Insertion sequences drive the emergence of a highly adapted human pathogen

Erwin Sentausa[1#], Pauline Basso[1+], Alice Berry[1], Annie Adrait[2], Gwendoline Bellement[1,2∞], Yohann Couté[2], Stephen Lory[3], Sylvie Elsen[1*] and Ina Attrée[1*]

[1]Université Grenoble Alpes, CNRS ERL5261, INSERM U1036, CEA, Laboratory Biology of Cancer and Infection, Bacterial Pathogenesis and Cellular Responses, Biosciences and Biotechnology Institute of Grenoble, 38000 Grenoble, France

[2]Université Grenoble Alpes, CEA, Inserm, BIG-BGE, 38000 Grenoble, France

[3]Department of Microbiology and Immunobiology, Harvard Medical School, USA

[#]Current address: Evotec ID (Lyon) SAS, Marcy l'Étoile, France

+Current address: Department of Microbiology and Immunology, University of California San Francisco (UCSF) School of Medicine, USA

∞Current address: Biozentrum, University of Basel, Basel, Switzerland

*Corresponding authors

Competing Interests: The authors declare no competing financial interests

KEYWORDS: *Pseudomonas aeruginosa*, adaptation, antibiotic resistance, bacterial infection, multiomics

Abstract

Taxonomic outliers of *Pseudomonas aeruginosa* of environmental origin have recently emerged as infectious for humans. Here we present the first genome-wide analysis of an isolate that caused fatal hemorrhagic pneumonia. We demonstrate that, in two sequential clones, CLJ1 and CLJ3, recovered from a patient with chronic pulmonary disease, insertion of a mobile genetic element into the *P. aeruginosa* chromosome affected major virulence-associated phenotypes and led to increased resistance to antibiotics used to treat the patient. Comparative proteome and transcriptome analyses revealed that this insertion sequence, ISL3, disrupted genes encoding flagellar components, type IV pili, O-specific antigens, translesion polymerase and enzymes producing hydrogen cyanide. CLJ3 possessed seven fold more IS insertions than CLJ1, some modifying its susceptibility to antibiotics by disrupting the genes for the outer-membrane porin OprD and the regulator of β-lactamase expression AmpD. In the *Galleria mellonella* larvae model, the two strains displayed different levels of virulence, with CLJ1 being highly pathogenic. This work reveals ISs as major players in enhancing the pathogenic potential of a *P. aeruginosa* taxonomic outlier by modulating both, the virulence and the resistance to antimicrobials, and explains the ability of this bacterium to adapt from the environment to a human host.

Introduction

Emerging infectious diseases caused by multidrug resistant bacteria represent a serious threat for human well-being and health. Several hundred novel pathologies caused by infectious agents have been reported during the last forty years (1). Environmental bacteria can adapt to a human host by acquisition of virulence determinants through chromosomal rearrangements due to mobile genetic elements, horizontal gene transfer, or small local sequence changes including single nucleotide substitutions, or horizontal gene transfer (2, 3).

The *Pseudomonas* genus is one of the most important groups of bacteria thriving in diverse environments capable of causing plant or animal diseases. Species such as *P. aeruginosa, Pseudomonas fluorescens*, and *Pseudomonas syringae* adversely impact human health and agriculture (4-6). *P. aeruginosa* is a particularly successful opportunistic pathogen found frequently in humid environments associated with human activities. In hospital settings, infections caused by multi-resistant *P. aeruginosa* strains present a real danger for elderly individuals, patients undergoing immunosuppressive therapies and those requiring treatment with invasive devices in Intensive Care Units. In addition to acute infections, *P. aeruginosa* is a common cause of chronic wound infections as well as long lasting respiratory infections of patients with cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD). During chronic infections the bacteria adapt to the particular host environment by changing metabolic pathways and synthesis of virulence-associated components (7). Additionally, in-patient evolution involves acquisition of loss-of-function mutations in genes of motility, antibiotic resistance, acute virulence and envelope biogenesis (8-12).

Recent massive whole genome sequencing allowed classification of clinical and environmental *P. aeruginosa* strains in three clades (13, 14). The pathogenic strategies of the three clades rely on different toxins. The two most populated clades inject the toxins, also referred to as effectors ExoS, ExoT, ExoY, and ExoU, directly into host cell cytoplasm through a complex molecular syringe using type III secretion system (T3SS) (15). The third clade is occupied by taxonomic outliers that lack all the genes encoding the effectors and the components of the T3SS. The first fully sequenced taxonomic outlier, PA7, was multi-drug resistant and non-virulent in an acute lung infection mouse model (16, 17). Other PA7-related strains were mainly of environmental origin (18, 19) or associated with both, acute (wounds and urinary tract) and chronic (CF and COPD) human infections (20-22). They recently emerged as highly virulent for humans potentially through the secretion of a pore-forming toxin Exolysin, ExlA (16, 21-23). The most pathogenic *exlA*$^{+}$ *P. aeruginosa* strain described up to date is the strain CLJ1 isolated at the University Hospital in Grenoble, France, from a COPD patient suffering from hemorrhagic pneumonia (16). In murine acute lung infection model, CLJ1-infected lungs featured an extensive damage to endothelial monolayers, bacteria transmigrated into the blood and disseminated into secondary organs without being detected by the immune system. This differs greatly from consequences observed by the T3SS^{+} strain PAO1 (16, 24).

To get insights into molecular determinants of pathogenesis expressed by the Exolysin-producing *P. aeruginosa* taxonomic outliers and to assess the extent of evolutionary adaptation during the course of infection, we performed a comprehensive comparative genome-wide study of two clonal variants, CLJ1 and CLJ3, isolated from the same patient

at different time-points during hospitalization. The gathered data demonstrated that mobile genetic elements belonging to the ISL3 family insertion sequences (IS), originally found in soil bacteria *Pseudomonas stutzeri* and *Pseudomonas putida*, shape the virulence traits and strategies employed by those strains to colonize and to adapt to the human host.

Materials and methods

Bacterial strains and culture conditions

P. aeruginosa strains used in this study are CLJ1 and CLJ3 (16). Bacteria were grown at 37°C in liquid Lysogeny Broth (LB) medium (10 g/L Bacto tryptone, 5 g/L yeast extract, 10 g/L NaCl) with agitation until the cultures reached an optical density at 600 nm (OD_{600}) of 1.0 unless indicated. For the assessment of c-di-GMP levels, pUCP22-p*cdrA*-*gfp*(ASV)^c was introduced into CLJ1 and CLJ3 strains by electroporation, as previously described (25), and selected on LB agar plates containing 200 µg/mL of carbenicillin.

Genome analysis

The details on sequencing, assembly, annotation and comparison are described in Supplementary materials and methods. The assembly and annotation statistics for the CLJ1 and CLJ3 genomes are presented in Table S1. Functional annotation was performed on the Rapid Annotations based on Subsystem Technology (RAST) Genome Annotation Server version 2.0 (26) using Classic RAST annotation scheme and GLIMMER-3 gene caller. Manual curation was done based on the annotations of orthologous genes in PA7 and PAO1 strains from the Pseudomonas Genome Database (27). Circos 0.69-3 (28) was used to create multiomic data visualizations in Fig. 1 and

Fig. S1. The whole genome shotgun projects of CLJ1 and CLJ3 have been deposited at DDBJ/ENA/GenBank under the accessions PVXJ00000000 and PZJI00000000, respectively.

Identification of insertion sites for IS elements

CLJ-ISL3 insertion locations in the gaps between CLJ3 contigs were detected by checking for the inverted repeat sequences at the contigs' ends, while accounting for the shared gene synteny between CLJ1, PA7 and PAO1 genomes. We also used panISa version 0.1.0 (https://github.com/bvalot/panISa) with default parameters to search for ISs in CLJ3 reads, mapped using BWA-MEM algorithm from BWA version 0.7.15 (29) to PA7 and CLJ1 genomes, respectively.

Transcriptome

The RNA for RNA-Seq was prepared as described (30) from bacterial cultures grown in duplicates in LB to OD_{600} of 1. The preparations of the Illumina libraries and sequencing were done by standard procedures at the Biopolymer Facility, Harvard Medical School, Boston, USA. The analysis was done as described in Supplementary materials and methods.

Mass spectrometry-based quantitative proteomic analyses

Samples for proteomics were prepared and analysed by nanoliquid chromatography coupled to tandem mass spectrometry (Ultimate 3000 coupled to LTQ-Orbitrap Velos Pro, Thermo Scientific) as described previously (31) with slight modifications, as described in Supplementary materials and methods. Each fraction was controlled by western blot,

using appropriate antibodies. The protein content in total, membrane, and secretome proteomes of CLJ1 and CLJ3 were analyzed independently from the others. Statistical analyses were performed using ProStaR (32). In total, a list of 2 852 quantified proteins was obtained. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (33) partner repository with the dataset identifier PXD011105.

Phenotypic analyses

HCN production was assessed on induction plate containing arginine (HCN precursor) as previously described (22). To monitor the c-di-GMP levels, the fluorescence-based reporter plasmid pUCP22-p*CdrA-gfp*(ASV)^{c} was used (34). Bacteria carrying the plasmid were subcultured at OD_{600nm} of 0.05 in black 96-well plate with clear bottom, and incubated at 37°C and 60 rpm in the Fluoroskan reader. The fluorescent emission was measured every 15 min at 527 nm following an excitation at 485 nm for 6 h. Serum sensitivity was assessed by a protocol adapted from (35). Two different human sera from the Etablissement Français du Sang (EFS) were used in all experiments. Overnight cultures of CLJ1 and CLJ3 bacteria were pelleted at 3,000 rpm for 5 min and suspended in Hanks Balanced Salt Solution (HBSS, GIBCO) with 0.1% of gelatin and adjusted to 10^8 colony forming units (CFU) per mL. The bacteria (10^6) were incubated in presence of 10% of human serum in a final volume of 3 mL for 15 min or 30 min at 37°C under gentle agitation. A control with heat inactivated serum at 56°C for 30 min was done. CFU were determined at 0, 15, and 30 min by serial dilutions and spreading on LB plates.

159 **Infections of *Galleria mellonella* larvae**

160 The calibrated wax moth larvae *Galleria mellonella* were purchased from the French
161 company Sud-Est Appats (http://www.sudestappats.fr). Healthy, uniformly white larvae,
162 measuring around 3 cm were selected for infection. The bacteria were grown until the
163 OD_{600} of 1 and diluted in PBS to approximately 10^3 bacteria/mL. Insulin cartridges were
164 sterilized and filled with bacterial solutions. The larvae were injected with 10 µL of
165 bacterial suspensions using the insulin pen. The exact number of bacteria used in pricking
166 was obtained by spotting five times 10 µL with the pen on agar plates and by counting
167 colony-forming units after growth at 37°C. The infected animals were placed in petri
168 dishes and set at 37°C. The dead larvae were counted over indicated period. Twenty
169 larvae were used per condition and the experiment was performed twice.

170 **RT-qPCR**

171 To quantify selected transcripts, total RNA from 2.0 mL of cultures (OD_{600} of 1.0) was
172 extracted with the TRIzol Plus RNA Purification Kit (Invitrogen) then treated with DNase I
173 (Amplification Grade, Invitrogen). RT-qPCR was performed as described (36) with few
174 modifications as described in Supplementary materials and methods. The sequences of
175 primers were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-
176 bin/primer3plus/primer3plus.cgi/) and are given in Table S2.

177

178

179

180

Results

Analysis of the CLJ1 genome and its regions of genomic plasticity

Strain CLJ1, an antibiotic-sensitive *P. aeruginosa,* was isolated from the patient with necrotizing hemorrhagic pneumonia. Twelve days later, after initiating antibiotic therapy the patient condition worsened and at this time CLJ3, a multidrug resistant clonal variant, was isolated (16). We have also shown that CLJ1 is a cytotoxic strain that shares the main genomic features with the first fully sequenced antibiotic resistant taxonomic outlier PA7 (16, 17). Notably, CLJ1 lacks the entire locus encoding the T3SS machinery and the genes encoding all known T3SS effectors; however, it carries the determinant for the two-partner secretion pore-forming toxin, Exolysin. To initiate genome-wide studies on mechanisms conferring the specific phenotypes of CLJ1, we fully sequenced the genome of this strain and compared it to PA7. As expected, most of the core genes of CLJ1 are similar to PA7 (Fig. 1 and Fig. S1). However, the content and the distribution of several regions of genomic plasticity (RGPs) (37) are different. CLJ1 genome contains 15 regions that are absent from the PA7 genome and lacks 26 PA7-specific regions. Among differences between CLJ1 and PA7 genomes there is a putative integrated plasmid carrying several genes that confers aminoglycoside resistance, regions encoding two type I restriction-modification systems, a pyocin protein, and a mercury resistance system (Tables S3 and S4). All CLJ1 specific regions (CLJ-SR), except CLJ-SR14, are present in other *P. aeruginosa* strains, some being phylogenetically unrelated to the PA7 strain. Strikingly, CLJ-SR14 carries 55 genes, many predicted to encode proteins involved in metabolism and resistance to heavy metals (Table S5), suggesting environmental origin

of the strain. Modifications in three replacement islands (38, 39) can mediate the CLJ1-specific phenotype. The CLJ replacement island in RGP60 (harboring pilin and pilin modification genes) carries a group I pilin allele, unlike PA7 that has a group IV allele (40). CLJ1 encodes, within RGP9, a b-type flagellin as the principal component of its flagellum, while PA7 possesses a-type flagellin (17). Both, RGP9 containing the flagellin glycosylation genes and the replacement island in RGP31 bearing the O-specific antigen (OSA) biosynthesis gene cluster are further modified in CLJ strains (see below). While most of the genes within RGP7 (pKLC102-like island) are missing in CLJ1, the so-called Dit island previously found in a CF isolate of *P. aeruginosa* (37) is inserted in 5' region of RGP27 at the $tRNA^{Gly}$. The determinants encoded in this island provide the bacteria the ability to degrade aromatic diterpenes, which are tricyclic resin acids produced by wounded trees, and to use them as sole carbon and energy source (41). The Dit island is uncommon in the genomes of *P. aeruginosa* strains, but it is frequently present in different soil bacteria such as *Burkholderia xenovorans*, *P. fluorescens*, and *Pseudomonas mendocina*, further implying the environmental origin of the CLJ1 strain. Among potential virulence-associated genes, CLJ1 lacks one of the type 6 secretion system loci (*PSPA7_2884-2917*) encoding an injection machine for toxins active against both, prokaryotes and eukaryotic cells (42). The *plcH* and *plcR* genes, coding for the hemolytic phospholipase C precursor and its accessory protein respectively, are also absent from the CLJ1 genome. Whereas *cupA* fimbrial gene cluster (*CLJ1_2899-2903*; *PSPA7_3019-3023*) is present in the remodeled RGP23, the entire *cupE1-6* operon (*PSPA7_5297-5302*) encoding cell surface fimbriae involved in the maintenance of a biofilm structure (43) is missing.

Evidence of mobile genetic elements in the key pathogenic regions

Following infection, CLJ1 strain provokes striking damage to mouse lung tissues without an immunological response by the host (24). The sequential isolate CLJ3 was sampled from the same patient after a series of aggressive antibiotic treatments and displayed different cytotoxicity and antibiotic-sensitivity profiles compared to CLJ1 (16). The genomes of these two strains are almost identical, with a reciprocal best-hit average nucleotide identity (44) between them estimated to be 99.97% (Fig. 1). When analyzing the genomes of CLJ1 and CLJ3, the most striking feature is the presence of multiple copies of a 2,985-bp fragment corresponding to a mobile genetic element, absent from the reference PA7 genome. This sequence is 99% identical to ISL3 family of insertion sequences ISPst2 of *Pseudomonas stutzeri*, ISpu12 of *Pseudomonas putida* and IS1396 of *Serratia marcescens* found in the ISfinder database (45). However, besides coding for a transposase, this CLJ-ISL3 encodes a putative inner membrane protein with seven transmembrane helices from the permease superfamily cl17795 (with the conserved domain COG0701 (46)) and a putative transcriptional metalloregulator of the ArsR family (Fig. 2A). The element is flanked by a pair of 24-bp imperfect inverted repeat sequences GGGTATCCGGAATTTCTGGTTGAT (left inverted repeat, IRL) and GGGTATACGGATTTAATGGTTGAT (right inverted repeat, IRR). Examination of publicly available genome sequences showed that the complete CLJ-ISL3 is also present in a few other bacteria, including a multi-drug resistant *Acinetobacter baumannii* isolate recovered from bronchoalveolar lavage fluid (47).

