nature biotechnology

Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*

Nicolas Vodovar[1], David Vallenet[2], Stéphane Cruveiller[2], Zoé Rouy[2], Valérie Barbe[2], Carlos Acosta[1], Laurence Cattolico[2], Claire Jubin[2], Aurélie Lajus[2], Béatrice Segurens[2], Benoît Vacherie[2], Patrick Wincker[2], Jean Weissenbach[2], Bruno Lemaitre[1], Claudine Médigue[2] & Frédéric Boccard[1]

Pseudomonas entomophila **is an entomopathogenic bacterium that, upon ingestion, kills** ***Drosophila melanogaster*** **as well as insects from different orders. The complete sequence of the 5.9-Mb genome was determined and compared to the sequenced genomes of four** ***Pseudomonas*** **species.** ***P. entomophila*** **possesses most of the catabolic genes of the closely related strain** ***P. putida*** **KT2440, revealing its metabolically versatile properties and its soil lifestyle. Several features that probably contribute to its entomopathogenic properties were disclosed. Unexpectedly for an animal pathogen,** ***P. entomophila*** **is devoid of a type III secretion system and associated toxins but rather relies on a number of potential virulence factors such as insecticidal toxins, proteases, putative hemolysins, hydrogen cyanide and novel secondary metabolites to infect and kill insects. Genome-wide random mutagenesis revealed the major role of the two-component system GacS/GacA that regulates most of the potential virulence factors identified.**

Pseudomonas spp. are ubiquitous Gram-negative bacteria that colonize and survive in numerous ecological niches including soil, water and plant surfaces. This versatility is reflected by the sizes of their genomes, which contain large sets of genes involved in carbon source utilization and adaptation. In 2001, we isolated a bacterial strain closely related to the saprophytic soil bacterium *Pseudomonas putida*, *Pseudomonas entomophila,* which triggers a systemic immune response in *D. melanogaster* after ingestion[1]. *P. entomophila* is highly pathogenic for both *D. melanogaster* larvae and adults. Its persistence in larvae leads to a massive destruction of gut cells[1].

Entomopathogenic bacteria such as the Gram-negative bacteria *Photorhabdus luminescens*, *Xenorhabdus nematophilus*, *Yersinia pestis*, *Serratia marcescens* and *Serratia entomophila* and the Gram-positive bacterium *Bacillus thuringiensis* have developed different strategies to interact with and kill insects[2]. Some gene products derived from these bacteria as well as the bacteria themselves, have been used to generate biopesticides[3]. The ability of *P. entomophila* to orally infect and kill larvae of insect species belonging to different orders makes it a promising model for the study of host-pathogen interactions and for the development of biocontrol agents against insect pests. To unravel features contributing to *P. entomophila's* entomopathogenic properties, we have determined its complete genome sequence and performed a genome-wide screen for mutants affected in their ability to trigger an immune response and lethality in *D. melanogaster*.

RESULTS

Genome features and comparative genomics

The *P. entomophila* genome is composed of a single circular chromosome of 5,888,780 base pairs (**Fig. 1**). Among 5,169 coding sequences identified, 3,466 genes (67%) have been assigned a predicted function (**Table 1**). The *P. entomophila* genome is smaller than the six other *Pseudomonas* genomes that have been published (**Table 1**): the human opportunistic pathogen *P. aeruginosa* PAO1 (ref. 4), the three *P. syringae* pathovars[5–7], the plant commensal *P. fluorescens* Pf-5 (ref. 8) and the saprophytic soil bacterium *P. putida* KT2440 (ref. 9).

GC skew analysis and the predicted location of the origin of replication *oriC* near *dnaA* and of the chromosome dimer resolution *dif* site in PSEEN2780 revealed the presence of two replichores of similar size, contrary to the unbalanced replichores found in the genomes of *P. putida* KT2440 (ref. 10) and *P. aeruginosa* PAO1 (ref. 4) (see **Supplementary Fig. 1** online). BLAST comparisons of genomes from the five *Pseudomonas* representative species identified a set of 2,065 genes that constitutes the *Pseudomonas* core genome. Based on this analysis, we identified 1,002 genes unique to the *P. entomophila* genome. We found that, consistent with the close relatedness between *P. entomophila* and *P. putida*[1], 70.2% of *P. entomophila* genes (3,630) have orthologs in the *P. putida* genome, of which more than 96% are found in synteny (see **Supplementary Table 1** online). The smaller size of the *P. entomophila* genome compared to that of other *Pseudomonas* does not seem to originate from reductive evolution.

[1]Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette, France. [2]Genoscope, Centre National de Séquençage and CNRS-UMR8030, 2 rue Gaston Crémieux, 91057 Evry Cedex, France. Correspondence should be addressed to F.B. (boccard@cgm.cnrs-gif.fr).

Received 30 January; accepted 7 April; published online 14 May 2006; doi:10.1038/nbt1212

