UCRL-JRNL-213733

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Comparison of the complete genome sequences of Pseudomonas syringae pv. syringae B728a and pv. tomato DC3000

H. Feil, W. S. Feil, P. Chain, F. Larimer, G. DiBartolo, A. Copeland, A. Lykidis, S. Trong, M. Nolan, E. Goltsman, J. Thiel, S. Malfatti, J. E. Loper, A. Lapidus, J. C. Detter, M. Land, P. M. Richardson, N. C. Kyrpides, N. Ivanova, S. E. Lindow

July 19, 2005

Proceedings of the National Academy of Sciences of the United States of America

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000

Helene Feil[¶], William S. Feil[¶], Patrick Chain[§*], Frank Larimer[¢], Genevieve DiBartolo[*], Alex Copeland[*], Athanasios Lykidis[*], Stephen Trong[*], Matt Nolan[*], Eugene Goltsman[*], James Thiel[*], Stephanie Malfatti[§*], Joyce E. Loper[£], Alla Lapidus[*], John C. Detter[*], Miriam Land[¢], Paul M. Richardson[*], Nikos C. Kyrpides[*], Natalia Ivanova[*], and Steven E. Lindow[¶†]

[¶]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [*]Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598; [§]Lawrence Livermore National Laboratory, Livermore, California 94550; [£]U.S. Department of Agriculture, Agricultural Research Service, Corvallis, OR 97330; and [¢]Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

Contributed by Steven E. Lindow, June 10, 2005

Abbreviations: *Pss* B728a, *Pseudomonas syringae* pv. *syringae*; *Pst* DC3000, *Pseudomonas syringae* pv. *tomato* DC3000; IS, insertion sequence; REPs, repetitive extragenic palindromic sequences; TTSS, type III secretion system; Tfp, type IV pili; IAA, indole-3-acetic acid; ROS, reactive oxygen species; AHL, acyl-homoserine lactone; SP, syringopeptin; SR, syringomycin.

Data deposition: The sequence and annotation of *Pss* B728a have been deposited in the GenBank database (accession no. CP000075)

[†]To whom correspondence should be addressed. E-mail: icelab@socrates.berkeley.edu

ABSTRACT The complete genomic sequence of *Pseudomonas syringae* pathovar *syringae* B728a (*Pss* B728a), has been determined and is compared with that of *Pseudomonas syringae* pv. *tomato* DC3000 (*Pst* DC3000). The two pathovars of this economically important species of plant pathogenic bacteria differ in host range and other interactions with plants, with *Pss* having a more pronounced epiphytic stage of growth and higher abiotic stress tolerance and *Pst* DC3000 having a more pronounced apoplastic growth habitat. The *Pss* B728a genome (6.1 megabases) contains a circular chromosome and no plasmid, whereas the *Pst* DC3000 genome is 6.5 mbp in size, composed of a circular chromosome and two plasmids. While a high degree of similarity exists between the two sequenced Pseudomonads, 976 protein-encoding genes are unique to *Pss* B728a when compared to *Pst* DC3000, including large genomic islands likely to contribute to virulence and host specificity. Over 375 repetitive extragenic palindromic sequences (REPs) unique to *Pss* B728a when compared to *Pst* DC3000 are widely distributed throughout the chromosome except in 14 genomic islands, which generally had lower GC content than the genome as a whole. Content of the genomic islands vary, with one containing a prophage and another the plasmid pKLC102 of *P. aeruginosa* PAO1.

Among the 976 genes of *Pss* B728a with no counterpart in *Pst* DC3000 are those encoding for syringopeptin (SP), syringomycin (SR), indole acetic acid biosynthesis, arginine degradation, and production of ice nuclei. The genomic comparison suggests that several unique genes for *Pss* B728a such as ectoine synthase, DNA repair, and antibiotic production may contribute to epiphytic fitness and stress tolerance of this organism.

Pseudomonas syringae, a member of the gamma subgroup of the Proteobacteria, is a widespread bacterial pathogen of many plant species. The species *P. syringae* is subdivided into approximately 50 pathovars based on pathogenicity and host range. *P. syringae* is capable of producing a variety of different symptoms depending on the host species and site of infection. For example, it causes leaf spot diseases that defoliate plants such as tomato, bean and soybean, trunk cankers, and so-called "blast" diseases on fruit, nut and ornamental species. Considerable variation occurs both between and within different pathovars of *P. syringae* (1). Because of its importance as a plant pathogen, it has been the subject of much research, especially of its epidemiology and virulence mechanisms (2). *Pseudomonas syringae* pv. syringae (*Pss*) strain B728a is typical of most strains of this pathovar in that it exhibits a very pronounced epiphytic phase on plants. Such strains achieve and maintain large populations on healthy plants, where they are exposed to stressful conditions such as dryness and sunlight that are hostile to bacterial growth(2). Epiphytic *Pss* populations serve as inocula that can subsequently invade plants and initiate disease. *Pss* strains are distinct from many *P. syringae* strains, such as *P. syringae pv. tomato (Pst) strain* DC3000, that poorly colonize the exterior of plants; these strains may be considered "endophytes" based on their ability to multiply mostly within the plant (3). True epiphytes such as *Pss* B728a often reach surface populations of over 10^7 cells/g while strains such as *Pst* DC3000 seldom exceed 10^5 cells/g (2, 3). Thus, these strains might be considered to occupy different ends of the epiphytic/endophytic spectrum of plant colonization as described by Beattie and Lindow (4).

As a pathogen and an epiphyte, *Pss* B728a has evolved to exploit at least two distinct habitats: the leaf surface and apoplast. Because rapid changes in temperature, low water content, and incident solar radiation occur on leaf surfaces, it has been hypothesized that the epiphyte *Pss* B728a posseses more genes conferring environmental stress tolerance than the endophyte *Pst* DC3000 (4). *Pss* B728a also exhibits several traits such as ice nucleation activity and SR production (2) that are lacking in some other strains of *P. syringae* including *Pst* DC3000. As the most ice nucleation active bacterial species, *P. syringae* is responsible for inciting frost injury to frost sensitive plants that can supercool and avoid damaging ice formation if not colonized by ice nucleation active bacteria (2, 4). We present here a genomic comparison between strains *Pss* B728a and *Pst* DC3000 of *P. syringae* pathovars as well as between these strains and *P. aeruginosa* and *P. putida*, two additional Pseudomonads recently sequenced. These genomic comparisons provide insights into the evolutionary history and diverse life styles of the pseudomonads, including their association with the environment, plant or mammalian hosts.

Materials and Methods

Sequencing and Annotation Methods. *Pss* B728a was isolated from a snap bean leaflet in Wisconsin (5). The complete genome was sequenced at the Joint Genome Institute

(JGI) using a combination of 3kb and fosmid (40kb) libraries. Library construction, sequencing, finishing, and automated annotation steps used were as described (6). Predicted coding sequences (CDSs) were subjected to manual analysis using an Integrated Microbial Genomes (IMG) annotation pipeline (http://img.jgi.doe.gov). REPs were identified using the sequence identified for palindromic repeats in *P. putida* (7) and using the Blastn search against the *Pss* B728a genome sequence. IS elements were identified using the ISFinder database (www-is.biotoul.fr/).

Comparative Genomics. Comparative analysis of *Pss* B728a and *Pst* DC3000 was performed using a set of tools available in IMG; BLASTp cutoff scores of $E < 10^{-2}$ and 20% identity were used to identify unique genes in *Pss* B728a and *Pst* DC3000 and cutoff scores of $E < 10^{-5}$ and 30% identity were used to find the homologous genes between *Pseudomonas* spp. LexA-binding sites were initially identified using a consensus sequence for Gram-negative bacteria CTG-N_{10}-CAG, which was further refined based on the sequences found upstream of the known members of the LexA regulon.

Results and Discussion

1. Genome Features

The *Pss* B728a genome is composed of one circular chromosome of 6,093,698 bp (Fig. 1, Fig. 1 in Supporting Information on the PNAS web site) harboring 5,217 genes. The *Pss* B728a genome has 3840 genes with predicted functions, 1297 without known functions and 80 RNA genes (Table 1). Repetitive extragenic palindromic (REP) elements are 35bp sequences of highly conserved inverted repeats with the potential of forming a stem-loop structure (7). At least 375 REPs occur in *Pss* B728a (Fig. 1). We identified 12 sequences characterizing REPs in *Pss* B728a (Table 1 in Supporting Information). These repeated sequences are similar but distinct from the REPs found in *P. putida* (7).