Altogether, we found CLJ-ISL3 in six and forty copies in the genomes of CLJ1 and CLJ3, respectively. The presence of ISs in CLJ3 genome was predicted using a combination of

bioinformatics tools, analysis of genome syntheny and detection of inverted sequence (see Materials and Methods). In agreement with the clonal origin of the strains, all CLJ-ISL3 of CLJ1, but one, were located in the same location in CLJ3 genome (Fig. 1 and Table S6). Mass spectrometry (MS)-based quantitative proteomic analyses revealed a strong increase in the level of the transposase protein in CLJ3 compared to CLJ1, that might explain the IS expansion within its genome (Table S7). As mobile genetic elements greatly contribute to phenotypic modifications by changing the gene expression or by gene inactivation (49, 50), we examined their impact on global gene expression and particular phenotypes. To that aim, we compared CLJ1 and CLJ3 transcriptomes and proteomes using respectively RNA-Seq and MS-based quantitative proteomics. The proteomes of three different bacterial fractions (whole bacteria, total membranes, and secretomes) were analyzed. Stringent statistical analyses of extracted data revealed 77 differentially expressed genes/proteins between CLJ1 and CLJ3, both at levels of mRNA and protein (Fig. 1, Tables S7, S8 and S9). Out of these, 27 (35%) are of phage origin, while 32 (42%) are predicted to be localized in bacterial membranes, periplasm or secreted.

Contribution of CLJ-ISL3 to antibiotic resistance of the CLJ3 strain

The CLJ1 and CLJ3 strains are recent isolates obtained from the patient treated with high doses of different antibiotics without successfully eliminating the infection. CLJ1, isolated before the beginning of the antibiotic therapy, was sensitive to the tested antibiotics, while CLJ3 displayed resistance to most of the antibiotics administrated to the patient (16). To

gain insight into mechanisms of antibiotics resistance of CLJ3, we examined the genomic data for gene modifications that could explain the switch in phenotypes to selected antibiotics. We observed that several genes encoding proteins potentially conferring antibiotic resistance are modified by ISs.

As the patient was given different antibiotics of the β-lactam family (ticarcillin, carbapenem, cephalosporin…) and the CLJ3 isolate developed the resistance to all of them, we examined the status of the outer membrane porin OprD (CLJ1_4366) and the chromosomal-encoded AmpC β-lactamase (CLJ1_0728), the two major determinants of intrinsic resistance to this group of antibiotics. We found that *oprD* gene was interrupted by the CLJ-ISL3 (Fig. 2B) resulting in absence of the protein (Table S7). Loss-of-function mutations or deletions in *oprD* are commonly seen in isolates from patients undergoing treatment with imipenem or meropenem (51-54) making the cell envelope impermeable to these antibiotics. Additionally, the identical ISL3 element was found in the 5' portion of the *ampD* gene (Fig. 2C), encoding the recycling amidase responsible for the production of muropeptide regulators of *ampC* expression (55, 56). Consequently, although higher expression of *ampC* gene was not detected in RNA-Seq, the CLJ3 proteomes contained significantly higher amounts of the AmpC protein compared to those of CLJ1 (Table S7).

Two additional periplasmic proteins involved in peptidoglycan recycling and biosynthesis, the AmpDH3 amidase (CLJ1_5671) (57) and the lytic transglycosylase MltA-interacting protein MipA (CLJ1_3357) (58), were overrepresented in the CLJ3 proteome. In agreement, our datasets also pointed to the overexpression of corresponding genes, further confirmed by RT-qPCR (Fig. 2D). The overproduction of these proteins in a clinical strain suggests their role in adaptation of the strain to host through modulation of

peptidoglycan synthesis. However, the molecular mechanisms involved in their increased expression and the significance of their overexpression in bacterial persistence in the host need to be determined.

All *P. aeruginosa* strains carry genes for multiple efflux pumps. The CLJ1/PA7 clade possesses the locus encoding the MexXY-OprA efflux pump which is able to transport multiple antibiotics including fluoroquinolones, aminoglycosides, and certain cephalosporins (reviewed in (59)). However, when compared to CLJ1, there is a *ca*. 20kb deletion in the CLJ3 genome (with a loss of 22 genes corresponding to *PSPA7_3247-3268*) eliminating the entire transcriptional repressor *mexZ* gene and truncating the *mexX* (Fig. 2E). The deleted region is replaced by a copy of CLJ-ISL3 that is bordered by truncated *puuP* and *mexX*. One plausible explanation for the observed genomic arrangement of this region in CLJ3 is that the strain was derived from a yet un-identified clonal strain that had another copy of ISL3 in *mexX*. A recombination between these two IS elements resulted in the 20 kb deletion, simultaneously eliminating the repressor and a portion of *mexX* genes, making this efflux pump nonfunctional. The truncation of the *puuP* gene encoding putrescine importer/permease indicates that the CLJ-ISL3 sequence affects transport of polyamine, which have multiple roles in pathogen biology, including resistance to some antibacterial agents (60).

Modifications of the O-specific antigens (OSA) cluster due to ISs

OSA is a component of lipopolysaccharide (LPS) and an integral component of the *P. aeruginosa* cell envelope. The OSA biosynthetic gene cluster of CLJ in RGP31 is similar

316 in content to that of PA7, consequently both PA7 and CLJ1 belong to serotype O12 (22).
317 However, insertion of two different ISs alters the gene content and expression levels of
318 the encoded proteins (Fig. 3A). In CLJ1, the ISL3 element was found within the gene
319 encoding NAD-dependent epimerase/dehydratase (CLJ1_1762 corresponding to
320 PSPA7_1970), whereas in the corresponding CLJ3 locus two copies were present, one
321 disrupting *wbjL* and the second located in the intergenic region between *CLJ3_1919* and
322 *CLJ3_1920* (*PSPA7_1972*). Moreover, in CLJ3, two copies of another IS from the IS66
323 family were also found within the gene *wbjB (CLJ1_1771; CLJ3_1930)* and downstream
324 of *wbjM (CLJ1_1778; CLJ3_1936)*. These insertions negatively affect expression of these
325 genes, or mRNA stability, as their mRNA and protein levels were higher in CLJ1 than in
326 CLJ3 (Tables S7 and S8).
327 Furthermore, some other proteins encoded in the 3' part of the OSA cluster, i.e.
328 CLJ1_1767 (PSPA7_1975), WbjB (CLJ1_1771), WbjC (CLJ1_1772), WbjI (CLJ1_1773)
329 and WbjJ (CLJ1_1774) were more abundant in CLJ1 (Table S7). Analogous variations
330 of the OSA loci leading to variability of O-antigens have been reported (61, 62).
331 Interestingly, the previously described deletion at the *mexX-puuP* locus in CLJ3
332 encompasses also the *galU* gene (*CLJ1_3121*, Fig. 2E), encoding UDP glucose
333 pyrophosphorylase (63) that adds sugar moieties onto the inner core on lipid A, serving
334 as the anchor for the full length LPS. The LPS truncation due to absence of *galU* leads to
335 a so-called rough LPS, higher susceptibility to serum-mediated killing and reduced *in vivo*
336 virulence (64). The CLJ3 strain was non-agglutinable and sensitive to serum compared
337 to the CLJ1 strain (Fig. 3B). The *in vivo* selection for acquisition of a serum sensitivity
338 phenotype is unclear but it may reflect the adaptation of CLJ3 to the host environment

339 through resistance to phages and antibiotic challenges, as it has been reported for some
340 other clinical strains (65). Modification in LPS structures in *P. aeruginosa* strains
341 chronically infecting CF patients is considered as an adaptation mechanism to a more
342 "persistent" lifestyle, with the LPS molecule being less inflammatory (64). Thus, multiple
343 mechanisms contributed to final LPS structure in CLJ clones. Those events could have
344 provided to strains variable advantages during infection process such as resistance to
345 antimicrobials or altered recognition by the immune system.
346
347 **ISs determine the repertoire of surface appendices**
348 Inspection of CLJ1 and CLJ3 genome sequences showed that CLJ-ISL3 has strongly
349 affected the flagellar biosynthetic locus in both strains. The CLJ3 genome revealed an
350 organization and gene content of the flagellar locus similar to that found in PAO1, which
351 synthesizes a b-type flagellum (66), but with two copies of CLJ-ISL3 (Fig. 4A). The first
352 IS interrupts *flgL (CLJ1_4222)*, encoding the flagellar hook-associated protein that
353 enables the anchoring of the flagellum to the cell envelope (67), while the second IS
354 element is in *fgtA (CLJ1_4219)* coding for the flagellin glycosyl transferase. The two ISs
355 seem to have recombined in CLJ1 creating a deletion between truncated *fgtA* and *flgL*
356 genes (Fig. 4A). This recombination event suggests that CLJ1 is not the direct ancestor
357 of CLJ3, but that the two strains evolved from a common ancestor. The absence of
358 flagellum is in agreement with the non-motile phenotype of the strain observed during
359 eukaryotic cell infection and on soft agar (22). Bacterial flagella also modulate immune
360 response during the infection, because the major flagellar subunit flagellin binds to TLR5
361 and initiates the TLR-dependent signaling and activation of expression of pro-

362 inflammatory cytokines (68). The absence of assembled flagella in the isolates explains
363 the lack of detectable pro-inflammatory cytokine interleukin-1β (IL-1β) and TNF in
364 bronchoalveolar lavages of CLJ1-infected mice (23, 24). CLJ1 is also devoid of twitching
365 mobility (22); therefore we examined the genomic data for possible mutations in genes
366 encoding type IV pili (T4P) and found an insertion of CLJ-ISL3 within the 5' part of *pilM*
367 gene in both CLJ1 and CLJ3 genomes (Fig. 4B). This gene, encoding a cytoplasmic actin-
368 like protein, is the first gene of the *pilMNOPQ* operon reported to be essential for both
369 T4P biogenesis and twitching motility (69). Expression of the entire operon, likely due to
370 the polar effect of the IS element on the downstream genes, seems to be affected since
371 no proteins were detected by proteomic analysis, unlike many other Pil proteins encoded
372 in other operons (Table S7). This finding was intriguing as the action of the toxin Exolysin
373 relies greatly on T4P in another *exlA*+ strain, IHMA87 (70). Therefore, we examined CLJ1
374 proteomes for the presence of other putative adhesive molecules that may substitute for
375 the T4P function during host cell intoxication. Based on proteomic datasets, the CLJ1
376 strain synthesizes components of five two-partner secretion (TPS) systems whose
377 predicted secreted components are annotated as haemolysins/haemagglutinins,
378 including the Exolysin (CLJ1_4479) responsible for CLJ1 cytotoxicity. CdrA and CdrB
379 (CLJ1_4999 and CLJ1_5000) are significantly overrepresented in the CLJ3 strain as
380 revealed by proteomics and RNA-Seq analysis (Tables S7 and S8). In the PAO1 strain,
381 the adhesin CdrA is regulated by the secondary messenger cyclic-di-GMP (c-di-GMP),
382 and its expression is increased in biofilm-growing condition (71). We assessed the cellular
383 levels of c-di-GMP by using a reporter that is based on the transcriptionally fused c-di-
384 GMP-responsive *cdrA* promoter to a gene encoding unstable green fluorescent protein

(34). The p*cdrA-gfp* (ASV)^{C} monitoring plasmid clearly showed higher levels of the second messenger in CLJ3 during growth (Fig. 4C), suggesting that CLJ3 may have adapted to host conditions by switching to biofilm lifestyle. Another TpsA protein detected in CLJ secretomes is CLJ1_4560, a HMW-like adhesin recently named PdtA in the PAO1 strain. PdtA plays a role in *P. aeruginosa* virulence as demonstrated by using *Caenorhabditis elegans* model of infection (72). The production of five TPSs, including the protease LepA (CLJ1_4911) (73) and at least one contact-dependent inhibition protein Cdi (CLJ1_2745) (74, 75), in CLJ strains (Table S7) indicates that this family of proteins may play an important role during colonization and infection, but their respective contributions in adhesion, cytotoxicity, or inter-bacterial competition during infection process need to be investigated.

Other putative virulence factors, phages and metabolism

RNA-Seq results supported by proteomics data showed an increase in expression of the enzymes HcnB (CLJ1_2955) and HcnC (CLJ1_2954) in the CLJ1 strain (Table S9). The inspection of the *hcn* operon revealed that the CLJ-ISL3 element was inserted in the 5' part of the *hcnB* gene in CLJ3 genome (Fig. 5A), in agreement with absence of expression measured using RT-qPCR (Fig. 5B). The *hcn* genes encode the subunits of hydrogen cyanide (HCN) synthase that produce this toxic secondary metabolite (76). To detect the HCN produced by the strains, a paper impregnated with a reaction mixture containing Cu^{2+} ions was placed above bacteria seeded onto agar plates, a white-to-blue color transition being indicative of HCN in the gas phase. In agreement with proteomic data,

CLJ1 was able to produce HCN in higher quantities than CLJ3 and other *P. aeruginosa* strains from the *exlA*$^+$ collection (Fig. 5B, (22)). Many *P. aeruginosa* isolates from individuals with CF produce high levels of HCN (77) and the molecule has been detected in the sputum of *P. aeruginosa*-infected CF and bronchiectasis patients (78, 79). HCN is also a regulatory molecule, capable of inducing and repressing other genes (80), including a cluster of genes *PA4129-PA4134* in PAO1. This cluster (*CLJ1_0701-CLJ1_0708*) is expressed at higher levels in CLJ1 compared to CLJ3 as revealed by RNA-Seq and quantitative proteomics (Fig. 1, Tables S7 and S8) and confirmed for *ccoG2* and *ccoN4* genes by RT-qPCR (Fig. 5C). The *ccoN4* gene belongs to the *ccoN4Q4* operon, one of the two *ccoNQ* orphan gene clusters present in *P. aeruginosa* genome. Its upregulation in the cyanogenic CLJ1 is in agreement with the study showing that, although *P. aeruginosa* encodes a cyanide-insensitive oxidase CIO, isoforms of cbb_3-type cytochrome *c* oxidase containing CcoN4 subunit were produced in response to cyanide, exhibiting higher tolerance towards this poisonous molecule in low-oxygen conditions (81).

Examining the data for differential expression of putative virulence determinants indicated that CLJ3 overproduces Lipotoxin F (LptF, CLJ1_1186), an outer-membrane protein contributing to adhesion to epithelial A549 cells and known to activate host inflammatory response (82). The *lptF* gene is located in a putative operon together with a gene encoding a hypothetical protein predicted to be a lipoprotein located in the periplasm (CLJ1_1187) that was found upregulated in CLJ3 strain by proteomic analyses (Table S7). As *lptF* upregulation was not detected by initial RNA-Seq, we performed RT-qPCRs that demonstrated significant overexpression of both genes *lptF* and

430 *CLJ1_1187*/*CLJ3_1413* in CLJ3 (Fig. 5C) suggesting that Lipotoxin F together with
431 CLJ1_1187 contributed to bacterial adaptation to in-patient environments which is in
432 agreement with increased *lptF* expression in CF isolates (83).
433 Finally, prophages play important roles in *P. aeruginosa* physiology, adaptation and
434 virulence (38, 84, 85). In the CLJ genomes, there are at least eight regions related to
435 phages (Table S10) and five of them are different or absent from PA7 (Table S3).
436 Transcriptomic and proteomic approaches revealed that twenty-seven phage-related
437 proteins from *CLJ1_0539-0556* in RGP3 and *CLJ1_4296-4314*, including CLJ-SR11, are
438 significantly more abundant in CLJ3 compared to CLJ1 (Tables S7, S8 and S9).
439 Interestingly, the genes encoding bacteriocins, namely pyocins S2, S4, and S5, are
440 absent from CLJ strains. Nevertheless, the activator of pyocin biosynthetic genes, PrtN
441 (CLJ1_0535) (86), is more highly expressed in CLJ3 than in CLJ1. Indeed, the *prtN* gene
442 is located in a region encoding phage-related proteins which is activated in CLJ3 (Fig. 1,
443 Table S9). In both CLJ genomes, we also identified the genes for AlpR (CLJ1_4295,
444 CLJ3_4269) and AlpA (CLJ1_4296, CLJ3_4268), transcriptional regulators of a
445 programmed cell death pathway in PAO1 (87). Finally, *alpA* and all the genes of the
446 *alpBCDE* lysis cassette are highly expressed in CLJ3, and we noticed high propensity of
447 this strain for lysis (Fig. 1, Table S9).

