



HAL
open science

Determination of the two-component systems regulatory network reveals core and accessory regulations across *Pseudomonas aeruginosa* lineages

Julian Trouillon, Lionel Imbert, Anne-Marie Villard, Thierry Vernet, Ina Attrée, Sylvie Elsen

► To cite this version:

Julian Trouillon, Lionel Imbert, Anne-Marie Villard, Thierry Vernet, Ina Attrée, et al.. Determination of the two-component systems regulatory network reveals core and accessory regulations across *Pseudomonas aeruginosa* lineages. *Nucleic Acids Research*, 2021, 49 (20), pp.11476-11490. 10.1093/nar/gkab928 . hal-03463808

HAL Id: hal-03463808

<https://hal.science/hal-03463808v1>

Submitted on 19 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Determination of the two-component systems regulatory network reveals core and accessory regulations across *Pseudomonas aeruginosa* lineages

Julian Trouillon^{1,*}, Lionel Imbert^{2,3}, Anne-Marie Villard², Thierry Vernet², Ina Attrée¹ and Sylvie Elsen^{1,*}

¹Université Grenoble Alpes, CNRS, CEA, IBS UMR 5075, Team Bacterial Pathogenesis and Cellular Responses, 38044 Grenoble, France, ²Université Grenoble Alpes, CNRS, CEA, IBS UMR 5075, 38044 Grenoble, France and ³Université Grenoble Alpes, CNRS, CEA, EMBL, ISBG UAR 3518, 38044 Grenoble, France

Received July 28, 2021; Revised August 24, 2021; Editorial Decision September 23, 2021; Accepted September 28, 2021

ABSTRACT

Pseudomonas aeruginosa possesses one of the most complex bacterial regulatory networks, which largely contributes to its success as a pathogen. However, most of its transcription factors (TFs) are still uncharacterized and the potential intra-species variability in regulatory networks has been mostly ignored so far. Here, we used DAP-seq to map the genome-wide binding sites of all 55 DNA-binding two-component systems (TCSs) response regulators (RRs) across the three major *P. aeruginosa* lineages. The resulting networks encompass about 40% of all genes in each strain and contain numerous new regulatory interactions across most major physiological processes. Strikingly, about half of the detected targets are specific to only one or two strains, revealing a previously unknown large functional diversity of TFs within a single species. Three main mechanisms were found to drive this diversity, including differences in accessory genome content, as exemplified by the strain-specific plasmid in IHMA87 outlier strain which harbors numerous binding sites of conserved chromosomally-encoded RRs. Additionally, most RRs display potential auto-regulation or RR-RR cross-regulation, bringing to light the vast complexity of this network. Overall, we provide the first complete delineation of the TCSs regulatory network in *P. aeruginosa* that will represent an important resource for future studies on this pathogen.

INTRODUCTION

Transcription factors (TFs) are major actors in the regulation of gene expression. From their action on DNA and their interaction with RNA polymerase, other regulatory proteins or signal molecules, results the activation or repression of gene expression, thereby dictating cellular physiology (1). Therefore, the characterization of TFs and of their target genes constitutes a major goal across most fields of biological research. Although binding to DNA does not always imply transcriptional regulation, the function of TFs is generally studied using high-throughput DNA-binding assays such as ChIP-seq. Recently, a new *in vitro* cistromic approach called DAP-seq was developed for the *in vitro* large-scale analysis of TFs binding sites (TFBSs) and notably allowed the characterization of hundreds of TFs in the plant model organism *Arabidopsis thaliana* due to its high scalability when used in combination with cell-free protein expression (2,3). Since then, DAP-seq has been used by us and others to analyze the TFBS landscapes of different bacterial regulators and showed high sensitivity, allowing the delineation of key regulatory features (4–7). Although it is believed that transcription regulatory networks are highly versatile and adaptable (8–10), as shown by *in silico* analyses revealing potential inter- and intra-species TF functional variability (11), the vast majority of TF studies still focus on a single strain to assess a particular TF function.

The major human opportunistic pathogen *Pseudomonas aeruginosa* exhibits a high intrinsic resistance to antibiotics, a large arsenal of virulence factors and a great capacity to adapt to changing environments (12). This latter characteristic relies in particular on a considerable number of TFs which constitute nearly 9% of all proteins (~500) encoded by its genome (13–15). Among the different families of TFs are the response regulators (RRs), which together with their cognate signal-sensing histidine kinases (HKs) form the

*To whom correspondence should be addressed. Tel: +41 44 633 32 45; Email: jtrouillon@ethz.ch
Correspondence may also be addressed to Sylvie Elsen. Email: sylvie.elsen@ibs.fr
Present address: Julian Trouillon, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland.

two-component regulatory systems (TCSs) (16,17). TCSs are numerous in *P. aeruginosa*, about twice as abundant as in the model organism *Escherichia coli*, and those that have been studied were found to be pivotal for orchestrating important cellular processes such as antibiotic resistance or the so-called acute-to-chronic lifestyle switch (18). Recently, several attempts at the global characterization of TCSs relied on phenotypic characterization of loss-of-function mutants (19–22). However, while bringing new information on selected phenotypes, this approach does not globally address direct RR-target regulatory interactions. Among all RRs encoded in the PAO1 genome, 51 are predicted TFs and the binding sites of only eight of them (GacA, AlgR, PhoB, PhoP, CzcR, GltR, DsbR and BfmR) have been determined at a genome scale so far (23–28). Consequently, even though the TCSs regulatory network is thought to be a major driver of *P. aeruginosa* adaptability to different environments, including during infection (29), only a small proportion of the genes directly regulated by RRs is known to date, representing a major knowledge gap in the understanding of this bacterium. This is notably due to the fact that RRs are only active *in vivo* in response to specific signals that are unknown in most cases, making it challenging to assess their role, let alone determining their entire regulons (30). Additionally, some RRs seem to be involved in cross- or co-regulations (23–25), which represent pivotal features of the network topology and thus call for a large-scale investigation of this phenomenon.

P. aeruginosa is a fast-evolving bacterium with a large intraspecies genetic diversity, gifting this pathogen with numerous different ways of establishing infection and resisting treatments. Three major phylogenetic lineages have been identified among *P. aeruginosa* strains (31), each exhibiting many phenotypical specificities. Notably, strains from each lineage possess different major virulence factors, including the recently-discovered ExlBA two-partner secretion system in the PA7 lineage and the Type III Secretion System (T3SS) and different secreted toxins in the two others (32–35). While both *P. aeruginosa* large regulatory network and genetic diversity act as pivotal features for its success as a pathogen, most of *P. aeruginosa* TFs are still uncharacterized and the intraspecies variability in this regulatory network has been mostly ignored so far.

In this study, we leveraged the large scalability and *in vitro* aspect of DAP-seq to investigate the entire TCSs regulatory network across the three major *P. aeruginosa* lineages. To that aim, we report 342 DAP-seq experiments over 55 RRs and three strains, more than doubling the number of TFs with a genome-wide binding profile in *P. aeruginosa*.

MATERIALS AND METHODS

Homolog determination, response regulators identification and phylogenetic analyses

All sequences and annotations were retrieved from the Pseudomonas Genome database v18.1 (15). For inter-strains comparisons, homolog determination was performed by Reciprocal Best Blast Hit (RBBH) analysis (36) on the European Galaxy server (37) using all protein sequences from *P. aeruginosa* PAO1, PA14 and IHMA87 strains with minimum percentage alignment coverages of 90

and sequence identities of 50. Functional classification of predicted RRs with DNA-binding domains were obtained from PseudoCAP (38), the P2TF database (39) and the Pseudomonas Genome database (15), and manually pooled and curated. All candidate RRs were then verified by protein classification using Interpro 73.0 (40) and candidates with confirmed identification of both a RR receiver domain (REC) and a DNA-binding domain were selected, resulting in the final list of 55 RRs across the three strains (Supplementary Table S1). Pfam annotations from the Interpro search were used for DNA-binding domain sequence determination for all RRs, which were then used to generate a multiple alignment using MUSCLE (41). The resulting alignment was used to build a maximum-likelihood phylogenetic tree using MEGA X with 100 bootstraps which was visualized and annotated using iTOL v5 (42,43).

Plasmids and genetic manipulations

Primers are listed in Supplementary Table S3. For production of recombinant proteins, the 55 gene sequences were amplified by PCR using genomic DNA as matrix and appropriate primer pairs and then integrated by Sequence- and Ligation-Independent Cloning (SLIC) (44) in pIVEX2.4d (45) cut with NotI–SalI. PAO1 genomic DNA was used as template for the 51 RRs found in that strain, IHMA87 (3 RRs) and PA14 (1 RR) DNA were used for the 4 remaining ones. All plasmids were transformed into competent TOP10 *E. coli* cells (ThermoFisher Scientific) and then checked by sequencing (Eurofins).

Cell-free protein expression

Cell-free expression was performed for all constructs after magnesium concentration optimization. The 55 RRs were expressed as previously described (46) in a final volume of 100 μ l for each RR during 2 h at 23°C under gentle agitation, with a batch mode configuration. A total of 16 μ g ml⁻¹ of RR template DNA in pIVEX2.4d were added to a reaction mixture containing 1 mM of each of the 20 amino acids, 0.8 mM rNTPs (guanosine-, uridine-, and cytidine-5'-triphosphate ribonucleotides), 1.2 mM adenosine-5'-triphosphate, 55 mM HEPES, pH 7.5, 68 μ M folinic acid, 0.64 mM cyclic adenosine monophosphate, 3.4 mM dithiothreitol (DTT), 27.5 mM ammonium acetate, 2 mM spermidine, 80 mM creatine phosphate, 208 mM potassium glutamate, 14 mM magnesium acetate, 250 μ g ml⁻¹ creatine kinase, 27 μ g ml⁻¹ T7 RNA polymerase, 0.175 μ g ml⁻¹ tRNA and 40% S30 *E. coli* bacterial extract. Protein extracts were then clarified by centrifugation at 14 000 g for 20 min at 10°C and the resulting supernatants were used for DAP-seq.

