlighted in blue in Figure 1), or for which we have inferred such transcription based on inclusion within a putative operon demonstrating HrpL-dependent transcription, are listed in Table 2. A summary flow plan for the iterative bioinformatic and experimental methods that were used to identify Hrp promoters in DC3000 is shown in Figure 5.

DISCUSSION

We have used a combination of approaches involving microarray analysis, Gibbs sampling, hidden Markov modeling, and real time-PCR to comprehensively identify HrpL-responsive genes in *P. syringae* pv. *tomato* DC3000 and to develop a robust bioinformatic protocol for predicting Hrp promoters in all *P. syringae* genomes. We discuss below the methods we have developed and the implications of our findings in the context of other studies of virulence regulons and regulatory networks and the potential roles in pathogenesis of the genes in DC3000 that are regulated by HrpL.

Methodology for identifying regulons.

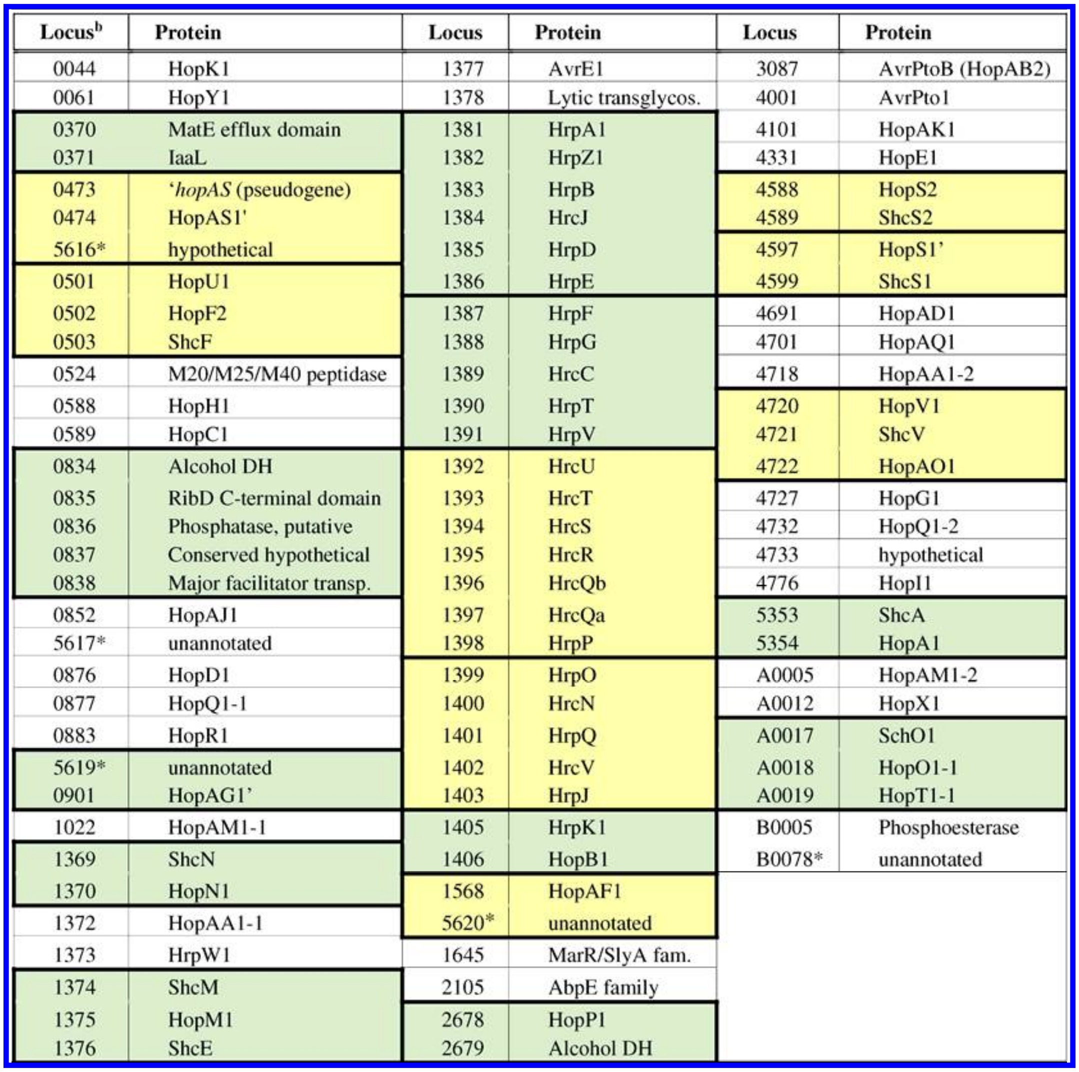
Although much previously was known about the HrpL regulon in DC3000, we wanted to ascertain the effectiveness of combining microarray expression data with clustering and bioinformatics techniques to identify putative regulons, without necessarily biasing such a search previously identified promoter sequences or coregulated genes. In this regard, our effort here has been rather successful; clustering analysis of the microarray expression data consistently reveals a well-correlated cluster that contains all of the known T3SS-related genes (as well as some others not so obviously related). Furthermore, Gibbs sampling of the entire set of upstream sequences implicated by the microarray consistently uncovers a promoter motif consistent with the Hrp promoter identified in previous work. The set of Hrp promoters discovered by the Gibbs sampling and subsequent HMM and PSWM scans is quite well correlated with the set of genes identified by clustering of expression profiles. The de novo reconstruction of the Hrp promoter motif did not use any prior knowledge of the motif (other than its approximate size expected on the basis of fact that HrpL is an ECF-type sigma factor). We expect a similar methodology to be applicable to the search for regulons of other sigma factors and transcription factors for which less is known experimentally. With discriminating HMM and PSWM models of Hrp promoter motifs in hand, we also can scan other sequenced genomes for candidate Hrp promoters. Application of this approach to *P. syringae* pv. *phaseolicola* 1448A is described in a companion article (Vencato et al. 2006). In addition, this methodology has suggested both further experimentation to clarify the HrpL regulon and further refinements to the original genome annotation by focusing our attention on bioinformatically significant promoters that were not immediately upstream of differentially expressed target genes. In several cases, we have been able to use such information to annotate new hypothetical proteins whose functions can be the focus of future study.

Comparisons with genome-enabled analyses of Hrp regulons in phytopathogenic bacteria.

A study analogous to ours used a 70-mer oligonucleotide-based microarray to identify genes differentially regulated in *Ralstonia solanacearum* wild-type and *hrpB* mutant strains (Occhialini et al. 2005). HrpB is an AraC-type positive activator of genes encoding the T3SS and its substrates and is important to the virulence of this bacterial wilt pathogen. The study by Occhialini and associates revealed that HrpB activates previously unknown candidate effectors as well as a variety of genes with functions unrelated to the T3SS. These included genes involved in chemotaxis and the metabolism of various low molecular weight compounds. The *R. solanacearum* Hrp regulon also included several transcriptional regulators as well as genes without an obvious HrpB-responsive promoter. Global study of the expression of *R. solanacearum* virulence genes also has been furthered by a genetic screen for HrpB-activated genes and by an in vivo expression technology analysis of genes that are differentially expressed in planta (Brown and Allen 2004; Mukaihara et al. 2004). These studies reveal overlapping sets of genes that are likely key to the virulence of *R. solanacearum*. In general, the HrpB regulon of *R. solanacearum* appears to be larger and more complex than the HrpL regulon of *P. syringae*.

Other experimental approaches also have provided evidence for HrpL-dependent regulation in DC3000, although in conditions involving *hrpL* overexpression in rich medium rather than an *hrpL* deletion in Hrp minimal medium (Chang et al. 2005; Fouts et al. 2002). We expect that some of the differences with previous work are due to these differences in the protocol. Regarding the two previous genetic screens, the reporter transposon screen of Fouts and associates (2002) was not intended to be saturating, whereas the more recent DFI screen of Chang and associates (2005) is more nearly so. The data from both genetic screens corroborate the bioinformatic analysis presented here: 35 operons with bioinformatically identified Hrp promoters were found to be HrpL dependent in both our *hrpL* deletion experiments and the earlier HrpL overexpression studies (Chang et al. 2005; Fouts et al. 2002). Ten operons not previously identified using HrpL overexpression were identified as differentially expressed in our Δ*hrpL*-based assays. Eleven operons were differentially expressed under HrpL overexpression conditions but not in the *hrpL* deletion conditions used here. Of those 11, 5 exhibited some differential expression in our microarray experiment (fold change ≥1.27) but failed to meet all of our criteria for significant differential expression; another 3 of the 11 operons were not printed on our microarray and were not subsequently tested via real time-PCR. The significance during infection conditions of those operons that respond in culture to HrpL only when overexpressed is an important question for future research.

Table 2. Genes in putative operons that are preceded by HrpL-responsive promoters[a]



Locus[b]	Protein	Locus	Protein	Locus	Protein
0044	HopK1	1377	AvrE1	3087	AvrPtoB (HopAB2)
0061	HopY1	1378	Lytic transglycos.	4001	AvrPto1
0370	MatE efflux domain	1381	HrpA1	4101	HopAK1
0371	IaaL	1382	HrpZ1	4331	HopE1
0473	'*hopAS* (pseudogene)	1383	HrpB	4588	HopS2
0474	HopAS1'	1384	HrcJ	4589	ShcS2
5616*	hypothetical	1385	HrpD	4597	HopS1'
0501	HopU1	1386	HrpE	4599	ShcS1
0502	HopF2	1387	HrpF	4691	HopAD1
0503	ShcF	1388	HrpG	4701	HopAQ1
0524	M20/M25/M40 peptidase	1389	HrcC	4718	HopAA1-2
0588	HopH1	1390	HrpT	4720	HopV1
0589	HopC1	1391	HrpV	4721	ShcV
0834	Alcohol DH	1392	HrcU	4722	HopAO1
0835	RibD C-terminal domain	1393	HrcT	4727	HopG1
0836	Phosphatase, putative	1394	HrcS	4732	HopQ1-2
0837	Conserved hypothetical	1395	HrcR	4733	hypothetical
0838	Major facilitator transp.	1396	HrcQb	4776	HopI1
0852	HopAJ1	1397	HrcQa	5353	ShcA
5617*	unannotated	1398	HrpP	5354	HopA1
0876	HopD1	1399	HrpO	A0005	HopAM1-2
0877	HopQ1-1	1400	HrcN	A0012	HopX1
0883	HopR1	1401	HrpQ	A0017	SchO1
5619*	unannotated	1402	HrcV	A0018	HopO1-1
0901	HopAG1'	1403	HrpJ	A0019	HopT1-1
1022	HopAM1-1	1405	HrpK1	B0005	Phosphoesterase
1369	ShcN	1406	HopB1	B0078*	unannotated
1370	HopN1	1568	HopAF1		
1372	HopAA1-1	5620*	unannotated		
1373	HrpW1	1645	MarR/SlyA fam.		
1374	ShcM	2105	AbpE family		
1375	HopM1	2678	HopP1		
1376	ShcE	2679	Alcohol DH		

[a] List of genes with bioinformatically-identified upstream Hrp promoters, for which there is experimental evidence for HrpL-dependent transcription or the gene is in a putative operon with other genes that are HrpL-dependent.