448

449 **CLJ-ISL3 inactivates the *imu* operon encoding translesion synthesis machinery**

450 Another location of the ISL3 element in the CLJ1 and CLJ3 genomes is within the *imu*
451 operon also named the “mutagenesis” cassette (Fig. 5D). The *imu* operon encodes the

ImuC polymerase (formerly called DnaE2) and other components of translesion synthesis (TLS), which can bypass lesions caused by DNA damage, and consequently, is mutagenic. The ImuC polymerase in *Pseudomonas* contributes to the tolerance to DNA alkylation agents (88), and the inactivation of the operon could limit accumulation of mutations. Interestingly, CF *P. aeruginosa* isolates frequently display a hyper-mutator phenotype, that is primary due to inactivation of *mutS* and has been previously suggested as being advantageous for bacterial adaptation to the CF lungs (89). The *mutS* gene is identical and intact in CLJ1 and CLJ3 and the predicted proteins differ by one amino acid at position 593 from MutS of PA7 (threonine in PA7, serine in CLJ1/CLJ3). Therefore, the physiological impact of inactivation of the translesion synthesis system is unclear. Two additional DNA-repair proteins, RecN (CLJ1_5150) and RecA (CLJ1_1263), were found overrepresented in CLJ3 at the transcriptome and proteome levels (Table S9), suggesting the interplay between different ways of defense mechanism against uncontrolled mutational rates induced by the hostile environment. We found more than 600 single nucleotide polymorphisms (SNPs) between CLJ1 and CLJ3. This is about six times higher compared to sequential isolates found in the same CF patients over 8.8 years in one study (10) and almost 10 times higher than mutations in a matched isolate pair, collected 7 and 1/2 years apart from a single patient in another study (12) . However, another study showed that some strains isolated from non-CF patients could share between 176 and 736 SNPs (9). Although we could not precisely account for the role of SNPs, they may contribute to differences in gene expression between the two isolates that could not be directly attributed to ISs.

Two CLJ clones show different pathogenic potential in *Galleria mellonella*

To assess the global virulence of the two strains in a whole organism, we used wax moth *Galleria* infection model, and followed the survival of infected larvae following inoculation with different strains. Under the same infection conditions, the CLJ1 strain was found more virulent than the PA7 strain, while CLJ3 was unable to kill *Galleria* larvae (Fig. 5E). This result shows that CLJ3, while gaining resistance to antimicrobials, has lost its virulence potential, in agreement with our previous observation that CLJ3 is sensitive to serum (Figure 3B) and less cytotoxic due to the loss of the ability to secrete Exolysin (16). More than 40 insertion sites of ISL3 have been detected in the genome of CLJ3, most within or upstream of genes encoding hypothetical proteins or putative regulators (Table S6) and some of them may have influenced fitness of the CLJ3 strain in the *Galleria* model of infection. Thus, we found that ISs have greatly contributed to pathogenicity of the CLJ1 strain and to multi-drug resistance of CLJ3. As the coexistence of the two isolates in the patient lungs is possible, we can conclude that ISs have determined the global success of the CLJ lineage in establishing fatal infection.

Discussion

P. aeruginosa strains belonging to the group of taxonomic outliers are abundant in humid environments (18, 19, 90, 91) and are considered to be innocuous, based on a previous study (92). Here we present the results of a multiomics approach applied to two recent clinical isolates, CLJ1 and CLJ3, belonging to the same group of taxonomic outliers. This work gave insights into genome-wide modifications that provided bacteria with the

497 weapons for a successful colonization and dissemination in the human host. We found
498 that the genomes of those strains are highly dynamic and evolved within the patient due
499 to high number of ISs. Global traits necessary for pathogenicity and survival in the host,
500 i.e. motility, adhesion and resistance to antimicrobials, are modulated by the CLJ-ISL3
501 element. The later isolate CLJ3 acquired resistance toward antibiotics provided to the
502 patients during hospitalization, with some ISs directly affecting components playing a role
503 in resistance. Compared to the early colonizer CLJ1, the CLJ3 strain also displays higher
504 intracellular levels of c-di-GMP, higher expression of the biofilm-associated adhesion
505 protein CdrA, and has lost part of the LPS by the deletion of the *galU* gene; these are all
506 features of adaptation to the so-called “chronic” lifestyle. Moreover, using bioinformatics
507 screens we found 35 additional IS elements in the CLJ3 genome and these very likely
508 contribute to other phenotypical changes not assessed in this study. In addition to
509 previously known virulence determinants, numerous differentially expressed genes were
510 annotated as “hypothetical” by using automated annotation technology RAST and their
511 contribution in *P. aeruginosa* adaptation to human host should be further explored.

512 Although the exact origin of the ISL3 in the CLJ lineage is unknown, we can speculate on
513 its acquisition from an environmental bacterial species present in the common microbial
514 community. Indeed, the GC content of the CLJ1-ISL3 is 54.8%, while the rest of the *P.*
515 *aeruginosa* genome is 66.6%, suggesting recent acquisition by a horizontal transfer. The
516 initial disruption of the genes encoding flagellar components and pili may have allowed
517 the CLJ strain to overcome the human immune defenses. Increased capacity to secrete
518 the pore-forming toxin Exolysin gave the strain additional advantage to further damage
519 the host epithelium and endothelium tissues and disseminate.

Previous studies indicated that the contribution of ISs to the adaptation of *P. aeruginosa* to CF environment is low with limited transposition events during chronic infection (93), which is in contrast to what was found in *in vitro* evolution experiment in *Escherichia coli* (94, 95). Search in the Pseudomonas data base (27) with the nucleotide sequence of the CLJ1-ISL3 fragment revealed altogether 12 strains with 100% identical fragments; 8 from Copenhagen University Hospital (96) and the others isolated from tertiary care hospital's intensive care units in Seattle (97). The exact positions of the ISL3 element in those genomes and phenotypes related to disrupted gene(s) are currently unknown, but this knowledge would give insights on whether and how the ISL3 participated in colonization process and adaptation of those strains to a particular infectious niche. More recent genomic characterization of environmental *P. aeruginosa* isolates from dental unit waterlines, showed that in addition of the OSA loci, the IS*Pa*11 fragment altered genes of two master regulators, LasR and GacS, supporting the idea of ecological adaptive potential of *P. aeruginosa* by mobile elements (48). Therefore, in addition to small nucleotide changes in pathoadaptive genes, mobile genetic elements drive the emergence of phenotypic traits leading to adaptation of *P. aeruginosa* to niche the bacteria encounter and undoubtedly take part in strain-specific pathogenicity. The contribution of ISs to the increase of the bacterial pathogenic arsenal is likely underestimated due to still limited availability of closed bacterial genomes. New DNA sequencing technologies accessing genome fragments of tens and even hundreds of kb should reveal the global impact of mobile elements in bacterial evolution.

Author's contributions

ES performed all presented bioinformatics analyses of the data. PB, AB, GB, YC and AA performed experiments. SL provided materials. SE and IA designed the research and performed experiments. YC, SE and IA analyzed the data. SE and IA wrote the manuscript. All authors participated in preparing figures and tables and revised the final version of the manuscript.

Acknowledgements

The authors thank Maria Guillermina Casabona for initial proteomic experiments on CLJ strains and Yves Vandenbrouck (CEA, Grenoble), Frédéric Boyer (UGA, Grenoble) and François Lebreton (HMS, Boston) for advice on genome analysis. Special thanks go to Mylène Robert-Genthon and Florian Chenavier for performing RT-qPCRs and PCRs and to Peter Panchev for carrying out *Galleria* infections. We are grateful to Tim Tolker-Nielsen for the generous gift of pUCP22-p*CdrA-gfp*(ASV)^{c} plasmid.

This work was supported by grants from Agence Nationale de la Recherche (ANR-15-CE11-0018-01), the Laboratory of Excellence GRAL (ANR-10-LABX-49-01) and the Fondation pour la Recherche Médicale (Team FRM 2017, DEQ20170336705). Proteomic experiments were partly supported by the Proteomics French Infrastructure (ANR-10-INBS-08-01 grant). We further acknowledge support from CNRS, INSERM, CEA, and University Grenoble Alpes. Work in SL's laboratory was supported by a grant from the Cystic Fibrosis Foundation. AB received a Ph.D. fellowship from the CEA Irtelis program.

The funding bodies have no roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Supplementary information is available at the journal's website

Competing Interests: The authors declare no competing financial interests

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008 Feb 21;451(7181):990-3.

2. Bartoli C, Roux F, Lamichhane JR. Molecular mechanisms underlying the emergence of bacterial pathogens: an ecological perspective. Molecular plant pathology. 2016 Feb;17(2):303-10.

3. Diard M, Hardt WD. Evolution of bacterial virulence. FEMS microbiology reviews. 2017 Sep 1;41(5):679-97.

4. Aujoulat F, Roger F, Bourdier A, Lotthe A, Lamy B, Marchandin H, et al. From environment to man: genome evolution and adaptation of human opportunistic bacterial pathogens. Genes. 2012 Mar 26;3(2):191-232.

5. Silby MW, Winstanley C, Godfrey SA, Levy SB, Jackson RW. Pseudomonas genomes: diverse and adaptable. FEMS microbiology reviews. 2011 Jul;35(4):652-80.

6. Xin XF, Kvitko B, He SY. Pseudomonas syringae: what it takes to be a pathogen. Nature reviews Microbiology. 2018 Feb 26.

585 7. Cullen L, McClean S. Bacterial Adaptation during Chronic Respiratory Infections.
586 Pathogens. 2015 Mar 02;4(1):66-89.
587 8. Folkesson A, Jelsbak L, Yang L, Johansen HK, Ciofu O, Hoiby N, et al. Adaptation
588 of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective.
589 Nature reviews Microbiology. 2012 Dec;10(12):841-51.
590 9. Hilliam Y, Moore MP, Lamont IL, Bilton D, Haworth CS, Foweraker J, et al.
591 Pseudomonas aeruginosa adaptation and diversification in the non-cystic fibrosis
592 bronchiectasis lung. The European respiratory journal. 2017 Apr;49(4).
593 10. Marvig RL, Damkiaer S, Khademi SM, Markussen TM, Molin S, Jelsbak L. Within-
594 host evolution of Pseudomonas aeruginosa reveals adaptation toward iron acquisition
595 from hemoglobin. mBio. 2014 May 6;5(3):e00966-14.
596 11. Marvig RL, Johansen HK, Molin S, Jelsbak L. Genome analysis of a transmissible
597 lineage of pseudomonas aeruginosa reveals pathoadaptive mutations and distinct
598 evolutionary paths of hypermutators. PLoS genetics. 2013;9(9):e1003741.
599 12. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, et
600 al. Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis
601 patients. Proceedings of the National Academy of Sciences of the United States of
602 America. 2006 May 30;103(22):8487-92.
603 13. Kos VN, Deraspe M, McLaughlin RE, Whiteaker JD, Roy PH, Alm RA, et al. The
604 resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility.
605 Antimicrobial agents and chemotherapy. 2015 Jan;59(1):427-36.
606 14. Thrane SW, Taylor VL, Freschi L, Kukavica-Ibrulj I, Boyle B, Laroche J, et al. The
607 Widespread Multidrug-Resistant Serotype O12 Pseudomonas aeruginosa Clone

608 Emerged through Concomitant Horizontal Transfer of Serotype Antigen and Antibiotic
609 Resistance Gene Clusters. mBio. 2015 Sep 22;6(5):e01396-15.
610 15. Hauser AR. The type III secretion system of Pseudomonas aeruginosa: infection
611 by injection. Nature reviews Microbiology. 2009 Sep;7(9):654-65.
612 16. Elsen S, Huber P, Bouillot S, Coute Y, Fournier P, Dubois Y, et al. A type III
613 secretion negative clinical strain of Pseudomonas aeruginosa employs a two-partner
614 secreted exolysin to induce hemorrhagic pneumonia. Cell host & microbe. 2014 Feb
615 12;15(2):164-76.
616 17. Roy PH, Tetu SG, Larouche A, Elbourne L, Tremblay S, Ren Q, et al. Complete
617 genome sequence of the multiresistant taxonomic outlier Pseudomonas aeruginosa PA7.
618 PloS one. 2010 Jan 22;5(1):e8842.
619 18. Selezska K, Kazmierczak M, Musken M, Garbe J, Schobert M, Haussler S, et al.
620 Pseudomonas aeruginosa population structure revisited under environmental focus:
621 impact of water quality and phage pressure. Environmental microbiology. 2012
622 Aug;14(8):1952-67.
623 19. Wiehlmann L, Cramer N, Tummler B. Habitat-associated skew of clone abundance
624 in the Pseudomonas aeruginosa population. Environmental microbiology reports. 2015
625 Dec;7(6):955-60.
626 20. Huber P, Basso P, Reboud E, Attree I. Pseudomonas aeruginosa renews its
627 virulence factors. Environmental microbiology reports. 2016 Jul 18.
628 21. Reboud E, Basso P, Maillard AP, Huber P, Attree I. Exolysin Shapes the Virulence
629 of Pseudomonas aeruginosa Clonal Outliers. Toxins. 2017 Nov 9;9(11).

630 22. Reboud E, Elsen S, Bouillot S, Golovkine G, Basso P, Jeannot K, et al. Phenotype
631 and toxicity of the recently discovered exlA-positive Pseudomonas aeruginosa strains
632 collected worldwide. Environmental microbiology. 2016 Oct;18(10):3425-39.
633 23. Basso P, Wallet P, Elsen S, Soleilhac E, Henry T, Faudry E, et al. Multiple
634 Pseudomonas species secrete exolysin-like toxins and provoke Caspase-1-dependent
635 macrophage death. Environmental microbiology. 2017 Oct;19(10):4045-64.
636 24. Bouillot S, Munro P, Gallet B, Reboud E, Cretin F, Golovkine G, et al.
637 Pseudomonas aeruginosa Exolysin promotes bacterial growth in lungs, alveolar damage
638 and bacterial dissemination. Scientific reports. 2017 May 18;7(1):2120.
639 25. Choi KH, Kumar A, Schweizer HP. A 10-min method for preparation of highly
640 electrocompetent Pseudomonas aeruginosa cells: application for DNA fragment transfer
641 between chromosomes and plasmid transformation. Journal of microbiological methods.
642 2006 Mar;64(3):391-7.
643 26. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST
644 Server: rapid annotations using subsystems technology. BMC genomics. 2008 Feb
645 8;9:75.
646 27. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. Enhanced
647 annotations and features for comparing thousands of Pseudomonas genomes in the
648 Pseudomonas genome database. Nucleic acids research. 2016 Jan 4;44(D1):D646-53.
649 28. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos:
650 an information aesthetic for comparative genomics. Genome research. 2009
651 Sep;19(9):1639-45.

652 29. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
653 MEM. arXiv. 2013:1303.3997.
654 30. Han K, Tjaden B, Lory S. GRIL-seq provides a method for identifying direct targets
655 of bacterial small regulatory RNA by in vivo proximity ligation. Nature microbiology. 2016
656 Dec 22;2:16239.
657 31. Casabona MG, Vandenbrouck Y, Attree I, Coute Y. Proteomic characterization of
658 Pseudomonas aeruginosa PAO1 inner membrane. Proteomics. 2013 Aug;13(16):2419-
659 23.
660 32. Wieczorek S, Combes F, Lazar C, Giai Gianetto Q, Gatto L, Dorffer A, et al.
661 DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery
662 proteomics. Bioinformatics. 2017 Jan 1;33(1):135-6.
663 33. Vizcaino JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016
664 update of the PRIDE database and its related tools. Nucleic acids research. 2016 Jan
665 4;44(D1):D447-56.
666 34. Rybtke MT, Borlee BR, Murakami K, Irie Y, Hentzer M, Nielsen TE, et al.
667 Fluorescence-based reporter for gauging cyclic di-GMP levels in Pseudomonas
668 aeruginosa. Applied and environmental microbiology. 2012 Aug;78(15):5060-9.
669 35. Zlosnik JE, Gunaratnam LC, Speert DP. Serum susceptibility in clinical isolates of
670 Burkholderia cepacia complex bacteria: development of a growth-based assay for high
671 throughput determination. Frontiers in cellular and infection microbiology. 2012;2:67.
672 36. Berry A, Han K, Trouillon J, Robert-Genthon M, Ragno M, Lory S, et al. cAMP and
673 Vfr Control Exolysin Expression and Cytotoxicity of Pseudomonas aeruginosa Taxonomic
674 Outliers. Journal of bacteriology. 2018 Jun 15;200(12).