Pss B728a shares 4273 genes with *Pst* DC3000 and has 976 genes with no counterparts in *Pst* DC3000 (Table 2). *Pst* DC3000 has 5608 protein-encoding genes while *Pss* B728 has 5090; therefore, *Pst* Dc3000 has 518 more protein encoding genes than *Pss* B728a, most of which are transport and regulatory genes that are among the many duplicated genes in the *Pst* DC3000 genome (8). In addition, *Pst* DC3000 has 28 unique type III secretion system (TTSS) effectors (http://www.pseudomonas-syringae.org/pst_func_gen2.htm) and unique genes associated with coronatine biosynthesis. The remaining unique *Pst* DC3000 genes are categorized as encoding hypothetical or conserved hypothetical proteins. The vast majority of genes on the two plasmids of *Pst* DC3000 are not present in the *Pss* B728a genome. The genome of *Pss* B728a shares with *Pst* DC3000 most of the many genes whose functions are associated with transcriptional regulation, signal transduction, transcription, gene replication, and membrane transport (Fig. 2 in Supporting Information). The distribution of genes in various functional categories in *Pss* B728a and *Pst* DC3000 is similar. There are 9 major rearrangements of the *Pss* B728a genome with respect to that of *Pst* DC3000 (Fig. 2). Since these two genomes each exhibit the expected G + C skews, we assume that the

rearrangements are not recent and that these two *P. syringae* pathovars diverged long ago.

Mobile genetic elements. The genome of *Pss* B728a has 16 IS elements, far fewer than the number found in *Pst* DC3000. *Pst* DC3000 has many more copies of each IS element than *Pss* B728a. The IS elements were found to be more frequent within the regions containing genes of unknown functions within the *Pst* DC3000 genome (9). Two IS elements (Psyr0106 and 0747) of *Pss* B728a are not present in *Pst* DC3000 (Table 2 in Supporting Information).

Genomic islands. Many of the 976 unique genes of *Pss* B728a when compared to *Pst* DC3000 are found in 14 genomic islands (Table 3 in Supporting Information). In particular, a genomic island for a prophage with many genes from phage LambdaSO from *Shewanella oneidensis* (10) occurs in *Pss* B728a. This region is flanked by tRNAs and therefore could have been horizontally acquired. Another genomic island inserted at tRNA-Lys-2 (position 1604394 to 1724175 on the *Pss* B728a genome) is related to the conjugative plasmid/genomic island pKLC102 of *P. aeruginosa* PAO1 (11). Similarly to pKLC102, it carries recombination and replication machinery and a *pil* operon encoding type IV thin sex pili. This island also encodes determinants of copper and arsenic resistance, none of which has homologs in Pst DC3000. As recently reported for *P. fluorescens* Pf-5 (12), REPs are largely absent from the genomic islands (Fig. 1). The lack of REPs suggests that these islands have a different evolutionary history from the rest of the genome.

2. Lifestyle

Siderophores. Siderophores are low molecular weight, high-affinity iron(III) chelators that are exported from the cell and chelate iron in the extracellular milieu; the iron-siderophore complex is then transported into the cell via specific, TonB-dependent receptor proteins, thereby providing iron for cellular functions. The fluorescent pseudomonads are characterized by the production of pyoverdines, a diverse class of siderophores containing a chromophore, which is responsible for the UV fluorescence typifying the group, linked to a small peptide of varied length and composition that is synthesized nonribosomally (13). In *Pss* B728a, as in *Pst* DC3000 (14), genes required for pyoverdine biosynthesis and uptake are present in a single gene cluster (Psyr1944-1961). Genes encoding for the biosynthesis and uptake of a second siderophore, probably yersinibactin, which are present in the genome of *Pst* DC3000 (13), are absent from *Pss* B728a. *Pss* B728a, however, has a second siderophore gene cluster composed of 14 genes with predicted biosynthetic, efflux, uptake, or regulatory functions. The six biosynthetic genes are closely related to genes for the biosynthesis of achromobactin (15), a citrate siderophore produced by *Pectobacterium chrysanthemi* (16) and apparently also by *E. carotovora* pv. *atroceptica* (17). Linked to the biosynthetic genes are genes predicted to encode a TonB-dependent outer membrane receptor and four periplasmic or cytoplasmic membrane proteins also related to those functioning in achromobactin transport in *P. chrysanthemi* (15), as well as FecI/FecR homologs typically involved in regulation of siderophore-mediated iron acquisition (18). Multiple siderophores are

required for full virulence of the plant pathogen *P. chrysanthemi* (15), and the capacity to draw from two distinct siderophore systems may also contribute to virulence or ecological fitness of *Pss* B728a.

UV resistance and tolerance to reactive oxygen species (ROS). To withstand UV-induced DNA damage, bacteria have developed a number of DNA repair mechanisms including photoreactivation, base excision repair (BER), nucleotide excision repair (NER), and damage bypass. DNA polymerase V, encoded by *rulAB* in *P. syringae,* is an error-prone polymerase responsible for translesion synthesis on damaged DNA templates. This enzyme allows bacteria to cope with otherwise lethal effects of UV radiation, albeit at the price of an increased mutation rate.

Pss B728a is far less sensitive than *Pst* DC3000 to UV irradiation (19). Analysis of the genes involved in UV and ROS resistance in *Pst* DC3000 and *Pss* B728a (Table 4 in Supporting Information) reveals that they share a number of pathways for detoxification of ROS and DNA repair, including photorepair, BER and NER. Nevertheless, there are some differences between the two genomes that may explain the higher UV tolerance of *Pss* B728a. Firstly, *Pss* B728a has five catalase isozymes, whereas *Pst* DC3000 has only three. Presumably, this would make the ROS quenching system of *Pss* B728a more efficient, which in turn could result in higher resistance to UVA irradiation. Secondly, *Pss* B728a has two copies of the *rulAB* operon. It has also an orphan *rulA* gene and remnants of two *rulAB* operons, one of them containing full-length *rulA* and truncated *rulB*, and the other containing two pseudogenes. The *rulAB* pseudogenes are highly similar to the full-length *rulAB* genes and appear to be the result of duplication and subsequent inactivation. In contrast, *Pst* DC3000 has fewer copies of *rulAB*, with a single *rulA* on plasmid A and one chromosomal *rulA* and *rulB*, which do not form an operon.

The difference in UV resistance between *Pss* B728a and *Pst* DC3000 could also be due in part to differential expression of the multiple *rulAB* genes in *Pss* B728a. In order to be fully functional, *rulAB* must be expressed in a highly coordinated fashion in response to DNA damage. In most bacteria, synchronized induction of various DNA repair pathways is mediated via LexA, which represses genes by binding to a 16-mer consensus sequence CTG-N_{10}-CAG (20). Analysis of the potential LexA-binding sites in the *Pss* B728a genome shows that both *rulAB* operons are likely to be members of the LexA regulon (Table 5 in Supporting Information). The finding of a LexA box with one mismatch upstream of the *rulAB* operon containing a full-length *rulA* gene and a *rulB* pseudogene suggests recent inactivation of this operon. No LexA boxes were detected in front of the other vestigial *rulAB* operon or the orphan *rulA* gene. Existence of multiple vestigial *rulAB* operons may reveal fluidity in mutagenic DNA repair systems in *Pss* B728a and suggests that its activity is regulated by changing the dosage of *rulAB* genes. The distribution of LexA-binding sites in *Pst* DC3000 tells a different story: while both copies of *rulA* belong to the LexA regulon, the *rulB* gene is devoid of a LexA box. Although the possibility exists that *rulB* is regulated by LexA indirectly via some unknown transcriptional regulator, it is likely that the *rulAB* system in *Pst* DC3000 is not fully functional due to insufficient expression of *rulB*.

Sequence variation within the *rulA* gene appears to have a significant impact upon UV tolerance of the host strain (21). A phylogenetic analysis of the *rulA* alleles found in

different pseudomonads (Fig. 3 in Supporting Information) shows that all *rulA* genes in *Pss* B728a and the chromosomal copy of *rulA* in *Pst* DC3000 (PSTO2121) are not clustered with the majority of *rulA* genes in other strains of *P. syringae*. The presence of the novel *rulAB* variants in highly UV-resistant *Pss* B728a calls for further experimental investigation.

Intra- and interspecies communication. Quorum sensing allows a population of a given bacterial species to synchronize behaviors that might be futile or detrimental unless done as a concerted action. Pseudomonads mediate quorum sensing by secretion of various types of autoinducers including *N*-acyl-homoserine lactones (AHLs), diketopiperazines and furanosyl borate diester (AI-2). However, the repertoire of autoinducers produced by *Pss* B728a appears to be limited to AHLs. *N*-AHL synthase AhlI and the regulator AhlR were identified previously in *Pss* B728a (22) (AhlI-AhlR, Psyr1621-1622), and the single predominant AHL produced by this strain has been shown to co-migrate on a chromatogram with a *N*-3-oxo-hexanoyl-L-homoserine lactone standard (23). Another potential AHL synthase similar to acyltransferases of the PlsC family is encoded by Psyr0009, which has 82% sequence identity with the HdtS protein of *P. fluorescens* F113 (24). HdtS is capable of producing three AHLs, including N-(3-hydroxy-7-cis-tetradecenoyl) homoserine lactone, a molecule also known as the *Rhizobium leguminosarum* small bacteriocin as a consequence of its growth-inhibiting properties. Homologs of both AHL synthases are present in the *Pst* DC3000 genome.