DAP-seq experimental procedure

Fragmented genomic DNA libraries were prepared exactly as previously described (5) using the purified genomic DNA of *P. aeruginosa* PAO1, PA14 or IHMA87 strains. DAP-seq was performed as previously described (5) with some modifications. Briefly, 100 μ l of cell-free soluble protein extracts (corresponding to 0.5–2 μ g of proteins) were diluted

in 100 μ l of Binding Buffer (sterile PBS supplemented with 10 mM $MgCl_2$, 0.01% Tween 20 and 50 μ M acetyl phosphate) containing 10 μ l of pre-washed (3 times) magnetic cobalt beads (Dynabeads His-Tag Isolation and Pulldown – Invitrogen) in 96-well plates and incubated for 40 min at room temperature on a rotating wheel. The bead–protein complexes were then washed six times in 200 μ l of Binding Buffer, including a transfer to a new 96-well plate before the last wash, and then resuspended in 80 μ l of Binding Buffer containing 50 ng of adaptor-ligated gDNA libraries and further incubated on a rotating wheel at room temperature for 1 h. The bead–protein–DNA complexes were then washed six times in 200 μ l of Binding Buffer, including a transfer to a new 96-well plate before the last wash. Beads were then resuspended in 30 μ l of sterile 10 mM Tris–HCl pH 8.5, and incubated for 10 min at 98°C for elution. After incubation, samples were placed on ice for 5 min, beads were then magnetically removed and the released DNA was used for PCR amplification as previously described (2) using a different indexed pair of primers for each sample. PCR products were pooled for sequencing using 5 μ l of each sample in pools of up to 104 samples. Library pools were then purified using SPRIselect beads at a 1:1 ratio. The quality of each library pool was assessed using High Sensitivity DNA chips on an Agilent Bioanalyzer and additional bead purifications were performed in case of excess amounts of primer dimers. Negative control experiments were done using a pIVEX2.4d vector expressing an untagged GFP protein for each DAP-seq 96-well plate, genome and sequencing pool in duplicates. All DAP-seq experiments were performed in duplicates.

Sequencing & primary data analysis

Sequencing was performed at the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif – <http://www.i2bc.paris-saclay.fr>) using an Illumina NextSeq500 instrument for a total of 4 High Output runs. An average of 2.6 million single-end 75-bp reads per sample were generated with >95% of reads uniquely aligning to the corresponding genome using Bowtie2 (47). Peak calling was done using MACS2 (48) on all uniquely aligned reads with a *P*-value threshold of 0.0001 for each duplicate against a pool of the two corresponding negative control samples. Only peaks found in both replicates were then selected using the Intersect tool from BEDTools (49) with a minimum overlap of 50% on each compared peaks. Peaks that passed the Intersect selection were then filtered by Irreproducible Discovery Rate (IDR) with a IDR threshold of 0.005 between replicates using IDR Galaxy version 2.0.3 on the European Galaxy server (37,50), yielding the final list of reproducible peaks.

For genomic coverage visualization, coverage bedgraph files obtained from MACS2 callpeak used with the –SPMR option were used to generate enrichment tracks using MACS2 bdgcmp in linear scale FE mode for each replicate against the corresponding control samples. Enrichment tracks were averaged between duplicates and used for data visualization.

For DNA motif discovery, the 100-bp central region of reproducible peaks from all three genomes were used for

each RR using MEME-ChIP (51) with default settings. For DNA motif scanning, MEME-ChIP output files were used to find motifs in selected sequences using FIMO (52).

Transcription regulatory network inference

For inference of gene targets from DAP-seq peak locations, since transcriptional units (TUs) and transcriptional start sites (TSSs) experimental annotations are only available for less than a third of all genes in PAO1 and PA14 (53,54) and not at all in IHMA87, we optimized a common *in silico* approach to perform genome-wide definition of promoter regions in all three strains through (i) TUs annotation and (ii) promoter region sizing. The two corresponding parameters—the operon prediction score threshold and the size of upstream promoter regions—were modified over a range of 18 different combinations used for complete target inference over the full dataset and for assessment of TFBS recovery against the entire RegulonDB TFBS database (Supplementary Figure S1) (55). The best-performing parameter pair was then chosen as allowing the highest recovery of known TFBS (>95%) while keeping the number of predicted TUs close to PA14 experimental number (0.994:1 ratio). Consequently, TUs were defined by genome-wide operon prediction for the three PAO1, PA14 and IHMA87 strains using Operon-mapper (56) with a minimum score threshold of 0.9. Promoter regions were then defined as the –400 bp to +20 bp region based on the translational start of the first gene of each transcription unit. Peaks with summit position in promoter regions were then identified using the Intersect tool from BEDTools (49) and assigned to the corresponding gene or operon for generation of the network and inference of regulatory interactions. In case of peaks found in overlapping promoter regions, the peak was assigned to the closest gene start, but additional overlapping regions are still reported in the final peak output files as additional columns on the peak row (Dataset S2), to allow for manual investigation.

ChIP-seq data analysis

Peak lists from ChIP-seq experiments with PhoB, GacA, CzcR, GltR, PhoP, DsbR and BfmR were used to retrieve all peak summit locations (23–28). Peak summit positions were averaged between replicates and used to infer gene targets using the genome-wide promoter regions defined above with the Intersect tool from BEDTools (49) on the corresponding genomes: PAO1 (24–26,28) or PA14 (23,27). For DNA motif determination, peak regions were extended by 50 bp on each side of the summits and retrieved using the GetFastaBed tool from BEDTools (49). The obtained sequences were then used for motif analysis using MEME-ChIP (51) with default settings, with the exception of the GltR, DsbR and PA14 PhoB ChIP-seq datasets for which minimum fold-change thresholds of 10, 5 and 5, respectively, had to be applied to peaks to obtain significantly enriched motifs, as also described in the original articles.

Network and functional enrichment analysis

Network analyses were performed on Cytoscape 3.8 (57), using either GLay clustering (58) or yFiles Hierarchical

clustering (59). Functional annotations were retrieved from the Pseudomonas database (15) and GO functional enrichment analyses were performed using DAVID v6.8 (60).

RESULTS

DAP-seq allows the investigation of the TCSs regulatory network

P. aeruginosa possesses one of the highest numbers of TCSs in bacteria, with 72 predicted RRs in PAO1, including 51 with DNA-binding domains (Figure 1A), which span five different RR families with specific domain architectures (Figure 1B) (29). To investigate the TCSs regulatory network across the species, we selected three strains, the reference strains PAO1 and PA14, and the human urinary tract isolate IHMA879472 (IHMA87), which each represents one of the three main *P. aeruginosa* phylogenetic lineages (31). Across these three strains, there is a total of 55 unique DNA-binding RRs, 48 of which are present in all three strains (Figure 1C; Supplementary Table S1). To investigate the *P. aeruginosa* TCSs regulatory network, we expressed all 55 RRs fused to a N-terminal polyhistidine tag in *E. coli*-derived cell-free extracts and used them for DAP-seq with the fragmented genomes of PAO1, PA14 and IHMA87 in duplicates, resulting in 342 independent DAP-seq experiments, including controls. Since RRs usually need to be phosphorylated to be active, we used acetyl phosphate to increase chances of *in vitro* RR activation as previously done (4,7), which can allow auto-phosphorylation of the RR receiver domain in some cases and thus alleviates the need of knowing the activating signal to identify RRs binding sites. We overall identified binding sites in all three strains for the vast majority ($n = 49$) of the 55 RRs (Supplementary Table S1, Dataset S1). Six RRs did not yield reproducible binding sites on the three genomes (Supplementary Table S2), including RRs from the two smaller subfamilies (Figure 1D), the LytTR- and ActR-like (each composed of only one RR: AlgR and RoxR, respectively), and thus did not seem to be active in our *in vitro* conditions. As expected for TFBSs, reproducible peaks were highly enriched in intergenic regions (Figure 1E). In order to ensure correct inference of gene targets from binding site locations and thus generate the TCSs regulatory network, transcriptional units (TUs) and promoter regions were defined based on TFBS recovery performance in comparison to experimentally obtained transcription start sites (TSSs) in PA14 (54) and known TFBSs relative position from RegulonDB (55) to optimize the sensitivity-specificity trade-off in our results (Supplementary Figure S1, Methods). Peaks allowing target inference were mostly found in intergenic regions and, to a lesser extent, at the 3'-end of TUs (Figure 1F), matching the usual position of promoters (54,55). Furthermore, peaks found within putative promoter regions with a previously defined TSS in PA14 (54) were centrally enriched at the core promoter (in the first 50–100 bp upstream of TSS), as expected for TFBSs (Figure 1G).

Overall, this approach identified an average of 1511 target TUs carrying a RR binding site in their promoter region per strain, encompassing 2263–2549 genes, or about 40% of all genes in each strain (Figure 1J, Dataset S2). To facilitate data exploration, we also provide a single table where all

interactions are summarized and searchable by gene name (Supplementary Table S2); further details such as fold enrichment and exact peak location can then be found in each RR DAP-seq results file (Dataset S1, Dataset S2). Interestingly, only 646 TUs, or 41–46% of all target TUs in each strain, exhibit a RR binding site in their promoter region in all three strains (Figure 1H), suggesting a high intra-species RR functional variability. As expected, PAO1 and PA14 show the highest overlap in identified target TUs. We also found large differences in number of targets between RRs, varying from seemingly very specialized RRs showing a very small number of targets (<10) to more global regulators with up to 285 inferred targets on average (Figure 1I). Additionally, 40–45% of targets were found to have more than one RR binding site on their promoter region (Figure 1K), suggesting a high proportion of co-regulation in the TCSs network. The analysis of the enriched DNA regions allowed the determination of DNA binding motifs for most RRs (Figure 2A). The motifs often showed similarities between RRs with phylogenetically close DNA-binding domains, as seen for several groups of RRs (Figure 2B). Surprisingly, some RRs with seemingly different DNA-binding domains also showed very high similarity in DNA binding motifs, such as KdpE and CpxR (Figure 2C). Overall, DAP-seq allowed the near-complete determination of the TCSs regulatory network across the three tested strains and revealed several interesting key features of the network, as detailed below.

To confront our datasets with the existing knowledge on *P. aeruginosa* RRs, we first compared it to known high-confidence targets, mostly from EMSA evidences. The comparison to this set extracted from 22 previous studies and composed of 59 known binding sites concerning 21 RRs showed that the vast majority were effectively retrieved, often resulting in very highly enriched peaks (Supplementary Figure S2), further validating our approach. We then compared our results to the previously reported ChIP-seq studies on RRs for which we detected binding sites (PhoB, GacA, GltR, PhoP, CzcR, DsbR and BfmR) (23–28). Here again, there was a strong overlap in detected binding sites between the DAP-seq data and the six different ChIP-seq studies, as well as highly similar identified DNA motifs (Supplementary Figure S3). The only exception was GacA; while our results revealed several new putative binding sites in addition to its two universally-recognized targets, *rsmY* and *rsmZ* (61), only three targets overlapped with the previous ChIP-seq experiment, which however did not identify *rsmZ* as a GacA target (Supplementary Figure S3C). Globally, our DAP-seq results confirmed known targets and expanded the landscape of known RRs binding sites.