675 37. Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, et al.
676 Dynamics of Pseudomonas aeruginosa genome evolution. Proceedings of the National
677 Academy of Sciences of the United States of America. 2008 Feb 26;105(8):3100-5.
678 38. Kung VL, Ozer EA, Hauser AR. The accessory genome of Pseudomonas
679 aeruginosa. Microbiology and molecular biology reviews : MMBR. 2010 Dec;74(4):621-
680 41.
681 39. Klockgether J, Cramer N, Wiehlmann L, Davenport CF, Tummler B. Pseudomonas
682 aeruginosa Genomic Structure and Diversity. Frontiers in microbiology. 2011;2:150.
683 40. Kus JV, Tullis E, Cvitkovitch DG, Burrows LL. Significant differences in type IV pilin
684 allele distribution among Pseudomonas aeruginosa isolates from cystic fibrosis (CF)
685 versus non-CF patients. Microbiology. 2004 May;150(Pt 5):1315-26.
686 41. Smith DJ, Park J, Tiedje JM, Mohn WW. A large gene cluster in Burkholderia
687 xenovorans encoding abietane diterpenoid catabolism. Journal of bacteriology. 2007
688 Sep;189(17):6195-204.
689 42. Sana TG, Berni B, Bleves S. The T6SSs of Pseudomonas aeruginosa Strain PAO1
690 and Their Effectors: Beyond Bacterial-Cell Targeting. Frontiers in cellular and infection
691 microbiology. 2016;6:61.
692 43. Giraud C, Bernard CS, Calderon V, Yang L, Filloux A, Molin S, et al. The PprA-
693 PprB two-component system activates CupE, the first non-archetypal Pseudomonas
694 aeruginosa chaperone-usher pathway system assembling fimbriae. Environmental
695 microbiology. 2011 Mar;13(3):666-83.
696 44. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.
697 DNA-DNA hybridization values and their relationship to whole-genome sequence

698 similarities. International journal of systematic and evolutionary microbiology. 2007
699 Jan;57(Pt 1):81-91.
700 45. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference
701 centre for bacterial insertion sequences. Nucleic acids research. 2006 Jan 1;34(Database
702 issue):D32-6.
703 46. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE:
704 functional classification of proteins via subfamily domain architectures. Nucleic acids
705 research. 2017 Jan 4;45(D1):D200-D3.
706 47. Castro-Jaimes S, Salgado-Camargo AD, Grana-Miraglia L, Lozano L, Bocanegra-
707 Ibarias P, Volkow-Fernandez P, et al. Complete Genome Sequence of a Multidrug-
708 Resistant Acinetobacter baumannii Isolate Obtained from a Mexican Hospital (Sequence
709 Type 422). Genome announcements. 2016 Jun 23;4(3).
710 48. Vincent AT, Freschi L, Jeukens J, Kukavica-Ibrulj I, Emond-Rheault JG, Leduc A,
711 et al. Genomic characterisation of environmental Pseudomonas aeruginosa isolated from
712 dental unit waterlines revealed the insertion sequence ISPa11 as a chaotropic element.
713 FEMS microbiology ecology. 2017 Sep 1;93(9).
714 49. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic
715 impact and diversity. FEMS microbiology reviews. 2014 Sep;38(5):865-91.
716 50. Vandecraen J, Chandler M, Aertsen A, Van Houdt R. The impact of insertion
717 sequences on bacterial genome plasticity and adaptability. Critical reviews in
718 microbiology. 2017 Nov;43(6):709-30.
719 51. Martinez E, Perez JE, Marquez C, Vilacoba E, Centron D, Leal AL, et al. Emerging
720 and existing mechanisms co-operate in generating diverse beta-lactam resistance

721 phenotypes in geographically dispersed and genetically disparate Pseudomonas
722 aeruginosa strains. J Glob Antimicrob Resist. 2013 Sep;1(3):135-42.
723 52. Sun Q, Ba Z, Wu G, Wang W, Lin S, Yang H. Insertion sequence ISRP10
724 inactivation of the oprD gene in imipenem-resistant Pseudomonas aeruginosa clinical
725 isolates. International journal of antimicrobial agents. 2016 May;47(5):375-9.
726 53. Vassilara F, Galani I, Souli M, Papanikolaou K, Giamarellou H, Papadopoulos A.
727 Mechanisms responsible for imipenem resistance among Pseudomonas aeruginosa
728 clinical isolates exposed to imipenem concentrations within the mutant selection window.
729 Diagnostic microbiology and infectious disease. 2017 Jul;88(3):276-81.
730 54. Wolkowicz T, Patzer JA, Kaminska W, Gierczynski R, Dzierzanowska D.
731 Distribution of carbapenem resistance mechanisms in Pseudomonas aeruginosa isolates
732 among hospitalised children in Poland: Characterisation of two novel insertion sequences
733 disrupting the oprD gene. J Glob Antimicrob Resist. 2016 Dec;7:119-25.
734 55. Lindberg F, Lindquist S, Normark S. Inactivation of the ampD gene causes
735 semiconstitutive overproduction of the inducible Citrobacter freundii beta-lactamase.
736 Journal of bacteriology. 1987 May;169(5):1923-8.
737 56. Perez-Gallego M, Torrens G, Castillo-Vera J, Moya B, Zamorano L, Cabot G, et
738 al. Impact of AmpC Derepression on Fitness and Virulence: the Mechanism or the
739 Pathway? mBio. 2016 Oct 25;7(5).
740 57. Lee M, Artola-Recolons C, Carrasco-Lopez C, Martinez-Caballero S, Hesek D,
741 Spink E, et al. Cell-wall remodeling by the zinc-protease AmpDh3 from Pseudomonas
742 aeruginosa. Journal of the American Chemical Society. 2013 Aug 28;135(34):12604-7.

743 58. Vollmer W, von Rechenberg M, Holtje JV. Demonstration of molecular interactions
744 between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the
745 scaffolding protein MipA of Escherichia coli. The Journal of biological chemistry. 1999
746 Mar 5;274(10):6726-34.
747 59. Li XZ, Plesiat P, Nikaido H. The challenge of efflux-mediated antibiotic resistance
748 in Gram-negative bacteria. Clinical microbiology reviews. 2015 Apr;28(2):337-418.
749 60. Shah P, Swiatlo E. A multifaceted role for polyamines in bacterial pathogens.
750 Molecular microbiology. 2008 Apr;68(1):4-16.
751 61. Lam JS, Taylor VL, Islam ST, Hao Y, Kocincova D. Genetic and Functional
752 Diversity of Pseudomonas aeruginosa Lipopolysaccharide. Frontiers in microbiology.
753 2011;2:118.
754 62. Raymond CK, Sims EH, Kas A, Spencer DH, Kutyavin TV, Ivey RG, et al. Genetic
755 variation at the O-antigen biosynthetic locus in Pseudomonas aeruginosa. Journal of
756 bacteriology. 2002 Jul;184(13):3614-22.
757 63. Dean CR, Goldberg JB. Pseudomonas aeruginosa galU is required for a complete
758 lipopolysaccharide core and repairs a secondary mutation in a PA103 (serogroup O11)
759 wbpM mutant. FEMS microbiology letters. 2002 May 07;210(2):277-83.
760 64. Priebe GP, Dean CR, Zaidi T, Meluleni GJ, Coleman FT, Coutinho YS, et al. The
761 galU Gene of Pseudomonas aeruginosa is required for corneal infection and efficient
762 systemic spread following pneumonia but not for infection confined to the lung. Infection
763 and immunity. 2004 Jul;72(7):4224-32.

764 65. Cabot G, Zamorano L, Moya B, Juan C, Navas A, Blazquez J, et al. Evolution of
765 Pseudomonas aeruginosa Antimicrobial Resistance and Fitness under Low and High
766 Mutation Rates. Antimicrobial agents and chemotherapy. 2016 Jan 4;60(3):1767-78.
767 66. Arora SK, Bangera M, Lory S, Ramphal R. A genomic island in Pseudomonas
768 aeruginosa carries the determinants of flagellin glycosylation. Proceedings of the National
769 Academy of Sciences of the United States of America. 2001 Jul 31;98(16):9342-7.
770 67. Ince D, Sutterwala FS, Yahr TL. Secretion of Flagellar Proteins by the
771 Pseudomonas aeruginosa Type III Secretion-Injectisome System. Journal of
772 bacteriology. 2015 Jun 15;197(12):2003-11.
773 68. Feuillet V, Medjane S, Mondor I, Demaria O, Pagni PP, Galan JE, et al.
774 Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. Proceedings
775 of the National Academy of Sciences of the United States of America. 2006 Aug
776 15;103(33):12487-92.
777 69. Tammam S, Sampaleanu LM, Koo J, Manoharan K, Daubaras M, Burrows LL, et
778 al. PilMNOPQ from the Pseudomonas aeruginosa type IV pilus system form a
779 transenvelope protein interaction network that interacts with PilA. Journal of bacteriology.
780 2013 May;195(10):2126-35.
781 70. Basso P, Ragno M, Elsen S, Reboud E, Golovkine G, Bouillot S, et al.
782 Pseudomonas aeruginosa Pore-Forming Exolysin and Type IV Pili Cooperate To Induce
783 Host Cell Lysis. mBio. 2017 Jan 24;8(1).
784 71. Borlee BR, Goldman AD, Murakami K, Samudrala R, Wozniak DJ, Parsek MR.
785 Pseudomonas aeruginosa uses a cyclic-di-GMP-regulated adhesin to reinforce the
786 biofilm extracellular matrix. Molecular microbiology. 2010 Feb;75(4):827-42.

787 72. Faure LM, Garvis S, de Bentzmann S, Bigot S. Characterization of a novel two-
788 partner secretion system implicated in the virulence of Pseudomonas aeruginosa.
789 Microbiology. 2014 Sep;160(Pt 9):1940-52.
790 73. Kida Y, Higashimoto Y, Inoue H, Shimizu T, Kuwano K. A novel secreted protease
791 from Pseudomonas aeruginosa activates NF-kappaB through protease-activated
792 receptors. Cellular microbiology. 2008 Jul;10(7):1491-504.
793 74. Melvin JA, Gaston JR, Phillips SN, Springer MJ, Marshall CW, Shanks RMQ, et al.
794 Pseudomonas aeruginosa Contact-Dependent Growth Inhibition Plays Dual Role in Host-
795 Pathogen Interactions. mSphere. 2017 Nov-Dec;2(6).
796 75. Mercy C, Ize B, Salcedo SP, de Bentzmann S, Bigot S. Functional Characterization
797 of Pseudomonas Contact Dependent Growth Inhibition (CDI) Systems. PloS one.
798 2016;11(1):e0147435.
799 76. Blumer C, Haas D. Iron regulation of the hcnABC genes encoding hydrogen
800 cyanide synthase depends on the anaerobic regulator ANR rather than on the global
801 activator GacA in Pseudomonas fluorescens CHA0. Microbiology. 2000 Oct;146 (Pt
802 10):2417-24.
803 77. Carterson AJ, Morici LA, Jackson DW, Frisk A, Lizewski SE, Jupiter R, et al. The
804 transcriptional regulator AlgR controls cyanide production in Pseudomonas aeruginosa.
805 Journal of bacteriology. 2004 Oct;186(20):6837-44.
806 78. Carroll W, Lenney W, Wang T, Spanel P, Alcock A, Smith D. Detection of volatile
807 compounds emitted by Pseudomonas aeruginosa using selected ion flow tube mass
808 spectrometry. Pediatric pulmonology. 2005 May;39(5):452-6.

809 79. Ryall B, Davies JC, Wilson R, Shoemark A, Williams HD. Pseudomonas
810 aeruginosa, cyanide accumulation and lung function in CF and non-CF bronchiectasis
811 patients. The European respiratory journal. 2008 Sep;32(3):740-7.
812 80. Frangipani E, Perez-Martinez I, Williams HD, Cherbuin G, Haas D. A novel
813 cyanide-inducible gene cluster helps protect Pseudomonas aeruginosa from cyanide.
814 Environmental microbiology reports. 2014 Feb;6(1):28-34.
815 81. Hirai T, Osamura T, Ishii M, Arai H. Expression of multiple cbb3 cytochrome c
816 oxidase isoforms by combinations of multiple isosubunits in Pseudomonas aeruginosa.
817 Proceedings of the National Academy of Sciences of the United States of America. 2016
818 Oct 24.
819 82. Firoved AM, Ornatowski W, Deretic V. Microarray analysis reveals induction of
820 lipoprotein genes in mucoid Pseudomonas aeruginosa: implications for inflammation in
821 cystic fibrosis. Infection and immunity. 2004 Sep;72(9):5012-8.
822 83. Damron FH, Napper J, Teter MA, Yu HD. Lipotoxin F of Pseudomonas aeruginosa
823 is an AlgU-dependent and alginate-independent outer membrane protein involved in
824 resistance to oxidative stress and adhesion to A549 human lung epithelia. Microbiology.
825 2009 Apr;155(Pt 4):1028-38.
826 84. Fortier LC, Sekulovic O. Importance of prophages to evolution and virulence of
827 bacterial pathogens. Virulence. 2013 Jul 1;4(5):354-65.
828 85. Winstanley C, Langille MG, Fothergill JL, Kukavica-Ibrulj I, Paradis-Bleau C,
829 Sanschagrin F, et al. Newly introduced genomic prophage islands are critical
830 determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas
831 aeruginosa. Genome research. 2009 Jan;19(1):12-23.

832 86. Matsui H, Sano Y, Ishihara H, Shinomiya T. Regulation of pyocin genes in
833 Pseudomonas aeruginosa by positive (prtN) and negative (prtR) regulatory genes.
834 Journal of bacteriology. 1993 Mar;175(5):1257-63.
835 87. McFarland KA, Dolben EL, LeRoux M, Kambara TK, Ramsey KM, Kirkpatrick RL,
836 et al. A self-lysis pathway that enhances the virulence of a pathogenic bacterium.
837 Proceedings of the National Academy of Sciences of the United States of America. 2015
838 Jul 7;112(27):8433-8.
839 88. Jatsenko T, Sidorenko J, Saumaa S, Kivisaar M. DNA Polymerases ImuC and
840 DinB Are Involved in DNA Alkylation Damage Tolerance in Pseudomonas aeruginosa and
841 Pseudomonas putida. PloS one. 2017;12(1):e0170719.
842 89. Oliver A, Canton R, Campo P, Baquero F, Blazquez J. High frequency of
843 hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science. 2000
844 May 19;288(5469):1251-4.
845 90. Boukerb AM, Decor A, Ribun S, Tabaroni R, Rousset A, Commin L, et al. Genomic
846 Rearrangements and Functional Diversification of lecA and lecB Lectin-Coding Regions
847 Impacting the Efficacy of Glycomimetics Directed against Pseudomonas aeruginosa.
848 Frontiers in microbiology. 2016;7:811.
849 91. Boukerb AM, Marti R, Cournoyer B. Genome Sequences of Three Strains of the
850 Pseudomonas aeruginosa PA7 Clade. Genome announcements. 2015 Nov 19;3(6).
851 92. Hilker R, Munder A, Klockgether J, Losada PM, Chouvarine P, Cramer N, et al.
852 Interclonal gradient of virulence in the Pseudomonas aeruginosa pangenome from
853 disease and environment. Environmental microbiology. 2015 Jan;17(1):29-46.

