Global transcription factors analyses reveal hierarchy and synergism of regulatory networks and master virulence regulators in *Pseudomonas aeruginosa*

Jiadai Huang[1#], Yue Sun[1#], Fang Chen[1#], Shumin Li[2], Liangliang Han[1], Jingwei Li[1], Zhe He[1], Canfeng Hua[1], Chunyan Yao[1], Tianmin Li[1], Beifang Lu[1], Yung-Fu Chang[3], Xin Deng[1,4,5*]

[1] Department of Biomedical Sciences, City University of Hong Kong, 83 Tat Chee Rd, Kowloon Tong, Hong Kong, China

[2] Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong, China

[3] Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, United States.

[4] Shenzhen Research Institute, City University of Hong Kong, Shenzhen, Guangdong, China

[5] Lead contact

[*]To whom correspondence should be addressed: Dr. Xin Deng. Tel: +852 3442 5693; Fax: +852 3442 0549; E-mail: xindeng@cityu.edu.hk.

[#] These authors contributed equally to the paper as the first authors.

Abstract

The transcription factor regulatory network in *Pseudomonas aeruginosa* is complex and involves multiple regulators that respond to various environmental signals and physiological cues by regulating gene expression. However, the biological functions of at least half of its 373 putative transcription factors (TFs) remain uncharacterised. Herein, chromatin immunoprecipitation sequencing (ChIP-seq) was used to investigate the binding sites of 172 TFs in the *P. aeruginosa* PAO1 strain. The results revealed 81,009 significant binding peaks in the genome, more than half of which were located in the promoter regions. To further decode the diverse regulatory relationships among TFs, a hierarchical network was assembled into three levels: top, middle, and bottom. Thirteen ternary regulatory motifs revealed flexible relationships among TFs in small hubs, and a comprehensive co-association atlas was established, showing the enrichment of seven core associated clusters. Twenty-four TFs were identified as the master regulators of virulence-related pathways. The pan-genome analysis revealed the conservation and evolution of TFs in *P. aeruginosa* complex and other species. A Web-based database combining existing and new data from ChIP-seq and the high-throughput systematic evolution of ligands by exponential enrichment was established for searching TF-binding sites. This study provides important insights into the pathogenic mechanisms of *P. aeruginosa* and related bacteria and is expected to contribute to the development of effective therapies for infectious diseases caused by this pathogen.

Keywords: *Pseudomonas aeruginosa*, transcription factor, ChIP-seq

Introduction

Transcription factors (TFs) are key regulators of transcription and thus play crucial roles in mediating multiple biological pathways and events in eukaryotes and prokaryotes. Based on their DNA-binding domains (DBDs), TFs can be classified into different families, such as the LysR, AraC, and LuxR families[1]. TFs can activate and repress the expression of a set of genes in response to various environmental and/or specific signal triggers[2,3]. Several approaches can be used to study TFs, including chromatin immunoprecipitation sequencing (ChIP-seq)[4] and ChIP-exo[5] in *in vivo* research and DNA affinity purification sequencing (DAP-seq)[6] and high-throughput systematic evolution of ligands by exponential enrichment (HT-SELEX)[7] in *in vitro* research. ChIP-seq is a powerful *in vivo* technique that enables more accurate genome-wide mapping of TF–DNA interactions than *in vitro* methods because TFs may interact with other co-regulators in an environment-specific pattern, thereby altering binding preferences *in vivo*[8,9].

These sequencing approaches have been used to illustrate the biological functions of TFs in several bacterial species, including *Mycobacterium tuberculosis*, *Vibrio cholerae*, *Salmonella enterica*, *Escherichia coli*, and *Clostridium thermocellum*[10-15]. In addition, several databases and models have been constructed based on experimental validation or machine learning for studying TFs in bacteria, such as CollecTF, PredicTF, PRODORIC, DBTBS, and RegulomePA[16-20]. However, the functions and spatiotemporal interactions of most bacterial TFs remain to be clarified.

The gram-negative bacterium *Pseudomonas aeruginosa* is a major human opportunistic pathogen that grows ubiquitously. *P. aeruginosa* tends to cause infection in burn victims, patients with cystic fibrosis (CF), and individuals hospitalised for long periods[21,22]. It utilises versatile pathways to exert its virulence, such as biofilm formation, quorum sensing (QS), type III and type VI secretion systems (T3SS and T6SS, respectively), motility, siderophore production, oxidative stress resistance, and antibiotic resistance, all of which are under the control of a complicated TF regulatory network (TRN). The metabolic changes regulated by the TRN enhance the pathogen's

adaptability under different infection and survival conditions.