The core TCSs regulatory network

The comparison of DAP-seq results between strains allowed the determination of the core TCSs regulatory network for the 48 conserved RRs, composed of 634 conserved RR-target interactions that were found in all three strains (Supplementary Table S2). The clustering analysis of this core network resulted in the identification of 12 regulatory modules, each containing 1 to 7 RRs (Figure 3A). Notably, two larger modules including five (RocA1, RocA2,

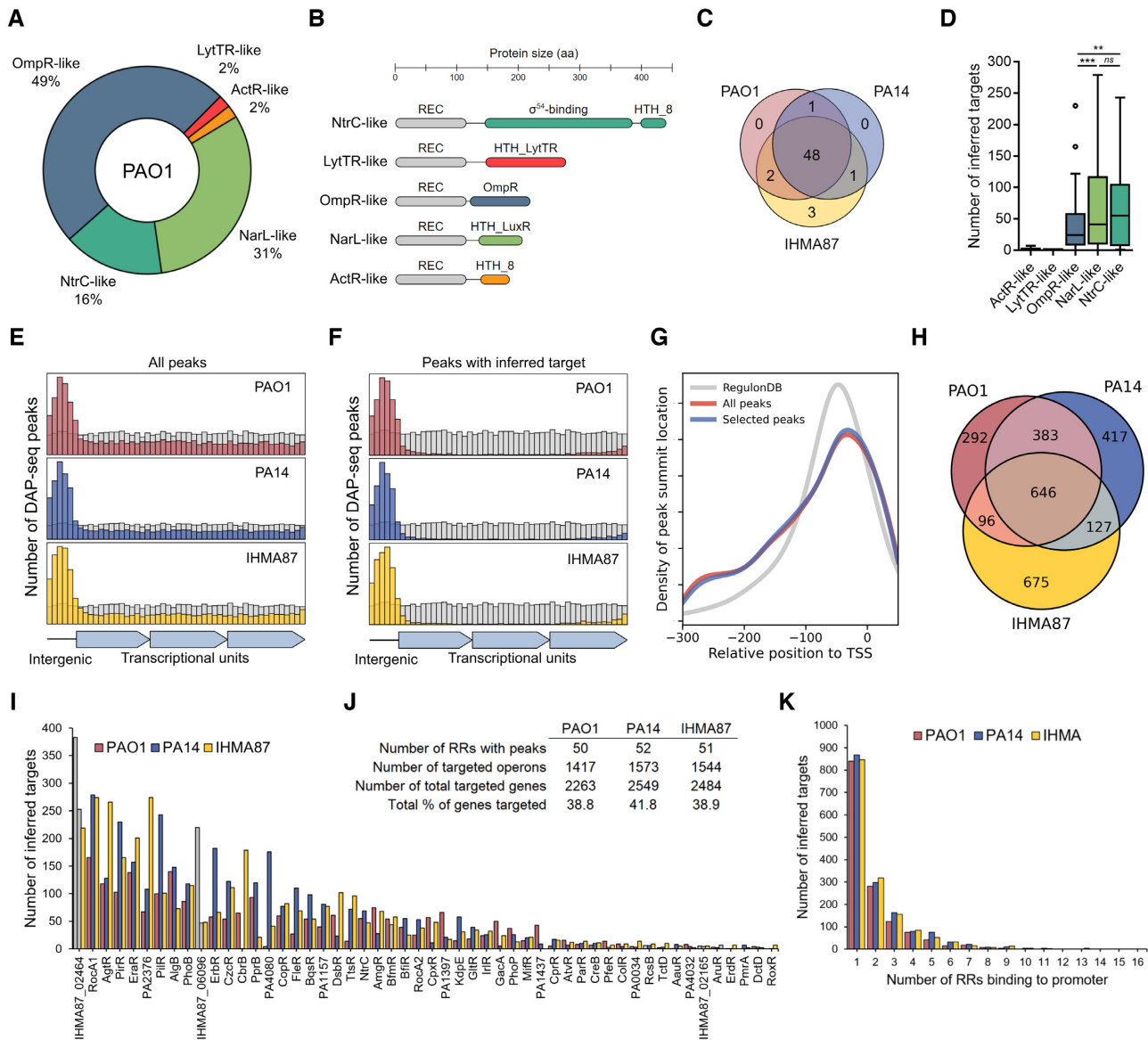


Figure 1. Overview of DAP-seq results. (A) Proportion of the five different RR subfamilies with a DNA-binding domain in *P. aeruginosa* PAO1. (B) Corresponding schematic representations of protein domain organizations. (C) Repartition of all DNA-binding RRs across the three strains PAO1, PA14 and IHMA87. (D) Number of inferred targets per RR families. Statistical significance was assessed using two-tailed *t* test (*P*-value < 0.01 [**] or 0.001 [***], *ns*: not significant). (E, F) Location of DAP-seq peak summits for all reproducible peaks (E) and all peaks that allowed target inference (F). Summits location in intergenic or inside of transcriptional units was assessed and rescaled to the genome's intergenic/intragenic ratio (10.66–10.95% intergenic). Grey bins represent random distribution. (G) Density plot of peak summit location relative to known TSSs in PA14. The density of TFBSs location from RegulonDB is used as reference. (H) Repartition of RR inferred targets between strains. (I) Number of inferred targets for each RR with detected reproducible peaks. Grey bars represent non-physiological combinations (i.e. BfmR DAP-seq on the PA14 genome while PA14 does not possess *bfmR*). (J) Summary table of DAP-seq results. (K) Number of RR binding sites detected per target promoter.

AgtR, TrsR and PA0034) and seven RRs (PirR, AmgR, KdpE, CpxR, PA1157 and GacA), showed high levels of gene target co-regulations across numerous biological processes. This analysis revealed both known (i.e. RocA1 and RocA2) and new (i.e. AmgR, PA1157, KdpE, CpxR and PirR) groups of RRs co-regulating common target genes. Some modules were composed of RRs with similar DNA motifs (Figure 2B, C), such as with RocA1, RocA2, AgtR and PA0034. These similarities in DNA motifs probably explain the larger proportion of shared targets between these TFs as they might share binding sites on the corresponding

promoters. Numerous gene targets encoding key virulence factors were found to be part of the core network. For example, key operons encoding bacterial motility appendages such as the type IV pili and flagellum emerge in several clusters (Supplementary Figure S4A, B). Moreover, several key regulatory genes were also targeted, including the two non-coding RNAs RsmY and RsmZ as well as their cognate RNA-binding protein RsmA (Figure 3A), all three being responsible for the regulatory switch between motile and sessile lifestyles (12), and thus representing a pivotal node in *P. aeruginosa* regulatory network. Strikingly, six RRs were

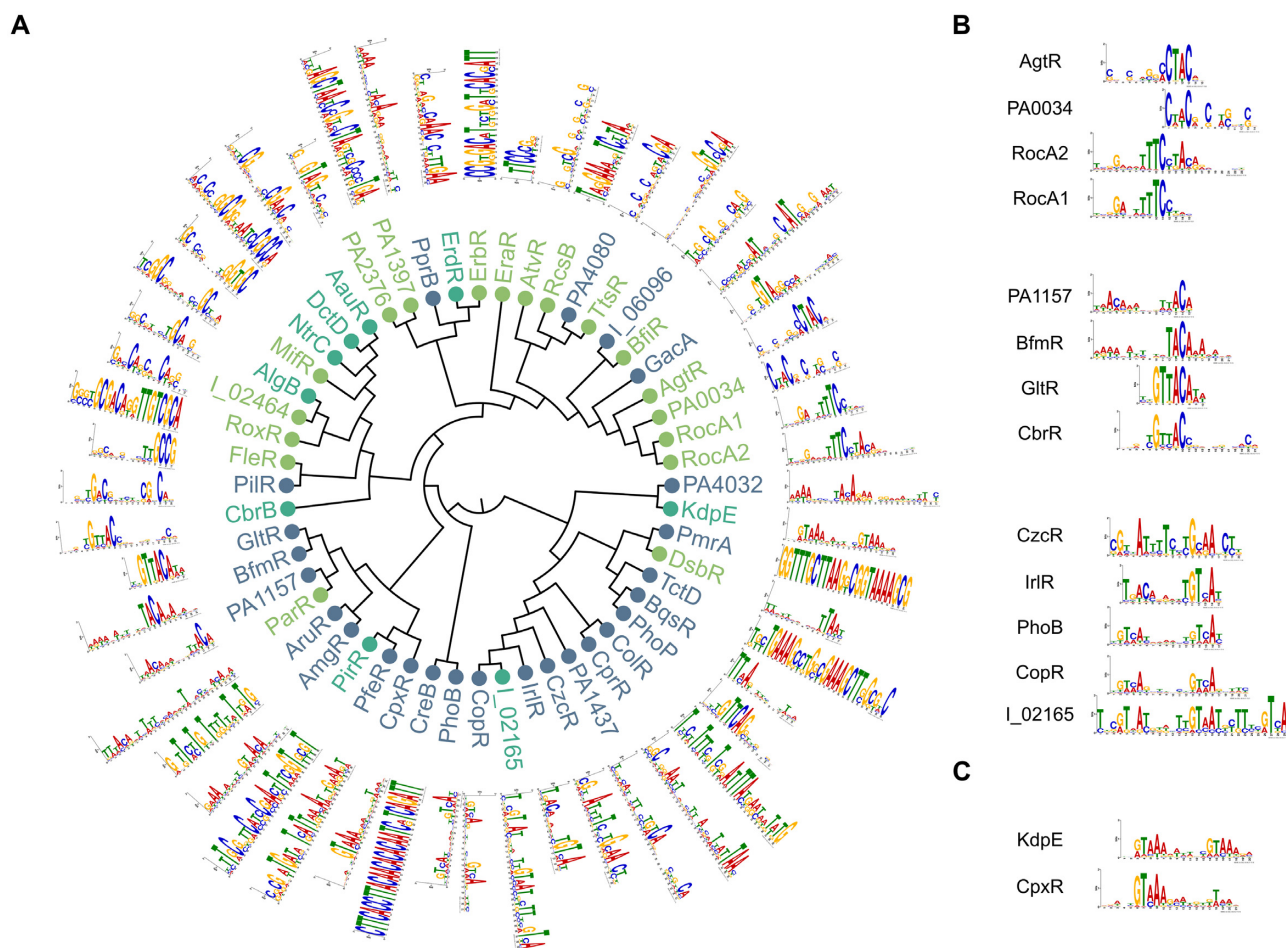


Figure 2. The global view of RRs DNA binding motifs. (A) Maximum-likelihood phylogenetic tree of 52 RR DNA-binding domains. The different OmpR, NarL and NtrC subfamilies are highlighted in blue, green and cyan, respectively. The DNA-binding motif found in peak regions using MEME-ChIP is shown for each RR. RR names with the 'IHMA87_xx' format were shortened to 'Lxx'. (B) Three groups of RRs that are neighbors in the tree and display similar DNA-binding motifs. (C) Two RRs with phylogenetically distant DNA-binding domains but sharing similar DNA-binding motifs.

found to bind to at least one promoter of these three genes in all strains (Supplementary Figure S4C), including the known regulator of RsmY and RsmZ, GacA.