[b] Genes are listed in order of their PSPTO locus number identifiers, with putative operons boxed with heavy black lines. Genes are presented in the order of their PSPTO number along the DC3000 chromosome, with plasmid-borne effector genes (PSPTO numbers preceded with letter A or B) being presented last. New gene calls included in our updated GenBank submission are noted with asterisks and placed in their appropriate operons. Designations for T3SS substrates follow current nomenclature recommendations (Lindeberg et al. 2005). Operons located on the plus strand are shaded green and those on the minus strand are shaded yellow. Abbreviations in protein designations: DH = dehydrogenase, facil. = facilitator, fam. = family, transglycos. = transglycosylase, and transp. = transporter.

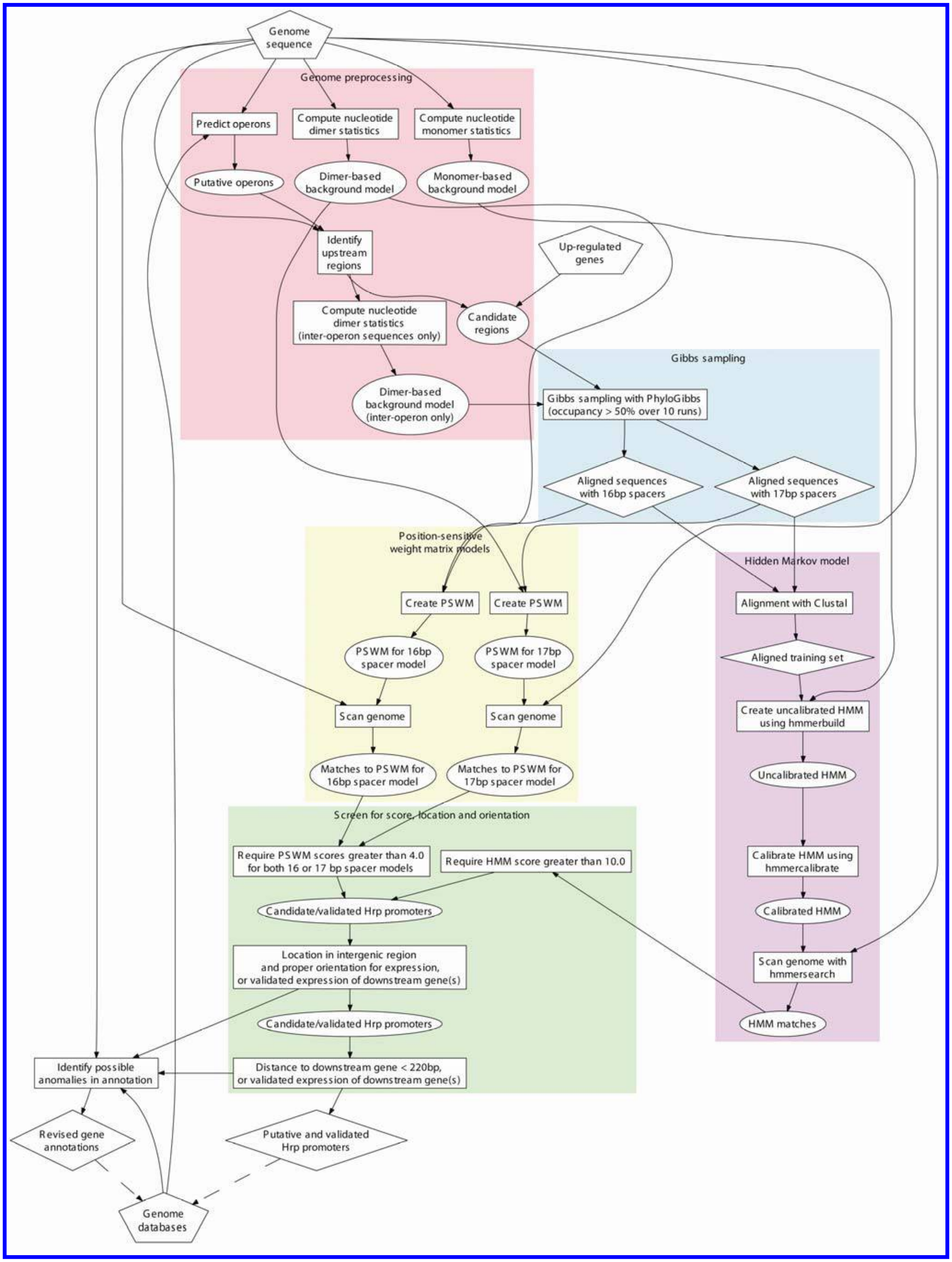


Fig. 5. Bioinformatic workflow represented as a directed graph. The major processing modules are grouped and highlighted by color. The nodes representing primary inputs, primary outputs, processing steps, and intermediate datasets are denoted with pentagons, diamonds, squares and ellipses, respectively. Feedback pathways that connect primary outputs to primary inputs are indicated by dashed arrows.

Implications for regulatory networks and evolution.

Our search for regulatory binding motifs upstream of differentially expressed genes found no consistent motifs other than the Hrp promoter presumably bound by HrpL. The Hrp promoter signal is rather well conserved in DC3000, with a significant number of binding motifs in the genome, so perhaps our methods were able to detect that strong signal but missed weaker or fuzzier ones. On the other hand, the lack of other regulatory motifs upstream of the T3SS suggests that regulation of the T3SS is simply and directly controlled only by HrpL in Hrp minimal media. Comparisons between different *P. syringae* pathovars and other pseudomonads reveal that there is a great deal of genomic shuffling and transposition in the segments of the genome associated with the T3SS and its effectors (Feil et al. 2005; Joardar et al. 2005a,b). This shuffling appears to be associated with the needs of host specificity, which is largely what differentiates various pathovars. We hypothesize that adaptations to changes in host susceptibility targets and defense surveillance systems are countered by a dynamic process of effector gene horizontal acquisition and inactivation. This may make advantageous the simple and direct regulation of the T3SS through a single sigma factor rather than through the combinatorial control of multiple factors. In the absence of control by other factors, the timing of the synthesis and delivery of T3SS components may be governed by promoter affinity levels (e.g., as has been hypothesized for flagellar biosynthesis) (Kalir et al. 2001), a scenario that requires further investigation.

Relationship between expression patterns and potential biological roles of HrpL-regulated genes.

Most of the genes in the strongly up-regulated DEG-UP cluster 1 are associated with Hrp promoters with an HMM2 score >10, and most of these genes encode components of the T3SS and its substrates (Fig. 1; Table 2). The induction kinetics of the HrpL regulon observed here are consistent with that previously observed in *P. syringae* pv. *phaseolicola* using real time-PCR and genes encoding the T3SS and four substrate proteins (Thwaites et al. 2004). The genes in one subcluster are activated particularly rapidly. These encode the HrpA1 pilus protein, the HrpW1 and HrpZ1 harpins, the HrpK1 putative translocator, and the duplicated HopAM1 encoding genes (the “1” in each Hrp name follows the new nomenclature recommendations for T3SS substrates of Lindeberg and associates [2005]). Of these, HrpA1 and HrpK1 are involved in the initial deployment of the extracellular T3SS machinery. The rapid induction of the harpin and HopAM1 genes suggest that these are among the first proteins to travel the Hrp T3SS. Several of the HrpL-activated genes in DC3000 may be subject to more complex regulation and their relationship with the T3SS is not obvious. We will discuss several examples below.

Phytotoxin and phytohormone genes. The *corR* and *cfl* genes control the biosynthesis of the phytotoxin coronatine (Bender et al. 1999). Coronatine is a methyl-jasmonate mimic whose activity antagonizes the plant salicylic acid-dependent defense pathway (He et al. 2004; Zhao et al. 2003). *P. syringae* pv. *tomato* DC3000 coronatine-deficient mutants are reduced in virulence in *Arabidopsis* (Brooks et al. 2004; Mittal and Davis 1995). The *corR* gene (PSPTO4704) is preceded by an Hrp promoter with an HMM2 score >10, and it appears to be in the same operon as *hopAQ1* (PSPTO4703). The *corR* gene was not differentially expressed in our microarray analysis, which employed a Δ*hrpL* mutant, but it was identified as HrpL responsive in the DFI analysis, which used overexpression of *hrpL*. The coronatine biosynthesis *cfl* genes are not linked with any discernable Hrp promoter but, nevertheless, are activated by HrpL, presumably through HrpL-activated CorR (Fouts et al. 2002). The significance of HrpL activation of the *iaaL* indoleacetetate lysine ligase gene (PSPTP0371) is unclear. As noted before, DC3000 carries genes necessary to produce the phytohormone indole acetic acid (IAA) and also *iaaL*, whose product would be expected to produce inactive IAA conjugates (Buell et al. 2003). The ability to produce IAA is widespread among *P. syringae* pathovars and other phytopathogenic bacteria; however, only DC3000 among the sequenced *P. syringae* pathovars carries *iaaL*. (Feil et al. 2005; Joardar et al. 2005a). Similar to *corR*, *iaaL* is expressed differentially in response to overexpression of *hrpL* but not in wild-type DC3000 in experiments involving the Δ*hrpL* mutant for comparison.

Twin-arginine substrates. Both PSPTOB0005 and PSPTO3648 were recently identified as secreted by the twin-arginine (TAT) pathway in DC3000 (Bronstein et al. 2005). Both were identified as being differentially expressed in the microarray, although we have since realized that both genes should not have been on the array because of their high sequence identity (approximately 90% nucleotide similarity). Because PSPTOB0005 has an upstream Hrp promoter and PSPTO3648 does not, we assume that the strong differential expression observed for both was due to the expression of PSPTOB0005 and subsequent cross-hybridization. Nonetheless, because Bronstein and associates (2005) established a connection between the TAT and T3SS pathways, confirmation of HrpL-dependent expression for one of those substrates is intriguing.

Myo-inositol metabolism. The region PSPTO3500 to PSPTO3492 includes several operons that encode enzymes and transporters related to myo-inositol metabolism. None of these has a detectable Hrp promoter but all are tightly clustered in a subgroup of group 3 (Fig. 1). This subgroup is characterized by strong HrpL-dependent upregulation 3 h after the shift to inducing conditions followed by rapid downregulation. In contrast, genes nominally associated with other aspects of phytate and inositol phosphate metabolism (e.g., PSPTO1419, PSPTO3047, and PSPTO3241) are not differentially expressed. It is known that phosphoinositide-specific phospholipase C is transiently expressed by soybean cells during infection by *P. syringae* pv. *glycinea* with large concomitant changes in the concentrations of inositol polyphosphate species within host cells (Shigaki and Bhattacharyya 2000). This class of phospholipases also has been implicated in other biotic and abiotic stress responses in *Arabidopsis* (Ton et al. 2005). Therefore, it is possible that Hrp-dependent regulation of the inositol-related catabolic genes in the region PSPTO3500 to PSPTO3492 may have a role in bacterial nutrition or defense suppression during infection.