Two S2-type pyocins (Psyr0309-0310 and 4651-4652) were found in the *Pss* B728a genome, only one of which (Psyr0310) has an ortholog in *Pst* DC3000. S2 pyocins produced by *P. aeruginosa* cause death of susceptible bacteria by DNA breakdown due to endonuclease activity; in addition, they are able to inhibit phospholipid synthesis under iron-limited conditions (25). Pyocin expression in *P. aeruginosa* can be induced by treatments that cause DNA damage, but it is independent of LexA repressor (26). However, one of the S-type pyocins in *Pss* B728a is predicted to be a member of the LexA regulon (Table 5 in Supplementary Information).

Copper and antibiotic resistance. Many *P. syringae* strains are resistant to copper and streptomycin, attributable to the wide use of copper- and streptomycin-based bactericides (27). Genes responsible for both copper and streptomycin resistance are present in the genome of *Pss* B728a. Copper resistance genes, including the *copABCD* operon and a *copRS* two-component regulatory system (Psyr1493-1498), reside in a genomic island at tRNA-Lys; these proteins are 92-96% identical to the plasmid-encoded CopABCDRS proteins found in other strains of *P. syringae* (28). A copper-exporting ATPase and associated transcriptional regulator (Psyr0653-0654) might also contribute to copper resistance of *Pss* B728a. This locus is similar to the *cueAR* system found in *P. putida* (29). While this locus is present in *Pst* DC3000, orthologs of *copABCDRS* genes are absent, possibly explaining the lower copper tolerance of *Pst* DC3000 when compared to *Pss* B728a (Lindow, unpublished).

A streptomycin resistance transposon, Tn5393a, which carries a *strA-strB* determinant (Psyr2669-2670) is found in *Pss* B728a, but not in *Pst* DC3000. This cassette resides near the large cluster of genes coding for biosynthesis of lipodepsipeptide toxins and the siderophore achromobactin (Fig. 3). Interestingly, this cluster also includes genes

for modification of lipidA with 4-amino-4-deoxy-L-arabinose (Psyr2689-2696), which confers resistance to cationic antimicrobial peptides and antibiotics, such as polymyxin. The genome of *Pst* DC3000 contains only remnants of the lipid A modification genes, indicating that they were lost by this strain.

Ice nucleation and anti-freeze proteins. A distinctive feature of *Pss* B728a is the presence of an ice nucleation gene (Psyr1608), which is inserted in an exchangeable locus (Fig. 4 in Supporting Information). This outer-membrane ice nucleation protein acts as an ice embryo, a template for the formation of ice crystals (30). *Pss* B728a also has an unlinked gene encoding an anti-freeze protein (Psyr0931) homologous to AfpA of *P. putida* GR12-2 (31). Anti-freeze proteins are secreted into the medium, where they inhibit the growth of external ice by adsorbing onto the ice surface and lowering the temperature at which it can grow. The putative antifreeze gene is found in an operon with 2 glycosyltransferase genes (Psyr0929-0930); 3 genes encoding components of the type I secretion system (Psyr0933-0935) are found on the opposite strand. Most likely these genes participate in glycosylation and secretion of anti-freeze protein. It is possible that the ice nucleation activity of *P. syringae* strains, which are quantitatively quite variable, is modulated by the activity of such antifreeze proteins. Orthologs of ice nucleation and anti-freeze proteins are absent from the genome of *Pst* DC3000.

Osmotolerance. Many bacteria respond to decreased water availability by accumulating compatible solutes, which protect enzymes and stabilize membranes. Some of the compatible solutes accumulated by Pseudomonads include betaine, ectoine and hydroxyectoine, *N*-alpha-acetylglutaminylglutamine amide, mannitol, and glucosylglycerol (32). Both *Pss* B728a and *Pst* DC3000 make betaine from choline due to the presence of choline dehydrogenase and betaine-aldehyde dehydrogenase (Psyr4732 and 4733, respectively). In addition to betaine, *Pss* B728a might also be capable of producing ectoine. Ectoine is synthesized from L-aspartate-4-semialdehyde via a 3-step pathway with L-2,4 diaminobutyrate and N-acetyl-L-2,4 diaminobutyrate as intermediates. A gene with strong similarity to ectoine synthase (Psyr0334) is present in *Pss* B728a but no homologs of *ectB* and *ectA*, which catalyze the first and second steps of the pathway, were found nearby. None of these genes are present in *Pst* DC3000. It is possible that this "orphan" ectoine synthase represents a remnant of an *ectABC* operon that is no longer functional. Alternatively, *N*-acetyl-L-2,4-diaminobutyrate may be produced in *Pss* B728a by a non-orthologous set of enzymes, which do not reside in an operon with ectoine synthase. Indeed, four homologs of a gene encoding L-2,4-diaminobutyrate aminotransferase are present in *Pss* B728a. While three of them most likely participate in biosynthesis of siderophores and phytotoxins, the physiological role of the fourth gene is unknown. There is no close homolog of the acetyltransferase, *ectA*, in the genome; however, it belongs to a large family of N-acetyltransferases. Thirty nine representatives of this family were found in *Pss* B728a, eight of which have no ortholog in *Pst* DC3000, and it is possible that one of these proteins is capable of catalyzing acetylation of L-2,4-diaminobutyrate. This concept is supported by the presence of "orphan" ectoine synthases in the genomes of other microbes that have been sequenced. If an ectoine synthase-like enzyme is functional in *Pss* B728a, it could explain the higher osmotolerance observed in this strain in culture, as compared to *Pst* DC3000 (33), and

also contribute to higher stress tolerance of *Pss* B728a on plants (2, 3). Work to address this hypothesis is underway.

Metabolism. *Pss* B728a shares with *Pst* DC3000 most pathways of central metabolism, such as glycolysis, gluconeogenesis, tricarboxylic acid cycle, pentose phosphate shunt, etc. Differences between the two strains in propionate, tricaproin, L-ascorbate, L-histidine, D-tartrate, erythritol, and L-lactate utilization have been reported (34). However, most of these differences should be attributed to the differences in transcriptional regulation, since the genes responsible for utilization of these substrates are present in both genomes. Unlike *Pst* DC3000, *Pss* B728a has a gene coding for L-lactate dehydrogenase (Psyr0908), which might explain the difference in L-lactate utilization.

3. Virulence

Because *Pss* B728a and *Pst* DC3000 differ both in host range as well as the apparent intimacy with which they interact with host plants, it is not surprising that substantial differences in known virulence determinants are found between these two pathovars. A total of 298 virulence genes conferring a myriad of different functions have been described in *Pst* DC3000 (8) and *Pss* B728a shares many of these genes. In addition, a large set of putative virulence genes are present only in one of the two strains. Pss B728a harbors unique genes such as a cellulase family protein (Psyr4600), a pectate lyase (Psyr0852), a xylanase (Psyr4508), and a type I secretion system (Psyr3075-77).

Phytotoxins. *Pss* is known to produce a number of phytotoxins, which contribute significantly to the virulence of this strain (35). *Pss* produces two classes of lipodepsipeptides: the syringopeptins (SP) and the lipodepsinonapeptides (including syringomycin (SR), syringostatin or pseudomycin). Usually each strain secretes a single type of SP and one or two lipodepsinonapeptides (36). SPs are potent necrosis-inducing phytotoxins, whereas SR enhances bacterial virulence. Both toxins also display strong antibacterial and antifungal activity. *Pss* B728a is known to synthesize two SPs and SR (37) and gene clusters for both phytotoxins (38, 39) as well as a gene encoding an ABC transporter for export of both metabolites (40) are present in the genome of *Pss* B728a, but not in *Pst* DC3000. They are part of a larger cluster (Fig. 3), which also includes genes for siderophore biosynthesis and uptake, streptomycin resistance, arginine degradation, and genes for the modification of lipid A with aminoarabinose. This cluster of 180,637 bp accounts for 2.95 % of the *Pss* B728a genome and appears to be an important virulence determinant of this strain.

Pss strains are also capable of producing a family of peptide derivatives called syringolins. Syringolins have no known impact on the interaction of bacteria with their host plants, but they are recognized by non-host plants, where these peptides activate defense-related genes and induce resistance to fungal pathogens (41). Orthologs of the genes participating in biosynthesis and export of syringolin A (42) were identified in the *Pss* B728a genome (Psyr1701-1706), but not in *Pst* DC3000.