The functional enrichment analysis of the core network pointed out to two main overrepresented molecular functions among target genes: 'porin activity' and 'DNA-binding' (Figure 3B). Indeed, conserved RR binding sites were found in the promoter of 10 out of the 27 genes annotated with porin activity (Figure 3C). Two of these genes, encoding the OprO and OprP porins, were found with the highest numbers of RRs binding to their promoter regions in all three strains (3 and 5, respectively), suggesting the existence of an important regulatory node on these two neighboring genes (Figure 3D). While it was shown that PhoB regulates both genes (23), our results show that four additional RRs seem to be involved in the regulation of one or two of these porins. These include CzcR and CopR which share similar DNA-binding motifs with PhoB (Figure 2B), suggesting that they either compete for binding at these promoters or act under different conditions. Interestingly, OprO and OprP participate in phosphate uptake in *P. aeruginosa* (62) and while TCSs rely on phosphorylation

for signal transduction and RRs activation, this result suggests the existence of several regulatory feedback loops from RRs onto phosphate uptake, which are probably important for correct TCS regulatory response (63).

The second and third most enriched molecular functions in target genes of the core TCSs regulatory network were DNA-binding and transcription factor activity, respectively (Figure 3B). Indeed, 82 TF-encoding genes, including 17 RRs and five sigma factors, are conserved RR targets (Figure 3E). Numerous major TFs were found targeted by several RRs, including the two quorum-sensing regulators, RhlR and LasR (Supplementary Figure S4D, E), and the virulence regulator MvfR (PqsR), which was the TF with the highest number of conserved RR binding sites in its promoter (Figure 3F). The manual investigation of RR binding sites location on *mvfR* promoter suggested two activating (KdpE and BqsR) and two repressing (PhoB and AmgR) interactions (Figure 3G) and illustrates the ability of DAP-seq to precisely delineate protein–DNA interactions. Another example was the binding of PilR to the promoter of the *erfA* gene, encoding the repressor of the ExlBA two-partner secretion system in the IHMA87/PA7-like lineage

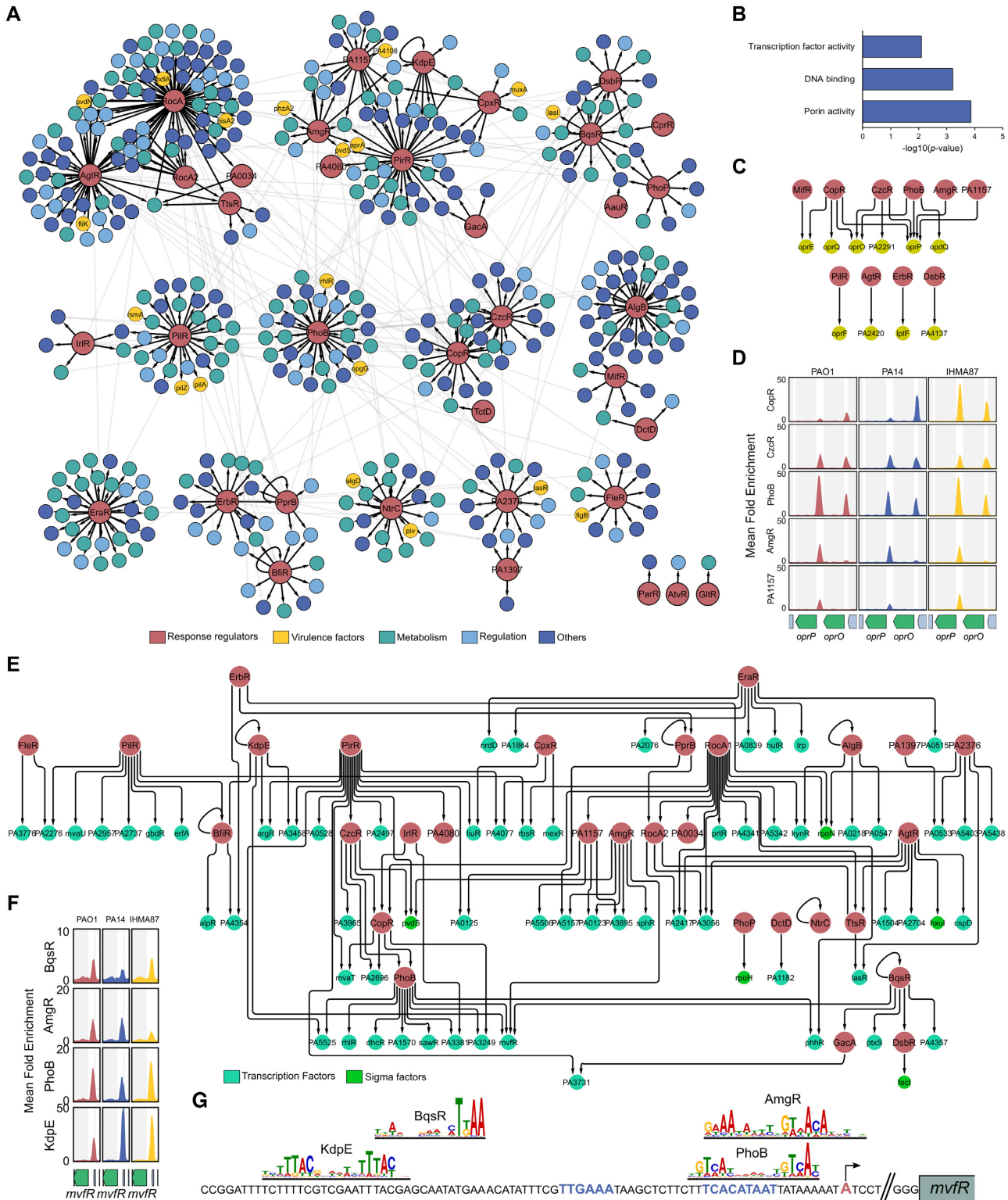


Figure 3. The core TCSs regulatory network. (A) Graph diagram of the core TCSs regulatory network. Cluster identification was performed using Glay (58). Each cluster is shown as an independent module of the graph with black edges. Inter-cluster interactions are shown as light grey edges. RRs are colored in red and target TUs are colored depending on groups of COG predicted functions. The ‘Virulence factor’ category comprises genes annotated as such in the Pseudomonas Database (15). For operons, the name of the first gene is given. (B) Molecular Function GO Term enrichment analysis of the target genes of the core TCSs regulatory network. All GO terms with P -value < 0.01 are shown. (C) Graph diagram of all core interactions involving target genes annotated with the GO Term ‘Porin activity’. (D) Enrichment coverage tracks of DAP-seq against negative controls are shown for the five RRs with binding sites on the promoters of *oprP* and/or *oprO* in all three genomes. (E) Graph diagram of all core interactions involving target genes with a DNA-binding domain (15). Each target node represents a TF-encoding gene found in target TUs from the core regulatory network. (F) Enrichment coverage tracks of DAP-seq in *mvfR* region. (G) Schematic view of *mvfR* promoter. RR motifs are shown at the location where they were identified by MEME. The TSS position experimentally determined in PA14 is shown in red and by a black arrow, predicted -10 and -35 boxes are in blue.

(Supplementary Figure S4F) (5). Additionally, the *exlBA* operon itself was found to be a target of PilR in IHMA87 (Dataset S2). Interestingly, PilR is the known activator of type IV pili (64) that promote ExlBA-dependent cytotoxicity through host cell-bacterial contact (65). These two co-operating virulence factors might thus be co-regulated by PilR, revealing a new function for this RR in the *exlBA*⁺ IHMA87/PA7-like lineage.

Overall, these results demonstrate that the core TCSs regulatory network seems to be highly enriched in regulatory functions. With both feed forward loops for signal amplification and a high number of regulated TFs, it appears that the core regulatory topology of the network is conserved and place RRs relatively high in the hierarchy of *P. aeruginosa* regulatory network. Such a conserved and robust set of interactions between TFs co- and cross-regulating each other might allow for overall conservation of network topology and stable grafting of new strain- or lineage-specific functions in the accessory network, as explored below.

Response regulators show functional variability between strains of *P. aeruginosa*

The pool of genes with a RR binding site on their promoter regions seems to vary between the three tested strains (Figure 1H). When looking at the 48 RRs present in all three strains, there is a difference between the global pool of genes targeted by at least one RR, even if not the same, in all three strains (Figure 4A) and the pool of conserved RR-target interactions (Figure 4B). This means that some target genes ($n = 114$) harbor RR binding sites in their promoter in all three strains, but from different RRs. One such example is the *fleQ* gene, encoding the major c-di-GMP-sensing biofilm regulator (66). PirR was found to bind to *fleQ* promoter in PAO1 and PA14 while RocA1 and RocA2 bound to this promoter in IHMA87 (Figure 4C). This difference could be explained by the absence or presence of the respective RR DNA-binding motifs in the *fleQ* promoter across the three strains (Figure 4D), hinting at a potential large diversity of regulation for genes with divergent promoters in our dataset. In the light of these observed differences between strains, we investigated the conservation of binding events and found various levels of conservation across all RRs and strains (Figure 4E). Interestingly, about 50% and 70% of genes targeted by RRs in only one or two strains, respectively, are conserved in more strains than that (Figure 4F), revealing that the absence/presence of genes explains only partially the regulatory differences between strains, while others are probably due to differences at the promoter level, as seen for *fleQ* (Figure 4C, D). This result highlights the necessity of conducting comparative approaches, even between closely related strains.

Another source of regulatory differences between strains is the difference of RRs content itself. In the three strains tested here, seven RRs are present in only one or two strains (Figure 1C). As seen for PFeR, which is absent in PA14 but present in PAO1 and IHMA87, where it binds to the *pfeA* promoter (Dataset S2). Additionally, AlgB was found to bind to the *pfeA* promoter specifically in PA14, potentially representing a backup mechanism for the lost regulatory in-

teraction (Dataset S2). Other such examples are the three IHMA87-specific RRs, especially IHMA87_02464 which harbors one of the highest number of inferred targets, with 219 in IHMA87 (Figure 1I). Among these, about 70% are genes also present in PAO1 and PA14, and IHMA87_02464 was found to be able to bind to the majority of them on the genomes of these two strains in our DAP-seq experiments (Dataset S2). Consequently, if this TF were to be acquired in one of these two strains, as common between bacteria through horizontal gene transfer (10), it would already have numerous binding sites and regulatory targets.