Sigma factors and other transcription factors. The microarray experiment implicated several sigma factors and transcription factors in the set of differentially expressed genes. Specifically, the sigma factors PSPTO0537 (RpoD), PSPTO1565 (RpoS), and PSPTO2298 (SigW) all were upregulated, although none appear to be under direct control of HrpL (i.e., they do not have upstream Hrp promoters). However, PSPTO1645 (SlyA/MarR family) is both upregulated and preceded by an Hrp promoter. In addition, the negative regulator PSPTO1391 (HrpV) also was upregulated and would appear to be under direct control of HrpL as part of the *hrcC* operon. Conversely, PSPTO1209 (a sigma-70 family protein) and PSPTO4224 (sigma-24) were found to be downregulated in the microarray, as were the transcriptional regulators PSPTO1379 (*hrpR*) and PSPTO1758 (in the TetR family). *hrpR* is not preceded by an Hrp promoter or expected to be expressed in an HrpL-dependent manner. Rather, HrpR is a sigma-54 enhancer-binding protein that activates *hrpL* expression (Xiao et al. 1994). Interestingly, real time-PCR analysis of *P. syringae* pv. *phaseolicola* revealed that *hrpR* expression increased relatively weakly in Hrp minimal medium but much more strongly in planta (Thwaites et al. 2004). However, the focus of this work is on genes downstream of HrpL, rather than upstream, in the Hrp regulatory cascade. An obvious open question involves the regulation of all other genes differentially expressed in our microarray that do not have Hrp promoters. Although we have not found compelling evidence for other regulatory binding motifs, effort focused on the set of sigma factors and transcription factors which were themselves differentially expressed may provide insights into the regulation of those other genes.

Conclusions.

The expression of HrpL affects many genes in *P. syringae* pv. *tomato* DC3000, including several with no obvious role in the T3SS. Some of these, such as the coronatine biosynthesis genes, are known to have a role in pathogenesis; however, we think that all of these genes are high-priority candidates for future work by the *P. syringae* research community. As an example, we report in the accompanying article by Vencato and associates (2006) that PSPTO2105 (ApbE-family protein) and PSPTO0834 (alcohol dehydrogenase) mutants are partially reduced in their ability to grow in *Arabidopsis* leaves and that both of these genes have homologs in *P. syringae* pv. *phaseolicola* 1448A that also are preceded by Hrp promoters and activated by HrpL. In summary, we hope that this exhaustive bioinformatic and experimental analysis of the HrpL regulon of *P. syringae* pv. *tomato* DC3000 will provide a solid foundation for future functional analysis of this key virulence regulon in *P. syringae*.

MATERIALS AND METHODS

Microarray design, fabrication, and use.

P. syringae strains were grown in King's B (KB) (King et al. 1954) or *hrp*-derepressing fructose minimal medium (Hrp MM) (Huynh et al. 1989). DNA sequences corresponding to putative genes were identified using a draft version of the DC3000 genome. Primers were selected using Primer 3.0 (Rozen and Skaletsky 2000) to exhibit a melting temperature between 55 and 65°C, a GC content of 50 to 65%, and a length of 18 to 25 nucleotides. Amplicon lengths were constrained to range from 300 to 800 bp. Design was unsuccessful using these criteria for 527 genes and they were omitted from the array. These included two T3SS-related genes, *hrpF* (PSPTO1387) and *hrpT* (PSPTO1390). In cases where groups of genes were highly similar (>60 bp with >80% nucleotide identity), we chose one representative gene in each set in order to minimize cross-hybridization. All transposase genes were excluded due to their high abundance. The final array contained 5,033 unique amplified DNAs representing at least 4,843 of 5,673 predicted genes. Differences between the final version of the genome and the version used for primer design resulted in the inclusion of a small number of duplicated genes on the array as well as some sequences that are not annotated as genes.

The 5,033 sequences were PCR amplified from purified *P. syringae* pv. *tomato* DC3000 genomic DNA. PCR reactions were analyzed using 96-well Ready-To-Run precast agarose gels (Amersham Biosciences, Piscataway, NJ, U.S.A.), and >99% were determined to have successfully yielded a single product. PCR products were purified using a Genesis RSP200 Liquid Handler (TECAN, Maennedorf, Switzerland) and 384-well filter plates (S384PCR10; Millipore, Bedford, MA, U.S.A.), vacuum dried, and resuspended in 10 ml of Pronto! spotting solution (Corning, Corning, NY, U.S.A.). Samples were spotted onto γ-amino-propyl-silane coated UltraGAPS slides (Corning) using a MicroGrid Pro arrayer (BioRobotics, Cambridge) with 32 MicroSpot2500 printing pins. PCR products were fixed to the modified glass slides by treatment with 250 mJ of UV irradiation followed by a 2-h incubation at 85°C. An additional 806 PCR-control (blank) spots, containing primerless PCR reactions, were included on the array, as well as 137 spots containing only spotting solution (empty) for a total of 6,144 spots. Following fabrication, slides were stored in a dust-free plexiglass chamber (approximately 21°C, 0% relative humidity) in protective storage boxes (Corning). To test the microarray, a single-replicate pilot experiment was conducted comparing genomic DNA from wild-type DC3000 with that from a strain containing a deletion in the *hrp*/*hrc* pathogenicity island (Fouts et al. 2003). The results correctly identified >95% of the genes deleted in the mutant strain (data not shown).

Two strains of DC3000, a wild-type strain and a strain containing a deletion in *hrpL* (UNL-134-1), were used for gene expression profiling. DC3000 UNL-134-1 has a 2-kb deletion affecting *hrpL*. Bacterial cells grown overnight at 28°C in KB containing rifampicin at 50 μg/ml were harvested by centrifugation at 4,000 rpm for 5 min and resuspended in Hrp MM (Huynh et al. 1989) to an optical density at 600 nm (OD_{600}) of approximately 0.5. Cell cultures then were divided in 10-ml aliquots and harvested at 0, 2, 3, 4, 5, 6, 7, and 9 h after incubation at 18°C. Cell pellets were resuspended in 200 μl of RNAprotect (Qiagen, Valencia, CA, U.S.A.) prior to storing at –20°C. Three independent time course experiments were performed.

Total RNA was isolated using RNeasy midi kits (Qiagen) following the manufacturer's protocol. The optional on-column DNA digestion was performed. Isolated total RNA (7 μg) was reverse transcribed in the presence of Cy3 or Cy5 dCTP using the ChipShot labeling system (Promega Corp., Madison, WI, U.S.A.) according to the manufacturer's protocol. cDNA of a given time point was labeled with either Cy3 or Cy5, such that each time point was differentially labeled in two of the three biological replicates. Labeled cDNAs were mixed following an incomplete block design (Chu et al. 2002; Churchill 2002; Kerr and Churchill 2001), and purified using QIAquick PCR purification spin columns (Qiagen). Purified labeled cDNAs were dried and resuspended in 70 μl of Pronto! Universal Hydridization Solution (Corning). Meanwhile, microarray slides were prepared for hybridization according to the Pronto! systems protocol, but skipping the sodium borohydride wash step. Hybridization was carried out under lifter slips (Erie Scientific, Portsmouth, NH, U.S.A.) for 14 to 16 h at 42°C. Slides were washed according to the Pronto! systems protocol, and then were dried immediately by centrifuging at 750 rpm for 5 min.

Microarray slides were scanned using a ScanArray5000 two-channel laser confocal microarray scanner (GSI Lumonics, Billerica, MA, U.S.A.) and the associated ScanArray software (v. 3.1; Packard BioChip Technologies, Boston, MA, U.S.A.) to obtain fluorescence images. Numerical representations of the fluorescence signals were obtained from the fluorescence images using the ImaGene software package (v. 5.6; Biodiscovery Inc., El Segundo, CA, U.S.A.). Numerical data were adjusted to account for hybridization inconsistencies across the surfaces of individual slides by implementing the spatial lowess algorithm (Cui et al. 2002) in the R environment for statistical computing. Data then were imported into the GeneSpring microarray data analysis environment (v. 6.1; Silicon Genetics, Redwood City, CA, U.S.A.), where the default intensity-dependent normalization was applied. The data set was analyzed for genes showing an HrpL-dependent expression pattern. To be assigned to this group, a gene was required to have data for at least one time point with i) an average spot fluorescent intensity >4,000, ii) a Student's *t* test *P* value <0.05, and iii) a wild-type:Δ*hrpL* or a Δ*hrpL*:wild-type ratio >1.4.

Real time-PCR.

All RNA extractions were performed using RNase-free tubes, and filter tips and equipment were treated with RNAzap (Ambion, Austin, TX, U.S.A.). real time-PCR-grade water (Ambion) also was used for all procedures. A 2-ml overnight culture of each strain was grown in KB supplemented with rifampicin at 50 μg/ml and kanamycin at 50 μg/ml. The cultures were reinoculated into 10 ml of KB media supplemented with rifampicin at 50 μg/ml and kanamycin at 50 μg/ml to an OD_{600} of 0.1 and allowed to grow to an OD_{600} of 0.2. Cells then were harvested from 1 ml of culture by centrifugation at 13,000 × *g* for 5 min and flash frozen in liquid nitrogen. Total RNA was prepared from cell pellets using the SV Total RNA Isolation System (Promega Corp.) according to the manufacturer's instructions. An additional DNase I digestion step was performed with DNA-*free* (Ambion) according to the rigorous DNase treatment included in the manufacturer's instructions. This additional DNase step was necessary to remove trace DNA contamination in the RNA preparation. RNA was purified further using the RNeasy Minikit (Qiagen) following the manufacturer's instructions. Quantification and sample integrity of RNA samples was assessed using spectrophotometer readings at an absorbance of 260 nm (A_{260}) and the $A_{260}/_{\mathrm{A280}}$ ratio, respectively.

Real-time PCR was performed by using the ABI 7000 Sequence Detection System (Applied Biosystems) and iTaq SYBR Green Supermix with Rox (Biorad) following the manufacturer's protocols. Total RNA (100 ng) extracted from strains of *P. syringae* pv. *phaseolicola* 1448a was reverse transcribed in a thermocycler using a cDNA synthesis kit (Biorad) according to the manufacturer's instructions. Then, 1 μl of the resulting total cDNA population was mixed with 0.3 μM concentrations of each primer and 12.5 μl of master mix in a 25-μl final volume. The PCR assay was carried out with one cycle at 95°C for 2.5 min followed by 35 cycles of 95°C for 15 s and 60°C for 30 s. The amount of fluorescence that resulted from the incorporation of the SYBR Green dye into double-stranded DNA was measured at the end of each cycle to determine the kinetics of PCR for each sample. DNA contamination and the formation of primer dimers were assessed by using controls lacking reverse transcriptase and template, respectively. The production of nonspecific products was determined by the dissociation protocol included in the software provided with the ABI 7000 real-time PCR machine. The resulting threshold cycle (Ct) values were calculated by the ABI 7000 software and analyzed using the relative standard curve method (separate tubes) described in ABI User Bulletin no. 2. In each strain, the Ct values of each gene tested were normalized to the Ct values of two housekeeping genes, *gyrA* and *gap1,* separately, and these values then were averaged to obtain relative expression data for each gene.