854 93. Rau MH, Marvig RL, Ehrlich GD, Molin S, Jelsbak L. Deletion and acquisition of
855 genomic content during early stage adaptation of Pseudomonas aeruginosa to a human
856 host environment. Environmental microbiology. 2012 Aug;14(8):2200-11.
857 94. Gaffe J, McKenzie C, Maharjan RP, Coursange E, Ferenci T, Schneider D.
858 Insertion sequence-driven evolution of Escherichia coli in chemostats. Journal of
859 molecular evolution. 2011 Apr;72(4):398-412.
860 95. Raeside C, Gaffe J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN, et al.
861 Large chromosomal rearrangements during a long-term evolution experiment with
862 Escherichia coli. mBio. 2014 Sep 9;5(5):e01377-14.
863 96. Marvig RL, Sommer LM, Molin S, Johansen HK. Convergent evolution and
864 adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nature
865 genetics. 2015 Jan;47(1):57-64.
866 97. Roach DJ, Burton JN, Lee C, Stackhouse B, Butler-Wu SM, Cookson BT, et al. A
867 Year of Infection in the Intensive Care Unit: Prospective Whole Genome Sequencing of
868 Bacterial Clinical Isolates Reveals Cryptic Transmissions and Novel Microbiota. PLoS
869 genetics. 2015 Jul;11(7):e1005413.
870
871
872
873
874
875

Figure Legends

Fig. 1 Comparison of the CLJ1 and CLJ3 genome, transcriptome and proteome. The overview of the three genomes, including the reference genome of the PA7 strain, is shown on the left, while the image on the right is an enlarged genomic segment at the hcn locus with a more detailed description of the data. The red bar charts indicate that the gene or protein is more expressed in CLJ1, whereas the green bars show higher expression in CLJ3; darker tone indicates statistically significant expression difference between the two strains (False Discovery Rate < 0.05). The labels linked to the outmost ring show the genes that are differentially expressed in both, RNA-Seq and at least one of the proteomic datasets. The protein subcellular localization (outmost ring) is colored as in the Pseudomonas Genome Database (30). The CLJ-ISL3 insertions are indicated by red, green and blue triangles, depending on their presence in CLJ1, CLJ3 or both strains, respectively.

Fig. 2 Insertions of CLJ-ISL3 into genes encoding determinants of antibiotic susceptibility. **A** Representation of the 2,985-bp CLJ-ISL3 IS with three genes and inverted repeats at its ends. **B** Representation of the *oprD* disruption in CLJ3 by the CLJ-ISL3 insertion sequence. **C** Location of CLJ-ISL3 in *ampD* in CLJ3. **D** Analysis of the relative expression of *ampDH3* and *mipA* in CLJ1 and CLJ3 strains by RT-qPCR. The bars indicate the standard error of the mean. **E** Gene organization of the region in PA7 and CLJ1 that is absent in CLJ3. CLJ-ISL3 is found in *puuP* in CLJ1 while, both *puuP* and *mexX* are disrupted in CLJ3, with the connecting sequence replaced by CLJ-ISL3 (see the text).

898 **Fig. 3** ISs within OSA loci and impact on serum sensitivity. **A** Comparison of the OSA
899 region in PA7 and CLJ strains. CLJ-ISL3 is inserted in one location in CLJ1 and at two
900 sites in CLJ3, while IS66 is present in two copies in CLJ3. Correspondence between non-
901 annotated genes is depicted with connecting lines. **B** Kinetics of serum killing of CLJ1
902 and CLJ3 strains. Bacteria were incubated with human sera for indicated time, diluted
903 and spotted on agar plates. Colony forming units were counted after incubation at 37°C.

904

905 **Fig. 4** Modifications in components associated with surface appendices. **A** Gene
906 organization of *flgL* region in PAO1 and the CLJ1 strain. Two CLJ-ISL3, interrupting *fgtA*
907 and *flgL*, were found in the CLJ3 genome that probably recombined in CLJ1, leaving only
908 one copy of the IS. **B** Representation of the *pilMNOPQ* operon. The CLJ-ISL3 in *pilM*
909 gene was found in CLJ1 and CLJ3. **C** Synthesis of the adhesin CdrA is linked to high
910 intracellular c-di-GMP levels. The c-di-GMP levels in CLJ1 and CLJ3 strains were
911 monitored using the pcdrA-gfp(ASV)c plasmid. Fluorescence was measured every 15 min
912 for 6 h of growth. The error-bars indicate the standard deviation.

913

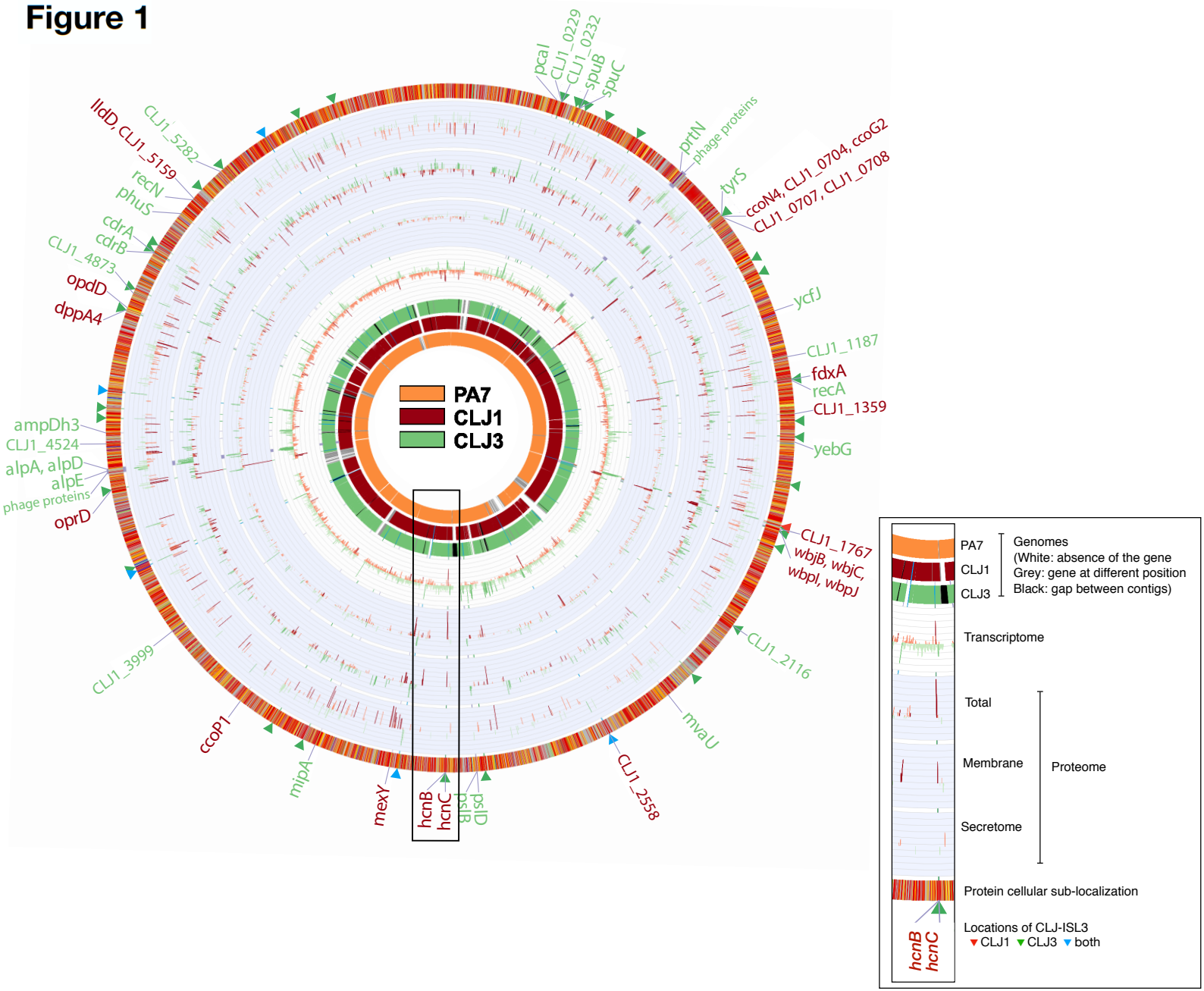
914 **Fig. 5** Expression of virulence factors and pathogenicity. **A** The *hcn* operon in CLJ1 and
915 insertion of CLJ-ISL3 within the *hcnB* gene in CLJ3. **B** Relative expression of *hcnB*
916 measured by RT-qPCR and hydrogen cyanide production in indicated strains. **C** Analysis
917 of the relative expression of *ccoG2*, *ccoN4*, *lptF*, and CLJ1_*1187* in CLJ1 and CLJ3
918 strains by RT-qPCR. The bars indicate the standard error of the mean. **D** The *imuABC*
919 operon and insertion of CLJ-ISL3. **E** Survival of *Galleria* following injection of different

920 strains. Twenty larvae were infected with 5-10 bacteria (estimated from CFU counts) and
921 their survival was followed over indicated period by counting the dead ones. PBS was
922 injected as mock control.

923

924

Figure 1


pcaI
CLJ1_0229
CLJ1_0232
spuB
spuC
prtN
phage proteins
tyrS
ccoN4, CLJ1_0704, ccoG2
CLJ1_0707, CLJ1_0708
ycfJ
CLJ1_1187
fdxA
recA
CLJ1_1359
yebG
CLJ1_1767
wbjB, wbjC,
wbpI, wbpJ
CLJ1_2116
mvaU
CLJ1_2558
hcnB
hcnC
pslB
pslD
mexY
mipA
ccoP1
CLJ1_3999
oprD
phage proteins
alpE
alpA, alpD
CLJ1_4524
ampDh3
dppA4
opdD
CLJ1_4873
cdrB
cdrA
phuS
recN
CLJ1_5282
lldD, CLJ1_5159
PA7
CLJ1
CLJ3
PA7
CLJ1
CLJ3
Genomes
(White: absence of the gene
Grey: gene at different position
Black: gap between contigs)
Transcriptome
Total
Membrane
Secretome
Proteome
Protein cellular sub-localization
Locations of CLJ-ISL3
▼CLJ1 ▼CLJ3 ▼both
hcnB
hcnC

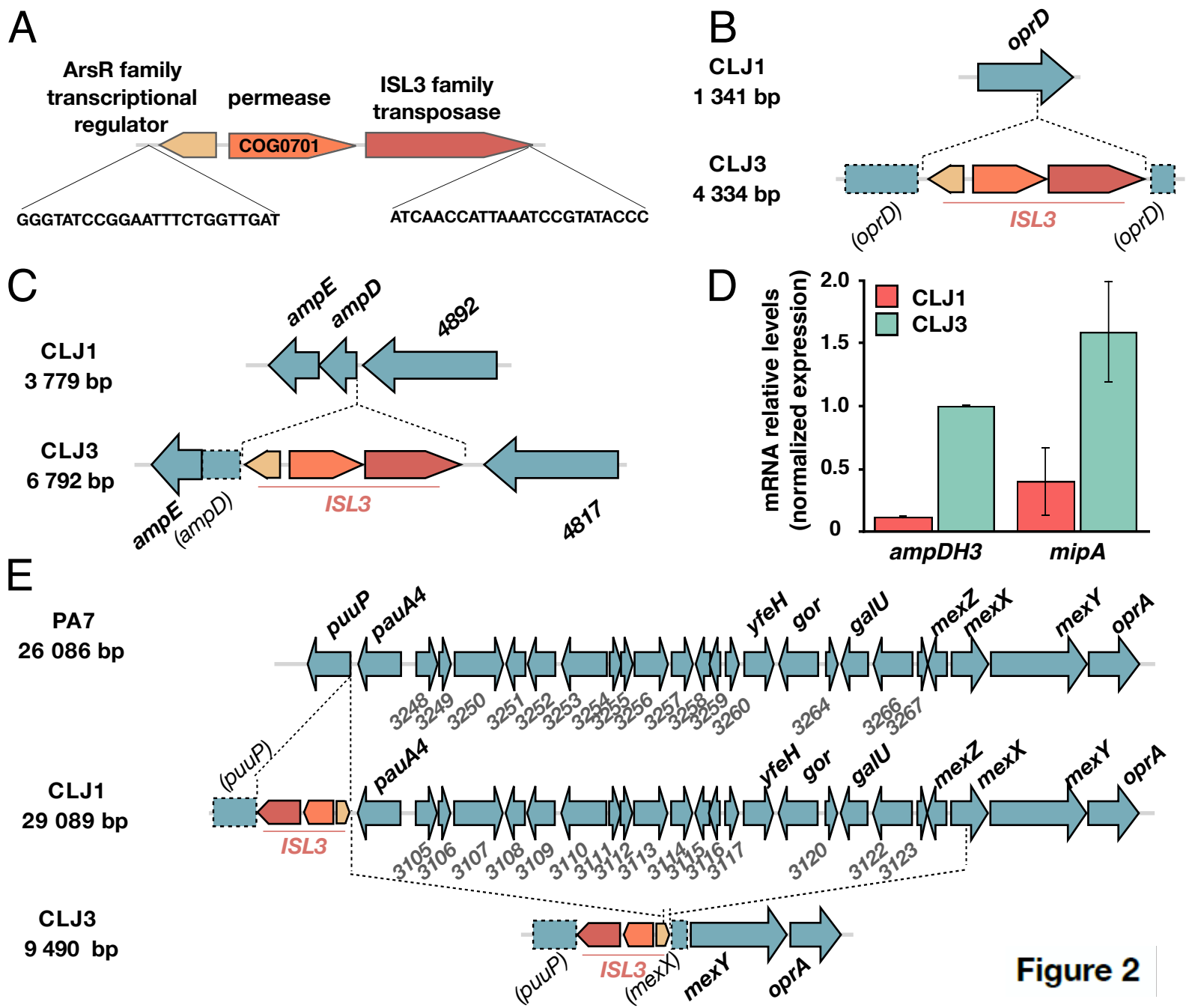
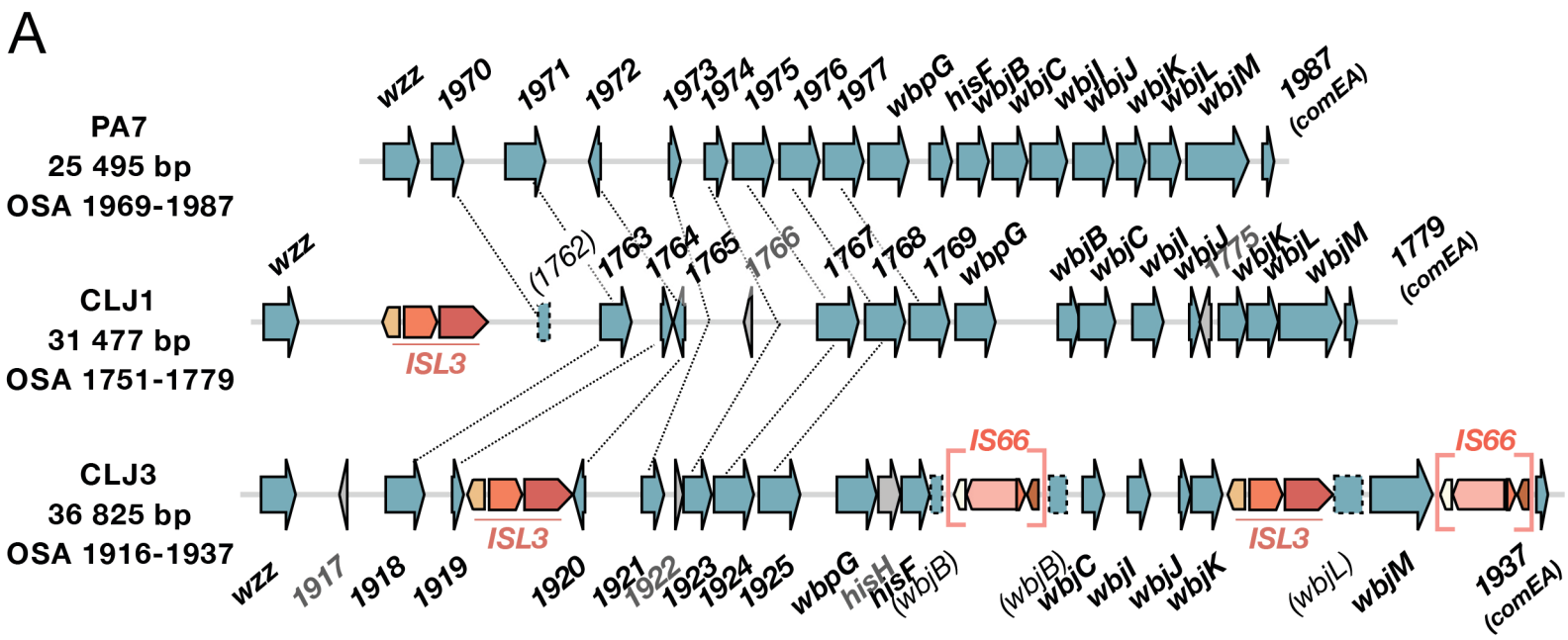