A third mechanism of regulatory diversification stems from differences in target genes conservation. We found that 50% and 30% of genes targeted in one or two strains, respectively, were part of the accessory genome and absent from the genomes of the other strains (Figure 4F). Nearly half of these cases were found in IHMA87 (Figure 4A), reflecting its higher phylogenetic distance compared to PAO1 and PA14. While *P. aeruginosa* is a highly diverse species with a core genome representing only 1.2% of all genes in the pangenome (33), our results highlight how far we still are from apprehending TFs functions across bacterial species. One important difference often stressed between the three lineages is the mutual exclusion between the T3SS and ExlBA secretion systems, defining the pathogenic strategy of the given strain (35). Strikingly, we found 13 RRs binding to the *exlBA* promoter (Figure 4G), representing one of the most targeted promoters in this strain. While four of these binding peaks did not encompass the corresponding RR DNA motif (Figure 4H), nine of them did, revealing this promoter as a very highly connected regulatory node, especially since two additional TFs are already known to bind to it and to regulate *exlBA* (5,67). Furthermore, the genes encoding various T3SS components, notably the T3SS master regulator ExsA, were targeted by numerous RRs (Figure 4I-L). Four RRs were found to bind to the same locations in the major T3SS gene cluster in both PAO1 and PA14 (Figure 4I), with additional RRs binding there in PA14 (Figure 4J). Additionally, the two genes encoding the strain-specific toxins, which are found at different loci, *exoS* in PAO1 and *exoU* in PA14, also showed RR binding events on their promoters (Figure 4K, L). On its own, this T3SS-related example across these two strains shows all three types of regulatory differences, with differences in RR content (Figure 4L), target gene content (Figure 4K, L) and DNA-RR affinity (Figure 4J). Overall, in accordance with the recent predictions that the T3SS-related genes might be targeted by as many as 26 TFs in PAO1 (68), we see that the regulation of these two major virulence factors seems highly connected, involving numerous TFs to allow for a precise and tightly controlled expression.

The IHMA87 plasmid represents a hub of regulatory network integration events

One notable feature of the IHMA87 strain is that it carries a 185 kb plasmid specific to that strain (5), named pIHMA87, which represents a great opportunity to explore regulatory plasticity and regulatory network grafting. Indeed, the acquisition of this plasmid by IHMA87, bringing 200 new genes, potentially required the wiring of some of these genes

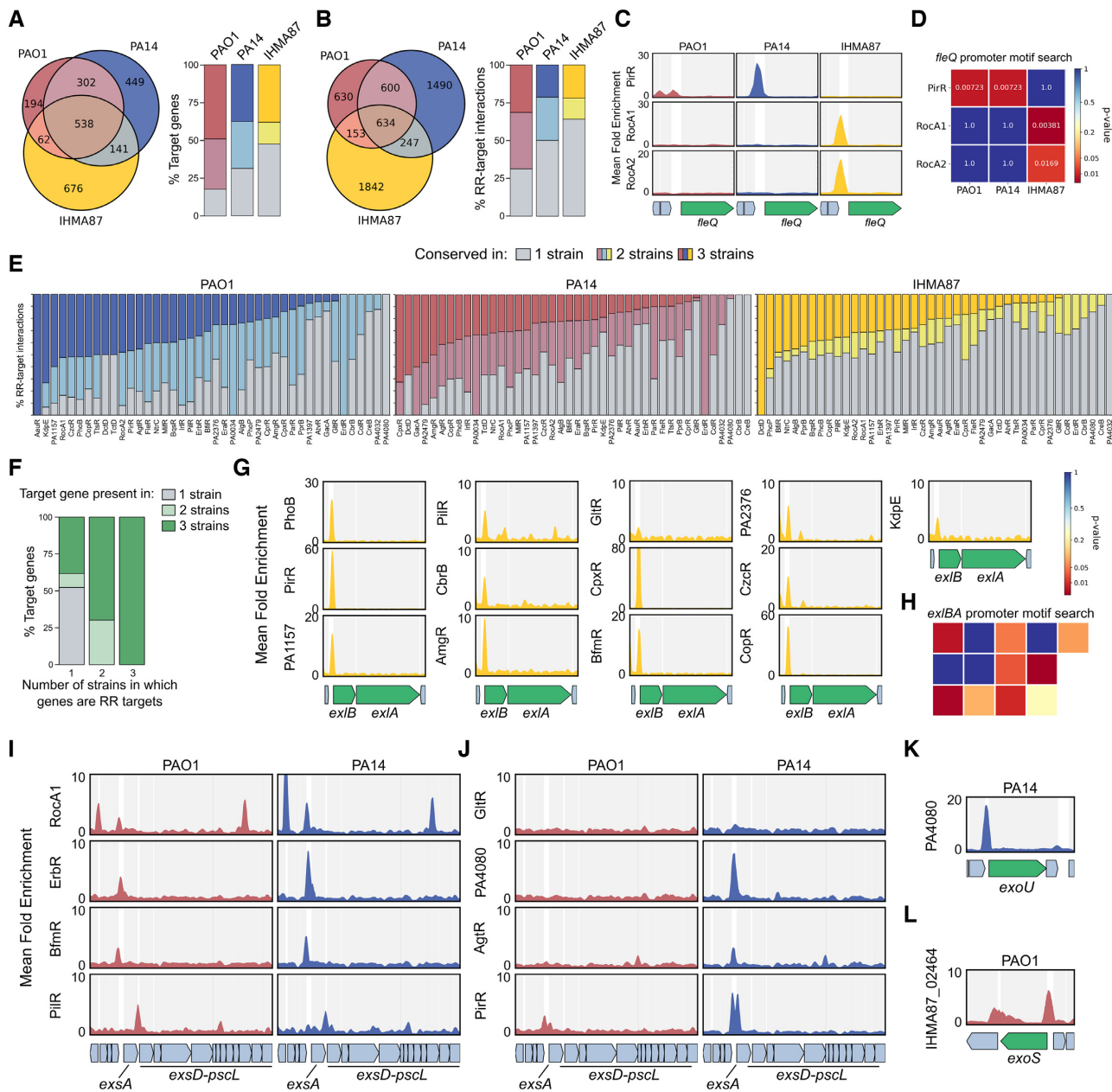


Figure 4. The accessory TCSs regulatory network. (A, B) Repartition of the total pool of gene targets (A) or RR-target interactions (B) between strains. (C) Enrichment coverage tracks of DAP-seq against negative controls for the three RRs with binding sites on the promoter of *fleQ* in all three genomes. (D) DNA-binding motif presence for the three RRs shown in (C) in *fleQ* promoter region in all three genomes, as shown as a heatmap of *P*-value obtained from motif searches using FIMO (52). (E) Proportion of RR-target interactions detected in one, two or three genomes for all genomes and conserved RRs with detected targets. RRs are sorted by number of targets found in three genomes. (F) Conservation of genes presence across the three strains for RR target genes, depending on the number of strains in which they are targeted. (G) Enrichment coverage tracks of DAP-seq against negative controls for the 13 RRs with binding sites on the promoter of *exlBA* in IHMA87. (H) DNA-binding motif presence for the 13 RRs shown in (G) in *exlBA* promoter region in IHMA87, as shown as a heatmap of *P*-value obtained from motif searches using FIMO (52). The RR order and repartition in the heatmap follow the organization of panel G. (I–L) Enrichment coverage tracks of DAP-seq against negative controls for the RRs with binding sites on T3SS-related genes in PAO1 and PA14, for conserved (I) or PA14-specific (J) T3SS-related RR-target interactions for shared genes, or for strain-specific genes in PA14 (K) or PAO1 (L).

to the pre-existing IHMA87 regulatory network, as often predicted for horizontally acquired genes (10). Strikingly, we found RR binding sites on the promoter regions of 73 of the pIHMA87 genes, involving 36 RRs (Figure 5A, B), revealing a high level of regulatory grafting to the regulatory network. Some RRs notably exhibited numerous binding sites encompassing their DNA motif on pIHMA87 (Figure 5C), showing the potential immediate addition of the corresponding genes to their regulons due to the presence of their binding sites on the acquired plasmid. While the vast majority of pIHMA87 genes are completely uncharacterized and often code for proteins with poor functional annotations, it is interesting to see that the target gene with the highest number (9) of RR binding sites in its promoter is *IHMA87_06166*, which encodes a predicted type IV secretion system conjugation protein (Figure 5D). One of the main functions of this type IV secretion system is to mediate the transfer of plasmid DNA by conjugation (69), suggesting that this key function in plasmid conjugation might have to be tightly regulated for the acquisition and conservation of the pIHMA87 plasmid.

The response regulators network redundancy and complexity

The present analysis revealed two interesting features of the TCSs regulatory network: (i) highly connected nodes representing target genes with numerous RR binding events on their promoter and (ii) a high proportion of RR-RR cross-regulation. As exemplified for some genes above, we found that numerous target genes exhibited a high number of RR binding sites in their promoter, with 52 genes harboring >5 RR binding sites on average across the three strains (Supplementary Table S2), going as high as 16 different RRs binding to one promoter.

Among the most targeted genes are uncharacterized genes (*PA0123* and *PA3601*) but mostly well-characterized genes involved in bacterial survival and virulence (i.e. *flp*, encoding the type IVb pilin, the *nuoA-N* operon involved in the electron transport chain, *zipA*, encoding the ZipA cell division protein or the non-coding RNA RsmY) (Supplementary Figure S5A). The gene encoding the alkaline protease AprA, a key secreted virulence factor important for immune evasion (70), stood out with the highest average number (13) of RRs binding sites (Supplementary Figure S5B). The second and third most targeted genes code for the uncharacterized transcription factor PA0123 and the Xcp type II secretion system, involved in the secretion of several exoproteins including the ToxA and LasB toxins (71). All three corresponding promoters were found with >10 RR binding events on average, representing the most connected target nodes of the network. While PA0123 is considered uncharacterized, it was found to bind to the promoter of *lasR* *in vitro* and its overexpression induces strong cell growth inhibition (72), suggesting a potential important role for this TF and calling for further characterization of its precise role. Overall, it appears that some genes requiring tight regulation exhibit complex promoter organization, allowing the potential coordination of an unusually high number of TFs acting on a single transcriptional unit.