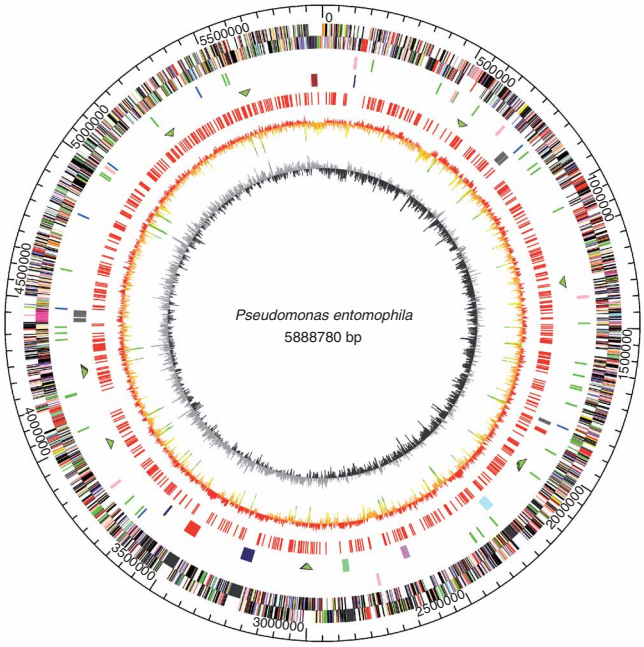

Figure 1 Circular representation of the *P. entomophila* genome. The outer scale indicates coordinates in base pairs (bp). Circles 1 and 2 (from outside to inside) show predicted coding regions transcribed clockwise and counterclockwise, respectively. Coding sequences are color coded by role categories: salmon, amino acid biosynthesis; light blue, biosynthesis of cofactors, prosthetic groups and carriers; light green, cell envelope; red, cellular processes; brown, central intermediary metabolism; yellow, DNA metabolism; green, energy metabolism; purple, fatty acid and phospholipid metabolism; violet, mobile and extrachromosomal element functions; pink, protein synthesis and fate; orange, purines, pyrimidines, nucleosides and nucleotides; navy blue, regulatory functions and signal transduction; lime green, secondary metabolite biosynthesis; gray, transcription; teal, transport and binding proteins; black, unknown function and hypothetical proteins. Circle 3 shows rRNA genes in salmon, tRNA genes in green and miscellaneous RNA genes in blue. Circle 4 shows transposase genes, putative prophages and gene clusters encoding secondary metabolites coded by colored symbols as follows: green arrowheads, transposases; gray, putative prophages; red, pyoverdine synthesis; light blue, cluster involved in lipopeptide II biosynthesis; violet, acinetobactin-like siderophore synthesis; light green, cluster involved in lipopeptide III biosynthesis; navy blue, cluster and isolated genes involved in lipopeptide I biosynthesis; pink, hydrogen cyanide production; brown, polyketide synthesis. Circle 5 shows the distribution of REPs. These repeats are scattered all over the genome and were found either as single elements, in paired elements or in clusters of up to six elements in alternating orientation. Circle 6 shows G+C in relation to the mean G+C in a 1,000-bp window. Circle 7 shows GC skew in a 1,000-bp window.

Indeed the 50 genes of *P. entomophila* present in other *Pseudomonas* but absent from *P. putida* belong to functional classes as diverse as the 34 genes of *P. putida* present in other *Pseudomonas* but absent from *P. entomophila*. Furthermore, comparison of gene contents in *P. entomophila* and *P. putida* indicates that the higher number of species-specific genes in *P. putida* (1,774 versus 1,539) largely results from the presence of a higher number of paralogous genes (**Fig. 2** and **Supplementary Table 2** online). Comparison of the chromosome structures of *P. entomophila* and *P. putida* KT2440 and scatter plot analysis of syntenic regions of the two strains revealed frequent genetic inversions that reverse the genomic sequence symmetrically across *oriC* as observed in other bacterial genera[11] (**Fig. 2** and **Supplementary Fig. 2** online). The same rearrangement profile was observed when comparing the *P. entomophila* genome with those of other *Pseudomonas* spp., even though the levels of orthology and of synteny were lower (see **Supplementary Table 1** and **Supplementary Fig. 2** online). A search for repetitive extragenic palindromic sequences (REPs) identified 943 REPs similar to those found in the genomes of *P. putida* KT2440 (ref. 12) and *P. fluorescens* Pf-5 (ref. 8). The genome of *P. entomophila* has been remodeled by genetic mobile elements and bacteriophage insertions considerably less than the genomes of other environmental pseudomonads such as *P. putida* KT2440 and *P. syringae* pv. tomato DC3000 (**Fig. 1**). Particularly notable are three clustered prophages related to FluMu phage, a pyocin-like phage and a lambdoid phage; they are inserted between *recA* and *mutS*, as observed for FluMu phage in *P. fluorescens* Pf-5 genome. Also of particular interest are two putative prophages inserted in genes encoding 4.5S RNA and tmRNA, respectively. The genome of *P. entomophila* contains only nine genes encoding transposase-like proteins including three that are remnant or inactive. Unlike the genomes of *P. putida* KT2440 and *P. syringae* pv. *tomato* DC3000, the genome of *P. entomophila* is devoid of type II introns.

Toxins against insects

We used several criteria to uncover genes that may contribute to the entomopathogenic properties of *P. entomophila*: specificity to the *P. entomophila* genome, localization within genomic islands that suggest recent lateral acquisitions (based on break of the synteny, GC content and absence of REPs) and similarity to genes associated with virulence in other systems (**Table 2**).

Particularly striking are three genes absent from other *Pseudomonas* genomes that encode proteins related to insecticidal toxin complexes that have been found only in entomopathogenic enterobacteria such as *Photorhabdus luminescens*, *Serratia entomophila*, *Xenorhabdus nematophilus* or in *Yersinia* spp.[13,14]. Three basic types of genetic

Table 1 General features of genomes of representative *Pseudomonas* species

General features	*Pe*	*Pp*[a]	*Pf*[a]	*Pa*[a]	*Pst*[a]
Size (Mb)	5.9	6.2	7.1	6.3	6.4
GC (%)	64.2	61.6	63.3	66.6	58.4
Nb CDS	5169	5420	6144	5570	5615
Coding (%)	89.1	87.7	88.8	89	86.8
rRNA operon	7	7	5	4	5
tRNA	78	74	71	63	63
Protein with predicted function (%)	67.1	65.8	62.2	54.2	61.0
Proteins without predicted function					
Conserved hypothetical proteins (%)	25.3	19.1	32.5	13.8	17.0
Hypothetical proteins (%)	7.5	15.1	5.3	31.9	22.0

[a]The distributions of ORFs for the published chromosomes are derived from the original annotation. These numbers, particularly those of hypothetical and conserved hypothetical proteins, may be different from numbers obtained with updated BLAST searches and annotations. Features of the genomes of *P. syringae* pv. syringae B728a (6.1 Mb) and *P. syringae* pv. phaesolicola 1448A (5.9 Mb) are not indicated. CDS, coding sequences; *Pe*, *P. entomophila*; *Pa*, *Pseudomonas aeruginosa*; *Pp*, *Pseudomonas putida*; *Pf*, *Pseudomonas fluorescens* Pf-5; *Pst*, *Pseudomonas syringae* pv. tomato DC3000.