Operon analysis and construction of candidate regulatory regions.

Assignment to putative operons used a modification of the procedure described by Collado-Vides and associates (Moreno-Hagelsieb and Collado-Vides 2002; Salgado et al. 2000). The DC3000 genome was segmented into putative transcription units based on log likelihood ratios. In practice, all intraoperon intergenic separations are ≤60 bp. The set of candidate regulatory regions consisted of the regions from –265 to –15 bases relative to the transcription start site of the first gene in the putative transcription unit, unless the end of an upstream gene was reached before the –265 position was reached, in which case the candidate regulatory region was terminated there. Two manual curations to the operon data were made based on previous experimental evidence characterizing the HrpC operon (Deng et al. 1998) and the HrpK operon (Alfano et al. 2000; Petnicki-Ocwieja et al 2005). In addition, for two genes of interest for which separations between gene exceeded 50 bp but were less than 60 bp, two putative operon starts were included: for gene PSPTO0713, sequence upstream of both PSPTO0713 and PSPTO0706 was included in the sampling set; and, for PSPTO0875, sequence upstream of both PSPTO0875 and PSPTO0874 was included. The first 15 bases upstream were stripped off to avoid including recurring ribosome binding sites in the sequence data, which can interfere with the detection of other regulatory motifs. Preliminary Gibbs sampling revealed nearly 150 identical bases upstream of PSPTO1022 (HopAM1-1) and PSPTOA0005 (HopAM1-2); because such sequence identity can unduly influence the identification of clusters within the Gibbs sampler, the sequence upstream of PSPTOA0005 was dropped from the sampling set. Accounting for putative operons sharing upstream sequence and the other modifications described above, the sampling set comprised 97 upstream sequence regions totaling 18,374 nucleotide bases.

Clustering of differential expression profiles.

The cDNA microarray expression clustering package developed by de Hoon and associates (de Hoon et al. 2004) was used through the Biopython interface provided (de Hoon et al. 2003). We performed clustering on the expression profiles of the 194 DEG-ALL genes using several different distance metrics. (In all cases, the default method for linkage clustering—pairwise maximum-linkage clustering—was used.) Clustering on the full set of 194 DEG-ALL genes was done using the uncentered correlation metric, applied to the logarithms of the expression fold changes. Clustering separately on the subset of 119 upregulated genes in DEG-UP and 76 downregulated genes in DEG-DOWN used the raw expression fold change data, with different correlation-based metrics (e.g., Pearson, Spearman rank, and Kendall's tau). In general, the correlation-based clustering methods performed better than distance-based methods. All of the correlation-based clustering methods gave roughly comparable results, especially with regard to the tight clustering of the known T3SS components. We have observed that there often is sizable variation of measured expression levels within a given operon; therefore, it would seem that distance-based metrics may not be especially useful for clustering without appropriate care.

Gibbs sampling.

We use Gibbs sampling techniques to do de novo motif identification in the set of candidate regulatory regions derived from the preliminary operon analysis; that is, to search for common regulatory motifs in the upstream regions of apparently coregulated genes as revealed by microarray expression data. Gibbs sampling was done using the Phylogibbs program by Siddharthan and associates (Siddharthan et al. 2005). Phylogibbs attempts to cluster common sequence motifs that are consistent with a weight matrix model of the motifs, and track the stability of those clusters through the repeated application of Monte Carlo moves which rearrange the organization of motifs and clusters. We have run Phylogibbs for a variety of upstream sequence sets, including DEG-ALL, DEG-UP, and DEG-DOWN. We report here results from sampling upstream of the genes in DEG-UP, having run 10 realizations of the Gibbs sampler, looking for promoter-sized motifs of a width 33 bases, and examining the resulting tracked clusters. For a given run, Phylogibbs reports occupancy statistics for each sequence motif (window) within a given cluster; these occupancies reflect the fraction of sampled configurations for which a sequence is attached to a cluster. We have averaged the occupancy fractions over the 10 runs, assigning an occupancy of 0 if a sequence does not appear in a given cluster in a given run. From the Gibbs sampler and the resulting occupancy statistics, we have identified a set of strongly co-clustered sequences for use in further analysis.

HMMs.

The HMMer 2.3.2 package (Eddy 1998) was used for all HMM searches. The null model for the entire *P. syringae* pv. *tomato* DC3000 genome was created by computing the single nucleotide frequencies from both strands of the complete genomic sequences for the main chromosome (NCBI RefSeq accession NC_004578) and both native plasmids, pDC3000A (NCBI RefSeq accession NC_004633) and pDC3000B (NCBI RefSeq accession NC_004632) using the program compseq from the EMBOSS suite (Rice et al. 2000). When necessary, the submodels identified by Gibbs sampling differing primarily in the separations between the nominal "–35" and "–10" regions were combined by aligning the sequences using ClustalX (Thompson et al. 1997). The HMMs were constructed from the aligned sequences and the null model using the default parameter settings using hmmbuild, then calibrated with hmmcalibrate by sampling 5×10^4 randomly generated sequences with a mean length of 500 bases and standard deviation of 100 bases. The resulting calibrated HMM was used to search both strands of the three genomic sequences listed above using hmmsearch with a score of 10.0 (an E value cutoff of 1×10^{-3}). This cutoff value was determined empirically by examining the matches generated by all of the models and attempting to retain as many matches as possible in intergenic regions with orientations consistent with promoter activity while minimizing the total number of matches occurring in annotated coding regions, in intergenic regions but with orientation inconsistent with promoter activity, and in intergenic regions between convergently transcribed genes.

WM models.

As a secondary, alternative bioinformatic model, the set of co-clustered sequences identified for the Gibbs sampler also has been analyzed with PSWMs, which typically do not allow for gaps; therefore, promoter sequences with different separations between the –35 and –10 boxes must be separated into distinct groups. For the Hrp-promoter-related motifs identified by the Gibbs sampler, we have generated two WMs for separations of length 16 and 17 bp, respectively. The resulting WMs then were used to scan the entire DC3000 genome to look for high-scoring motifs. PSWM-based methods provide no intrinsic cutoff threshold; therefore, a background model was constructed to assess statistical significance. Noncoding regions in DC3000 were assembled and analyzed for the frequency of monomer (N = 1) and dimer (N = 2) motifs, and 100,000 random sequences, drawn from a distribution consistent with the N = 1 and N = 2 statistics, were generated. Raw PSWM scores were converted to scaled scores reflecting the number of standard deviations above the mean of the background score distribution. Motifs with scores above 4.0 for both the 16- and 17-bp spacer models were included as candidates.

ACKNOWLEDGMENTS

This work was supported by the NSF Plant Genome Research Program Cooperative Agreement DBI-0077622. C. R. Myers and G. B. Martin acknowledge support from the United States Department of Agriculture–Agricultural Research Service under Specific Cooperative Agreements 5819074428 and 5819071139, respectively. C. R. Myers wishes to thank R. Siddharthan and E. Siggia for use of, and consultation regarding, their Phylogibbs program.

LITERATURE CITED

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

Arnold, D., Pittman, A., and Jackson, R. 2003. Pathogenicity and other genomic islands in plant pathogenic bacteria. Mol. Plant Pathol. 4:407-420.

Bender, C. L., Alarcon-Chaidez, F., and Gross, D. C. 1999. *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiol. Mol. Biol. Rev. 63:266-292.

Bronstein, P. A., Marrichi, M., Cartinhour, S., Schneider, D. J., and Delisa, M. P. 2005. Identification of a twin-arginine translocation system in *Pseudomonas syringae* pv. *tomato* DC3000 and its contribution to pathogenicity and fitness. J. Bacteriol. 187:8450-8461.

Brooks, D., Hernandez-Guzman, G., Koek, A. P., Alarcon-Chaidez, F., Sreedharan, A., Rangaswarmy, V., Penaloza-Vasquez, A., Bender, C. L., and Kunkel, B. N. 2004. Identification and characterization of a well-defined series of coronatine biosynthetic mutants of *Pseudomonas syringae* pv. *tomato* DC3000. Mol. Plant-Microbe Interact. 17:162-174.

Brown, D. G., and Allen, C. 2004. *Ralstonia solanacearum* genes induced during growth in tomato: An inside view of bacterial wilt. Mol. Microbiol. 53:1641-1660.

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M. J., Haft, D. H., Nelson, W. C., Davidsen, T., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Vanaken, S. E., Feldblyum, T. V., D'Ascenzo, M., Deng, W.-L., Ramos, A. R., Alfano, J. R., Cartinhour, S., Chatterjee, A. K., Delaney, T. P., Lazarowitz, S. G., Martin, G. B., Schneider, D. J., Tang, X., Bender, C. L., White, O., Fraser, C. M., and Collmer, A. 2003. The complete sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186.

Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M., and Dangl, J. L. 2005. A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. Proc. Natl. Acad. Sci. U.S.A. 102:2549-2554.

Chu, T. M., Weir, B., and Wolfinger, R. 2002. A systematic statistical linear modeling approach to oligonucleotide array experiments. Math. Biosci. 176:35-51.

Churchill, G. A. 2002. Fundamentals of experimental design for cDNA microarrays. Nat. Genet. 32(Suppl.):490-495.

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. 2004. WebLogo: A sequence logo generator. Genome Res. 14:1188-1190.

Cui, X., Kerr, K., and Churchill, G. 2002. Transformations for cDNA microarray data. Stat. Apps. Gen. Mol. Biol. 2:1-20.

de Hoon, M. J., Imoto, S., Nolan, J., and Miyano, S. 2004. Open source clustering software. Bioinformatics 20:1453-1454.

de Hoon, M. J. L., Chapman, B., and Friedberg, I. 2003. Bioinformatics and computational biology with Biopython. Genome Inf. 14:298-299.

Deng, W.-L., Preston, G., Collmer, A., Chang, C.-J., and Huang, H.-C. 1998. Characterization of the *hrpC* and *hrpRS* operons of *Pseudomonas syringae* pathovars syringae, tomato, and glycinea and analysis of the ability of *hrpF*, *hrpG*, *hrcC*, *hrpT,* and *hrpV* mutants to elicit the hypersensitive response and disease in plants. J. Bacteriol. 180:4523-4531.

Eddy, S. R. 1998. Profile hidden Markov models. Bioinformatics 14:755-763.