Some pathovars of *P. syringae* produce chlorosis-inducing phytotoxins, such as coronatine and phaseolotoxin. Based on the absence of chlorotic halos around lesions

induced by *Pss* B728a, this strain is presumed to lack the genes for biosynthesis of chlorosis-inducing toxins. Indeed, it lacks the genes for coronatine biosynthesis found in *Pst* DC3000. However, homologs of all six genes from one of the eight phaseolotoxin biosynthesis loci, *phtE* (43, 44), were found in *Pss* B728a (Psyr2549-2555). Orthologs of these six genes are absent from *Pst* DC3000. One additional gene, annotated in GenBank as phaseolotoxin synthetase (45), is a fragment of a pyoverdine sidechain peptide synthetase IV. It is likely that the *phtE* locus participates in biosynthesis of one of the phaseolotoxin precursors and *Pss* B728a lacks other loci for biosynthesis of phaseolotoxin itself.

The *Pss* B728a genome contains a number of genes encoding for non-ribosomal peptide synthases of unknown specificity. Two of them (Psyr2576 and Psyr2577) have orthologs in *Pst* DC3000 that appear to be controlled by the sigma factor HrpL (45). Another, (Psyr3722), has an ortholog in Pst DC3000 that is truncated and therefore likely to be non-functional or to produce a different peptide. A third gene cluster includes three putative non-ribosomal peptide synthases (Psyr1792-1795) with orthologs in *P. fluorescens* Pfo-1 but not in *Pst* DC3000. One of the subunits of this synthase is a hybrid non-ribosomal peptide synthase/polyketide synthase. An operon encoding subunits of a type I polyketide synthase is present in the genome of *Pss* B728a. These genes have some similarity to the *pks* operon in *Bacillus subtilis* 168, which makes a polyketide compound of unknown structure. Thus, *Pss* B728a is able to produce a wide repertoire of secondary metabolites; however, their biological activity and physiological role has yet to be determined.

Secreted factors. *Pss* B728a has 27 TTSS effectors, five of which are not found in *Pst* DC3000 (http://www.pseudomonas-syringae.org/pst_func_gen2.htm). These five include the four unique effectors identified by Greenberg and Vinatzer (46), and another putative unique effector (Psyr1890). This effector bears 41% identity in its C-terminal part to the type III effector HopPtoH, and 28% identity over its entire length to a non-LEE-encoded type III effector D from *Citrobacter rodentium*. We identified two additional putative TTSS effectors common to both organisms namely *avrF1* and *shcZ3* (Psyr1187 and Psyr1225, respectively). *Pst* DC3000 has 28 more TTSS effectors than does *Pss* B728a, having a total of 55 effectors (http://www.pseudomonas-syringae.org/pst_func_gen2.htm). The distinct complement of effectors present in each strain should prove helpful in deciphering the common and unique functions of these proteins.

Auxin. Bacterial IAA can have either a stimulatory or inhibitory effect on plant growth depending on its concentration and the plant tissue exposed (47). Most pathovars of *P. syringae* produce IAA (48), but the amounts produced vary greatly between strains. *Pss* B728a produces substantially higher levels of IAA than *Pst* DC3000 in most culture media (T. Powell, unpublished) and IAA production is known to contribute to virulence and epiphytic fitness of Pss B728a (47). *Pss* B728a has an operon for biosynthesis of IAA, which includes tryptophan monooxygenase, *iaaM* (Psyr1536) and indoleacetamide hydrolase *iaaH* (Psyr1537), and genes first described in *P. syringae* pv. savastanoi (49). The genes in the operon are more than 90% identical to those in the IAA operon of *Pss* Y30 (47). Although *Pst* DC3000 has no orthologs of *iaaM* and *iaaH*, it produces IAA (8)

possibly due to PSPTO0517 and PSPTO0518, which have low sequence similarity to tryptophan monooxygenase and indoleacetamide hydrolase, respectively. Orthologs of these genes are also found in *Pss* B728a.

Pss B728a might additionally produce IAA via a pathway in which indole-3-acetaldoxime, not tryptophan, is the initial precursor. Indole-3-acetaldoxime is a common intermediate in two plant pathways of IAA biosynthesis, and in cruciferous plants it also serves as a branching point in biosynthesis of IAA (49), indole phytoalexins (such as camalexin), and indole glucosinolates (50). Both phytoalexins and glucosinolates are involved in plant defense. A chromosomal cluster found in *Pss* B728a includes a homolog of *Rhodococcus globerulus* aldoxime dehydratase (Psyr0006), which is active on both aliphatic and arylaliphatic aldoximes and a nitrilase (Psyr0007). These two enzymes would catalyze conversion of indole-3-acetaldoxime into indole-3-acetonitrile and IAA, respectively, thus redirecting the flux of indole-3-acetaldoxime towards IAA rather than phytoalexins and glucosinolates (provided that the plant host produces any of those). This would circumvent one of the plant defense mechanisms and might even inhibit the plant hypersensitive response (51). An ortholog of nitrilase is present in *Pst* DC3000, but an aldoxime dehydratase is apparently a pseudogene disrupted by an IS element. Further metabolism of auxin also appears to differ between *Pss* B728a and *Pst* DC3000. A gene encoding indoleacetate-lysine ligase, which metabolizes IAA to an amino acid conjugate devoid of auxin activity, is present in *Pst* DC3000 (PSPTO371), but not in *Pss* B728a.

Pss B728a and *Pst* DC3000 differ markedly from one another with respect to host range and manner of association with host plants. A large number of gene differences that likely contribute to such differential behavior have been described here. The challenge remains to associate these genes to specific traits of the bacterial pathogens. As more strains of *P. syringae* are sequenced, comparative genomic analyses should prove useful in deciphering virulence and behavioral traits of this important species of phytopathogenic bacteria.

We thank the members of The Joint Genome Institute who contributed to the work. Funding for the work described in this study was provided by the Department of Energy, Office of Biological and Environmental Research.

1. Sawada, H., Suzuki, F., Matsuda, I. & Saitou, N. (1999) *J. Mol. Evol.* **49,**627-644.
2. Hirano, S. S. & Upper, C. D. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 624-653.
3. Boureau, T., Routtu, J., Roine, E., Taira, S. & Romantschuk, M. **(**2002) *Mol. Plant Pathology* **3**,451-460.
4. Beattie, G. A. & Lindow, S. E. (1999) *Phytopathology* **89,** 353-359.
5. Loper, J. E. & Lindow, S. E. (1987) *Phytopathology* **77**, 1449-1454.
6. Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., et al. (2003) *J. Bacteriol.* **185**, 2759-2773.
7. Aranda-Olmedo, I., Tobes, R., Manzanera, M., Ramos, J. L. & Marques, S. (2002) *Nucleic Acids Res.* **30**, 1826-1833.

8. 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., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10181-10186.
9. Jensen, L. J., Skovgaard, M., Sicheritz-Ponten, T., Hansen, N. T., Johansson, H., Jorgensen, M. K., Kiil, K., Hallin, P. F. & Ussery, D. (2004) p. 158. in: *Pseudomonas, Life Style and Molecular Architecture*, ed. Ramos, J. L. (Kluwer Academic, New York) **1**, pp. 835.
10. Heidelberg, J. F., Paulsen, I. T., Nelson, K. E., Gaidos, E. J., Nelson, W. C., Read, T. D., Eisen, J. A., Seshadri, R., Ward, N., Methe, B., *et al.* (2002) *Nature Biotech.* **20**, 1118 – 1123.
11. Klockgether, J., Reva, O., Larbig, K. & Tummler, B. (2004) *J. Bacteriol.* **186**, 518-534
12. Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q., Madupu, R. *et al.* (2005) *Nature Biotech.* (in press).
13. Meyer, J. M. (2000) *Arch. Microbiol.* **174**, 135-142.
14. Ravel, J. & Cornelis, P. (2003) *Trends Microbiol.* **11**, 195-200.
15. Franza, T., Mahé, B. & Expert, D. (2005) *Mol. Microbiol.* **55**, 261–275.
16. Muzinger, M., Budzikiewicz, H., Expert, D., Enard, C. & Meyer, J. M. (2000) *Naturforsch* **55**, 328-332.
17. Bell, K. S., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J., Holeva, M. C., Thomson, N. R., Bentley, S. D., Churcher, L. J. C., Mungall, K., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11105-1110.
18. Braun, V., Mahren, S., and Ogierman, M. (2003) Curr. Opin. Microbiol. **6**, 173–180.
19. Jacobs, J. L., Carroll, T. L. & Sundin, G. W. (2005) *Microbial Ecology* **49**, 104-113.
20. Erill, I., Escribano, M., Campoy, S. & Barbe, J. (2003) *Bioinformatics*. **19**, 2225-2236.
21. Sundin, G. W., Jacobs, J. L. & Murillo, J. (2000) *Appl. Environ. Microbiol.* **66**, 5167-5173.
22. Quinones, B., Pujol, C.J., and Lindow, S.E. (2004). *Molec. Plant-Microbe Interact.* **17,**521-531.
23. Cha, C., Gao, P., Chen, Y-C., Shaw, P. D. & Farrand, S. K. (1998) *Mol. Plant Microbe Interact.* **11**, 1119-1129.
24. Laue, B. E., Jiang, Y., Chhabra, S. R., Jacob, S., Stewart, G. S., Hardman, A., Downie, J. A., O'Gara, F. & Williams, P. (2000) *Microbiol.* **146**, 2469-2480
25. Michel-Briand, Y. & Baysse, C. (2002) *Biochimie*. **84**, 499-510.
26. Matsui, H., Sano, Y., Ishihara, H. & Shinomiya, T. (1993) *J. Bacteriol.* **175**, 1257-1263.
27. Sundin, G. W. & Bender, C. L. (1993) *Appl. Environ. Microbiol*. **59**, 1018-1024.
28. Mill, S. D., Jasalavich C. A. & Cooksey D. A. (1993) *J. Bacteriol*. **175**, 1656-1664.
29. Adaikkalam, C. & Swarup, S. (2002) *Microbiology*. **148**, 2857-2867.
30. Green, R. L. & Warren. G. J. (1985) *Nature* **317:**645-648.
31. Muryoi, N., Sato, M., Kaneko, S., Kawahara, H., Obata, H., Yaish, M. W. F., Griffith, M. & Glick, B. R. (2004) *J. Bacteriol*. **187**, 5661-5671.
32. Kets, E. P. W., Galinski, E. A., de Wit, M., de Bont, J. A. M. & Heipieper, H. J. (1996) *J. Bacteriol*. **178**, 6665-6670.