QS, a major virulence pathway in *P. aeruginosa*, is a bacterial cell–cell communication mechanism involving the expression of many virulence genes. Its regulation occurs hierarchically, comprising the interconnected *las*, *rhl*, and *pqs* systems[23,24]. Through its transcriptional regulator, LasR, and in collaboration with its corresponding autoinducer, 3-oxo-C_{12}-HSL (encoded by *lasI*), the *las* system triggers the activation of the *rhl* and *pqs* systems[23,25]. In addition, *P. aeruginosa* attaches to and colonises surfaces primarily using two forms of motility: (i) swarming and swimming powered by flagella, which are multicellular and individual cell movements, respectively, and (ii) twitching powered by type 4 pili (T4P)[26], wherein T4P promote surface attachment by twitching to pull the cell closer to the attachment sites[27]. The motility of *P. aeruginosa* is thus the initial step in its pathogenic process and is associated with the upregulation of virulence and the induction of host defence. Overall, the virulence-related pathways, including QS and motility, are controlled by the sophisticated TRN in *P. aeruginosa*.

Despite the TRN being essential for understanding how bacterial genes are regulated synergistically, only a limited number of binding targets of TFs have been identified in *P. aeruginosa* over the past decades. For example, ChIP-seq and RNA sequencing (RNA-seq) co-analysis has been used to map a *P. aeruginosa* genomic regulatory network (PAGnet) of 20 key virulence-related TFs, while HT-SELEX has been used to characterise the DNA-binding specificities of 182 TFs[28]. DAP-seq has been used to reveal a regulatory network of 55 response regulators (RRs) of two-component systems in *P. aeruginosa*, 51 of which are from the PAO1 strain[29]. However, almost half of the TFs remain to be annotated with biological functions, and their downstream targets and upstream regulators remain unknown. To fill this knowledge gap, we performed ChIP-seq of 172 TFs **(Table S1)** from *P. aeruginosa*, which revealed a hierarchical and co-association regulatory network among TFs and newly identified virulence-related regulators. We also established a Web-based database to store and search TF-binding patterns together with ChIP-seq and HT-SELEX data,

which is available to the public. Our findings and tools will significantly contribute to future mechanistic studies on and a comprehensive understanding of the TRN in *P. aeruginosa* and other bacterial species.

Results

ChIP-seq of 172 uncharacterised TFs reveals a global transcriptional atlas of *P. aeruginosa.*

In this study, 373 proteins in *P. aeruginosa* were annotated as TFs with DNA-binding activity based on the existing annotations in the *Pseudomonas* Genome Database[30] **(Fig. 1A)**. To profile the whole TF–DNA binding landscape and construct a comprehensive regulatory network of *P. aeruginosa*, ChIP-seq experiments were performed on 172 TFs to determine the TF-binding sites (TFBSs) **(Fig. 1A)**. Libraries were constructed using vesicular stomatitis virus glycoprotein tagging, and raw sequence reads were mapped to the PAO1 genome using bowtie2 (Version 2.3.4.1)[31]. Using MACS2[32], 81,009 significant binding peaks with a cut-off p-value of 0.001 were identified. Subsequently, the R package ChIPpeakAnno[33] was used to define the peak locations and find the nearest genes. The peak location can be divided into six features: upstream, overlapStart, inside, overlapEnd, downstream, and includeFeature. The percentages of binding peak locations for each TF indicated that more than half of the TFs preferentially bind to the upstream (upstream intergenic regions) and overlapStart (overlapped with the transcription start sites [TSSs] of genes) regions **(Fig. 1B)**. Apart from these regions, some TFs bind to the inside regions (coding regions of genes). Few TFs were found to preferentially bind to the rest of the peak location features.

To further provide insights into the functional roles of different TFs in gene regulation, we analysed the peak distance to the TSS, which revealed several patterns **(Fig. 1C)**. First, most of the TFs exhibited peak distributions that were concentrated in the proximal region around the TSS. Some TFs, such as ArsR and AraC, showed a broader distribution of peak distance, indicating that these TFs might also be active at positions farther from the TSS. Furthermore, certain TFs, including CRP and DeoR, displayed distinct bimodal or multimodal distributions, while other TFs, such as Fis and AraC, showed smoother and more narrowly focused distributions, suggesting their different binding preferences. The number of peaks of each TF and its DBD

types are summarised in a treemap in **Fig. 1D**. TFs in the LysR and AraC families accounted for almost half of all of the peaks. PA2718, classified in the MerR family, yielded the highest number (2,480) of binding peaks, followed by the TF PA0756 (2,080 binding peaks) from the OmpR family. Furthermore, eight TFs showed no more than 10 peaks, and the TFs PA4074 and PA4806 had only two peaks each. To preliminarily identify the different regulatory functions of TFs via their different binding preferences, we performed Gene Ontology (GO) functional enrichment analysis for each TF and presented the top 10 GO terms **(Fig. 1E)**. The significantly enriched GO terms (BH-adjusted $P < 0.05$) for all of the 172 TFs analysed by ChIP-seq are listed in **Table S2**, and these terms suggest divergent functions of the TFs. Of these functional categories, most were related to metabolism, such as transcription, translation, ribosome structure, and GTPase activity, while some of the functional categories were associated with virulence, including T4P, O antigen biosynthesis, and biofilm formation **(Fig. 1E)**.

Hierarchical networks of TFs based on pairwise interactions.