Feil, H., Feil, W. S., Chain, P., Larimer, F., Dibartolo, G., Copeland, A., Lykidis, A., Trong, S., Nolan, M., Goltsman, E., Thiel, J., Malfatti, S., Loper, J. E., Lapidus, A., Detter, J. C., Land, M., Richardson, P. M., Kyrpides, N. C., Ivanova, N., and Lindow, S. E. 2005. Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000. Proc. Natl. Acad. Sci. U.S.A. 102:11064-11069.

Fouts, D. E., Abramovitch, R. B., Alfano, J. R., Baldo, A. M., Buell, C. R., Cartinhour, S., Chatterjee, A. K., D'Ascenzo, M., Gwinn, M. L., Lazarowitz, S. G., Lin, N.-C., Martin, G. B., Rehm, A. H., Schneider, D. J., van Dijk, K., Tang, X., and Collmer, A. 2002. Genomewide identification of *Pseudomonas syringae* pv. *tomato* DC3000 promoters controlled by the HrpL alternative sigma factor. Proc. Natl. Acad. Sci. U.S.A. 99:2275-2280.

- Fouts, D. E., Badel, J. L., Ramos, A. R., Rapp, R. A., and Collmer, A. 2003. A *Pseudomonas syringae* pv. *tomato* DC3000 Hrp (type III secretion) deletion mutant expressing the Hrp system of bean pathogen *P. syringae* pv. *syringae* 61 retains normal host specificity for tomato. Mol. Plant-Microbe Interact. 16:43-52.
- Guo, M., Chancey, S. T., Tian, F., Ge, Z., Jamir, Y., and Alfano, J. R. 2005. *Pseudomonas syringae* type III chaperones ShcO1, ShcS1, and ShcS2 facilitate translocation of their cognate effectors and can substitute for each other in the secretion of HopO1-1. J. Bacteriol. 187:4257-4269.
- He, P., Chintamanani, S., Chen, Z., Zhu, L., Kunkel, B. N., Alfano, J. R., Tang, X., and Zhou, J.-M. 2004. Activation of a COI1-dependent pathway in Arabidopsis by *Pseudomonas syringae* type III effectors and coronatine. Plant J. 37:589-602.
- Hirano, S. S., and Upper, C. D. 2000. Bacteria in the leaf ecosystem with emphasis on Pseudomonas syringae-a pathogen, ice nucleus, and epiphyte. Microbiol. Mol. Biol. Rev. 64:624-653.
- Huynh, T. V., Dahlbeck, D., and Staskawicz, B. J. 1989. Bacterial blight of soybean: Regulation of a pathogen gene determining host cultivar specificity. Science 245:1374-1377.
- Innes, R. W., Bent, A. F., Kunkel, B. N., Bisgrove, S. R., and Staskawicz, B. J. 1993. Molecular analysis of avirulence gene *avrRpt2* and identification of a putative regulatory sequence common to all known *Pseudomonas syringae* avirulence genes. J. Bacteriol. 175:4859-4869.
- Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., Daugherty, S. C., DeBoy, R., Durkin, A. S., Giglio, M. G., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S., Haft, D. H., Creasy, T., Davidsen, T., Zafar, N., Zhou, L., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C. M., Chatterjee, Cartinhour, S., Schneider, D. J., Mansfield, J., Collmer, A., and Buell, C. R. 2005a. Whole genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals sequence divergence among pathovars in genes involved in virulence and mobile genetic elements. J. Bacteriol. 187:6488-6498.
- Joardar, V., Lindeberg, M., Schneider, D. J., Collmer, A., and Buell, C. R. 2005b. Lineage specific regions in *Pseudomonas syringae* pv. *tomato* DC3000. Mol. Plant Pathol. 6:53-64.
- Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M. G., and Alon, U. 2001. Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. Science 292:2080-2083.
- Kerr, M. K., and Churchill, G. A. 2001. Experimental design for gene expression microarrays. Biostatistics 2:183-201.
- King, E. O., Ward, M. K., and Raney, D. E. 1954. Two simple media for the demonstration of pyocyanin and fluorescin. J. Lab. Clin. Med. 44:301-307.
- Lindeberg, M., Stavrinides, J., Chang, J. H., Alfano, J. R., Collmer, A., Dangl, J. L., Greenberg, J. T., Mansfield, J. W., and Guttman, D. S. 2005. Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. Mol. Plant-Microbe Interact. 18:275-282.
- Mittal, S., and Davis, K. R. 1995. Role of the phytotoxin coronatine in the infection of *Arabidopsis thaliana* by *Pseudomonas syringae* pv. *tomato*. Mol. Plant-Microbe Interact. 8:165-171.
- Moreno-Hagelsieb, G., and Collado-Vides, J. 2002. A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics 18(Suppl. 1):S329-36.
- Mukaihara, T., Tamura, N., Murata, Y., and Iwabuchi, M. 2004. Genetic screening of Hrp type III-related pathogenicity genes controlled by the HrpB transcriptional activator in *Ralstonia solanacearum*. Mol. Microbiol. 54:863-75.
- Occhialini, A., Cunnac, S., Reymond, N., Genin, S., and Boucher, C. 2005. Genome-wide analysis of gene expression in Ralstonia solanacearum reveals that the hrpB gene acts as a regulatory switch controlling multiple virulence pathways. Mol. Plant-Microbe Interact. 18:938–949.
- Peñaloza-Vázquez, A., Preston, G. M., Collmer, A., and Bender, C. L. 2000. Regulatory interactions between the Hrp type III protein secretion system and coronatine biosynthesis in *Pseudomonas syringae* pv. tomato DC3000. Microbiology 146:2447-2456.
- Petnicki-Ocwieja, T., van Dijk, K., and Alfano, J. R. 2005. The *hrpK* operon of *Pseudomonas syringae* pv. tomato DC3000 encodes two proteins secreted by the type III (Hrp) protein secretion system: HopB1 and HrpK, a putative type III translocator. J. Bacteriol. 187:649-663.
- Preiter, K., Brooks, D. M., Penaloza-Vazquez, A., Sreedharan, A., Bender, C. L., and Kunkel, B. N. 2005. Novel virulence gene of *Pseudomonas syringae* pv. *tomato* strain DC3000. J. Bacteriol. 187:7805-7814.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16:276-277.
- Rohmer, L., Guttman, D. S., and Dangl, J. L. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. Genetics 167:1341-1360.
- Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132:365-386.
- Salgado, H., Moreno-Hagelsieb, G., Smith, T. F., and Collado-Vides, J. 2000. Operons in *Escherichia coli*: Genomic analyses and predictions. Proc. Natl. Acad. Sci. U.S.A. 97:6652-6657.
- Shen, H., and Keen, N. T. 1993. Characterization of the promoter of avirulence gene D from *Pseudomonas syringae* pv. tomato. J. Bacteriol. 175:5916-5924.
- Shigaki, T., and Bhattacharyya, M. K. 2000. Decreased inositol 1,4,5-trisphosphate content in pathogen-challenged soybean cells. Mol. Plant-Microbe Interact. 13:563-567.
- Siddharthan, R., Siggia, E. D., and van Nimwegen, E. 2005. PhyloGibbs: A Gibbs sampling motif finder that incorporates phylogeny. PLoS Comput. Biol. 1:534-556.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.
- Thwaites, R., Spanu, P. D., Panopoulos, N. J., Stevens, C., and Mansfield, J. W. 2004. Transcriptional regulation of components of the type III secretion system and effectors in *Pseudomonas syringae* pv. *phaseolicola*. Mol. Plant-Microbe Interact. 17:1250-1258.
- Ton, W. J., Jakab, G., Toquin, V., Flors, V., Iavicoli, A., Maeder, M. N., Metraux, J.-P., and Mauch Mania, B. 2005. Dissecting the beta-aminobutyric acid induced priming phenomenon in *Arabidopsis*. Plant Cell 17:987-999.
- Vencato, M., Tian, T., Alfano, J. R., Buell, C. R., Cartinhour, S., DeClerck, J., Guttman, D. S., Joardar, V., Lindeberg, M., Mansfield, J., Myers, C. R., Collmer, A., and Schneider, D. J. 2006. The Hrp (type III secretion system) regulon and effector genes of *Pseudomonas syringae* pv. *phaseolicola* 1448A. Mol. Plant-Microbe Interact. 19:1193-1206.
- Xiao, Y., Heu, S., Yi, J., Lu, Y., and Hutcheson, S. W. 1994. Identification of a putative alternate sigma factor and characterization of a multicomponent regulatory cascade controlling the expression of *Pseudomonas syringae* pv. syringae Pss61 *hrp* and *hrmA* genes. J. Bacteriol. 176:1025-1036.
- Xiao, Y., and Hutcheson, S. 1994. A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in *Pseudomonas syringae*. J. Bacteriol. 176:3089-3091.
- Zhao, Y., Thilmony, R., Bender, C. L., Schaller, A., He, S. Y., and Howe, G. A. 2003. Virulence systems of *Pseudomonas syringae* pv. *tomato* promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. Plant J. 36:485-499.
- Zwiesler-Vollick, J., Plovanich-Jones, A. E., Nomura, K., Brandyopadhyay, S., Joardar, V., Kunkel, B. N., and He, S. Y. 2002. Identification of novel *hrp*-regulated genes through functional genomic analysis of the *Pseudomonas syringae* pv. *tomato* DC3000 genome. Mol. Microbiol. 45:1207-1218.

AUTHOR-RECOMMENDED INTERNET RESOURCES

BioPythonwebsite: www.biopython.org
EMBOSS, The European Molecular Biology Open Software Suite: emboss.sourceforge.net
Institute of Genetics and Molecular and Cellular Biology ClustalX windows interface: www-igbmc.u-strasbg.fr/BioInfo/ClustalX
Profile hidden Markov model software: hmmer.wustl.edu (Eddy 1998)
Pseudomonas-Plant Interaction website: pseudomonas-syringae.org
The R Project for statistical computing: www.r-project.org
WebLogo website: weblogo.berkeley.edu

Supplementary Table S1. Primers for SYBR Green real-time PCR experiments: sequences are 5’ to 3’.