33. Wright, C. A. & Beattie, G. A. (2004) *Proc. Natl. Acad. Sci. USA*. **101**, 3269-3274.
34. Sands, D. C., Schroth, M. N. & Hildebrand, D. C. (1970) *J. Bacteriol.* **101**, 9-23.
35. Bender, C. L., Alarcon-Chaidez, F., & Gross, D. C. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 266-292.
36. Lu, S-E., Soule, J. D., & Gross, D. C. (2003) *Appl. Environ. Microbiol.* **69,** 7273-7280.
37. Grgurina, I., Mariotti, F., Fogliano, V., Gallo, M., Scaloni, A., Iacobellis, N. S., Lo Cantore, P., Mannina, L., van Axel Castelli, V., Greco, M. L. & Graniti, A. (2002) *Biochim. Biophys. Acta*. **1597**, 81-89.
38. Scholz-Schroeder, B. K., Hutchison, M. L., Grgurina, I. & Gross, D. C. (2001a) Mol. Plant Microbe Interact. **14**, 336-348.Greenberg, J. T. & Vinatzer, B. A. (2003) *Cur. Op. Microbiol.* **6,** 20-28.
39. Guenzi E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. (1998) *J. Biol. Chem.* **273**, 32857-32863.
40. Scholz-Schroeder, B. K., Soule, J. D., Lu, S. E., Grgurina, I. & Gross, D. C. (2001b) *Mol. Plant Microbe Interact.* **14**, 1426-1435.
41. Waspi, U., Schweizer, P. & Dudler, R. (2001) *Plant Cell.* **13**, 153-161.
42. Amrein, H., Makart, S., Granado, J., Shakaya, R., Schneider-Pokorny, J., & Dudler, R. (2004) *Mol. Plant Microbe Interact.* **17**, 90-97.
43. Zhang, Y., Rowley, K. B. & Patil, S. S. (1993) *J. Bacteriol.* **175**, 6451-6458.
44. Zhang, Y. X. & Patil, S. S. (1997) *Mol. Plant Microbe Interact.* **10**, 947-960.
45. 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. & Collmer, A. (2002). *Proc. Natl. Acad. Sci. USA*. **99**, 2275-2280.
46. Greenberg, J. T. & Vinatzer, B. A. (2003) *Cur. Op. Microbiol.* **6,** 20-28.
47. Mazzola, M. & White, F. F. (1994) *J. Bacteriol.* **176**, 1374-1382.
48. Glickmann, E., Gardan, L., Jacquet, S., Hussain, S., Elasri, M., Petit, A. & Dessaux, Y. (1998) *Mol. Plant Microbe Interact.* **11**, 156-162.
49. Cohen, J. D., Sloven, J. P. & Hendrickson, A. M. (2003) *Trends Plant Sci.* **8**, 197-199.
50. Glawischnig, E., Hansen, B. G., Olsen, C. E. & Halkier, B. A. (2004) *Proc. Natl. Acad. Sci.* USA. **101**, 8245-8250.
51. Robinette, D. & Matthysse, A. G. (1999) *J. Bacteriol.* **172**, 5742-5749.

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

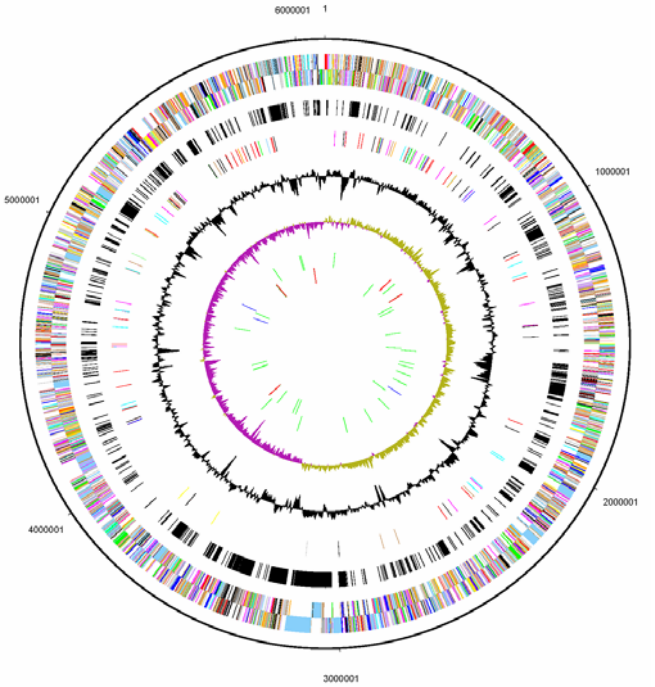


Fig. 1. Chromosome of *Pss* B728a with the outermost circle depicting ORFs on the positive strand and the second circle depicting ORFs on the negative strand. The ORFs are color-coded based on the major grouping of role categories as follows: hypothetical proteins (black); conserved hypothetical proteins (light grey); nucleotide metabolism and transport (light red); translation, ribosomal structure and biogenesis (yellow); signal transduction (lavender); transcription (magenta); lipid metabolism (cyan); carbohydrate metabolism and transport (blue); coenzyme metabolism (pink); energy metabolism (green); DNA replication, recombination and repair (red); general function prediction (brown); cellular processes (sky blue); amino acid metabolism and transport (orange); and structural RNA (pale green). The third circle represents the ORFS with no orthologs in Pst DC3000 ($E < 10^{-5}$0. The fourth circle represents the palindromic REPs color-coded as follows: family 1 (dark grey); family 2 (red); family 3 (green); family 4 (blue); family 5 (cyan); family 6 (magenta); family 7 (yellow); family 8 (pale green); family 9 (sky blue); family 10 (orange); family 11 (brown); and family 12 (light grey). The fifth and sixth circles represent GC content and skew respectively. The seventh and eighth circles represent the tRNA (green), rRNA (red) and misc_RNA (blue).

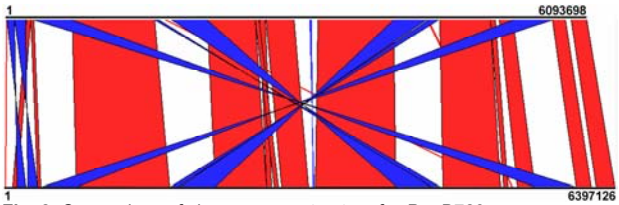


Fig. 2. Comparison of chromosome structure for *Pss* B728a versus *Pst* DC3000. Axes represent the portions coded for in the order in which they occur in the chromosomes. Top axis, B728a; bottom axis, DC3000. Red represents colinear regions on the + strand of both genomes, Blue represents colinear regions that have been inverted in one of the two genomes. The Display was generated by using Artemis (www.sanger.ac.uk/software/Artemis/).