A
ArsR family transcriptional regulator
permease
ISL3 family transposase
COG0701
GGGTATCCGGAATTTCTGGTTGAT
ATCAACCATTAAATCCGTATACCC
B
CLJ1
1 341 bp
oprD
CLJ3
4 334 bp
(oprD)
ISL3
(oprD)
C
CLJ1
3 779 bp
ampE
ampD
4892
CLJ3
6 792 bp
ampE
(ampD)
ISL3
4817
D
mRNA relative levels (normalized expression)
2.0
1.5
1.0
0.5
0
CLJ1
CLJ3
ampDH3
mipA
E
PA7
26 086 bp
puuP
pauA4
3248
3249
3250
3251
3252
3253
3254
3255
3256
3257
3258
3259
3260
yfeH
gor
3264
galU
3266
3267
mexZ
mexX
mexY
oprA
CLJ1
29 089 bp
(puuP)
ISL3
pauA4
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
yfeH
gor
3120
galU
3122
3123
mexZ
mexX
mexY
oprA
CLJ3
9 490 bp
(puuP)
ISL3
(mexX)
mexY
oprA

Figure 2


A
PA7
25 495 bp
OSA 1969-1987
wzz
1970
1971
1972
1973
1974
1975
1976
1977
wbpG
hisF
wbjB
wbjC
wbjI
wbjJ
wbjK
wbjL
wbjM
1987
(comEA)
CLJ1
31 477 bp
OSA 1751-1779
wzz
(1762)
1763
1764
1765
1766
1767
1768
1769
wbpG
wbjB
wbjC
wbjI
wbjJ
1775
wbjK
wbjL
wbjM
1779
(comEA)
ISL3
CLJ3
36 825 bp
OSA 1916-1937
IS66
IS66
ISL3
ISL3
wzz
1917
1918
1919
1920
1921
1922
1923
1924
1925
wbpG
hisH
hisF
(wbjB)
(wbjB)
wbjC
wbjI
wbjJ
wbjK
(wbjL)
wbjM
1937
(comEA)

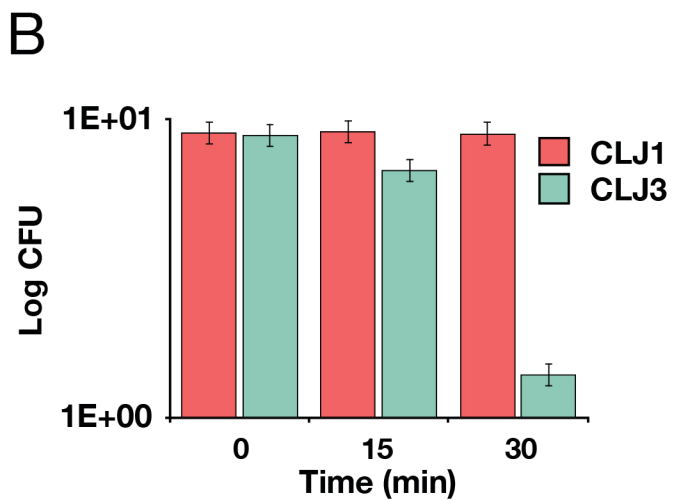

B
1E+01
1E+00
Log CFU
CLJ1
CLJ3
0
15
30
Time (min)

Figure 3

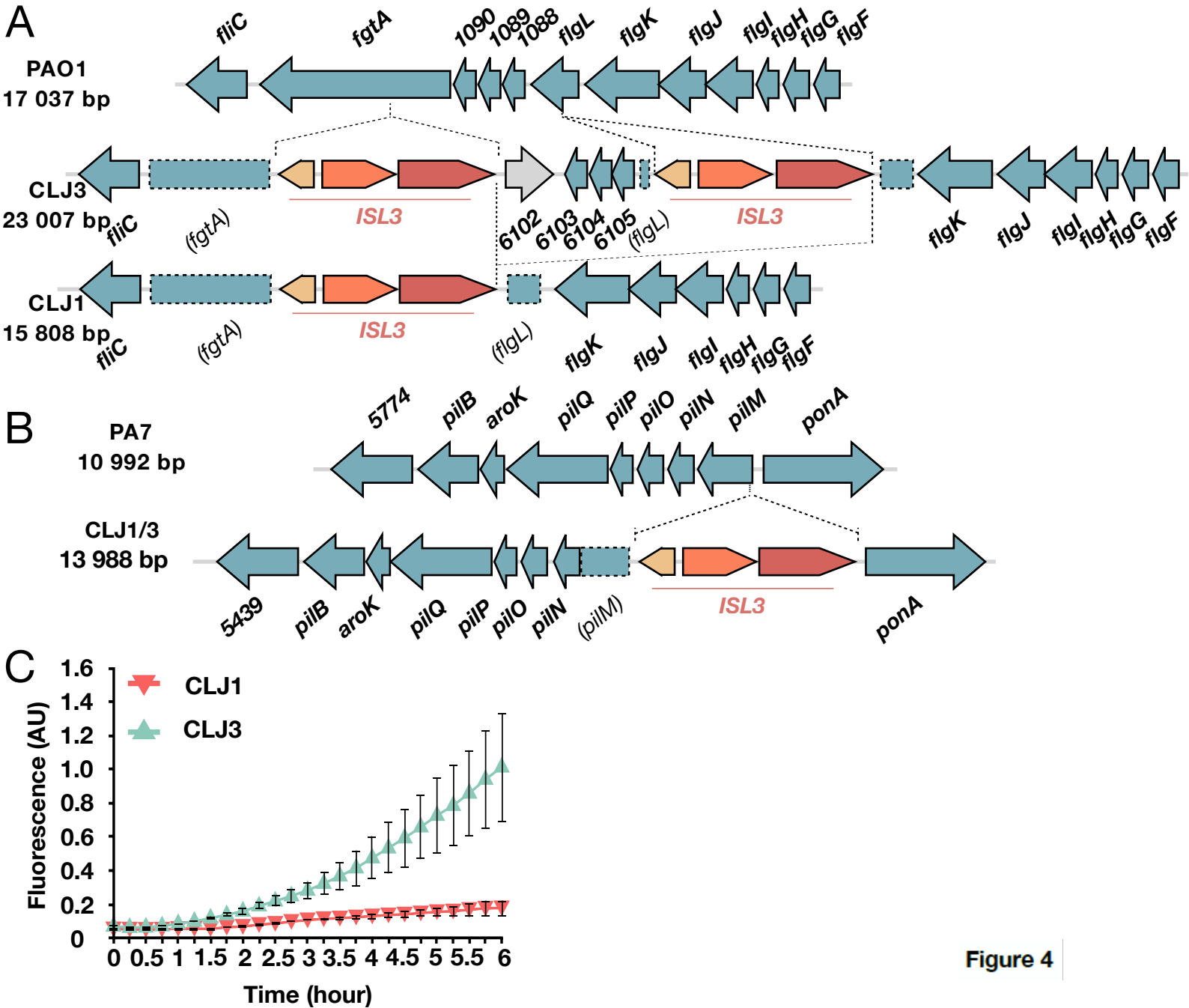
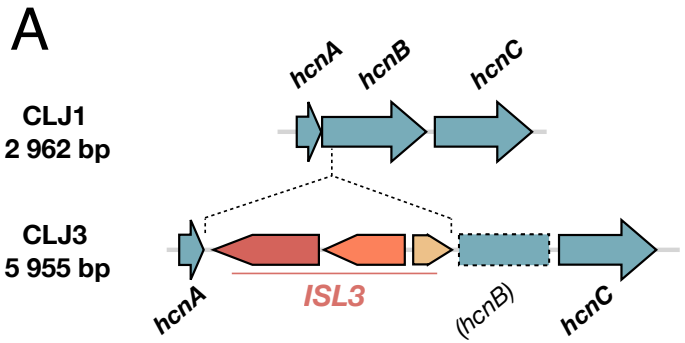
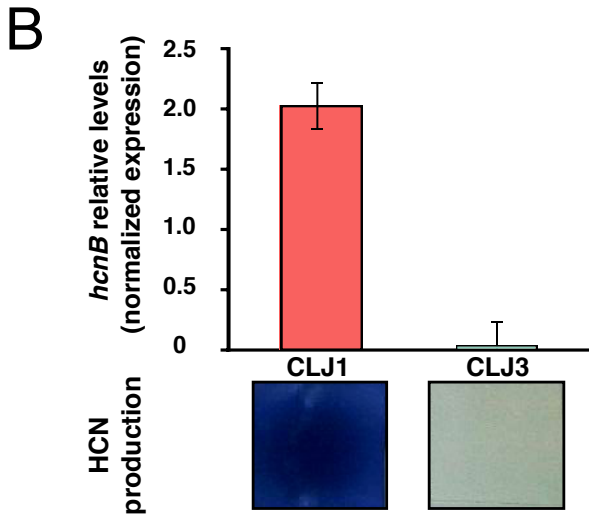
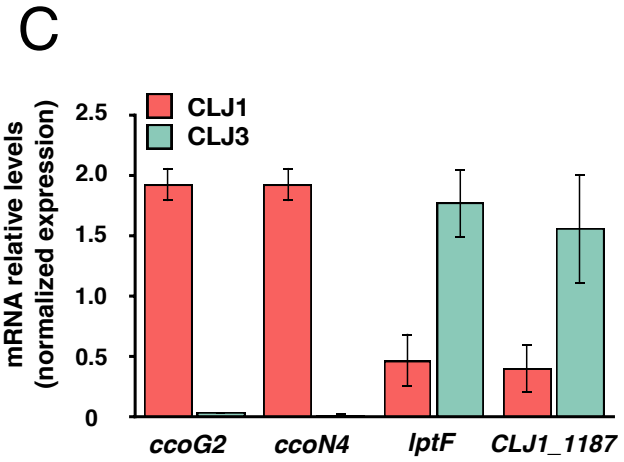

A
PAO1
17 037 bp
fliC
fgtA
1090
1089
1088
flgL
flgK
flgJ
flgI
flgH
flgG
flgF
CLJ3
23 007 bp
fliC
(fgtA)
ISL3
6102
6103
6104
6105
(flgL)
ISL3
flgK
flgJ
flgI
flgH
flgG
flgF
CLJ1
15 808 bp
fliC
(fgtA)
ISL3
(flgL)
flgK
flgJ
flgI
flgH
flgG
flgF
B
PA7
10 992 bp
5774
pilB
aroK
pilQ
pilP
pilO
pilN
pilM
ponA
CLJ1/3
13 988 bp
5439
pilB
aroK
pilQ
pilP
pilO
pilN
(pilM)
ISL3
ponA
C
CLJ1
CLJ3
Fluorescence (AU)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
Time (hour)

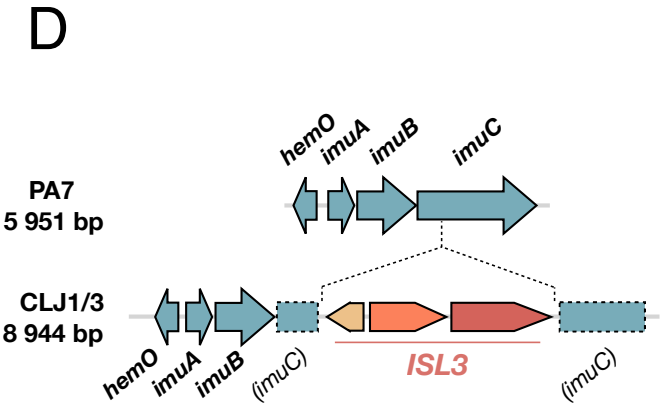
Figure 4

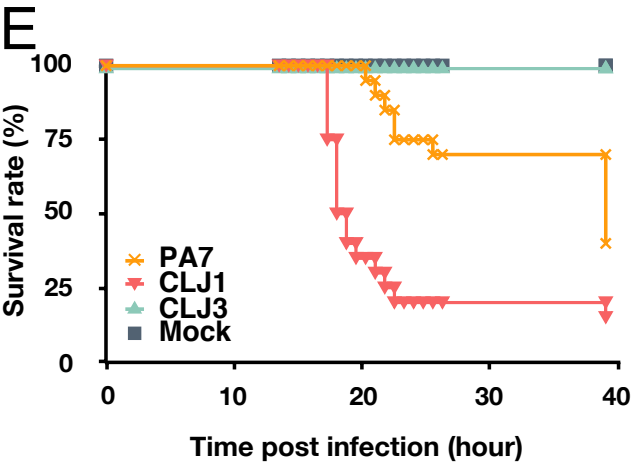
Figure 5


A
CLJ1
2 962 bp
hcnA
hcnB
hcnC
CLJ3
5 955 bp
hcnA
ISL3
(hcnB)
hcnC


B
hcnB relative levels
(normalized expression)
2.5
2.0
1.5
1.0
0.5
0
CLJ1
CLJ3
HCN
production


C
mRNA relative levels
(normalized expression)
CLJ1
CLJ3
2.5
2.0
1.5
1.0
0.5
0
ccoG2
ccoN4
lptF
CLJ1_1187


D
PA7
5 951 bp
hemO
imuA
imuB
imuC
CLJ1/3
8 944 bp
hemO
imuA
imuB
(imuC)
ISL3
(imuC)


E
Survival rate (%)
100
75
50
25
0
PA7
CLJ1
CLJ3
Mock
0
10
20
30
40
Time post infection (hour)

Supporting Information

Insertion sequences drive the emergence of a highly adapted human pathogen

Erwin Sentausa[1#], Pauline Basso[1+], Alice Berry[1], Annie Adrait[2], Gwendoline Bellement[1,2∞], Yohann Couté[2], Stephen Lory[3], Sylvie Elsen[1]* and Ina Attrée[1]*

*Corresponding authors: ina.attree-delic@cea.fr; sylvie.elsen@cea.fr.

The supplementary information includes:

Supplementary Materials and Methods

Supplementary References

Supplementary Figure S1

Supplementary Tables S1 to S10

Supplementary Materials and Methods

Genome sequencing, assembly, annotation and comparison

The DNA extracted from the CLJ1 bacteria was sequenced on Illumina HiSeq 2000 systems at the Beijing Genomics Institute, China with a 2×50 base-pairs (bp) paired-end mode, generating 12,333,368 reads with a genome coverage of 90-100x. PacBio (Base Clear, Leiden, Netherlands) technology was also used, providing 114,707 reads. The Illumina reads were trimmed to remove low quality sequences (limit = 0.05) and ambiguous nucleotides (maximum two nucleotides allowed) and assembled *de novo* using CLC Genomics Workbench 9.0 (Qiagen, Aarhus, Denmark). The resulting contigs were then combined with the PacBio reads using the "Join Contigs" function of the CLC Genome Finishing Module version 1.6.1 (Qiagen). Contigs or scaffolds consisting of fewer than 100 reads were filtered out and BLAST (1) searches were performed to check and remove those that show no match to *Pseudomonas* sequence in GenBank (2). CLJ3 genomic DNA was sequenced using Illumina MiSeq at the Biopolymers Facility, Harvard Medical School, Boston, USA, with 150-bp single-end runs, generating 2,495,173 reads. Quality trimming was done using CLC Genomics Workbench 9.0 using the same parameters as those used for CLJ1, followed by *de novo* assembly with minimal contig length of 1 kb. The order of both CLJ1 and CLJ3 scaffolds or contigs was determined based on the genome sequence of PA7 strain using the "move contigs" tool of the Mauve Genome Alignment Software version snapshot_2015-02-13 (3). Average nucleotide identity between genomes were estimated using ANI calculator tool on the enveomics collection server (4) with minimum alignment length, identity, and number of 700 bp, 70%, and 50, respectively, using 1 000 bp window size and 200 bp step size. SNPs between

CLJ1 and CLJ3 genomes were detected by mapping trimmed CLJ3 reads to CLJ1 genome in CLC Genomics Workbench 9.0, followed by using Basic Variant Detection 1.71 tool in the Workbench with the following parameters: Ploidy = 1, Ignore positions with coverage above = 100 000, Restrict calling to target regions = Not set, Ignore broken pairs = Yes, Ignore non-specific matches = Reads, Minimum coverage = 10, Minimum count = 2, Minimum frequency (%) = 35.0, Base quality filter = Yes, Neighborhood radius = 5, Minimum central quality = 20, Minimum neighborhood quality = 15, Read direction filter = No, Relative read direction filter = Yes, Significance (%) = 1.0, Read position filter = No, Remove pyro-error variants = No. COG annotation was done using the WebMGA server (5) with E-value cutoff of 0.001. Subcellular localizations of PA7 proteins were retrieved from the Pseudomonas Genome Database, whereas those of CLJ1 and CLJ3 were predicted using PSORTb version 3.0.2 (6) and LocTree3 (7). Othologous genes among the strains were identified using OrthoMCL version 2.0.9 (8), with a BLASTp E-value cutoff of 1×10^{-5} and the default Markov cluster algorithm (MCL) inflation parameter of 1.5.