Target	**Forward (upstream) primer**	**Reverse (downstream) primer**
PSPTO0044	GCTTTTGATGCTCAACGCCT	GGTTTGAAGTGTCGTTGTTGGAG
PSPTO0474	TCCGAACGTCAGACCAATTTC	TCGATACCTGCTCACCTTCCTC
PSPTO0524	TCAGACGTTCGCTTACCTGGA	GCGCCTGCAATGCATAAAG
PSPTO0834	GAATACTGTGGCGTTTGCGAA	CATGTTTGTTGACACCTGCGA
PSPTO0869 Upstream	TAAGCAGCCCTGAGACTCTACGTA	TACCAAATCGTTCGTGGCC
PSPTO1394	AAGGCATGTTTCTGGTGGTGA	CTGAATCTGCATGAGTGCCTG
PSPTO1568	TGTGGATGAACGTCCTGCAA	TTTCTGCTTGATAAGCCGCTG
PSPTO1645	TCTGCTCGACAGCCTTGAAA	GGTGTCGCTGAGCAGAATCTTT
PSPTO2105	ACGTCGATCAACAGATGTCGC	CAAAACCGACGTGGACATGA
PSPTO2691	GGAACATCAGCGCGAAAAG	CTGTCAGCTGGACGATGTAAGC
PSPTO3576	AGCGAATGCTGATGGTGCT	AGAACATCATGCCGATGGC
PSPTO3648	CAACCATCAATACCTGATCTCGG	AGCGGATTCATGCTCTGCA
PSPTO4718	GGATTGCAGTGCAGACGTTCT	GCCAAAATCAACAGCACCCT
PSPTO4732	GGTCGTTTGATGATATCTGGCC	TTAAAGAGCAGTTTCGCCGC
PSPOT4733	ACTTCCTATGGCTCATTCGCC	TCGCCATCGAAACCATCAG
PSPTO4776	TTGTTCTCTCTGGGCAAGAAGC	TGGAGAAATAGGCATGGCG
PSPTOB0005	GTGTTCTACCGAAACCAGATGCAG	TTTCTGCGTAGTTGCCCATG
GyrA	AGGCAAGTATTTCCTGTCGCC	CTGGTACTCACCCAGCAGTTTTT
Gap-1	CCGCAAGGTGATTATCTCAGC	TGGAGATGATCTGGTGCGACT

This file was revised on December 18, 2006.

Supplementary Table S2. Computationally identified Hrp promoters in *P. syringae* pv. *tomato* DC3000 and their associated downstream targets. Each row corresponds to a Hrp promoter identified by scanning the complete DC3000 genome with the HMM2 and WM2 models (see text for details), so genes and/or operons can appear more than once if there are multiple putative promoter hits. Entries are sorted in descending order according to their HMM2 scores. Additional entries at the bottom of the table are included for which there is experimental evidence for HrpL-dependent expression under *hrpL* overexpression, but for which no discernable Hrp promoter is evident.

OPERON	EVID	START	STOP	S	GEO	DIST	HMM2	WM2	GC	PROTEIN ANNOTATIONS
588	CM	648424	648456	-	I/C	-212	26.1	7.2	44.6	HopH1
4599-4597	FM-CM	5192613	5192645	-	I/C	-62	24.2	7.1	42.4	T3SS chaperone ShcS1 – HopS1::ISPssy
4331	CM	4881097	4881129	+	I/C	-70	23.6	7.0	48.9	HopE1
877	CM	949826	949858	-	E/N	-60	23.3	6.1	52.2	HopQ1-1
2105	FCR	2279883	2279915	+	I/C	-143	22.8	6.3	59.8	ApbE-family protein
4753	0	5384479	5384511	+	E/N	-4706	22.5	6.2	55.4	hypothetical protein
4001	FCM	4515296	4515328	-	I/C	-68	22.3	6.6	45.7	AvrPto1
4733	0R	5361707	5361739	-	I/C	-85	22.2	6.8	52.2	hypothetical protein
3087	FCM	3470185	3470217	-	I/C	-86	21.2	6.7	53.3	HopAB2 (AvrPtoB)
883	FCM	954203	954235	+	I/C	-205	20.7	6.7	44.6	HopR1
1378	FCM	1519666	1519698	+	I/C	-59	20.7	5.8	56.5	membrane-bound lytic murein transglycosylase D, putative
1381-1382-1383-1384-1385-1386	CM-CM-M-M-M-M	1524204	1524236	+	I/C	-77	20.6	5.7	48.9	T3SS components HrpA1 - HrpZ1 - HrpB - HrcJ - HrpD - HrpE
589	FCM	649735	649767	-	E/N	-62	20.4	7.0	43.5	HopC1
876	FCM	946154	946186	+	I/C	-69	20.4	6.2	55.4	HopD1
1387-1388-1389-1390-1391	CN-M-M-N-M	1528184	1528216	+	I/C	-62	20.4	6.3	54.4	T3SS components HrpF - HrpG - HrcC - HrpT - HrpV

1403-1402-1401-1400-1399-1398-1397-1396-1395-1394-1393-1392	CM-M-M-0-M-M-M-M-M-M-M-N-0-0	1542621	1542653	-	I/C	-59	20.4	5.8	55.4	T3SS components HrpJ - HrcV - HrpQ - HrcN - HrpO - HrpP - HrcQa - HrcQb - HrcR - HrcS - HrcT - HrcU
5617*	NR	939413	939445	-	E/N	-96	20.2	6.8	45.7	conserved hypothetical protein (pseudogene)
5053	0	5751475	5751507	+	E/C	-487	19.8	5.5	56.5	hypothetical protein
2678-2679	FCM-N	2973825	2973857	+	I/C	-121	19.6	6.9	47.8	HopP1 - hypothetical protein
834-(835-836-837-838)	CR-C0-C0-0-0	905339	905371	+	I/C	-236	19.2	6.3	50.0	alcohol dehydrogenase, zinc-containing - ribD C-terminal domain protein - conserved domain protein - hypothetical protein - major facilitator family transporter
873	0	941100	941132	+	I/C	-919	19.2	5.9	44.6	amidinotransferase family protein
503-(502-501)	M-FC0-0	550602	550634	-	I/C	-65	18.8	6.3	39.1	T3SS chaperone ShcF - HopF2 - HopU1
4718	0R	5344375	5344407	-	I/C	-196	18.8	5.9	51.1	HopAA1-2
1372	FCM	1507652	1507684	+	I/C	-194	18.7	5.9	55.4	HopAA1-1
4589-4588	C0-CN	5186123	5186155	-	I/C	-61	18.7	5.3	57.6	T3SS chaperone ShcS2 – HopS2
A0005	FM	6595	6627	-	I/C	-66	18.4	5.8	48.9	HopAM1-2
1022	FCM	1116378	1116410	-	I/C	-66	18.4	5.8	48.9	HopAM1-1
1369-1370	FM-FCM	1504886	1504918	+	I/C	-63	18.4	6.1	41.3	conserved effector locus protein - HopN1
524	0R	572473	572505	+	I/C	-101	18.3	6.3	52.2	peptidase, M20 - M25 - M40 family
A0012	FM	16103	16135	-	I/C	-67	18.1	5.8	40.2	HopX1
1374-1375-(1376)	CM	1510881	1510913	+	I/C	-89	17.9	5.0	45.7	T3SS chaperone ShcM - HopM1 – T3SS chaperone ShcE
5353-5354	CM-FCM	6085756	6085788	+	I/C	-59	17.5	6.0	53.3	T3SS chaperone ShcA - HopA1

5620*-1568	0-CR	1731421	1731453	-	E/N	-288	17.0	5.8	43.5	hypothetic protein identified by sequence similarity - HopAF1
2856	0	3212151	3212183	-	I/C	-1634	16.8	5.1	48.9	site-specific recombinase, phage integrase family
4699	0	5328019	5328051	-	I/C	-854	16.8	5.6	35.9	non-ribosomal peptide synthetase, terminal component
4776	FCR	5418197	5418229	-	I/C	-56	16.5	6.0	52.2	HopI1
A0017-A0018-A0019	M-M-M	19658	19690	+	I/C	-61	16.4	5.7	56.5	T3SS chaperone ShcO1 - HopO1-1 - HopT1-1
1405-1406	CM	1543416	1543448	+	I/C	-64	16.4	6.3	51.1	HrpK1 - HopB1
61	FCM	82447	82479	+	I/C	-54	16.3	5.4	53.3	HopY1
5622*-1408-1409	0-FN-M(i)	1548389	1548421	+	I/C	-531	16.1	5.5	47.8	hypothetical protein - hypothetical protein – hypothetical protein
B00078*-B0005	P-MR	6894	6926	-	I/C	-820	16.0	6.4	45.7	unannotated protein - phosphoesterase family protein
4724-4725-4726#	N	5350034	5350066	+	I/C	-655	16.0	5.7	47.8	HopD (interruption-N) - IS52, transposase - Hop D (interruption-C)
[4579]	0	5176133	5176165	+	E/C	681	15.9	5.4	57.6	peptide chain release factor 3
852	FCM	921879	921911	-	I/C	-219	15.5	6.1	47.8	HopAJ1
1373	FCM	1510785	1510817	-	I/C	-153	15.4	6.1	55.4	HrpW1
1377	C0	1519570	1519602	-	I/C	-99	15.2	5.7	50.0	AvrE1
4101	FCM	4621129	4621161	+	I/C	-95	15.1	5.9	54.4	HopAK1
5616*-474-473	F0R-0	522444	522476	-	I/C	-309	14.7	6.2	50.0	conserved hypothetical protein - Hop AS1’ (interruption-N) - HopAS1 (interruption-C)
4344	0	4905189	4905221	+	E/C	-1456	14.6	5.0	64.1	insecticidal toxin protein, putative
4727	CM	5355224	5355256	-	I/C	-90	14.4	5.7	45.7	HopG1
4464	0	5024215	5024247	+	E/N	-1316	14.3	5.3	57.6	conserved hypothetical protein
4276	0	4820136	4820168	-	E/N	-2127	13.9	4.7	56.5	transcriptional regulator, LysR

										family
4722 – (4721 – 4720)	FC0	5348578	5348610	-	I/C	-86	13.8	5.1	47.8	HopAO1 – T3SS chaperone ShcV – HopV1
437	0	488341	488373	-	E/N	-6645	13.3	5.3	48.9	dihydrofolate reductase
44	CR	61504	61536	-	E/N	-113	13.2	5.4	51.1	HopK1
370-(371)	0-FC0	404752	404784	+	I/C	-114	13.2	5.5	52.2	MATE efflux family protein - indoleacetate-lysine ligase (iaaL)
4703	CN	5330688	5330720	+	I/C	-114	13.2	5.4	52.2	HopAQ1
1398-1397-1396-1395-1394-1393-1392	M-M-M-M-M-N-0-0	1536874	1536906	-	E/C	-870	13.1	5.4	57.6	T3SS components HrpP - HrcQa - HrcQb - HrcR - HrcS - HrcT - HrcU
[1707]	0	1876880	1876912	-	E/C	1104	12.9	5.3	60.9	L-sorbosone dehydrogenase
1841	0	2008405	2008437	+	E/N	-887	12.9	4.4	53.3	threonine aldolase, low-specificity
1996-1995-1994	0-0-N-0	2182088	2182120	-	E/C	-120	12.9	NA	66.3	copper-translocating P-type ATPase - cytochrome oxidase maturation protein, cbb3-type - membrane protein, putative
1370	FCM	1505219	1505251	+	E/C	-222	12.8	4.4	57.6	HopN1
2346	0	2600072	2600104	+	E/N	-1061	12.7	4.4	68.5	4-hydroxyphenylpyruvate dioxygenase, putative
1869	0	2044553	2044585	-	E/C	-1243	12.5	5.1	55.4	conserved hypothetical protein
5200	0	5908418	5908450	+	E/N	-9545	12.5	4.9	65.2	autotransporter, putative
1645	0	1802304	1802336	+	E/C	-722	12.3	4.8	62.0	transcriptional regulator, SlyA/MarR family
3019	0	3393410	3393442	+	E/C	-579	12.3	4.8	56.5	oxidoreductase, Gfo - Idh - MocA family
5522	0	6289194	6289226	+	E/N	-249	12.3	5.0	55.4	conserved hypothetical protein
[1268]	0	1391965	1391997	-	E/C	224	12.2	4.6	67.4	valyl-tRNA synthetase
4721	0	5346758	5346790	-	I/C	-60	12.2	5.6	43.5	hypothetical protein
4205	0	4733222	4733254	+	E/N	-3960	12.0	4.8	68.5	transcriptional regulator, GntR