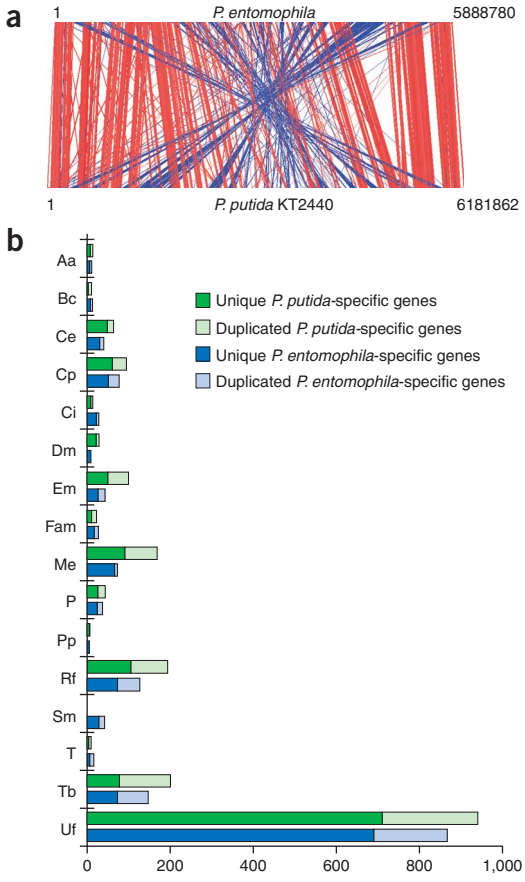

Figure 2 Comparison of the *P. entomophila* and *P. putida* genomes. (**a**) Regions of significant sequence identity between the nucleotide sequence of *P. entomophila* (top) and *P. putida* KT2440 (bottom). Colinear regions are connected by red lines and inverted regions by blue lines. The display was generated using Artemis Comparison Toll (freely available at http://www.sanger.ac.uk/Software/ACT/). (**b**) Specific gene content comparison of the genomes of *P. entomophila* and *P. putida* KT2440. Specific genes of *P. entomophila* (*Pe*) and of *P. putida* KT2440 (*Pp*) with no ortholog in the other species are indicated in blue and green respectively, and are classified according to role categories as described in **Figure 1**. Two genes were considered as orthologs when their products share more than 60% identity over more than 80% of their length. Duplicated genes indicated by light colors were detected by using a constraint of 35% identity over more than 80% of the length of the protein. Aa, amino acid biosynthesis; Bc, biosynthesis of cofactors, prosthetic groups and carriers; Ce, cell envelope; Cp, cellular processes; Ci, central intermediary metabolism; Dm, DNA metabolism; Em, energy metabolism; Fam, fatty acid and phospholipid metabolism; Me, mobile and extrachromosomal element functions; P, protein synthesis and fate; Pp, purines, pyrimidines, nucleosides and nucleotides; Rf, regulatory functions and signal transduction; Sm, secondary metabolite biosynthesis; T, transcription; Tb, transport and binding proteins; Uf, unknown function and hypothetical proteins.

elements encode insecticidal toxin complexes: *tcdA*-, *tcdB*- and *tccC*-like genes. The *P. entomophila* genome encodes three TccC-type insecticidal toxins (PSEEN2485, PSEEN2697, PSEEN2788) (see **Supplementary Fig. 3** online). In addition to these three insecticidal toxins, the *P. entomophila* genome, like that of *P. syringae*, encodes proteins more distantly related to TccC-type toxins (PSEEN701 and PSEEN702) and to TcdB-type toxins (PSEEN1172). The three *P. entomophila* insecticidal toxins likely play a major role in the pathogenicity of *P. entomophila* as TccC and TcdB proteins have been shown to have entomocidal activity[15,16], even though the molecular mechanisms remain to be characterized. These findings highlight the efficient spreading of toxin-complex gene homologs in insect-interacting soil bacteria belonging to different genera.

Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture by poorly defined mechanisms[17]. Contrary to the other *Pseudomonas* tested, *P. entomophila* secretes a strong diffusible hemolytic activity (see **Supplementary Fig. 4** online) that may also be involved in pathogenicity against *D. melanogaster*. We identified three genes unique to *P. entomophila* that may be responsible for this activity (**Table 2**). The gene encoding PSEEN3925, a putative repeats-in-toxin (RTX) protein, is clustered with genes encoding a type I secretion system. PSEEN0968 and PSEEN3843 are proteins related to outer membrane autotransporters that have been associated with virulence in other bacteria. A number of lipases have also been shown to confer hemolytic activity. The *P. entomophila* genome encodes four lipases that are absent from *P. putida* KT2440 and that may contribute to its hemolytic activity (PSEEN709, PSEEN1065, PSEEN2195, PSEEN3432). Interestingly, the gene encoding a lysophospholipase (PSEEN709) is found in a genomic islet associated with two genes encoding proteins related to insecticidal toxins.

Proteases constitute another important group of extracellular, biologically active substances that are thought to contribute to the virulence of bacterial species. *P. entomophila* encodes three serine proteases (PSEEN3027, PSEEN3028, PSEEN4433) and an alkaline protease (PSEEN1550) absent from *P. putida* KT2440. These four genes are located at synteny break points between the genomes of *P. entomophila* and other *Pseudomonas* spp. PSEEN1550 is the homolog of the alkaline protease AprA, which has been shown to be involved in various virulence processes among different species[18]. AprA likely plays a key role in virulence because pathogenicity is affected in mutants defective in PrtR, the predicted transcriptional regulator of *aprA* (see below).