One of the key observed features of the RR network was the high number of regulated TFs (Figure 3E), and partic-

ularly RRs. Indeed, numerous RRs were found to bind to the promoters of RR-encoding genes/operons, representing the RR cross-regulatory network (Supplementary Table S2, Figure 6A). In total, we identified 149 such RR-RR regulatory interactions, including 26 that were found in all three strains (Figure 6B). Surprisingly, the proportion of RR-RR interactions conserved in all three strains was lower than the average of all inferred interactions (Figure 4), suggesting a more strain-specific need for these cross-regulation events. There was notably a high number of inferred auto-regulations, confirming few examples underlined in previous studies, such as for BqsR, PprB and PhoB (23,73,74). Overall, some RRs seem to be more prone to either bind to many RR promoters (Figure 6C) or to have many RR binding sites on their promoter (Figure 6D). Among these, RocA1 targeted 12 RRs, including five in all three strains, placing this TF at the top of the TCS regulatory network hierarchy (Figure 6E). On the other hand, CzcR showed the highest number of RRs binding to its promoter region. Additionally, the contiguous *czcCBA* operon is in the top 10 most targeted transcriptional units globally (Supplementary Table S2, Supplementary Figure S5A), making the 512-bp *czcC-czcR* intergenic region the most targeted region, with a total of 22 RRs binding to it in at least one strain, including nine conserved in all three strains (Figure 6F). As for many such cases, while these regions are very similar between PA01 and PA14, they share only ~79% identity with the one in IHMA87, from which stems the vast majority of the non-conserved binding events. The *czcCBA* operon codes for an efflux pump responsible for the regulation of heavy metals concentration, which are the activating signal of the CzcRS TCS (75). Consequently, all RRs binding to this intergenic region could potentially regulate CzcR activity directly or indirectly, making CzcR a key downstream node of the RR network. Overall, these results highlight the high complexity of the TCS network, which contains numerous RR cross-regulation allowing for potentiation or repression of signal transduction and thus fine-tuning of gene expression.

DISCUSSION

Global health issues caused by bacterial pathogens are consistently increasing in scale and are now reaching alarming levels, mainly due to the development of antibiotic resistance (76) and host adaptation (77). Both mechanisms rely on acquisition or adaptation of regulatory functions, which mostly involve transcription factors. However, knowledge on bacterial TFs is still sparse; most of them are still uncharacterized even in the most studied organisms, and intra-species functional variability is hardly ever addressed, as the vast majority of studies are performed on a single well-established laboratory strain supposed to be representing its entire species. As a starting point to address the importance of this issue and to make our contribution to the needed characterization of *P. aeruginosa* TFs, we used DAP-seq to identify the binding sites of all RRs in three strains representing the three major *P. aeruginosa* lineages.

The main limitation when studying TCSs *in vivo* is the need of knowing the activating signal. As most signals are still unknown, this has considerably limited the ability of

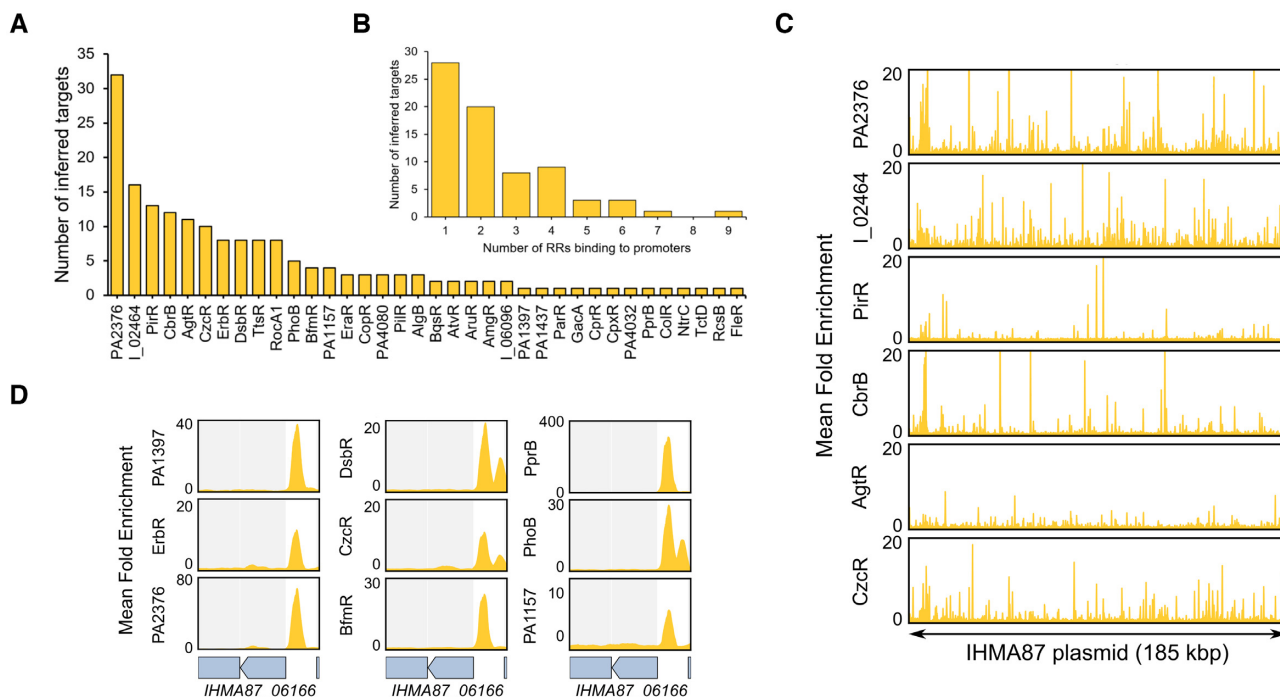


Figure 5. The integration of the pIHMA87 plasmid into the TCSs regulatory network. (A) Number of inferred targets on pIHMA87 per RR shown for all RRs with at least one plasmid binding site. RR names with the ‘IHMA87_xx’ format were shortened to ‘I.xx’. (B) Number of RR binding sites detected per target promoter on pIHMA87. (C) Enrichment coverage tracks of DAP-seq against negative controls for the six RRs with >10 binding sites in promoter regions on pIHMA87. (D) Enrichment coverage tracks of DAP-seq against negative controls for the nine RRs with a binding site on the promoter of IHMA87_06166 on pIHMA87.

previous studies to characterize RRs globally (30). In this context, the *in vitro* aspect of DAP-seq and use of phosphoryl group donor allowed us to circumvent this limitation and made the family-wide RR binding site investigation possible. While DAP-seq represents a powerful tool for the study of TFs, notably due to its higher scalability than ChIP-seq, it also comes with limitations. Indeed, as an *in vitro* assay working with individually purified proteins, DAP-seq might not fully picture the effect of TFs interacting with other molecules in the cell, such as TF–TF interactions, as previously seen for PhoB and TctD (23). More targeted work will be required to evaluate such interactions. A recent adaptation of the DAP-seq protocol was developed for the study of TF–TF interactions and their impact on DNA binding (78) and could notably allow such investigation. Additionally, in a similar way to what is observed for ChIP-seq using overexpressed TFs, some detected binding events might not completely reflect physiological interactions. For instance, while being present in all three strains, TtsR is an orphan RR in PAO1 and PA14, as its cognate HK, TtsS, is only present in the PA7/IHMA87 lineage (79). Consequently, it is unclear whether TtsR can be activated in PAO1 and PA14, potentially by other HKs, and thus bind to DNA in these strains. Finally, it is important to note that part of the observed differences might arise from technical variability, notably due to the intrinsic decreased accuracy of high-throughput sequencing and large scale approaches. One such example is the CbrB experiment on PA14 genome that seemed to have an unexpectedly low true positive recovery (Supplementary Figure S2) and over-

all number of binding sites detected (Supplementary Table S1). On the other hand, DAP-seq identifies the full catalogue of binding abilities of a TF since, unlike ChIP-seq, it is not limited to one specific growth condition, and thus can potentially identify binding events that might occur in conditions for which the use of ChIP-seq is challenging, such as during infection.

Overall, our work represents the first inter-strain comparison of TF function of this scale and reveals numerous differences in TF binding landscapes between three strains of the same species. The observed differences arose from at least three main mechanisms: (i) difference in TF contents, (ii) difference in target gene contents and (iii) difference in target promoter sequences. Across the three strains tested here, the IHMA87 strain displayed the highest number of strain-specific interactions. This was expected as this strain is phylogenetically more distant from the other two and thus harbors many gene and promoter sequence differences. In addition, one major difference is the presence of the strain-specific plasmid in the IHMA87 strain, which contains 200 genes absent in the other strains. We found that 73 of these genes harbor a RR binding site on their promoter. This result exemplifies the striking capacity of acquired genes to graft to the existing regulatory network in the recipient bacterium. Indeed, previous *in silico* analyses showed that horizontally acquired genes are more prone to be regulated by multiple TFs and to be added to existing regulons, either through their concomitantly-acquired promoters that already carry the corresponding TF binding sites or through evolution of the binding sites