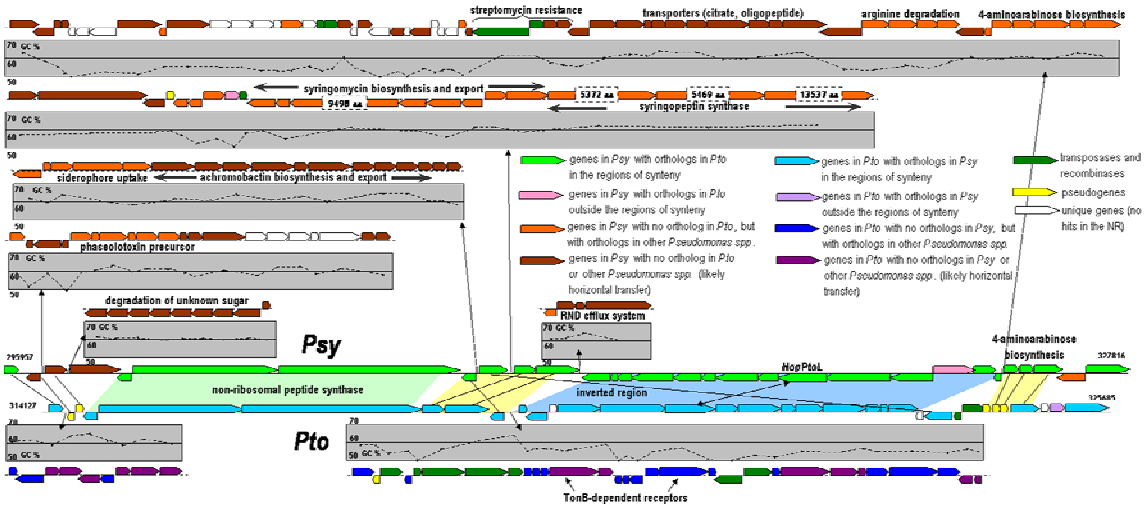


Fig 3. Detailed view of the non-ribosomal peptides and phytotoxins region of the *Pss* B728a genome. Comparison to the corresponding region for the *Pst* DC3000 are also shown.

TABLE 1. Characteristics of the *Pss* B728a genome

Genomic features	#	%
Size, bp	6093698	
DNA coding sequence	5393016	88.50
%GC content		59.23
General statistics		
Total no. of genes	5217	100.00
Protein coding genes	5137	98.47
RNA genes	80	1.53
rRNA genes	16	0.31
5S rRNA	6	0.12
16S rRNA	5	0.10
23S rRNA	5	0.10
tRNA genes	64	1.23
Genes with function prediction	3840	73.61
Genes without function prediction	1297	24.87
with similarity	1255	24.06
without similarity	42	0.81
Pseudogenes	47	0.90

Table 2. A) Unique genes: Number of genes in the column organism without a homolog in the row organism(s). B) Common genes: Number of genes in the column organism with a homolog in the row organism. *Pa* (*P. aeruginosa PAO1*), *Pp* (*P. putida KT2440*), *Pst* (*Pst* DC3000), *Pss* (*Pss* B728a). The calculations were based on a maximum E-value of 1e-5 and minimum identity of 60%.

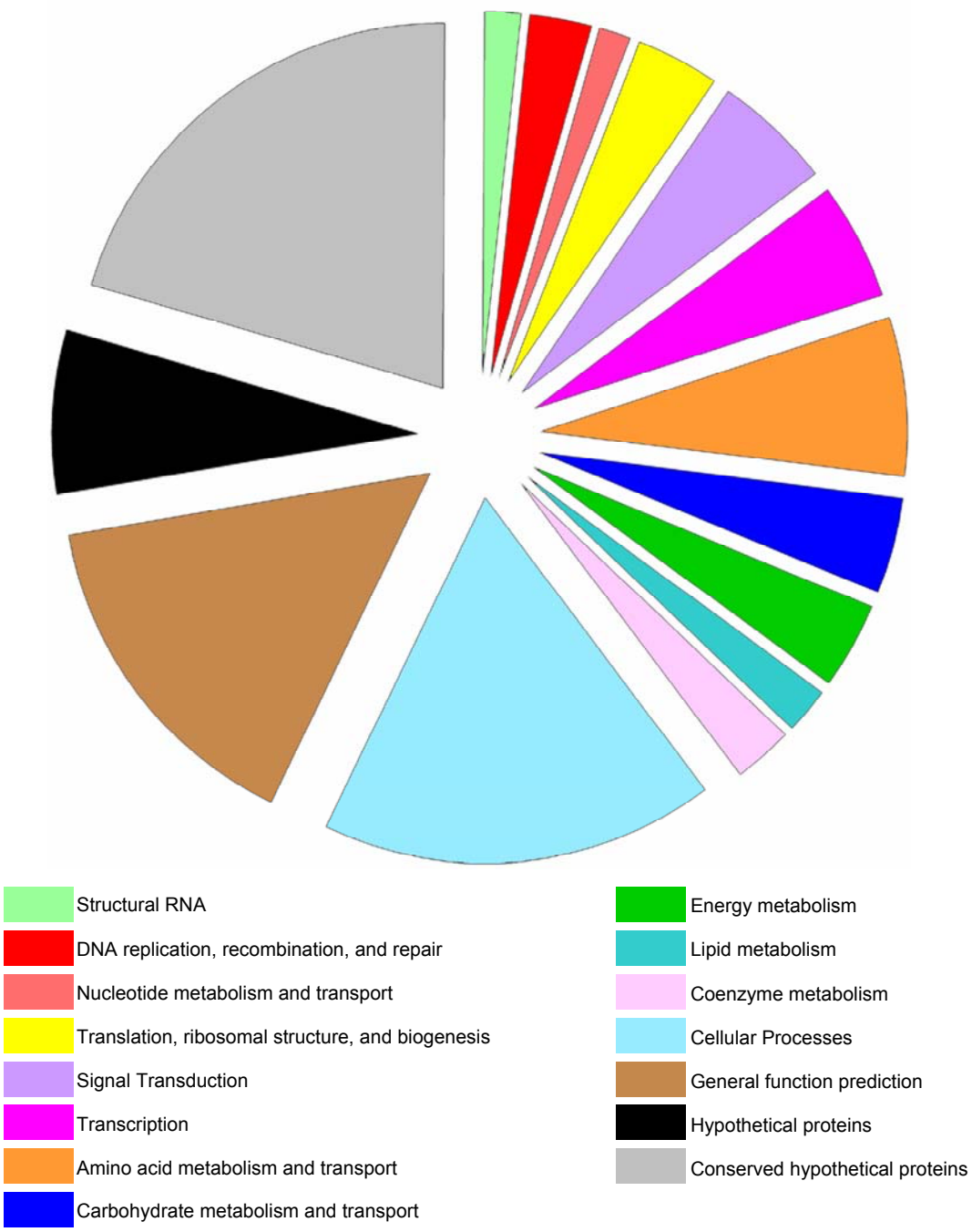
A.

	Pa	*Pp*	*Pst*	*Pss*
Pa		2927	3441	2969
Pp	3201		3017	2624
Pst	3449	2859		976
Pss	3444	2816	1335	
Pa+Pp			2773	2403
Pa+Pst		2383		913
Pa+Pss		2353	1248	
Pp+Pst	2938			884
Pp+Pss	2952		1260	
Pst+Pss	3351	2719		

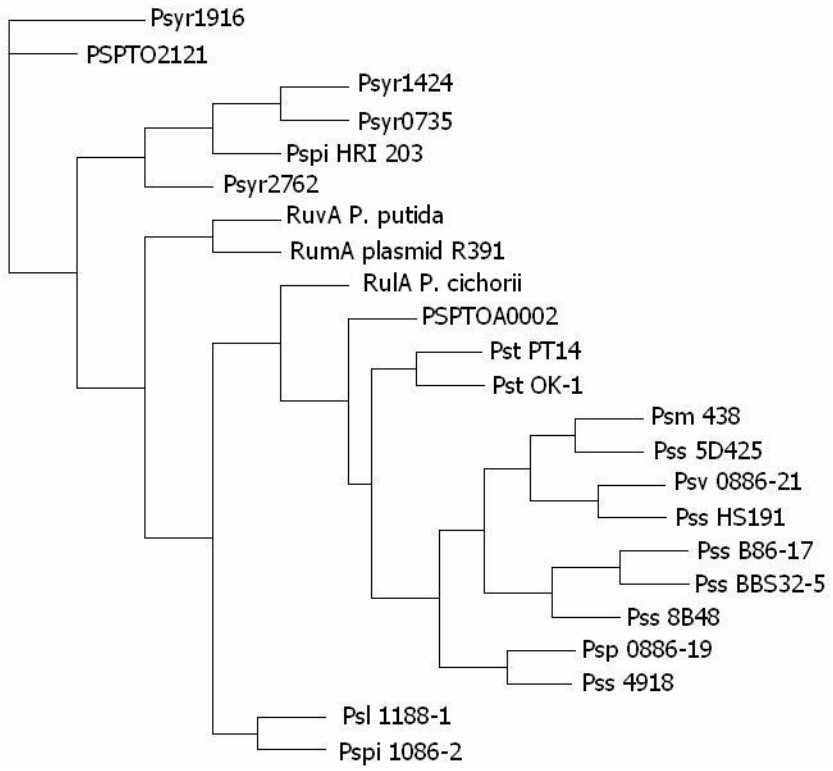
B.