Pathogenic bacteria rely on a variety of cell surface–associated virulence factors that allow adhesion to the host surface and promote effective colonization. Filamentous hemagglutinin-like adhesins are broadly important virulence factors in both plant and animal pathogens. The genome of *P. entomophila* encodes three proteins (PSEEN0141, PSEEN2177, PSEEN3946) that are predicted to be involved in adhesion and cluster with genes encoding type I or two-partner secretion system proteins (**Table 2**). We also noticed the presence of two putative autotransporter proteins with a pertactin-type adhesion domain.

Toxins against competitors

In addition to the putative toxins described above that may be crucial for its entomopathogenic properties, *P. entomophila* carries a number of genes specifying diverse traits that may be required not only for interaction with insects but also for its lifestyle in soil, aquatic or rhizosphere environments (see **Supplementary Fig. 5** online).

Fluorescent pseudomonads are characterized by the production of pyoverdines, a diverse class of siderophores containing a chromophore linked to a small peptide of varying length and composition synthesized by nonribosomal peptide synthases[19]. In *P. entomophila*, the two gene clusters that encode proteins required for pyoverdine biosynthesis and uptake (PSEEN1813-PSEEN1815 and PSEEN3224-3234) present a general organization similar to that found in other fluorescent pseudomonads[20]. We also identified a gene cluster responsible

Table 2 Gene/gene products potentially involved in *P. entomophila*-*D. melanogaster* interaction

Gene/gene product[a,b,c]	Function	Ps.[d]
Adhesion		
PSEEN0141[a]	Putative surface adhesion protein	54% PP0168[e]
PSEEN2177[a]	Putative filamentous hemagglutinin	51% PFL4237
PSEEN3946[a]	Putative filamentous hemagglutinin	41% PA0041
PSEEN3161	Putative autotransporter, pertactin-like protein	63% PP3069
PSEEN4310[a]	Putative autotransporter, pertactin-like protein	42% PSPTO2225
Proteases		
aprA [c]	Alkaline metalloprotease	72% PSPTO3332
PSEEN3027[b]	Putative autotransporter, SSP-h1 serine protease	68% PSPTO1650
PSEEN3028[b]	Putative autotransporter, serine protease	64% PA3535
PSEEN4433[a]	Putative subtilisin-like serine protease	Absent
Lipases		
PSEEN0709[b]	Lysophospholipase	76% PA2540
PSEEN1065[b]	Phospholipase C	62% PFL0888
PSEEN2195	Triacylglycerol lipase	64% Pf B52 (P21773)[g]
PSEEN3432[a,b]	Lipase class3	48% Pfo0149
Toxins		
hcnABC[c]	Hydrogen cyanide production	76% PA2193 (*hcnA*)
PSEEN0132/3332/3042-5[a,b]	Cluster involved in lipopeptide I biosynthesis	See text[h]
PSEEN2138-56[a,b]	Cluster involved in lipopeptide II biosynthesis	Absent
PSEEN2716-20[b]	Cluster involved in lipopeptide III biosynthesis	77% Pfo2266 (2717)
PSEEN5524-36[a,b]	Cluster involved in polyketide biosynthesis	Absent
PSEEN0701[a,b]	Protein related to TccC-type insecticidal toxin	Absent[f]
PSEEN0702[a,b]	Protein related to TccC-type insecticidal toxin	Absent[f]
PSEEN1172[a]	Protein related to TcdB-type insecticidal toxin	Absent[f]
PSEEN2485[a,c]	TccC-type insecticidal toxin	Absent
PSEEN2697[a,b,c]	TccC-type insecticidal toxin	Absent
PSEEN2788[a,b,c]	TccC-type insecticidal toxin	Absent
PSEEN3326[a,b]	Putative toxin (cytolethal distending toxin B domains)	Absent
PSEEN3925-9[a]	Putative RTX toxin and type I secretion system	Absent
Miscellaneous		
PSEEN0968[a,b]	Putative autotransporter with unknown passenger domain	Absent
PSEEN3843[a]	Putative autotransporter with unknown passenger domain	53% PSPTO0714
Noninfectious and nonlethal Tn5 derivatives[i]		
gacS$^{(5)}$	Sensor histidine kinase	88% PP1650
gacA$^{(2)}$	Response regulator, LuxR family	98% PP4099
bioC$^{(1)}$	Biotin biosynthesis	86% PP0365
PSEEN5207$^{(1)}$-8$^{(2)}$	Putative amino acid ABC transporter	97%/82% PP0283-2
PSEEN4425$^{(2)}$	CHP[j], CAIB/BAIF family	62% PFL4631
Infectious and nonlethal Tn5 derivatives[i]		
prtR$^{(3)}$	Transmembrane transcriptional regulator	74% PP2889
algR$^{(2)}$	Transcriptional regulator involved in alginate production	91% PP0185
PSEEN0132$^{(3)}$-3$^{(1)}$	NRPS loading protein, CHP (operonic)	59%/75% PSPTO5546-7
PSEEN0389$^{(1)}$	Putative chorismate mutase, operonic with *glnA*, *ntrBC*	44% PFL0385

[a]Gene products specific to *P. entomophila* and not found in other *Pseudomonas* species (constraint of 60% identity over more than 80% of the protein length).
[b]Unusual GC content (differing by more than 1 s.d. from the average GC) likely due to recent lateral transfer.
[c]Gene products or predicted domains associated with virulence in other systems.
[d]Sequence identity between the protein encoded by *P. entomophila* and the best BLAST hit among proteins from other *Pseudomonas*. PP, *P. putida* KT2440; PA, *P. aeruginosa* PA01; PSPTO, *P. syringae* pv. tomato DC3000; PFL, *P. fluorescens* Pf-5; Pfo *P. fluorescens* Pf0-1 and Pf, *P. fluorescens*.
[e]PSEEN0141 and PP0168 are aligned only on 67% of PP0168 length.
[f]<40% identity.
[g]TrEMBL accession number.
[h]This cluster and similarity with that of *P. fluorescens* Pf-5 are discussed in the **Supplementary Figure 5**.
[i]Superscripted numerals indicate the number of independent *Tn*5 insertions.
[j]Conserved hypothetical protein.