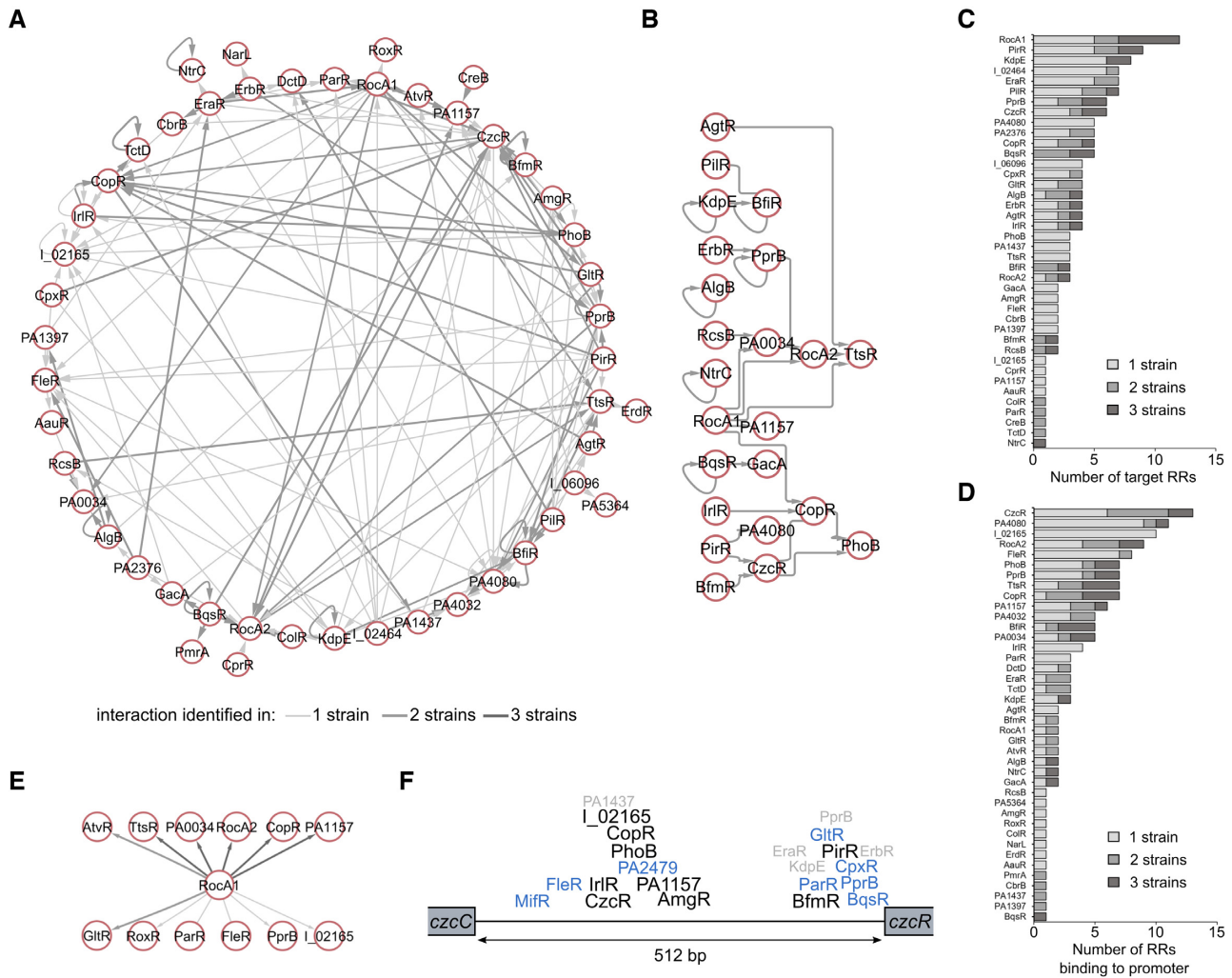


Figure 6. The RR cross-regulation network. (A) Graph diagram of all interactions involving RR target genes. Each node represents a DNA-binding RR found in target TUs from the TCS regulatory network. Edges thickness and darkness increase with the number of strains in which the interactions were identified. (B) Graph diagram of all RR-RR interactions found in all three strains. (C) Number of inferred RR targets for each RR. (D) Number of RRs binding to the promoters of each RR. (E) Graph diagram of all RR cross-regulation interactions involving RocA1 as source node. (F) Schematic representation of the binding events at the *czcC-czcR* intergenic region. The names of RRs that were found to bind to this region in one (grey), two (blue) or three strains (black) are indicated, centered at the average relative DAP-seq peak summit between strains. RR names with the ‘IHMA87.xx’ format were shortened to ‘Lxx’.

after acquisition (10). Although we cannot know which of the two happened for genes on IHMA87 plasmid, this mechanism plays a key role for phenotypic diversification and thus needs to be further investigated globally. In light of the huge genetic diversity across the wide *P. aeruginosa* species, and in fact across most bacterial species, it is obvious, from our results on only three strains, that TF functional diversity is omnipresent in bacteria and should be explored in much more details than what is currently done.

The functional characterization of TFs is a major goal that serves most biological fields of research. In *P. aeruginosa*, few major contributions towards this goal were recently published, mostly focused on virulence regulation (25,68). Similarly, the present work significantly contributes to the characterization of *P. aeruginosa* TFs and constitutes an important resource for the community. While most RRs had been studied with targeted approaches, often identify-

ing only one or few regulatory targets linked to a single studied phenotype, the complete direct regulons were still missing for ~85% of the RRs. In most of these cases, our results confirm existing knowledge and complete it. Additionally, we provide binding profiles for 10 so far completely uncharacterized predicted RRs, such as PA1157 and PA4080. Overall, we report new RR binding sites in 1511 promoters on average between the three strains. We believe that this dataset will serve the community and guide future studies for the delineation of regulatory and virulence key features of bacterial pathogens.

DATA AVAILABILITY

DAP-seq data files have been deposited on NCBI Gene Expression Omnibus (GEO) and can be accessed through GEO Series accession number GSE179001.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We acknowledge the RoBioMol (hosted by the Pneumococcus Group-IBS) and the Cell-free platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). This work is supported by the French National Research Agency in the framework of the “Investissements d’avenir” program (ANR-15-IDEX-02) and has been partially supported by CBH-EUR-GS (ANR-17-EURE-0003). We thank Dr Jérôme Boisbouvier and Dr François Parcy for initial discussions on cell-free protein expressions and DAP-seq, respectively. We acknowledge the High-throughput sequencing facility of I2BC for its sequencing and bioinformatics expertise. We further acknowledge support from CNRS, INSERM, CEA, and Grenoble Alpes University.

Author contributions: Conceptualization, J.T. and S.E.; Methodology, J.T.; Investigation, J.T., L.I., A.M.V., T.V. and S.E.; Formal analysis and Visualization, J.T.; Writing – Original Draft, J.T. and S.E.; Writing – Review & Editing, J.T., L.I., A.M.V., T.V., I.A. and S.E.; Funding Acquisition, S.E. and I.A..

FUNDING

French National Research Agency (ANR) in the framework of the ‘Investissements d’avenir’ program [ANR-15-IDEX-02]; ANR [ANR-15-CE11-0018-01]; Laboratory of Excellence GRAL financed within the Grenoble Alpes University Graduate School (Ecoles Universitaires de Recherche) CBH-EUR-GS [ANR-17-EURE-0003]; Fondation pour la Recherche Médicale [Team FRM 2017, DEQ20170336705]; Julian Trouillon received a Ph.D. fellowship from the French Ministry of Education and Research. Funding for open access charge: ANR in the framework of the ‘Investissements d’avenir’ program.

Conflict of interest statement. None declared.

REFERENCES

- Mejía-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet.*, **21**, 699–714.
- Bartlett, A., O’Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A. and Ecker, J.R. (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.*, **12**, 1659–1672.
- O’Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A. and Ecker, J.R. (2016) Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, **165**, 1280–1292.
- Garber, M.E., Rajeev, L., Kazakov, A.E., Trinh, J., Masuno, D., Thompson, M.G., Kaplan, N., Luk, J., Novichkov, P.S. and Mukhopadhyay, A. (2018) Multiple signaling systems target a core set of transition metal homeostasis genes using similar binding motifs. *Mol. Microbiol.*, **107**, 704–717.
- Trouillon, J., Sentausa, E., Ragno, M., Robert-Genthon, M., Lory, S., Attrée, I. and Elsen, S. (2020) Species-specific recruitment of transcription factors dictates toxin expression. *Nucleic Acids Res.*, **48**, 2388–2400.
- Trouillon, J., Ragno, M., Simon, V., Attrée, I. and Elsen, S. (2021) Transcription inhibitors with XRE DNA-binding and cupin signal-sensing domains drive metabolic diversification in *Pseudomonas*. *mSystems*, **6**, e00753-20.
- Zhang, Q., Huang, Q., Fang, Q., Li, H., Tang, H., Zou, G., Wang, D., Li, S., Bei, W., Chen, H. *et al.* (2020) Identification of genes regulated by the two-component system response regulator NarP of *Actinobacillus pleuropneumoniae* via DNA-affinity-purified sequencing. *Microbiol. Res.*, **230**, 126343.
- Lozada-Chavez, I. (2006) Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.*, **34**, 3434–3445.
- Perez, J.C. and Groisman, E.A. (2009) Evolution of transcriptional regulatory circuits in bacteria. *Cell*, **138**, 233–244.
- Price, M.N., Dehal, P.S. and Arkin, A.P. (2008) Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.*, **9**, R4.
- Galardini, M., Brilli, M., Spini, G., Rossi, M., Roncaglia, B., Bani, A., Chianciani, M., Moretto, M., Engelen, K., Bacci, G. *et al.* (2015) Evolution of intra-specific regulatory networks in a multipartite bacterial genome. *PLoS Comput. Biol.*, **11**, e1004478.
- Moradali, M.F., Ghods, S. and Rehm, B.H.A. (2017) *Pseudomonas aeruginosa* lifestyle: a paradigm for adaptation, survival, and persistence. *Front. Cell. Infect. Microbiol.*, **7**, 39.
- Rodrigue, A., Quentin, Y., Lazdunski, A., Méjean, V. and Foglino, M. (2000) Cell signalling by oligosaccharides. Two-component systems in *Pseudomonas aeruginosa*: why so many? *Trends Microbiol.*, **8**, 498–504.
- Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warriner, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M. *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature*, **406**, 959–964.
- Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A. and Brinkman, F.S.L. (2016) Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the Pseudomonas genome database. *Nucleic Acids Res.*, **44**, D646–D653.
- Mitrophanov, A.Y. and Groisman, E.A. (2008) Signal integration in bacterial two-component regulatory systems. *Genes Dev.*, **22**, 2601–2611.
- Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. *Annu. Rev. Biochem.*, **69**, 183–215.
- Valentini, M. and Filloux, A. (2016) Biofilms and cyclic di-GMP (c-di-GMP) signaling: lessons from *Pseudomonas aeruginosa* and other bacteria. *J. Biol. Chem.*, **291**, 12547–12555.
- Badal, D., Jayarani, A.V., Kollaran, M.A., Kumar, A. and Singh, V. (2020) *Pseudomonas aeruginosa* biofilm formation on endotracheal tubes requires multiple two-component systems. *J. Med. Microbiol.*, **69**, 906–919.
- Gellatly, S.L., Bains, M., Breidenstein, E.B.M., Strehmel, J., Reffuveille, F., Taylor, P.K., Yeung, A.T.Y., Overhage, J., Hancock, R.E.W., Gellatly, S.L. *et al.* (2018) Novel roles for two-component regulatory systems in cytotoxicity and virulence-related properties in *Pseudomonas aeruginosa*. *AIMS Microbiol.*, **4**, 173–191.
- Kollaran, A.M., Joge, S., Kotian, H.S., Badal, D., Prakash, D., Mishra, A., Varma, M. and Singh, V. (2019) Context-specific requirement of forty-four two-component loci in *Pseudomonas aeruginosa* swarming. *iScience*, **13**, 305–317.
- Wang, B.X., Cady, K.C., Oyarce, G.C., Ribbeck, K. and Laub, M.T. (2021) Two-component signaling systems regulate diverse virulence-associated traits in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.*, **87**, e03089-20.
- Bielecki, P., Jensen, V., Schulze, W., Gödeke, J., Strehmel, J., Eckweiler, D., Nicolai, T., Bielecka, A., Wille, T., Gerlach, R.G. *et al.* (2015) Cross talk between the response regulators PhoB and TctD allows for the integration of diverse environmental signals in *Pseudomonas aeruginosa*. *Nucleic Acids Res.*, **43**, 6413–6425.