	Pa	*Pp*	*Pst*	*Pss*
Pa		2365	2117	2122
Pp			2591	2534
Pst				4273
Pss				

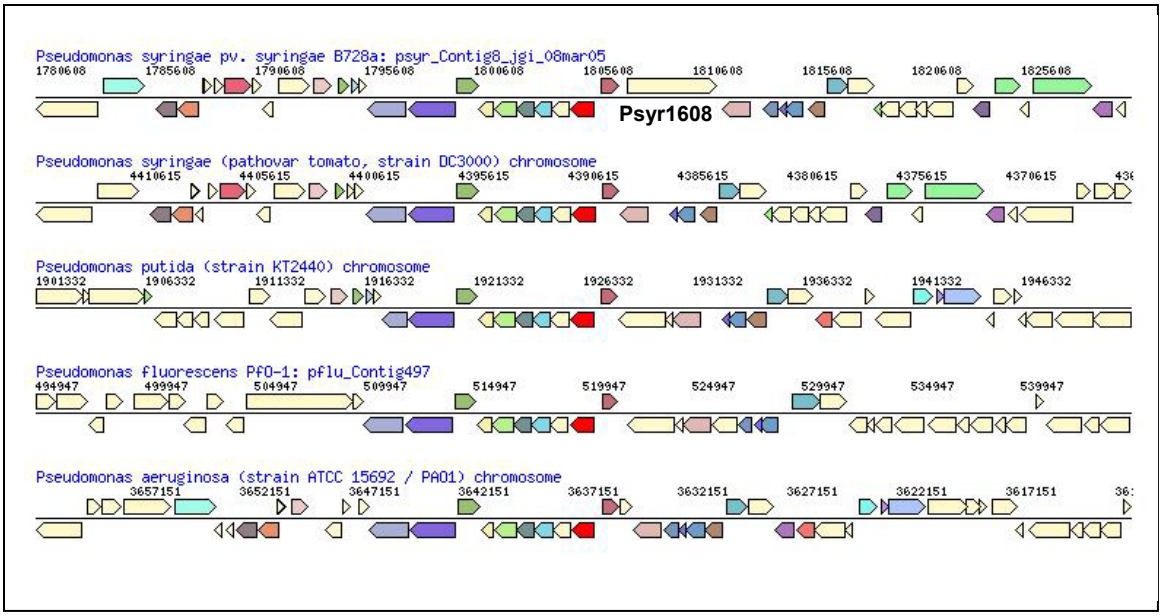
Supplemental Fig 1. Linear map of the *Pseudomonas syringae* pv. *syringae* B728a genome.



Supplemental Fig. 2. Pie chart graph representing the functional classification of annotated ORFs in Pss B728a. The categories were color-coded as follows: structural RNA (pale green); DNA replication, recombination, and repair (red); nucleotide metabolism and transport (light red); translation, ribosomal structure, and biogenesis (yellow); signal transduction (lavender); transcription (magenta); amino acid metabolism and transport (orange); carbohydrate metabolism and transport (blue); energy metabolism (green); lipid metabolism (cyan); coenzyme metabolism (pink); cellular processes (sky blue); general function prediction (brown); hypothetical proteins (black); and conserved hypothetical proteins (light grey).



Supplemental Figure 3. A cladogram of RulA proteins found in the *P. syringae* pv. *syringae* B728a (Psyr1916, Psyr1424, and Psyr2762) and *P. syringae* pv. *tomato* DC3000 (PSPTO2121 and PSPTOA0002) genomes compared to rulA sequences determined in (20), rulA from *P*. syringae pv. pisi HRI 203, rulA from *P. cichorii* 302959, rumA from plasmid R391, and RuvA protein from *P. putida* plasmid pWW0 (GenBank accession numbers CAB96972, AAO89163, AAC45831, and NP_542804, respectively). The two outliers (Psyr1916 and PSPTO2121) are >85% identical to each other, but have only ≈55% identity to the other RulA proteins. These "orphans" have no RulB gene next to them. Only full length RulA genes were included in this tree. A fifth RulA (Psyr1884) was excluded from this phylogenetic tree because it is a psuedogene (59aa out of 144). Phylogeny was inferred using a Protpars method of PHYLIP package (version 3.6, Felsenstein, J. 2004). Other abbreviations: **Pspi**, (*P. syringae* pv. *pisi*); **Pst**, (*P. syringae* pv. *tomato*); **Psm**, (*P. syringae* pv. *maculicola*); **Pss**, (*P. syringae* pv. *syringae*); **Psv**, (*P. syringae* pv. *savastanoi*); **Psp**, (*P. syringae* pv. *phaseolicola*); **Psl**, (*P. syringae* pv. *lachrymans*).



Supplemental Fig 4. Neighboring region of the *Pss* ice nucleating gene (Psyr1608, Bacterial ice-nucleation proteins octamer repeat) compared to conserved neighborhood regions in other Pseudomonads. Regions are aligned to Psyr1606 (ABC transporter, periplasmic substrate-binding protein, marked in red) of *Pss*. Notice the differences occurring in the region between the Bacterial regulatory protein (GntR, marked in brown) and the coding region for Aspartyl aminopeptidase (marked in pink).

Supplemental Table 1. Families of palidromic repetitice extragenic sequences (REPs) and numbers in *Pss* B728a and *Pst* DC3000.

Rep #															# in *Pss*	# in *Pst*
1a	g	c	g	a	a	c	t	t	g	t	t	c	g	c	105	8
1b	g	c	g	a	a	c	a	a	g	t	t	c	g	c	98	8
2a	g	c	g	g	a	c	a	a	g	t	c	c	g	c	26	18
2b	g	c	g	g	a	c	t	t	g	t	c	c	g	c	18	23
3a	g	c	g	a	g	c	a	a	g	c	t	c	g	c	44	5
3b	g	c	g	a	g	c	t	t	g	c	t	c	g	c	44	3
4	g	c	g	g	c	c	t	t	g	g	c	c	g	c	2	0
5	g	c	g	a	c	c	g	c	g	g	t	c	g	c	1	0
6	g	c	g	c	c	c	a	a	g	g	g	c	g	c	1	0
7	g	c	g	a	c	c	a	a	g	g	t	c	g	c	1	0
8a	g	t	g	a	g	c	t	t	g	c	t	c	a	c	21	0
8b	g	t	g	a	g	c	a	a	g	c	t	c	a	c	14	1
consensus	g	c	g	n	n	c	n	n	g	n	n	c	g	c		

All sequences are from the plus strand. Complementary sequences are designated with an a or b.

Supplemental Table 2. IS elements present in *P. syringae* pv. *syringae* B728a.

Element	Family	Copies[a]	Length[b]	Comments[c]
ISPsy9	IS3	5	1452	present in Pst
ISPsy2		1	1194	Psyr0092
ISPsy8	IS3	4	1432	present in Pst
ISPsy6	unclassified	(2)	1603	present in Pst
ISPsy5	IS66	(1)	2059	In Psyr1166
ISPsp1		(1)	1188	Psyr0092
ISPpu14		2	2383	Psyr4387, Psyr4914
ISPpu15		(1)	2041	Within Psyr1166
IS1384		(1)	1178	Psyr0092
ISPa10		(1)	1193	Psyr0092
ISPs1		2	251	Psyr0714, Psyr4887 present in Pst
IS222		2	191	Psyr3589, Psyr5039 present in Psm

[a]The number of full-length copies and the number of incomplete copies in parentheses.
[b]Size in base pairs of the element.
[c]Pst=*Pst* DC3000, Psm=*P. syringae* pv. *maculicola*

Supplemental Table 3. Genomic islands unique to *Pss* B728a when compared to *Pst* DC3000.