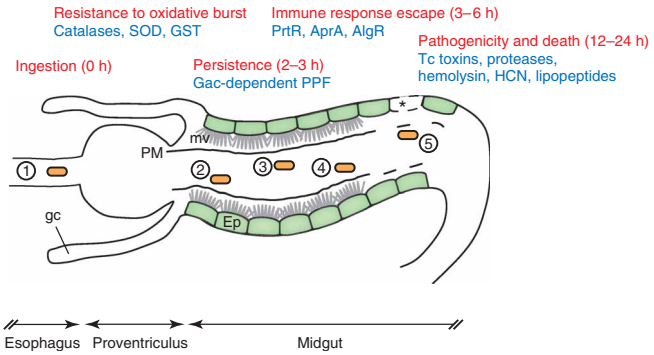

Figure 3 Steps in the interaction between *P. entomophila* and *D. melanogaster*. Five different steps are shown: 1. ingestion of *P. entomophila* through the esophagus; 2. resistance to oxidative stress in response to a oxidative burst in the gut; 3. persistence of *P. entomophila* in the gut; 4. escape from immune response effectors; 5. pathogenicity and lethal outcome of the interaction after important modifications of the midgut physiology including microvilli disruption, cell destruction (indicated by a *) and in some cases peritrophic matrix disorganization (indicated by a dashed line). Red indicates important steps in the infection process. Blue indicates newly identified proteins that could be involved at these steps in the process. Time scale is indicated in brackets. Ep, epithelial cell; mv, microvilli; PM, peritrophic matrix; gc, gastric cecum.

for the synthesis of a siderophore related to acinetobactin and containing a salicylamide moiety[21] (**Supplementary Fig. 5**).

Five gene clusters that direct the production of secondary metabolites have been identified (see **Supplementary Fig. 5**). PSEEN5520-PSEEN5522 are responsible for hydrogen cyanide production that is involved in *Caenorhabditis elegans* killing by *P. aeruginosa*[22] and in the suppression of soil-borne plant pathogens by certain *Pseudomonas* species[23]. The genome of *P. entomophila* contains four clusters of genes predicted to encode three different lipopeptides and a polyketide (**Table 2** and **Supplementary Fig. 5**).

Regulation of virulence revealed by a genome-wide mutagenesis

To directly identify factors that modulate the interaction between *P. entomophila* and *D. melanogaster*, we generated a Tn*5*-derived library of variants that were individually screened for their infectious and pathogenic properties. Among the 7,500 clones, we isolated 23 mutants whose growth was not affected and that displayed attenuated infectious and/or pathogenic properties (**Table 2**). Identification of the mini-Tn*5* insertion sites identified directly only a putative lipopeptide as a virulence factor. No other genes predicted to be virulence factors were identified, indicating a likely redundancy. By contrast, a number of insertions affected regulators that likely modulate the expression of such virulence factors. Seven independent insertions inactivated the two-component system GacS/GacA involved in the regulation of various processes, including virulence in different species, and resulted in the inability of these mutants to induce an immune response. *P. entomophila gac* mutants are defective in secretion of protease and hemolysin (data not shown) and do not persist in the gut of *D. melanogaster*[1], indicating the pivotal role of GacS/GacA in modulating the entomopathogenic properties of that strain. As observed in other *Pseudomonas* species[23], the GacS/GacA two-component system probably regulates *P. entomophila* virulence genes at a post-transcriptional level via the two identified small noncoding RsmY and RsmZ RNAs that alleviate post-transcriptional repression by RsmA and RsmE homologs. Three independent insertions in the *prtR* gene reduce the pathogenic properties of *P. entomophila* but retain the capacity to induce an immune response. In *P. fluorescens* LS107d2 (ref. 24), PrtR and PrtI regulate the transcription of the *aprA-inh-aprDEF* operon suggesting that *P. entomophila* relies on AprA protease to fully express its pathogenic properties in *D. melanogaster*. Two independent insertions that had the same consequences for the interaction with *D. melanogaster* have been found in *algR*. In *P. aeruginosa*, AlgR regulates a number of processes including fimbrial biogenesis, biofilm formation and cyanide production[25,26]. Altogether, genetic analysis indicates that GacA is a master regulator of the interaction and that PrtR and AlgR regulators, seem to play secondary roles in the infection process.

Metabolism, transport and regulation

The *P. entomophila* genome encodes most of the central metabolic pathways found in the other *Pseudomonas* including the pentose phosphate pathway, the Entner-Doudoroff pathway and the tricarboxylic acid cycle. Consistent with *Pseudomonas* metabolism, *P. entomophila* has an incomplete Embden-Meyerhof-Parnas pathway owing to the absence of 6-phosphofructokinase, and relies on a complete Entner-Doudoroff route for hexose utilization. The *P. entomophila* genome harbors several genes that encode hydrolytic activities such as chitinases, lipases and proteases as well as a set of 19 uncharacterized hydrolases, which are potentially involved in the degradation of polymers found in the soil. However, contrary to phytopathogenic strains such as *P. syringae*[5–7], the genome of *P. entomophila* is devoid of genes encoding enzymes capable of degrading plant cell walls. This is consistent with the observation that this species is not pathogenic for plants (M. Arlat, Institut National de la Recherche Agronomique, Castanet, France, personal communication).

The *P. entomophila* genome also contains determinants for the catabolism of various aromatic compounds (see **Supplementary Fig. 6** online) and long-chain carbohydrates. *P. entomophila* shares several gene clusters with *P. putida*[27] that are involved in the degradation of various classes of aromatic compounds including benzoate and quinate, 4-hydroxybenzoate, phenylacetaldehyde and phenylalkanoate as well as phenylalanine and tyrosine. The *P. entomophila* genome contains two additional catabolic gene clusters present in the genome of *P. aeruginosa* PAO1 that encode determinants for the degradation of 3-hydroxybenzoate through gentisate[28] and for the meta-cleavage of homoprotocatechuate[29,30].