24. Fan, K., Cao, Q. and Lan, L. (2021) Genome-wide mapping reveals complex regulatory activities of BfmR in *Pseudomonas aeruginosa*. *Microorganisms*, **9**, 485.
25. Huang, H., Shao, X., Xie, Y., Wang, T., Zhang, Y., Wang, X. and Deng, X. (2019) An integrated genomic regulatory network of virulence-related transcriptional factors in *Pseudomonas aeruginosa*. *Nat. Commun.*, **10**, 2931.
26. Xu, C., Cao, Q. and Lan, L. (2021) Glucose-binding of periplasmic protein GltB activates GtrS-GtrR two-component system in *Pseudomonas aeruginosa*. *Microorganisms*, **9**, 447.
27. Yang, B., Liu, C., Pan, X., Fu, W., Fan, Z., Jin, Y., Bai, F., Cheng, Z. and Wu, W. (2021) Identification of novel PhoP-PhoQ regulated genes that contribute to polymyxin B tolerance in *Pseudomonas aeruginosa*. *Microorganisms*, **9**, 344.
28. Yu, L., Cao, Q., Chen, W., Yang, N., Yang, C.-G., Ji, Q., Wu, M., Bae, T. and Lan, L. (2021) A novel copper-sensing two-component system for inducing Dsb gene expression in bacteria. *Sci. Bull.*, <https://doi.org/10.1016/j.scib.2021.03.003>.
29. Francis, V.I., Stevenson, E.C. and Porter, S.L. (2017) Two-component systems required for virulence in *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.*, **364**, fnx104.
30. Rajeev, L., Garber, M.E. and Mukhopadhyay, A. (2020) Tools to map target genes of bacterial two-component system response regulators. *Environ. Microbiol. Rep.*, **12**, 267–276.
31. Freschi, L., Jeukens, J., Kukavica-Ibrulj, I., Boyle, B., Dupont, M.-J., Laroche, J., Larose, S., Maaroufi, H., Fothergill, J.L., Moore, M. et al. (2015) Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. *Front. Microbiol.*, **6**, 1036.
32. Elsen, S., Huber, P., Bouillot, S., Couté, Y., Fournier, P., Dubois, Y., Timsit, J.-F., Maurin, M. and Attrée, I. (2014) A type III secretion negative clinical strain of *Pseudomonas aeruginosa* employs a two-partner secreted exolysin to induce hemorrhagic pneumonia. *Cell Host Microbe*, **15**, 164–176.
33. Freschi, L., Vincent, A.T., Jeukens, J., Emond-Rheault, J.-G., Kukavica-Ibrulj, I., Dupont, M.-J., Charette, S.J., Boyle, B. and Levesque, R.C. (2019) The *Pseudomonas aeruginosa* Pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol. Evol.*, **11**, 109–120.
34. Hauser, A.R. (2009) The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat. Rev. Microbiol.*, **7**, 654–665.
35. Huber, P., Basso, P., Reboud, E. and Attrée, I. (2016) *Pseudomonas aeruginosa* renews its virulence factors: *Pseudomonas aeruginosa* renews its virulence factors. *Environ. Microbiol. Rep.*, **8**, 564–571.
36. Cock, P.J.A., Chilton, J.M., Gruning, B., Johnson, J.E. and Soranzo, N. (2015) NCBI BLAST+ integrated into Galaxy. *GigaScience*, **4**, s13742-015-0080-7.
37. Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J. and Nekrutenko, A. (2020) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.*, **48**, W395–W402.
38. Brinkman, F.S.L., Hancock, R.E.W. and Stover, C.K. (2000) Sequencing solution: use volunteer annotators organized via Internet. *Nature*, **406**, 933–933.
39. Ortet, P., De Luca, G., Whitworth, D.E. and Barakat, M. (2012) P2TF: a comprehensive resource for analysis of prokaryotic transcription factors. *BMC Genomics*, **13**, 628.
40. Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S. et al. (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.*, **49**, D344–D354.
41. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
42. Kumar, S., Stecher, G., Li, M., Niyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.*, **35**, 1547–1549.
43. Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*, **47**, W256–W259.
44. Li, M.Z. and Elledge, S.J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods*, **4**, 251–256.
45. Martin, G. a., Kawaguchi, R., Lam, Y., DeGiovanni, A., Fukushima, M. and Mutter, W. (2001) High-yield, in vitro protein expression using a continuous-exchange, coupled transcription/translation system. *BioTechniques*, **31**, 948–953.
46. Imbert, L., Lenoir-Capello, R., Crublet, E., Vallet, A., Awad, R., Ayala, I., Juillan-Binard, C., Mayerhofer, H., Kerfah, R., Gans, P. et al. (2021) In vitro production of perdeuterated proteins in H₂O for biomolecular NMR studies. In: Chen, Y.W. and Yiu, C.-P.B. (eds). *Structural Genomics: General Applications, Methods in Molecular Biology*. Springer US, NY, pp. 127–149.
47. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
48. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
49. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
50. Li, Q., Brown, J.B., Huang, H. and Bickel, P.J. (2011) Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.*, **5**, 1752–1779.
51. Machanick, P. and Bailey, T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.
52. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
53. Gill, E.E., Chan, L.S., Winsor, G.L., Dobson, N., Lo, R., Ho Sui, S.J., Dhillon, B.K., Taylor, P.K., Shrestha, R., Spencer, C. et al. (2018) High-throughput detection of RNA processing in bacteria. *BMC Genomics*, **19**, 223.
54. Wurtzel, O., Yoder-Himes, D.R., Han, K., Dandekar, A.A., Edelheit, S., Greenberg, E.P., Sorek, R. and Lory, S. (2012) The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog.*, **8**, e1002945.
55. Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeda, D., García-Sotelo, J.S., Alquicira-Hernández, K., Muñoz-Rascado, L.J., Peña-Loredo, P. et al. (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.*, **47**, D212–D220.
56. Taboada, B., Estrada, K., Ciria, R. and Merino, E. (2018) Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, **34**, 4118–4120.
57. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
58. Su, G., Kuchinsky, A., Morris, J.H., States, D.J. and Meng, F. (2010) Glay: community structure analysis of biological networks. *Bioinformatics*, **26**, 3135–3137.
59. Wiese, R., Eiglsperger, M. and Kaufmann, M. (2004) In: yFiles - Visualization and Automatic Layout of Graphs.
60. Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C. et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.*, **35**, W169–W175.
61. Brencic, A., McFarland, K.A., McManus, H.R., Castang, S., Mogno, I., Dove, S.L. and Lory, S. (2009) The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Mol. Microbiol.*, **73**, 434–445.
62. Chevalier, S., Bouffartigues, E., Bodilis, J., Maillot, O., Lesouhaitier, O., Feuilloley, M.G.J., Orange, N., Dufour, A. and Cornelis, P. (2017) Structure, function and regulation of *Pseudomonas aeruginosa* porins. *FEMS Microbiol. Rev.*, **41**, 698–722.
63. Klein, A.H., Shulla, A., Reimann, S.A., Keating, D.H. and Wolfe, A.J. (2007) The intracellular concentration of acetyl phosphate in *Escherichia coli* is sufficient for direct phosphorylation of two-component response regulators. *J. Bacteriol.*, **189**, 5574–5581.
64. Ishimoto, K.S. and Lory, S. (1992) Identification of pilR, which encodes a transcriptional activator of the *Pseudomonas aeruginosa* pilin gene. *J. Bacteriol.*, **174**, 3514–3521.
65. Basso, P., Ragno, M., Elsen, S., Reboud, E., Golovkine, G., Bouillot, S., Huber, P., Lory, S., Faudry, E. and Attrée, I. (2017) *Pseudomonas aeruginosa* pore-forming exolysin and type IV pili cooperate to induce host cell lysis. *mBio*, **8**, e02250-16.

66. Baraquet,C., Murakami,K., Parsek,M.R. and Harwood,C.S. (2012) The FleQ protein from *Pseudomonas aeruginosa* functions as both a repressor and an activator to control gene expression from the pel operon promoter in response to c-di-GMP. *Nucleic Acids Res.*, **40**, 7207–7218.
67. Berry,A., Han,K., Trouillon,J., Robert-Genthon,M., Ragno,M., Lory,S., Attrée,I. and Elsen,S. (2018) cAMP and Vfr control exolysin expression and cytotoxicity of *Pseudomonas aeruginosa* taxonomic outliers. *J. Bacteriol.*, **200**, e00135-18.
68. Wang,T., Sun,W., Fan,L., Hua,C., Wu,N., Fan,S., Zhang,J., Deng,X. and Yan,J. (2021) An atlas of the binding specificities of transcription factors in *Pseudomonas aeruginosa* directs prediction of novel regulators in virulence. *eLife*, **10**, e61885.
69. Cascales,E. and Christie,P.J. (2003) The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.*, **1**, 137–149.
70. Bardoel,B.W., van Kessel,K.P.M., van Strijp,J.A.G. and Milder,F.J. (2012) Inhibition of *Pseudomonas aeruginosa* virulence: characterization of the AprA–AprI interface and species selectivity. *J. Mol. Biol.*, **415**, 573–583.
71. Durand,É., Bernadac,A., Ball,G., Lazdunski,A., Sturgis,J.N. and Filloux,A. (2003) Type II protein secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *J. Bacteriol.*, **185**, 2749–2758.
72. Longo,F., Rampioni,G., Bondi,R., Imperi,F., Fimia,G.M., Visca,P., Zennaro,E. and Leoni,L. (2013) A new transcriptional repressor of the *Pseudomonas aeruginosa* quorum sensing receptor gene lasR. *PLoS One*, **8**, e69554.
73. Bentzmann,S. de, Giraud,C., Bernard,C.S., Calderon,V., Ewald,F., Plésiat,P., Nguyen,C., Grunwald,D., Attree,I., Jeannot,K. *et al.* (2012) Unique biofilm signature, drug susceptibility and decreased virulence in *Drosophila* through the *Pseudomonas aeruginosa* two-component system PprAB. *PLOS Pathog.*, **8**, e1003052.
74. Kreamer,N.N., Costa,F. and Newman,D.K. (2015) The ferrous iron-responsive bqsrs two-component system activates genes that promote cationic stress tolerance. *mBio*, **6**, e02549.
75. Perron,K., Caille,O., Rossier,C., van Delden,C., Dumas,J.-L. and Köhler,T. (2004) CzcR-CzcS, a two-component system involved in heavy metal and carbapenem resistance in *Pseudomonas aeruginosa*. *J. Biol. Chem.*, **279**, 8761–8768.
76. Blair,J.M.A., Webber,M.A., Baylay,A.J., Ogbolu,D.O. and Piddock,L.J.V. (2015) Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.*, **13**, 42–51.
77. Sheppard,S.K., Guttman,D.S. and Fitzgerald,J.R. (2018) Population genomics of bacterial host adaptation. *Nat. Rev. Genet.*, **19**, 549–565.
78. Lai,X., Stigliani,A., Lucas,J., Hugouvieux,V., Parcy,F. and Zubieta,C. (2020) Genome-wide binding of SEPALLATA3 and AGAMOUS complexes determined by sequential DNA-affinity purification sequencing. *Nucleic Acids Res.*, **48**, 9637–9648.
79. Cadoret,F., Ball,G., Douzi,B. and Voulhoux,R. (2014) Txc, a new type II secretion system of *Pseudomonas aeruginosa* strain PA7, is regulated by the TtsS/TtsR two-component system and directs specific secretion of the CbpE chitin-binding protein. *J. Bacteriol.*, **196**, 2376–2386.