Coordinate in *Pss*[*]	Island no.	Starting position	Acquisition evidence	Putative function(s)	Ortholog in
Psyr0005-0008	GI1	6282		Indole aceic acid biosynthesis	
Psyr0092-0129	GI2	101677	tRNA	Hypothetical; membrane associated	
Psyr0734-0757	GI3	839680	tRNA	Hypothetical; Hypersensitive response	*P. syringae* pv. *pisi*
Psyr0767-0777	GI4	872972		Hypothetical; ABC transporter	*Rhizobium* sp.
Psyr0908-0940	GI5	1029403	tRNA	Glycosyl transfer	*P. putida* (KT2440)
Psyr1420-1541	GI6	1605214	tRNA	Hypothetical; RulB; TraG; Cu and As resistance; Pili	*P. syringae* pv. *Pisi*, *P. putida*, *P. aeruginosa* PAO1 Plasmid pKLC102
Psyr1702-1706	GI7	1911953		Non-ribosomal Peptide, syringolin	
Psyr1790-1796	GI8	2042710		Non-ribosomal Peptide, syringomycin	
Psyr1882-1932	GI9	2153993	tRNA	RulA;RulB; Transcriptional regulation	*P. cichorii*, *Photorhabdus luminescens* subsp. *lamondii*
Psyr2170-2197	GI10	2523343	tRNA	Transcriptional regulation;	*Bradirhizobium japonicum*
Psyr2545-3091	GI11	2960263	tRNA	Siderophore/achromobactin; Non-ribosomal Peptide, syringopeptin; Hypothetical; transport; arginine metabolism; Prophage LambdaSo Hypothetical; Transcriptional regulation Transcriptional regulation; transport Two component regulatory system; acriflavin resistance	*Erwinia chrysanthemi*, *P. aeruginosa*, *Schewanella oneidensis* *P. putida* *Erwinia carotovora* *Xanthomonas axonopodis* pv. *citri*
Psyr4308-4664	GI12	5133734		Polyketide synthesis Membrane association Phage FluMu Pyocin synthesis	*P. putida* KT2440 *P. aeruginosa*
Psyr4917-4928	GI13	5828952	tRNA	Bacteriophage K139; HP2	
Psyr4947-5000	GI14	5863519	tRNA	Hypothetical; Outer membrane association	*P. putida* KT2440

[*]The region of coding sequences given represents one or more unique genomic islands separated by coding sequences common to both genomes.

Supplemental Table 4. DNA repair and ROS resistance proteins in Pss B728a and Pst DC3000. Pseudogenes are marked with stars.

System	Pathway	Function	Gene in Pss B728a	Gene in Pst DC3000
DNA repair	Photorepair	Deoxyribopyrimidine photolyase	Psyr0961	PSPTO1121
	Nucleotide excision repair	Excinuclease ABC subunit A	Psyr4520	PSPTO0654
		Excinuclease ABC subunit B	Psyr1974	PSPTO2164
		Excinuclease ABC subunit C	Psyr2896	PSPTO3023
	Base excision repair	Endonuclease III (Nth)	Psyr3888	PSPTO4149
		Formamidopyrimidine-DNA glycosylase	Psyr4761	PSPTO5342
		A/G-specific adenine glycosylase MutY	Psyr4901	PSPTO5342
		Uracil-DNA glycosylase	Psyr3970	PSPTO4236
		DNA-3-methyladenine glycosylase I	Psyr0013	PSPTO0183
		DNA-3-methyladenine glycosylase II	Psyr3366	
	Mismatch repair	DNA mismatch repair protein MutS	Psyr1376	PSPTO4058
		DNA mismatch repair protein MutL	Psyr0570	PSPTO4944
		ATP-dependent DNA helicase UvrD	Psyr5065	PSPTO5516
		Exodeoxyribonuclease I	Psyr4020	PSPTO4316
		Exodeoxyribonuclease VII large subunit (xseA)	Psyr1259	PSPTO1446
		Exodeoxyribonuclease VII small subunit (xseB)	Psyr0606	PSPTO0700
		Exonuclease RecJ	Psyr1298	PSPTO1488
Translesion synthesis	Error-prone DNA polymerases	DNA polymerase II (polB)	Psyr2361	PSPTO2621
		DNA polymerase IV(DinB)	Psyr1397	PSPTO3990
		DNA polymerase V RulB	Psyr2761, Psyr1425, Psyr0736*, Psyr1885*	PSPTO0591

		DNA polymerase V subunit RulA	Psyr2762, Psyr1424, Psyr0735, Psyr1916, Psyr1884*	PSPTO2121, PSPTOA0002
General DNA recombination and repair proteins		RecA protein	Psyr1378	PSPTO4033
		RecX protein	Psyr1379	PSPTO4032
		Transcription-repair coupling factor	Psyr1896	PSPTO2101
		DNA polymerase I	Psyr0270	PSPTO0344
		DNA ligase (NAD)	Psyr1819	PSPTO3656
		DNA ligase (ATP), putative	Psyr3245, Psyr3873	PSPTO3464, PSPTO4135
ROS resistance	Catalases, peroxidase and SOD	Catalase KatA	Psyr4522	
		Catalase KatB	Psyr3353	PSPTO3582
		Catalase KatE	Psyr0280	PSPTO5263
		Manganese catalase KatN	Psyr5095	
		Catalase/peroxidase HPI	Psyr4208	PSPTO4530
		Mn superoxide dismutase	Psyr4152	PSPTO4459
		Fe superoxide dismutase	Psyr4059	PSPTO4368
		Cu-Zn superoxide dismutase	Psyr1154	PSPTO1338
	Bacterioferritins and Dps family proteins	Bacterioferritin	Psyr3897, Psyr4521	PSPTO4160, PSPTO0653
		Iron-binding protein Dps	Psyr1404	PSPTO3983
		Bacterioferritin family protein	Psyr4448	PSPTO4906
	Organic hydroperoxide resistance proteins	Organic hydroperoxide resistance protein Ohr (OsmC family)	Psyr0040, Psyr3627	PSPTO0152, PSPTO1767
		Peroxiredoxin AhpC	Psyr4877, Psyr2975	PSPTO5317, PSPTO3108
		Peroxiredoxin reductase AhpF	Psyr2974	PSPTO3107

Supplemental Table 5. Members of the LexA regulon in Pss B728a and Pst DC3000. Pseudogenes are marked with stars.

Function	Gene in Pss B728a	LexA box in Pss B728a	Ortholog in Pst DC3000	LexA box in DC3000
Confirmed members				
LexA repressor	Psyr3283	ctgtatataatcccag, ctgtataaaaagccag	PSPTO3510	ctgtatataatcccag, ctgtataaaaagccag
RecA protein	Psyr1378	ctgtctacttatacag	PSPTO4033	ctgtttacttatacag
RecN protein	Psyr4197	ctgtataaataaccag	PSPTO4507	ctgtataaatatccag
Exonuclease RecJ	Psyr1298	ctgcacagagaatcag	PSPTO1488	No LexA box
ATP-dependent helicase DinG	Psyr3856	ctgaaaatttatccag	PSPTO4120	ctgaaaatttatccag
RulAB operon	Psyr2762, Psyr2761	ctgtatatgcatacag	No ortholog	
RulAB operon	Psyr1424, Psyr1425	ctgtatatatatacag	No ortholog	
RulAB operon	Psyr0735, Psyr0736*	ctgtatatatatacac	No ortholog	
RulA protein	Psyr1916		PSPTO2121	ctgtatatgcatacag
RulA protein	No ortholog		PSPTOA0002	ctgtacgtctatacag
YebG protein	Psyr1033	ctgtataaaaacacag	PSPTO1197	ctgtataaaaacacag
Predicted members				
RecX protein	Psyr1379	operon with recA	PSPTO4032	operon with recA
DNA topoisomerase III (topB)	Psyr2244	ctgtatatccatacag	PSPTO2478	ctgtatatccatacag
Radical SAM family protein	Psyr2365	ctgtatatacaaacag	PSPTO2625*	ctgtacatccaaacag
Uracil-DNA glycosylase family protein	Psyr2364	operon with Psyr2365	PSPTO2624	operon with PSPTO2625
Hypothetical protein	Psyr3807	ctgtatataaagacag	No ortholog	
Hypothetical protein	Psyr3808	operon with Psyr3807	No ortholog	
Hypothetical protein	Psyr3809	operon with Psyr3807	No ortholog	
Pyocin S2	Psyr4651	ctgtatatatatacag	No ortholog	
Pyocin immunity protein	Psyr4652	operon with Psyr4652	No ortholog	

DNA adenine methylase family	No ortholog		PSPTO3388	ctgttcatatatacag
MoxR family	Psyr1721	ctgttcgccctgccag	PSPTO3759	ctgataaccctgccag
Conserved hypothetical protein	Psyr1720	operon with Psyr1721	PSPTO3760	operon with PSPTO3759
Conserved hypothetical protein	Psyr1719	operon with Psyr1721	PSPTO3761	operon with PSPTO3759
Willebrand factor type A domain protein	Psyr1718	operon with Psyr1721	PSPTO3762	operon with PSPTO3759
TPR domain protein	Psyr1717	operon with Psyr1721	PSPTO3763	operon with PSPTO3759
Conserved hypothetical protein	Psyr1716	operon with Psyr1721	PSPTO3764	operon with PSPTO3759