Consistent with the size of its genome, *P. entomophila* possesses more than 535 transporter-encoding genes. Remarkably, no genes encoding a type III or type IV secretion system, present in numerous Gram-negative bacterial pathogens[31], were found in *P. entomophila*. The high numbers of transcriptional regulators (more than 300) and genes whose products are involved in signal transduction suggests that *P. entomophila* is able to adapt to substantial substrate variations in its habitats.

The soil and entomopathogenic lifestyle of *P. entomophila*

The metabolic properties of *P. entomophila* predicted from its genome suggest that this strain is a ubiquitous, metabolically versatile bacterium that may colonize diverse habitats including soil, rhizosphere and aquatic systems as shown for *P. putida* KT2440. However, in contrast to *P. putida*, *P. entomophila* contains a number of genes that are predicted, or have been shown, to be important for virulence. The expression of these factors is under the control of the major regulator GacA and presumably allows this strain to exploit new niches and interact with various insects, particularly *D. melanogaster* (**Fig. 3**).

In *D. melanogaster*, an environment hostile for microbial colonization is maintained in the gut by secretion of antimicrobial factors such as lysozymes[32,33] and other digestive enzymes. Recently, it has been shown that a unique epithelial oxidative burst limits microbial proliferation in the gut[34]; resistance to oxidative stress might therefore be a prerequisite for *D. melanogaster* gut colonization. The *P. entomophila* genome encodes 40 proteins that are predicted to be involved in resistance to oxidative stress including four catalases, two superoxide dismutases, three hydroperoxide reductases and eleven glutathione-*S*-transferases. It is noteworthy that resistance to oxidative stress is probably not sufficient for colonization as other *Pseudomonas* species that possess a large repertoire of oxidant detoxifying proteins are not able to persist in the gut of *D. melanogaster*[1]. This assumption is further reinforced by the observation that *P. entomophila gacA* mutants were not less resistant to peroxide, hypochlorite or paraquat (data not shown). As the *P. entomophila-D. melanogaster* interaction is specific, *P. entomophila* infectivity likely involves the expression of a specific gene enabling this strain to persist in the *D. melanogaster* gut, as shown for the *Erwinia carotovora* Evf factor[35]. Because *P. entomophila* does not contain any *evf*-related genes, we cannot predict candidates for this putative persistence promoting factor (*ppf* in **Fig. 3**). Nonetheless, this gene is likely regulated by the GacS/GacA two-component system: the *gacA*::*Tn5* or *gacS*::*Tn5* mutants do not persist in the gut and *P. entomophila* cells are infectious only at stationary phase, concomitant with Gac activation of virulence genes (data not shown). It is striking to note that in both *P. entomophila* and *E. carotovora*[35], genes required to interact with *D. melanogaster* are under the control of global regulators, that is, Hor and GacA, respectively, revealing the branching of virulence genes in a complex regulatory network.

Infection of *D. melanogaster* by *P. entomophila* is accompanied by blockage of food-uptake[1]. This phenomenon is also observed in the interaction between *Serratia entomophila* and the grass grub *Costelytra zealandica* or between *Yersinia pestis* and the flea. The processes used to effect food blockage seem to be different in the two systems; *Y. pestis* relies on phospholipase synthesis and biofilm formation[36,37] whereas the mechanism used by *S. entomophila* remains unknown[38]. Genes responsible for the anti-feeding determinants of *S. entomophila* have a prophage origin and no related genes were identified in the genome of *P. entomophila*. Since *algR* mutants still provoke food-uptake blockage, biofilm formation is probably not essential for *D. melanogaster* infection by *P. entomophila*.

The persistence of *P. entomophila* in the larval gut triggers both a local and systemic immune response[1]. The *P. entomophila* level remains high in wild-type larvae, similar to that observed in a *relish* mutant unable to induce an immune response[1], suggesting that this strain is able to escape the *D. melanogaster* immune response. Biofilm formation might protect *P. entomophila* cells from immune effectors or persistence of bacteria might result from the degradation of effectors. The defects observed with *prtR* mutants indicated that AprA may degrade antimicrobial peptides, as indicated by recent *in vivo* studies[39], and consequently disable the immune response.

Twelve hours after *D. melanogaster* ingests the bacteria, physiological modifications to the fly caused by *P. entomophila* are dramatic and the expression of 205 *D. melanogaster* genes is modified[1] (**Fig. 3**). These changes probably result from the action of virulence factors such as proteases, hemolysins, insecticidal toxin-like proteins, secondary metabolites or hydrogen cyanide. However, lethality starts to be apparent after 16 h, indicating that this late gene expression will have no effect on the fatal outcome of the interaction.

DISCUSSION

The complete genome sequence of *P. entomophila* provides insight into this organism's entomopathogenic lifestyle. Combined with a genetic approach, it has revealed potential virulence factors along with regulators that modulate their expression. This study also revealed that *P. entomophila* is the first *Pseudomonas* strain to be pathogenic in a multicellular organism and at the same time to be devoid of a type III secretion system. Its potential to use various plant-derived compounds including aromatic molecules, and its antibiotic- and oxidative stress-resistance capacities suggest that *P. entomophila* is a commensal bacterium. As this strain is not a plant pathogen, it may have potential to control insects. Unexpectedly for an environmental isolate, *P. entomophila* has a genome that contains a limited number of bacteriophages and transposons. This may contribute to its relatively small size compared to other *Pseudomonas* genomes. Finally, the complete genome sequence of *P. entomophila* provides a framework for further studies to characterize its pathogenic properties and for a host-pathogen system in which both organisms are amenable to genetic and genomic analysis.