										family
[B0069]	N	59060	59092	+	E/C	602	11.8	4.7	59.8	traY protein
5089	0	5790510	5790542	+	E/N	-977	11.8	4.8	55.4	conserved hypothetical protein
5619* - 901	C0	981177	981209	+	I/C	-324	11.7	5.2	39.1	putative T3SS chaperone - candidate HopAG1::ISPssy
[4248]	0	4786717	4786749	+	E/C	765	11.7	5.0	50.0	3-hydroxyacyl-CoA-acyl carrier protein transferase
2696	0	2990246	2990278	-	I/C	-241	11.6	4.4	48.9	mutT - nudix family protein
3489-3488	M(i)-M	3939658	3939690	-	E/C	-317	11.6	4.4	59.8	sugar ABC transporter, ATP-binding protein - sugar ABC transporter, permease protein
932	0	1014826	1014858	-	E/N	-3955	11.5	4.9	55.4	sugar ABC transporter, permease protein
5057	0	5761609	5761641	-	I/C	-316	11.5	4.6	64.1	dihydroxy-acid dehydratase
4920	0	5569089	5569121	+	I/C	-269	11.4	5.1	51.1	transcriptional regulator, LysR family
119	0	142796	142828	+	E/C	-179	11.3	4.9	51.1	peptidase, M16 family
224-223-222-221	0	247050	247082	-	E/C	-681	11.3	4.9	62.0	diaminopimelate epimerase - conserved hypothetical protein - integrase - recombinase XerC - HAD-superfamily hydrolase
2130	0	2304328	2304360	-	E/C	-220	11.3	4.8	51.1	DNA-binding response regulator, LuxR family
3984	0	4489552	4489584	+	I/N	-1436	11.3	5.2	62.0	cold shock domain family protein
[4005]	0	4519000	4519032	+	E/C	418	11.3	4.7	55.4	hypothetical protein
1213	0	1329860	1329892	-	E/N	-783	11.2	4.5	54.4	transcriptional regulator, AraC family
[1842]	0	2012107	2012139	+	E/C	1667	11.2	4.5	54.4	alanyl-tRNA synthetase

2145-2144-2143-2142-2141-2140-2139-2138	0	2354845	2354877	-	E/N	-25784	11.2	5.1	51.1	iron-regulated membrane protein, putative - conserved hypothetical protein - conserved hypothetical protein - conserved hypothetical protein - cation ABC transporter, periplasmic cation-binding protein - cation ABC transporter, ATP-binding protein - cation ABC transporter, permease protein - ABC transporter, periplasmic substrate-binding protein, putative
3721	0	4199603	4199635	+	E/C	-1220	11.0	4.6	62.0	enoyl-(acyl-carrier-protein) reductase
[2362]	0	2615952	2615984	+	E/C	513	10.9	4.8	60.9	methionine aminopeptidase, type I
3270	N	3693598	3693630	+	E/N	-2442	10.9	4.2	57.6	conserved hypothetical protein
5623* - 4732	MR	5355530	5355562	+	I/N	-4324	10.9	4.5	55.4	HopH::IsPsy4 - HopQ1-2
665	0	714155	714187	+	E/N	-4377	10.7	4.4	68.5	ROK family protein
300	0	325510	325542	+	I/C	-255	10.6	4.2	53.3	succinate-semialdehyde dehydrogenase
2107	0	2287258	2287290	-	E/N	-3931	10.6	4.1	56.5	glycerophosphoryl diester phosphodiesterase, putative
4157-4158	0	4684462	4684494	+	E/N	-264	10.6	5.1	59.8	dihydroorotase, homodimeric type - ribonuclease T
4976	0	5649753	5649785	-	E/N	-1312	10.6	5.1	62.0	thiamin biosynthesis protein ThiC
[4996]	0	5673452	5673484	+	E/C	163	10.6	4.9	62.0	conserved domain protein
2431	0	2683181	2683213	+	E/C	-969	10.5	4.8	44.6	conserved hypothetical protein
3641	0	4099725	4099757	+	E/C	-639	10.5	4.0	64.1	endoribonuclease L-PSP family protein
4516.1	0	5085928	5085960	+	E/C	-255	10.5	4.6	54.4	tmRNA, putative [1]
[2151]	0	2358286	2358318	-	E/C	1194	10.4	4.4	60.9	TonB-dependent siderophore receptor, putative
4381	0	4943533	4943565	-	E/N	-353	10.4	4.8	65.2	conserved hypothetical protein

5045	0	5752184	5752216	-	E/N	-6388	10.4	4.9	66.3	twitching motility protein
2162	0	2377712	2377744	-	E/C	-533	10.3	4.3	55.4	hypothetical protein
4294-4295	0	4839209	4839241	+	I/C	-107	10.3	4.3	56.5	chaperone protein HscC - DnaJ domain protein
[4955]	0	5616670	5616702	+	E/C	881	10.3	4.3	57.6	rhodanese domain protein - phosphatidylserine decarboxylase
5396	0	6125242	6125274	+	I/N	-3202	10.3	4.2	66.3	dTDP-glucose 4,6-dehydratase
4507	0	5074481	5074513	+	E/N	-4119	10.2	4.9	64.1	DNA repair protein RecN
1621-1620-1619	0	1779969	1780001	-	E/N	-1902	10.1	5.0	63.0	conserved hypothetical protein - conserved hypothetical protein - hypothetical protein
[2014]	N	2201776	2201808	-	E/C	135	10.1	4.3	63.0	aerotaxis receptor
5150-5151-5152	0	5859272	5859304	+	E/N	-1729	10.1	4.6	68.5	ubiquinone - menaquinone biosynthesis methlytransferase UbiE - conserved hypothetical protein - ubiquinone biosynthesis protein UbiB
5353-5354	CM	6085463	6085495	+	E/C	-352	10.1	4.5	57.6	T3SS chaperone ShcA – HopA1
371	FC0	406209	406241	+	E/C	-114	10.0	4.8	54.4	indoleacetate-lysine ligase (iaaL)
675	0	731309	731341	-	I/C	-10	9.6	4.4/ 4.5	53.3	arylesterase
1209-1208	0	1325953	1325985	-	I/C	-177	9.1	3.8/ 4.3	47.8	RNA polymerase sigma-70 factor – regulatory protein, putative
4691	C0	5305220	5305252	+	I/C	-583	8.7	3.8/ 4.4	47.8	HopAD1
1278	0	1405112	1405144	-	I/C	-44	7.7	4.2/ 4.9	56.5	sensory box protein/response regulator
1708-1709-1710-1711-1712-1713-1714-1715-1716-1717	0	1878314	1878346	+	I/C	-127	7.1	3.9/ 3.8	64.1	cobO operon: cob(I)yrinic acid a,c-diamide adenosyltransferase – nitroreductase family protein – cobD – cobC – cobQ – cobP – cobT - alpha-ribazole-5'-phosphate

										phosphatase, putative - cobS
2497-2498	N	2762679	2762711	-	I/C	-54	6.9	4.1/ 4.6	46.7	hypothetical protein – hypothetical protein
4071	0	4577328	4577360	-	I/C	-39	6.4	3.9/ 4.2	52.2	hypothetical protein
2936	0	3300862	3300894	-	I/C	-52	6.1	4.5/ 3.8	46.7	LamB/YscF family protein
2525	0	2787908	2787940	+	E/N	-341	<6.0	4.7/ 4.1	59.8	peptide ABC transporter, permease protein
2790	0	3108233	3108265	+	E/C		<6.0	4.0/ 4.2	38.0	hypothetical protein
4964	0	5630188	5630220	-	I/C	-41	<6.0	4.1/ 4.0	53.3	conserved hypothetical protein
401	0	445291	445323	-	I/C	-233	<6.0	3.9/ 3.9	60.9	iron-sulfur cluster-binding protein, Rieske family
1672	0	1843567	1843599	+	I/C	-47	<6.0	4.0/ 3.9	63.0	DNA-binding response regulator
5618*	N	922923	922955	-	I/C		<6.0	4.1/ 4.5	46.7	HopAT1 (frameshift)
1407	CN	NO HRP PROMOTER								pseudogene not annotated as CDS
4680	C0	NO HRP PROMOTER								coronafacic acid synthetase, ligase component (cfl)
4681-4682-4683-4684-4685-4686-4687-4688-4689-4690	F-N-N-N-N-F-N-N-N-N	NO HRP PROMOTER								coronafacic acid (cfa) synthetase components cfa1 through cfa7 - hypothetical protein - cfa8- cfa9

Entries are color coded as follows:

BLUE: experimental evidence for HrpL-dependent expression both under conditions of *hrpL* deletion in Hrp minimal medium as examined here, and *hrpL* overexpression in rich medium, as previously reported (Chang *et al.*, 2005; Fouts *et al.*, 2002).

GREEN: experimental evidence for HrpL-dependent expression under conditions of *hrpL* deletion in Hrp minimal medium as examined here, but not under *hrpL* overexpression in rich medium, as previously reported (Chang *et al.*, 2005; Fouts *et al.*, 2002).

ORANGE: experimental evidence for HrpL-dependent expression under *hrpL* overexpression in rich medium, as reported (Chang *et al.*, 2005; Fouts *et al.*, 2002), but not under conditions of *hrpL* deletion in Hrp minimal medium as examined here.