Transcriptome analysis

The preparations of the Illumina libraries and sequencing were done by standard procedures at the Biopolymer Facility, Harvard Medical School, Boston, USA. Illumina HiSeq was used for the sequencing with 50-bp single-end runs, generating 3,225,727 and 21,633,018 reads for CLJ1 and 56,804,547 and 2,139,035 reads for CLJ3. For each replicate, raw RNA-Seq read trimmings were done in CLC Genomics Workbench 9.0 using the same parameters as for genome sequencing reads, and the trimmed reads were mapped to the annotated CLJ1 genome using the RNA-Seq analysis tool. Total

number of reads mapped to the genes were incorporated into a tabular format and analyzed using the DESeq2 differential expression analysis pipeline (9). Differentially expressed genes between CLJ1 and CLJ3 were identified using a 5% False Discovery Rate (FDR).

Proteomics

Sample preparation. Overnight cultures of CLJ1 were diluted to an OD_{600} 0.1 in 30 mL, CLJ3 was left a night at room temperature and then the cultures were incubated at 37°C under shaking to an OD_{600} 0.8. At this point, 30 mL of bacterial cultures were centrifuged at 6,000 rpm, 4°C for 10 min, and supernatants were filtered with 0.22 µm filters. Total membranes separation. Pellets were re-suspended in 1 mL of 10 mM Tris-HCl, 20 % sucrose, pH8 buffer supplemented with protease inhibitors cocktail (PIC, Roche, Basel, Switzerland) and were disrupted by sonication. Unbroken bacteria were removed by a centrifugation at 8,000 rpm for 10 min at 4°C. Total membrane fraction was obtained by ultracentrifugation at 200,000 g for 1 h at 4°C, and the pelleted fraction was washed twice with 1 mL of 10 mM Tris-HCl, 20 mM $MgCl_2$, pH8, supplemented with PIC, and re-suspended in 500 µL of the same buffer. Supernatant fraction. Proteins from supernatants were precipitated by a TCA-sarkosyl method (0.5% final volume of sarkosyl and 7.5% final volume of TCA) for 2 h on ice and centrifuged at 12,000 rpm for 15 min. Pellets were washed twice with tetrahydrofuran, and re-suspended in 50 µL of loading buffer. Prepared samples, in triplicates, were then analyzed by SDS-PAGE and immunoblotting using antibodies directed against 3 synthetic peptides of ExlA ((10), 1:1 000), anti-RpoA (1:5 000) as control for whole cell, anti-TagQ ((11), 1:10 000) as control for the membrane fraction and anti-DsbA (1:2 000) as control for the periplasm fraction. Secondary

antibodies used were HRP-coupled anti-rabbit (1:50 000) and anti-mouse (1:5 000) (Sigma-Aldrich). Western blots were developed using Luminata Crescendo Western HRP (Millipore) substrate.

Mass spectrometry-based quantitative proteomic analyses. Extracted proteins were prepared as described in (12). Briefly, proteins were stacked in the top of a SDS-PAGE gel (NuPAGE 4-12%, ThermoFisher Scientific), stained with Coomassie blue (R250, Bio-Rad) before in-gel digestion using modified trypsin (Promega, sequencing grade). Resulting peptides were analysed by nanoliquid chromatography coupled to tandem mass spectrometry (Ultimate 3000 coupled to LTQ-Orbitrap Velos Pro, Thermo Scientific) using a 120-min gradient (2 analytical replicates per biological replicate). RAW files were processed using MaxQuant (13) version 1.5.3.30. The protein content in total, membrane, and secretome proteomes of CLJ1 and CLJ3 were analyzed independently from the others. Spectra were searched against the homemade CLJ database and the frequently observed contaminants database embedded in MaxQuant. Trypsin was chosen as the enzyme and 2 missed cleavages were allowed. Peptide modifications allowed during the search were: carbamidomethylation (C, fixed), acetyl (Protein N-ter, variable) and oxidation (M, variable). Minimum peptide length was set to 7 amino acids. Minimum number of peptides, razor + unique peptides and unique peptides were all set to 1. Maximum false discovery rates - calculated by employing a reverse database strategy - were set to 0.01 at peptide and protein levels. Intensity-based absolute quantification values iBAQ (14) were calculated from MS intensities of unique+razor peptides. Statistical analyses were performed using ProStaR (15). Proteins identified in the reverse and contaminant databases, proteins only identified by site and proteins exhibiting less than

3 iBAQ values in one condition were discarded from the list. After log2 transformation, intensity values were normalized by median centering before missing value imputation (replacing missing values by the 2.5 percentile value of each column); statistical testing was conducted using *limma* t-test. Differentially recovered proteins were sorted out using a log2(fold change) cut-off of 2 and a FDR threshold on remaining p-values of 1% using the Benjamini-Hochberg procedure. The different lists were then combined together. In total, this allows us to end up with a list of 2 852 quantified proteins.

RT-qPCR

Yield, purity and integrity of RNA were evaluated on Nanodrop and by agarose gel migration. Complementary DNA synthesis was carried using 3 µg of RNA with SuperScript III First-Strand Synthesis System (Invitrogen) with or without SuperScript III RT enzyme to assess the absence of genomic DNA. The CFX96 Real-Time system (BioRad) was used to amplify the cDNA and the quantification was based on use of SYBR green fluorescent molecules. cDNA was incubated with 5 µL of Gotaq qPCR master mix (Promega) and reverse and forward specific primers at a final concentration of 125 nM in a total volume of 10 µL. Cycling parameters of the real time PCR were 95°C for 2 min, 40 cycles of 95°C for 15 s and 60°C for 45 s, and finally a melting curve from 65°C to 95°C by increment of 0.5°C for 5 s to assess the specificity of the amplification. To generate standard curves, serial dilutions of cDNA pool of the CLJ strains were used. The experiments were performed with three biological samples for each strain, in duplicate, and the results were analyzed with the CFX manager software (BioRad). The relative expression of mRNAs was calculated using the ΔΔCq method relative to *rpoD* reference Cq values. The graph was represented with the SEM (Standard Error of the Mean).

Statistical analysis was carried out using a non-parametric Mann-Whitney U test. The *p*-values < 0.05 were considered as significant.

Supplementary References

1. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997 Sep 1;25(17):3389-402.

2. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic acids research. 2016 Jan 4;44(D1):D67-72.

3. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome research. 2004 Jul;14(7):1394-403.

4. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 2016;4:e1900v1.

5. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC genomics. 2011 Sep 7;12:444.

6. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010 Jul 1;26(13):1608-15.

153 7. Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, et al. LocTree3
154 prediction of localization. Nucleic acids research. 2014 Jul;42(Web Server issue):W350-
155 5.
156 8. Li L, Stoeckert CJ, Jr., Roos DS. OrthoMCL: identification of ortholog groups for
157 eukaryotic genomes. Genome research. 2003 Sep;13(9):2178-89.
158 9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion
159 for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.
160 10. Elsen S, Huber P, Bouillot S, Coute Y, Fournier P, Dubois Y, et al. A type III
161 secretion negative clinical strain of Pseudomonas aeruginosa employs a two-partner
162 secreted exolysin to induce hemorrhagic pneumonia. Cell host & microbe. 2014 Feb
163 12;15(2):164-76.
164 11. Casabona MG, Robert-Genthon M, Grunwald D, Attree I. Defining Lipoprotein
165 Localisation by Fluorescence Microscopy. Methods in molecular biology. 2017;1615:65-
166 74.
167 12. Casabona MG, Vandenbrouck Y, Attree I, Coute Y. Proteomic characterization of
168 Pseudomonas aeruginosa PAO1 inner membrane. Proteomics. 2013 Aug;13(16):2419-
169 23.
170 13. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized
171 p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature
172 biotechnology. 2008 Dec;26(12):1367-72.
173 14. Swearingen MC, Mehta A, Mehta A, Nistico L, Hill PJ, Falzarano AR, et al. A novel
174 technique using potassium permanganate and reflectance confocal microscopy to image

biofilm extracellular polymeric matrix reveals non-eDNA networks in Pseudomonas aeruginosa biofilms. Pathogens and disease. 2016 Feb;74(1):ftv104.

15. Wieczorek S, Combes F, Lazar C, Giai Gianetto Q, Gatto L, Dorffer A, et al. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. Bioinformatics. 2017 Jan 1;33(1):135-6.

Fig. S1 Details of whole genome comparison between PA7 and CLJ strains.

The outer ring shows all the genes in the strains colored according to their COG (Clusters of Orthologous Groups) functional categories as listed on the bottom, and the three other rings represent the PA7 (orange), CLJ3 (green), and CLJ1 (red) genomes, respectively. White color indicates that the gene is absent from the genome, grey color indicates that the gene or its homolog is present in a different position in the genome, while black color in the CLJ rings represents gaps between contigs. The inner labels show the CLJ-specific regions and the outer labels are the PA7-specific regions (represented in RGPs or PSPA7 locus numbers) described in Additional file 2: Table S1 and Table S2, respectively. (PDF 6909 KB)

Table S1 Genome assembly and annotation statistics (DOCX 17Ko)

Table S2 RT-qPCR primers used in this study (DOCX 16Ko)

196 **Table S3** Specific regions of CLJ compared to PA7 (DOCX 17Ko)

197 **Table S4** Specific regions of PA7 compared to CLJ (DOCX 18Ko)

198 **Table S5** Genes in CLJ-SR14 (DOCX 18Ko)

199 **Table S6** Prediction of CLJ-ISL3 insertions between contigs of CLJ3 (XLSX 14Ko)

200 **Table S7** Differential proteomic analysis between CLJ1 and CLJ3 in total proteome,
201 membrane proteome, and secretome proteome (XLSX 376Ko)

202 **Table S8** Differential gene expression analysis between CLJ1 and CLJ3 (XLSX 436Ko)

203 **Table S9** List of genes/proteins that are statistically significantly differentially expressed
204 between CLJ1 and CLJ3 in both, RNA-Seq and in at least one of the proteomic datasets
205 (XLSX 24K0)

206 **Table S10** CLJ phage-related region (DOCX 16Ko)

207

Table S1. Genome assembly and annotation statistics

	CLJ1	**CLJ3**
Number of contigs/scaffolds	68	135
Average coverage (×)	98.3	96.9
Contig/scaffold size (bp)	201-847201	1025-297138
Average contig/scaffold size (bp)	95801	47065
Contig/scaffold N50 (bp)	456227	97517
Total genome size (bp)	6514448	6353726
G+C content (%)	66.6	66.6
Protein coding sequences	6259	6107
tRNA	62	57
rRNA	9	3

Table S2. RT-qPCR primers used in this study

Name	5'-3' sequence	Amplicon size
qPCR-hcnB-F	TACGGTGATCTGCCGTTGTG	152 bp
qPCR-hcnB-R	GATCGCTGCAATAGCCGATG	
qPCR-pvdE-F	CCAATCCCGAACCCTACCTG	184 bp
qPCR-pvdE-R	TTCTCGCCGACGATGAAGAG	
qPCR-ampDH3-F	GCGACAACCTCAACGACACC	165 bp
qPCR-ampDH3-R	CATTCTTCGGCGTCATGTCC	
qPCR-2958-F2	CGGCATCGAGCACTGCTACT	152 bp
qPCR-2958-R2	CCAGTTCGTGGCCGATGAT	
qPCR-ccoG2-F	GCTGGACCTGGAAAGCCTGT	160 bp
qPCR-ccoG2-R	GGCGTCGTAGGAAACGATCA	
qPCR-ccoN4-F	TGGGCAATACCACCACCAAG	171 bp
qPCR-ccoN4-R	CGGTGGTCAGGATGAACGAG	
oprD-up	GGGTTCATCGAAGACAGCAG	139 bp
oprD-up	TGCCTTGGGTGAAGCCGGATT	
uvrD-up	CATATCCTGGTGGACGAGTTCC	160 bp
uvrD-down	CGCTGAACTGCTGGATGTTCTC	
rpoD-qPCR F1	GCG-CAA-CAG-CAA-TCT-CGT-CT	177 bp
rpoD-qPCR R1	ATC-CGG-GGC-TGT-CTC-GAA-TA	

Table S3. Specific regions of CLJ compared to PA7

Region	RGP	CLJ1 locus	Number of genes	Features
CLJ-SR1	RGP66	0487-0494	8	Phage-related
CLJ-SR2	RGP29*	2128-2214	87	PAGI-2-like island
CLJ-SR3	RGP28	2316-2322	6	Mobile element proteins
CLJ-SR4	RGP27	2415-2573	160	Dit island
CLJ-SR5	RGP72	2579-2583	5	D-galactonate catabolism
CLJ-SR6	RGP26	2615-2621	7	Phage-related; lytic enzymes
CLJ-SR7	RGP23	2917-2924	7	Mobile element proteins; possible DNA helicase
CLJ-SR8		3303-3309	7	ABC transporter ATP-binding protein
CLJ-SR9		3607-3616	9	
CLJ-SR10	RGP15	3774-3784	11	Threonine dehydratase; ABC transporter proteins
CLJ-SR11		4314-4301	14	AlpBCDE-lysis cassette, phage-related
CLJ-SR12		4838-4876	39	Pyrroloquinoline quinone synthesis proteins
CLJ-SR13		5192-5203	10	Phage related; accessory cholera enterotoxin
CLJ-SR14		5811-5871	55	Heavy metal resistance
CLJ-SR15		6156-6161	6	Phage-related

RGP: region of genomic plasticity, *CLJ strains do not share any genes with PA7 in the RGP

Table S4. Specific regions of PA7 compared to CLJ

PSPA7 numbering	Number of genes	RGP	Features
0069-0139	71	RGP63*	Type I restriction-modification system; mercury resistance cluster
0263-0267	5		Sulfate ester transport system
0270-0274	5		
0278-0283	6		*tonB2-exbB1-exbD1* operon
0355-0369	15	RGP64*	Phage-related
0776-0788	12	RGP4*	Phage-related
1586-1593	8		Polymyxin and cationic antimicrobial peptide resistance cluster
1678-1685	8		Sulfur starvation utilization operon
2109-2112	4	RGP70	Probable transposase
2363-2435	73	RGP56*	Phage-related
2515-2526	10	RGP28	
2790-2795	6	RGP25	Hemagglutinins
2886-2908	23		Hcp secretion island-3 encoded type VI secretion system (H3-T6SS)
2911-2917	7		Methionine ABC transporters; monooxygenases
2930-2935	6		Alkanesulfonate assimilation; nitrate and nitrite ammonification
3034-3071	37	RGP23	Pyocin killing protein; phage-related
3695-3747	53	RGP75*	Conjugal transfer protein cluster; resistance genes; transcriptional regulators
4281-4289	9	RGP9	Rieske family iron-sulfur cluster-binding protein
4427-4530	103	RGP7	Type IV B pilus protein cluster
5053-5075	23	RGP78	Phage-related
5143-5160	17	RGP60*	Phage-related
5297-5302	6		Fimbrial chaperone/usher pathway E operon
5324-5357	34	RGP42	Mobile element proteins; Streptomycin phosphotransferase
5364-5378	15	RGP42	Phage-related
5708-5718	11		Arginine:pyruvate transaminase; 2-ketoarginine decarboxylase
6033-6063	31	RGP79*	Type I restriction-modification system

RGP: region of genomic plasticity, *PA7 does not share any genes with CLJ strains in the RGP.