METHODS

Genome sequencing, assembly and annotation. The complete genome sequence of *P. entomophila* L48 was determined using the whole-genome shotgun method (10× coverage, using two plasmid libraries and one BAC library to order contigs). Finishing was performed by PCR amplification from contigs extremities. After a first round of annotation, regions of lower quality as well as regions with putative frameshifts were resequenced from PCR amplification of the dubious regions. Using the AMIGene software (annotation of microbial genes)[40], a total of 5,279 CoDing Sequences were predicted (and assigned a unique identifier prefixed with "PSEEN"), and submitted to automatic functional annotation: exhaustive BLAST searches against the UniProt databank were performed to determine significant homology. Protein motifs and domains were documented using the InterPro databank. In parallel, genes coding for enzymes were classified using the PRIAM software[41]. TMHMM vs2.0 was used to identify transmembrane domains[42], and SignalP 3.0 was used to predict signal peptide regions[43]. Finally, tRNAs were identified using tRNAscan-SE[44]. Sequence data for comparative analyses were obtained from the NCBI databank (RefSeq section). Putative orthologs and synteny groups (that is, conservation of the chromosomal colocalization between pairs of orthologous genes from different genomes) were computed between *P. entomophila* and all the other complete genomes as previously described[45]. Manual validation of the automatic annotation was performed using the MaGe (Magnifying Genomes) interface, which allows graphic visualization of the *P. entomophila* annotations enhanced by a synchronized representation of synteny groups in other genomes chosen for comparisons[45]. All the data (that is, syntactic and functional annotations, and results of comparative analysis) were stored in a relational database, called EntomoScope. This database is publicly available via the MaGe interface at http://www.genoscope.cns.fr/agc/mage/.

Bacterial mutagenesis and screening. Random mutagenesis was performed by biparental mating using *P. entomophila*[1] and *Escherichia coli* S17.1-λ*pir*[46] carrying the pUT-*Tn*5-*Tc* suicide plasmid as previously described[47]. A total of 7,500 TcR colonies obtained from several independent conjugations were screened individually as previously described[35]. Transconjugants that displayed attenuated virulence were subjected to several secondary screenings by natural infection as previously described[1]. Insertion sites were determined using two different methods. First, genomic DNA was digested by *Pst*I or *Not*I/*Pst*I and ligated into pUC18 and pBlueScript, respectively. Clones that contained the mini-transposon and its flanking sequences were selected by plating the *E. coli* BW25142 transformants on tetracycline (10 μg/ml). One flanking region was sequenced from the *Tc* gene using the oligonucleotide (Tc-F) 5′-TCGTCGACA AGCTTCGG-3′. Some insertion sites were determined by reverse PCR method. Genomic DNA was digested by either *Pst*I or *Eag*I, self-ligated and amplified using the oligonucleotides Tc-F and 5′-AGATCTGATCAAGAGACAT-3′

for *Pst*I-digested DNA or 5′-GGCGGCCCTATACCTTGTCTG-3′ (Tet-end) and 5′-CATAATGGGGAAGGCCAT-3′ for *Eag*I-digested DNA, respectively. One flanking region was sequenced using the oligonucleotides Tc-F or Tet-end. Insertion sites were confirmed by amplifying the region overlapping the insertion site. Southern blot analysis was carried out to verify that the selected clones only carried a single copy of the transposon.

Accession numbers. The *P. entomophila* nucleotide sequence and annotation data have been deposited in the EMBL databank under accession number CT573326.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was supported by CNRS (Programme Séquençage à grande échelle), by IFR115 and by MRT/ACI IMPBio 2004 'MicroScope.' We thank Célia Floquet and Camille Jourlain for technical assistance, Matthieu Arlat for plant assays and helpful discussions, Alexandra Gruss, Linda Sperling and Sean Kennedy for critical reading of the manuscript, Olivier Espéli for expert annotation. N.V. was supported by a doctoral fellowship from the Association Vaincre la Mucoviscidose and the Association pour la Recherche sur le Cancer.