GRAY: no experimental evidence for HrpL-dependent expression

For each Hrp promoter, the following items are identified:

- OPERON: the predicted downstream target(s) of the promoter. Operon members enclosed in parentheses are assumed to be part of the stated operon because of experimental evidence, but which were not identified so by our computational operon prediction method. Operon members enclosed in square brackets are not considered to be actual downstream targets but are included for the sake of completeness; these represent genes in which the predicted Hrp promoter is embedded, with no plausible downstream target that can be identified (i.e., no subsequent genes in a predicted operon which could be controlled by the promoter).
  - NOTES:
    - * PSPTO numbers annotated with an asterisk refer to new gene calls that we are making based on this work. These new PSPTO identifiers are included in the updated GenBank file we are submitting in conjunction with this paper.
    - # entry 4724-4725-4726 is presumed vestigial, although the promoter may indeed be functional; 4725 is a transposase which appears to have inserted into, and disrupted, a copy of HopD1, which is composed of the fragments in 4724 and 4726.
- EVID: summary of experimental evidence for HrpL-dependent expression, listing an entry for each member of the predicted operon, respectively; if only one entry is listed for a polycistronic operon, that evidence label applies to every member of the operon
  - M = exhibited differential expression (up-regulation) on our Microarray
    - M(i) = = exhibited differential expression (up-regulation) on our Microarray, but just missed our spot intensity cutoff of 4000 units; reducing that cutoff to 3900 units included two extra spots, associated with PSPTO1409 and PSPTO3489. For further evidence of expression of PSPTO1409, see Petnicki-Ocwiega et al. (2005).
  - R = exhibited differential expression (up-regulation) in our RT-PCR assays
  - C = listed in Chang et al. (2005) Table 2 with nonzero reads in functional screen for DC3000

  - F = listed in Fouts et al. (2002), Table 2
  - 0 = on our Microarray, but did not exhibit differential expression
  - N = not on our Microarray
  - P = evidence for existence of peptide fragment via proteomics
- START: start coordinate of Hrp promoter
- STOP: stop coordinate of Hrp promoter
- S: strand of Hrp promoter
- GEO: geometry of the Hrp promoter hit with respect to the local genome
  - I = Intergenic (start of Hrp promoter is located between two predicted genes)
  - E = Embedded (start of Hrp promoter is located within a predicted coding region)
  - C = Coding: if hit is Embedded, then Hrp promoter is oriented in the coding direction of the gene in which it is embedded; if hit is Intergenic, then next downstream gene is on the same strand as the Hrp promoter
  - N = Noncoding: if hit is Embedded, then Hrp promoter is oriented oppositely to the coding direction of the gene in which it is embedded; if hit is Intergenic, then next downstream gene is on the opposite strand of the Hrp promoter\
- DIST: distance from the start of the Hrp promoter to the start of the predicted operon (NOTE: distance is calculated on the same strand as the Hrp promoter, even if the next downstream gene is on the opposite strand (e.g., Intergenic/Noncoding)); negative distances indicate the start of the Hrp promter is upstream of the start of the target operon; positive distances indicate the start of the Hrp promoter is downstream of the start of the gene; in this case, the Hrp promoter is embedded
- HMM2: score of Hrp promoter within the Hidden Markov Model level 2 model
- WM2: score of Hrp promoter within the Weight Matrix level 2 model (number of standard deviations above the mean)
- GC: percentage of G+C nucleotides in a window around and including the Hrp promoter, extending 40 bases upstream of the start of the Hrp promoter and 20 bases downstream of the stop
- PROTEIN ANNOTATIONS: GenBank annotations of the protein corresponding to each member of the predicted operon. Abbreviations in annotations: TT = Type III; prot = protein. T3SS substrate names are updated to reflect the new Hop nomenclature conventions (Lindeberg et al. 2005).

This file was revised on December 18, 2006.

Supplementary Table S3. Schedule for combination of cDNA samples prior to hybridization. Labeled cDNAs were mixed following an incomplete block design (Chu et al. 2002; Churchill 2002; Kerr and Churchill 2001) and then treated as described in Materials and Methods. (wt = wide type, mt = *hrpL* deletion mutant)

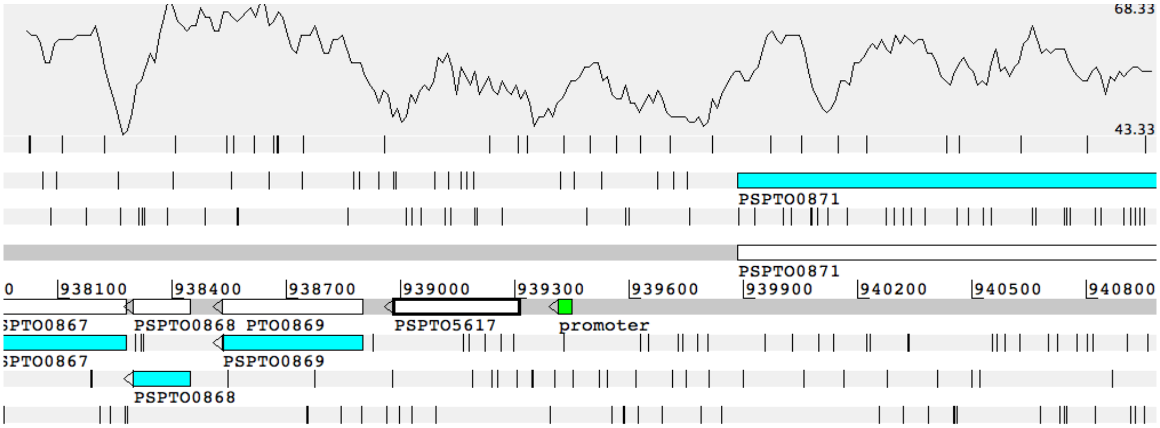
Biological Replicate	Slide	Cy3	Cy5
1	4	0 wt	3 mt
1	5	2 mt	9 wt
1	6	3 wt	5 mt
1	7	4 mt	2 wt
1	31	5 wt	7 mt
1	32	6 mt	4 wt
1	83	7 wt	12 mt
1	84	9 mt	6 wt
1	85	12 wt	0 mt
2	1	2 mt	9 wt
2	91	0 wt	3 mt
2	94	4 mt	2 wt
2	95	3 wt	5 mt
2	96	6 mt	4 wt
2	97	5 wt	7 mt
2	98	9 mt	6 wt
2	99	7 wt	0 mt
3	82	0 mt	5 wt
3	84	6 wt	2 mt
3	85	3 mt	7 wt
3	86	9 wt	4 mt
3	87	5 mt	0 wt
3	88	2 wt	6 mt
3	89	7 mt	3 wt
3	90	4 wt	9 mt

This file was revised on December 18, 2006.

Supplementary Table S4. Set of aligned Hrp promoter sequences used to train the HMM level 2 and PSWM level 2 bioinformatic models, as described in the text. PSPTO numbers indicate nearest annotated downstream target, and dashes in sequences indicate gaps inserted by the Clustal alignment process. The full set was used to train the HMM model. Separate WM models with different spacers between the -35 and -10 boxes were constructed; those sequences with gaps were used to train the 16bp spacer model, and those without gaps were used to train the 17bp model.

Target	Promoter sequence
PSPTO4599	TGGAACCGG-ACGAGGCTTTTACCACTCAATG
PSPTO0883	CGGAACCGA-TCCGGTTGCCTGGCCACTCAATT
PSPTO0876	TGGAACCCA-AGAGCCCTTGCGACCACACATTG
PSPTO1378	GGGAACCGT-AACGGCGAGCGTGCCACGTAGGG
PSPTO2678	GGGAACCGA-GTCACTCAGTGAACCACTCAGTT
PSPTO1405	TGGAACCAA-CTTGCACCTTCAACCACACAGTT
PSPTO1373	GGGAACCGG-TCGCTGCGCTTTGCCACTCACTT
PSPTO1369	GGGAACCGC-ATCACGTCTTGAACCACAGAGGA
PSPTO0061	AGGAACTCA-TCACCGCGAATCGCCACTCAGCA
PSPTOA0012	GGGAACTGA-CCCGGCCAAATGACGACATAGCT
PSPTO4101	CGGAACTCT-TTCCCTGCGCTTTCCACTCAGGG
PSPTO0589	GGGAACTGA-ACCGCTTATGAAACCACTCATTT
PSPTOA0017	TGGAACCTT-ACGGAGGTTCACGCCACCAAGTG
PSPTO1372	TGGAACCGTCAACCGATCCGGGACCACACAGCC
PSPTO4001	TGGAACCGATCCGCTCCCTATGACCACTCAAGT
PSPTO0877	TGAAACCGAAACGGCGTTGCTTGCCACACAGCA
PSPTO1388	TGGAACCGCTCGGCGGGTTTGCTCCACTCAAGG
PSPTO1403	GGGAACTGATCCGGGACCGTGACCCACTCAGCG
PSPTO3087	TGGAACTCTTTCCTGCTCTTTTGCCACACAGCG
PSPTO0503	GGGAACCTGATGCTGCTCAGTGACCACTCATAA
PSPTO0852	TGGAACCTCACGCTTAGTGATGACCACGCATAG
PSPTO4727	TGGAACTCTACGCTTGGCGATGACCACGTATGA
PSPTO1381	TGGAACCGATTCGCAGGCTGCTGCCACCTAGTG
PSPTO1374	AGGAACTGAAATGCCTATGCCTGCGACTCAGTG
PSPTO0588	TGGAACCGAATCCATCTCGAGGGCCACTCAAGG
PSPTO4331	TGGAACCGAATCCGCCTCAAAGTCCACACAAGC
PSPTO5353	TGGAACCGCCTCGAGCAGAGGCTCCACTCATTG
PSPTO1022	GGGAACCACATCATGGGTAAAAGCCACGAAGAG
PSPTO2394	CAAGACTCAGAGGTGCATGTATGCCACACAGTT
PSPTO4733	GGGAACCAAATCCGCCTCAAAGTCCACTCAAGC
PSPTO0869	TGGACCCGAATCCGTCTTAAACACCACTCAAGG
PSPTO4718	TGGAACCGTTAACGGCCCAGCGACTACACAGCT
PSPTOB0005	TGGAACCGA-TTCCGCTATCTGACCACCTACAC

This file was revised on December 18, 2006.



Supplementary Fig S1. Artemis screen shot of the six possible open reading frames and the revised annotation in the PSPTO0869-PSPTO0871 region. The upper pane, the local variation in GC content is plotted using a window of 120 base pairs. The low GC content of the region between PSTO0869 and PSPTO0871 (51%) relative to the genome as a whole (59%) complicates gene calling in this area. In the lower pane, the green box represents the location of the Hrp promoter apparently responsible for the HrpL-dependent expression of PSPTO0869, black vertical lines represent stop codons in each frame, white boxes represent genes or pseudogenes and blue boxes on the corresponding frame lines indicated CDS features. The initial genome annotation contained a gene call based solely on Glimmer 2 predictions, PSPTO0870, in the short open reading frame overlapping the Hrp promoter on the opposite strand. Extensive sequence analysis of the unusually long 5' UTR of PSPTO0869 revealed the existence of a pseudogene, PSPTO5617, on the basis of highly similar intact genes from *P. syringae* pv. *syringae* B728a (NCBI GI no. 63255405), *P. phaseolicola* pv. *phaseolicola* 1302A (NCBI GI no. 68637911), P. fluorescens Pf-5 (NCBI GI no. 70732033), and a number of *P. aeruginosa* pathovars. PSPTO0870 was deprecated on the basis of its exclusive reliance on a Glimmer 2 gene call, lack of conservation in other pseudomonads, and overlap with the Hrp promoter and N-terminal region of the PSPTO5617 pseudogene with support provided by comparative evidence.