Table S5. Genes in CLJ-SR14

PA7	CLJ1	CLJ3	RAST annotation
NA	CLJ1_5811	CLJ3_5608	MG(2+) CHELATASE FAMILY PROTEIN / ComM-related protein
NA	CLJ1_5814	CLJ3_5609	hypothetical protein
NA	CLJ1_5815	CLJ3_5610	hypothetical protein
NA	CLJ1_5817	CLJ3_0003	hypothetical protein
NA	CLJ1_5818	CLJ3_0004	Error-prone, lesion bypass DNA polymerase V (UmuC)
ns	CLJ1_5819	CLJ3_0005	hypothetical protein
ns	CLJ1_5820	CLJ3_0006	Mercuric ion reductase (EC 1.16.1.1)
ns	CLJ1_5821	CLJ3_0007	Mercuric transport protein, MerC
ns	CLJ1_5822	CLJ3_0008	Periplasmic mercury(+2) binding protein
ns	CLJ1_5823	CLJ3_0009	Mercuric transport protein, MerT
ns	CLJ1_5824	CLJ3_0010	Mercuric resistance operon regulatory protein
ns	CLJ1_5825	CLJ3_0011	hypothetical protein
NA	CLJ1_5826	CLJ3_0012	Sterol desaturase
NA	CLJ1_5827	CLJ3_0013	Transcriptional regulator, AraC family
NA	CLJ1_5828	CLJ3_0014	Lipoprotein signal peptidase (EC 3.4.23.36)
NA	CLJ1_5829	CLJ3_0015	hypothetical protein
NA	CLJ1_5830	CLJ3_0016	Cobalt-zinc-cadmium resistance protein CzcD
NA	CLJ1_5831	CLJ3_0017	COG3267: Type II secretory pathway, component ExeA (predicted ATPase)
NA	CLJ1_5832	CLJ3_0018	FIG131328: Predicted ATP-dependent endonuclease of the OLD family
ns	CLJ1_5833	CLJ3_0019	Mobile element protein
ns	CLJ1_5834	CLJ3_0020	Mobile element protein
NA	CLJ1_5836	CLJ3_6032	transcriptional regulator MvaT, P16 subunit, putative
ns	CLJ1_5837	CLJ3_6031	Gifsy-2 prophage protein
ns	CLJ1_5838	CLJ3_6030	Error-prone repair protein UmuD
ns	CLJ1_5839	CLJ3_6029	Error-prone, lesion bypass DNA polymerase V (UmuC)
NA	CLJ1_5840	CLJ3_6028	putative (L31491) ORF2; putative [Plasmid pTOM9]
NA	CLJ1_5841	CLJ3_6027	putative ORF1 [Plasmid pTOM9]
NA	CLJ1_5842	CLJ3_6026	NreA-like protein
NA	CLJ1_5843	CLJ3_6025	Inner membrane protein
NA	CLJ1_5844	CLJ3_6024	probable membrane protein YPO3302
NA	CLJ1_5845	CLJ3_6023	hypothetical protein
NA	CLJ1_5846	CLJ3_6022	hypothetical protein
NA	CLJ1_5847	CLJ3_6021	Chromate transport protein ChrA
NA	CLJ1_5848	CLJ3_6020	Chromate resistance protein ChrB
NA	CLJ1_5849	CLJ3_6019	Phage integrase family protein
NA	CLJ1_5851	CLJ3_6018	RuBisCO operon transcriptional regulator CbbR
NA	CLJ1_5852	CLJ3_6017	Phosphonate dehydrogenase (EC 1.20.1.1) (NAD-dependent phosphite dehydrogenase)
NA	CLJ1_5853	CLJ3_6016	Phosphonate ABC transporter permease protein phnE (TC 3.A.1.9.1)
NA	CLJ1_5854	CLJ3_6015	Phosphonate ABC transporter phosphate-binding periplasmic component (TC 3.A.1.9.1)
NA	CLJ1_5855	CLJ3_6014	Phosphonate ABC transporter ATP-binding protein (TC 3.A.1.9.1)
ns	CLJ1_5856	CLJ3_6013	FIG002188: hypothetical protein
ns	CLJ1_5857	CLJ3_6012	FIG067310: hypothetical protein
NA	CLJ1_5858	CLJ3_6011	hypothetical protein
ns	CLJ1_5859	CLJ3_6010	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)
NA	CLJ1_5860	gap	Enoyl-CoA hydratase (EC 4.2.1.17)
NA	CLJ1_5861	ns	Mobile element protein
NA	CLJ1_5862	ns	Mobile element protein
NA	CLJ1_5863	gap	transcriptional regulator, TetR family
NA	CLJ1_5864	CLJ3_5612	Error-prone, lesion bypass DNA polymerase V (UmuC)
NA	CLJ1_5866	ns	Mobile element protein

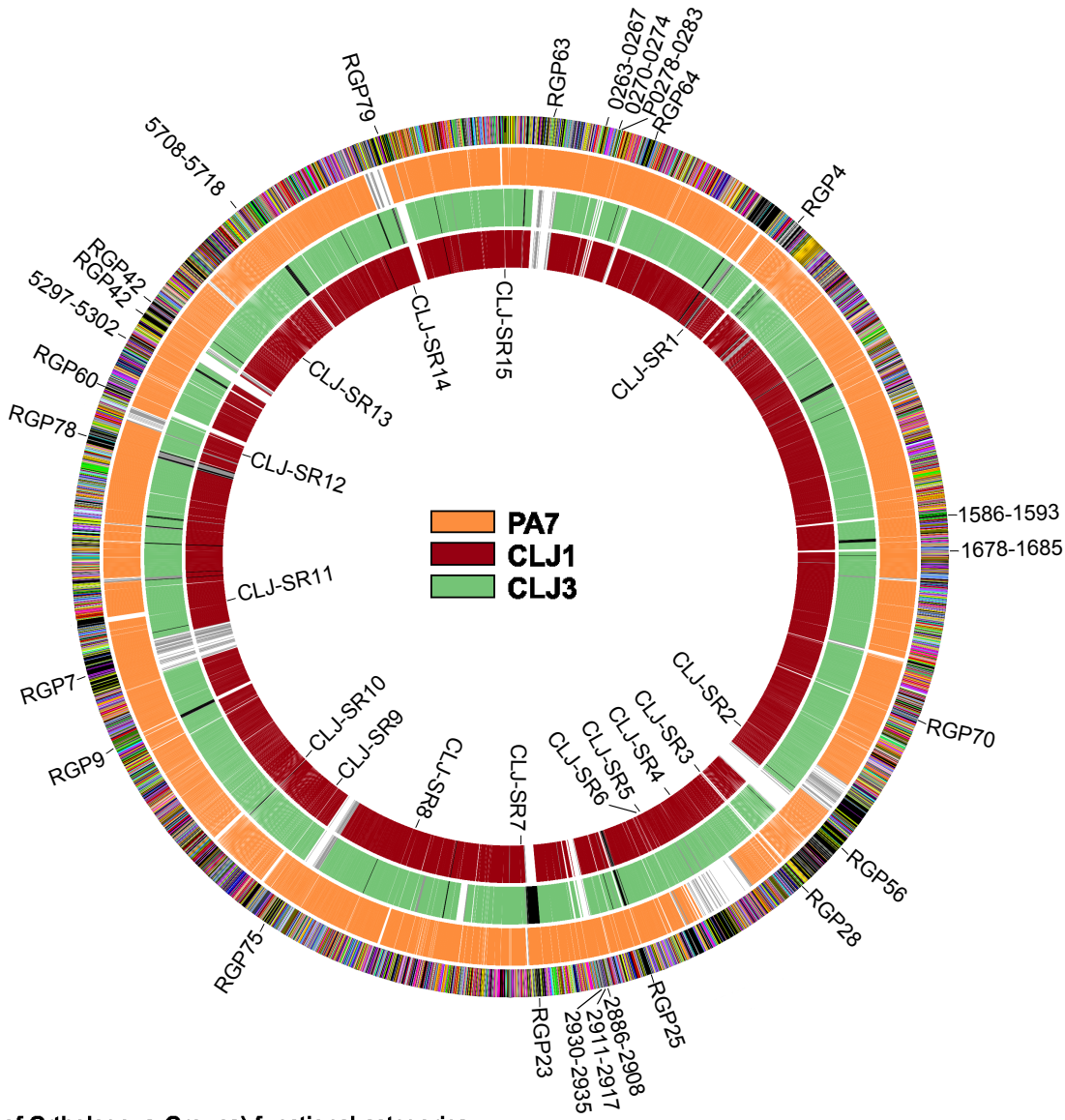
NA	CLJ1_5867	CLJ3_5613	hypothetical protein
NA	CLJ1_5868	CLJ3_5614	Mobile element protein
NA	CLJ1_5869	CLJ3_5615	hypothetical protein
NA	CLJ1_5870	CLJ3_5616	hypothetical protein
NA	CLJ1_5871	CLJ3_5617	hypothetical protein

NA: no orthologous gene is found in the corresponding genome, ns: orthologous gene(s) is/are found in other location(s) in the corresponding genome, gap: the gene location corresponds to a gap between contigs in the corresponding genome.

Table S10. CLJ phage-related regions.

CLJ1 locus	Region	Number of genes	Corresponding PA7 locus	Mobility gene present
0462-0497	RGP66 (including CLJ-SR1)	36	PSPA7_0678-0716	Integrase
0535-0557	RGP3	23	PSPA7_0754-0775	None
2607-2624	RGP26 (including CLJ-SR6)	18	PSPA7_2648-2661	Integrase
4295-4314	Including CLJ-SR11	20	PSPA7_4602-4606	None
4778-4791	RGP78	14	PSPA7_5040-5080	Integrase
5192-5203	CLJ-SR13	12	NA	Integrase
5758-5774	Including RGP6	17	PSPA7_4699-4703	Integrase
6156-6161	CLJ-SR15	6	NA	None

NA: not applicable



COG (Clusters of Orthologous Groups) functional categories

- RNA processing and modification
- Energy production and conversion
- Cell cycle control, cell division, chromosome partitioning
- Amino acid transport and metabolism
- Nucleotide transport and metabolism
- Carbohydrate transport and metabolism
- Coenzyme transport and metabolism
- Lipid transport and metabolism
- Translation, ribosomal structure and biogenesis
- Transcription
- Replication, recombination and repair
- Cell wall/membrane/envelope biogenesis
- Cell motility
- Posttranslational modification, protein turnover, chaperones
- Inorganic ion transport and metabolism
- Secondary metabolites biosynthesis, transport and catabolism
- Signal transduction mechanisms
- Intracellular trafficking, secretion, and vesicular transport
- Defense mechanisms
- Multiple classes
- General function prediction only
- Function unknown
- No COG

Figure S1

Table S6. Prediction of CLJ-ISL3 insertions between contigs of CLJ3, as described in Materials and Methods

Flanking contigs	Inverted repeat at contigs' end	Found by panISa	Truncated or deleted gene(s) CLJ1/PA7/PA01 gene numbers	Gene function/remarks	presence in CLJ1
006-007	Yes	Yes	CLJ1_0230/PSPA7_0328/PA0243	transcriptional regulator, TetR family	No
007-008	Only in 008	Yes	CLJ1_0262/PSPA7_0376/PA0285	GGDEF domain protein	No
019-020	Yes	Yes	CLJ1_0698/PSPA7_0954/PA4136	MFS family transporter	No
022-023	Yes	Yes	CLJ1_0863/PSPA7_1123/PA3985	Putative membrane protein	No
023-024	Only in 023	Yes	hupN/CLJ1_0909/PSPA7_1168/PA3940	DNA-binding protein HU-beta	No
026-063	Only in 026	No	CLJ1_3101/PSPA7_3246/PA2041, mexX/CLJ1_3125/PSPA7_3269/PA2019	mexX, puu permease, multidrug efflux/next to CLJ3 deletion	Yes
030-031	Only in 030	Yes	rhlBRL/CLJ1_1381-1383/PSPA7_1648-1650/PA3478-3476	quorum-sensing regulon	No
033-034	Yes	Yes	CLJ1_1423/PSPA7_1697/PA3429	putative epoxide hydrolase	No
035-036	Yes	Yes	CLJ1_1578/PSPA7_1845/PA3276	hypothetical protein	No
039-040	Yes	Yes	wbpL/CLJ1_1777/PSPA7_1935/PA3193	OSA region	No
045-046	Yes	Yes	CLJ1_2247/PSPA7_2441	hypothetical protein	No
060-061	Yes	Yes	hcnB/CLJ1_2955/PSPA7_3102/PA2194	HCN synthase	No
067-068	Yes	Yes	CLJ1_3400/PSPA7_3548/PA1759	transcriptional activator of maltose regulon, MalT	No
068-069	Yes	Yes	CLJ1_3511/PSPA7_3656/PA1617	putative AMP-binding protein	No
074-131	Yes	No	fgtA/CLJ1_4219/PSPA7_4280/PA1091	flagellum	Yes*
131-075	Yes	No	flgL/CLJ1_4222/PSPA7_4290/PA1087	flagellum	Yes*
075-076	Yes	Yes	oprD/CLJ1_4366/PSPA7_4550/PA0958	antibiotic flux	No
080-081	Only in 080	Yes	CLJ1_5754-5755/PSPA7_4791-4792/PA0732-0731	hypothetical proteins	No
083-084	Yes	Yes	imuC/dnaE2/CLJ1_4582-4585/PSPA7_4841/PA0669	mutagenesis	Yes
093-094	Yes	Yes	ampD/CLJ1_4891/PSPA7_5139/PA4522	AmpD, beta-lactamase expression regulator	No
096-097	Yes	Yes	CLJ1_5005/PSPA7_5275/PA4629	probable transmembrane protein	No
100-101	Yes	No	CLJ1_5197	Hypothetical protein in CLJ-SR13	No
104-105	Yes	Yes	pilM/CLJ1_5446/PSPA7_5781/PA5044	pili	Yes
105-106	Only in 105	Yes	dctQM/CLJ1_5570-5571/PSPA7_5907-5908/PA5168-5169	TRAP-type C4-dicarboxylate transport system small & large permease component	No
107-108	Only in 107	Yes	yjbQ/CLJ1_5806/PSPA7_6028/PA5286, amtB/CLJ1_5808/PSPA7_6029/PA52	hypothetical protein, ammonium transporter	No
077-048	Yes	No	None	in CLJ-SR4, also a gap in CLJ1, between CLJ1_2557&2558	Yes
010-011	Yes	Yes	None	between CLJ1_0366&0368/PSPA7_0480&0481	No
011-012	Yes	Yes	None	between CLJ1_3881&3880/PSPA7_0593&0594	No
042-043	Yes	Yes	None	between CLJ1_2116&2117/PSPA7_2324&2325	No
029-030	Yes	Yes	None	between CLJ1_1258&1259/PSPA7_1518&1519	No
008-009	Yes	Yes	None	between CLJ1_0273&0274/PSPA7_0387&0388	No
037-038	Yes	Yes	None	OSA region/between CLJ1_1764&1765/PSPA7_1971&1972	No
009-010	Yes	Yes	None	between CLJ1_0295&0296/PSPA7_0409&0410	No
058-059	Yes	Yes	None	between CLJ1_1706&1707/PSPA7_2982&2983	No
123-079	Yes	No	None	in RGP6, between CLJ1_5767&5768	No
079-080	Yes	Yes	None	between CLJ1_5718&5720/PSPA7_4757&4759	No
101-102	Yes	Yes	None	between CLJ1_5292&5293/PSPA7_5613&5614	No
041-042	Yes	Yes	None	between *xcp* Z&tRNAVal/CLJ1_1825&rna21/PSPA7_2038&2039/PA3095&3094.3	No
089-090	Yes	Yes	None	between roxS&hypothetical/CLJ1_4819&4820/PSPA7_5108&5109/PA4494&4495	No
097-098	Yes	Yes	None	In place of cupE1-6/between CLJ1_5026&5027/PSPA7_5296&5303/PA4647&4654	No

*** The two ISs probably recombined in CLJ1 creating a deletion between truncated *fgtA* and *flgL* genes**