AUTHOR CONTRIBUTIONS

N.V., V.B., P.W., B.S., J.W., B.L., C.M. and F.B. designed research; N.V., D.V., S.C., Z.R., V.B., C.A., L.C., C.J., A.L., B.V. and F.B. performed research; N.V., D.V., S.C., V.B., C.A., C.M. and F.B. contributed new reagents/analytic tools; N.V., D.V., S.C., Z.R., V.B., C.A., L.C., C.J., A.L., B.V., B.L., C.M. and F.B. analyzed data; and N.V. and F.B. wrote the paper.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Vodovar, N. *et al. Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species. *Proc. Natl. Acad. Sci. USA* **102**, 11414–11419 (2005).
2. Waterfield, N.R., Wren, B.W. & ffrench-Constant, R.H. Invertebrates as a source of emerging human pathogens. *Nat. Rev. Microbiol.* **2**, 833–841 (2004).
3. Chattopadhyay, A., Bhatnagar, N.B. & Bhatnagar, R. Bacterial insecticidal toxins. *Crit. Rev. Microbiol.* **30**, 33–54 (2004).
4. Stover, C.K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
5. Buell, C.R. *et al.* The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. USA* **100**, 10181–10186 (2003).
6. Feil, H. *et al.* Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *Proc. Natl. Acad. Sci. USA* **102**, 11064–11069 (2005).
7. Joardar, V. *et al.* Whole-genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.* **187**, 6488–6498 (2005).
8. Paulsen, I.T. *et al.* Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* **23**, 873–878 (2005).
9. Nelson, K.E. *et al.* Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799–808 (2002).
10. Weinel, C., Nelson, K.E. & Tummler, B. Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ. Microbiol.* **4**, 809–818 (2002).
11. Eisen, J.A., Heidelberg, J.F., White, O. & Salzberg, S.L. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* **1**, RESEARCH0011 (2000).
12. Aranda-Olmedo, I., Tobes, R., Manzanera, M., Ramos, J.L. & Marques, S. Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*. *Nucleic Acids Res.* **30**, 1826–1833 (2002).
13. Waterfield, N.R., Bowen, D.J., Fetherston, J.D., Perry, R.D. & ffrench-Constant, R.H. The tc genes of *Photorhabdus*: a growing family. *Trends Microbiol* **9**, 185–191 (2001).
14. Bowen, D. *et al.* Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* **280**, 2129–2132 (1998).
15. Joo Lee, P. *et al.* Cloning and heterologous expression of a novel insecticidal gene (tccC1) from *Xenorhabdus nematophilus* strain. *Biochem. Biophys. Res. Commun.* **319**, 1110–1116 (2004).
16. Waterfield, N., Hares, M., Yang, G., Dowling, A. & ffrench-Constant, R. Potentiation and cellular phenotypes of the insecticidal Toxin complexes of *Photorhabdus* bacteria. *Cell Microbiol.* **7**, 373–382 (2005).
17. Wilson, M., McNab, R. & Henderson, B.. *Bacterial Disease Mechanisms* (Cambridge University Press, Cambridge, UK, 2002).
18. Miyoshi, S. & Shinoda, S. Microbial metalloproteases and pathogenesis. *Microbes Infect.* **2**, 91–98 (2000).
19. Meyer, J.M. Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. *Arch. Microbiol.* **174**, 135–142 (2000).
20. Ravel, J. & Cornelis, P. Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol.* **11**, 195–200 (2003).
21. Mercado-Blanco, J. *et al.* Analysis of the *pmsCEAB* gene cluster involved in biosynthesis of salicylic acid and the siderophore pseudomonine in the biocontrol strain *Pseudomonas fluorescens* WCS374. *J. Bacteriol.* **183**, 1909–1920 (2001).
22. Gallagher, L.A. & Manoil, C. *Pseudomonas aeruginosa* PAO1 kills *Caenorhabditis elegans* by cyanide poisoning. *J. Bacteriol.* **183**, 6207–6214 (2001).
23. Haas, D. & Defago, G. Biological control of soil-borne pathogens by fluorescent pseudomonads. *Nat. Rev. Microbiol.* **3**, 307–319 (2005).
24. Burger, M., Woods, R.G., McCarthy, C. & Beacham, I.R. Temperature regulation of protease in *Pseudomonas fluorescens* LS107d2 by an ECF sigma factor and a transmembrane activator. *Microbiology* **146**, 3149–3155 (2000).
25. Lizewski, S.E. *et al.* Identification of AlgR-regulated genes in *Pseudomonas aeruginosa* by use of microarray analysis. *J. Bacteriol.* **186**, 5672–5684 (2004).
26. Whitchurch, C.B. *et al.* Phosphorylation of the *Pseudomonas aeruginosa* response regulator AlgR is essential for type IV fimbria-mediated twitching motility. *J. Bacteriol.* **184**, 4544–4554 (2002).
27. Jimenez, J.I., Minambres, B., Garcia, J.L. & Diaz, E. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 824–841 (2002).
28. Liu, D.Q., Liu, H., Gao, X.L., Leak, D.J. & Zhou, N.Y. Arg169 is essential for catalytic activity of 3-hydroxybenzoate 6-hydroxylase from *Klebsiella pneumoniae* M5a1. *Microbiol. Res.* **160**, 53–59 (2005).
29. Prieto, M.A., Diaz, E. & Garcia, J.L. Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: engineering a mobile aromatic degradative cluster. *J. Bacteriol.* **178**, 111–120 (1996).
30. Thotsaporn, K., Sucharitakul, J., Wongratana, J., Suadee, C. & Chaiyen, P. Cloning and expression of p-hydroxyphenylacetate 3-hydroxylase from *Acinetobacter baumannii*: evidence of the divergence of enzymes in the class of two-protein component aromatic hydroxylases. *Biochim. Biophys. Acta* **1680**, 60–66 (2004).
31. Hueck, C.J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
32. Hultmark, D. Insect lysozymes. *EXS* **75**, 87–102 (1996).
33. Regel, R., Matioli, S.R. & Terra, W.R. Molecular adaptation of *Drosophila melanogaster* lysozymes to a digestive function. *Insect Biochem. Mol. Biol.* **28**, 309–319 (1998).
34. Ha, E.M. *et al.* An antioxidant system required for host protection against gut infection in *Drosophila*. *Dev. Cell* **8**, 125–132 (2005).
35. Basset, A., Tzou, P., Lemaitre, B. & Boccard, F. A single gene that promotes interaction of a phytopathogenic bacterium with its insect vector, *Drosophila melanogaster*. *EMBO Rep.* **4**, 205–209 (2003).
36. Hinnebusch, B.J. *et al.* Role of Yersinia murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector. *Science* **296**, 733–735 (2002).
37. Darby, C., Ananth, S.L., Tan, L. & Hinnebusch, B.J. Identification of *gmhA*, a *Yersinia pestis* gene required for flea blockage, by using a *Caenorhabditis elegans* biofilm system. *Infect. Immun.* **73**, 7236–7242 (2005).
38. Hurst, M.R., Glare, T.R. & Jackson, T.A. Cloning *Serratia entomophila* antifeeding genes–a putative defective prophage active against the grass grub *Costelytra zealandica*. *J. Bacteriol.* **186**, 5116–5128 (2004).
39. Liehl, P., Blight, M., Vodovar, N., Boccard, F. & Lemaitre, B. Prevalence of local immune response against oral infection in a *Drosophila/Pseudomonas* infection model. *PLoS Pathog.*, in the press.
40. Bocs, S., Cruveiller, S., Vallenet, D., Nuel, G. & Medigue, C. AMIGene: Annotation of MIcrobial Genes. *Nucleic Acids Res.* **31**, 3723–3726 (2003).
41. Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* **31**, 6633–6639 (2003).
42. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
43. Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783–795 (2004).
44. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
45. Vallenet, D. *et al.* MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* **34**, 53–65 (2006).
46. Miller, V.L. & Mekalanos, J.J. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J. Bacteriol.* **170**, 2575–2583 (1988).
47. de Lorenzo, V., Herrero, M., Jakubzik, U. & Timmis, K.N. Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J. Bacteriol.* **172**, 6568–6572 (1990).