TF-coding genes can also serve as targets for other TFs, leading to interactions between the DNA-binding profiles of TFs. To clarify these relationships, lists of binding targets of the TFs obtained using ChIP-seq and HT-SELEX were used to assemble a system-level hierarchical regulatory network. After filtering by TFs that bind to promoter regions, 947 unique pairwise patterns between TFs were identified from a total of 13,375 promoter interactions.

We next defined a statistic index, h, based on the out-degree (O) and in-degree (I) of interactions to measure the direction and hierarchical level of TFs [$h = (O-I)/(O+I)$] in our established hierarchical regulatory network and thus identify hubs and information-flow bottlenecks[34]. The h value of each TF can reflect the direction and extent of information flow. A positive h indicates that the TF acts ‘upstream’, whereas a negative value suggests that it acts ‘downstream’. The higher the h value, the stronger the implication of the direction. A density plot of all of the TFs with their h values revealed that they can be approximately divided into three groups: top (h ∈

[0.75, 1]), middle (h ∈ [-0.75, 0.75]), and bottom (h ∈ [-1, 0.75]) levels **(Fig. S2A)**. The top-level TFs are those that tend to control many other TFs, while the bottom-level TFs are those that are more regulated by other TFs than acting as regulators themselves. The middle-level TFs are those that connect the top- and bottom-level TFs. Of the 231 TFs in this hierarchy, 100, 46, and 85 TFs were defined as top-, middle-, and bottom-level TFs, respectively. The complete hierarchical information of all of the TFs is provided in **Table S3**. To avoid extreme cases and aid visualisation, we summarised the hierarchical regulatory network by removing TFs for which ($O+I$) was more than 10 **(Fig. 2A and Fig. S2B)**. The network was visualised using Cytoscape software[35].

Additionally, we investigated the factors influencing the hierarchical level of TFs, such as DBD types (e.g., DBD families containing more TFs). TFs from the AraC, GntR, and TetR families tended to cluster at the top level. TFs from the MarR family and two-DBD-type TFs (i.e., TFs with two DBDs) were mostly found in the bottom level. Additionally, among the characterised TFs, 58% were clustered at the bottom level, probably because these TFs directly regulated the downstream genes involved in important phenotypes, such as AlgR, AmrZ, MvfR, FleQ, and VqsM. Overall, the organised hierarchical regulatory network profiling at the three levels showed complex pairwise interactions among TFs in *P. aeruginosa.*

Ternary regulatory motifs show flexible relationships among TFs in *P. aeruginosa*.

Apart from their global hierarchical regulatory structure, we investigated the constituent regulatory network motifs of TFs, which revealed small connectivity patterns associated with canonical functions[36]. Positive and negative feedback loops are prevalent in bacterial regulatory networks, which generally involve combinations of several genes to appropriately respond to stimuli. One of the regulatory patterns was a single auto-regulator that can self-regulate its expression or activity. Through analysis, nine auto-regulator TFs were identified **(Fig. 2B)**. Compared with non-auto-regulators, auto-regulators usually tend to be repressors, which play important roles in maintaining a steady state[37,38]. For example, the key T3SS activators PsrA and HrpL

can repress their own expression in *P. aeruginosa* and *P. savastanoi*, respectively[39-41].

In addition, the most fundamental regulatory pattern was the ternary TF motif. We computed all possible motifs, which are listed in **Fig. 2C** and **Table S4**. The basic triangular motifs were monodirectional regulatory structures with five different motifs. The most frequently observed motif (6,535 occurrences) was motif_2, wherein two different TFs co-regulate another TF. The remaining seven types of motifs, which contained toggle switches, occurred much less frequently than the basic motifs. The more complex the relationships among TFs, the lesser their occurrence. In particular, motif_10, motif_12, and motif_13, characterised by two or three mutually regulating TFs, were not found even once **(Fig. 2C)**. The high occurrence of different TFs in diverse motifs revealed a complex and flexible regulatory network in *P. aeruginosa*. Next, we constructed a basic motif_3 regulatory network based on potential TF–TF interactions to detect the regulatory preferences based on DBD types **(Fig. 2D).** Basic motif_3 was a typical hierarchical regulatory network in which TF1 regulates TF2, which in turn regulates TF3. We found that the AraC, Fis, and GntR families tended to harbour top-level TFs rather than middle- and bottom-level TFs, while the OmpR family tended to harbour TFs from middle- and bottom-levels more than those from the top level. Furthermore, the LysR and LuxR families revealed non-significant differences among these three TF levels **(Fig. 2D)**. Taken together, our integrated hierarchical regulatory profiling revealed a multifaceted TF–TF connection, reflecting the regulatory preferences of TFs and DBD types in *P. aeruginos*a.

Clustering of 103 TF-binding motifs.

According to their binding sites, a total of 103 binding motifs were determined. The TFs were clustered based on the positional weight matrix similarity of their motifs. A circular phylogenetic tree was constructed using hierarchical clustering of the pairwise similarity matrix **(Fig. S3 and Table S5)**. Each node in the tree represents a TF, colour-highlighted according to its DBD family classification. The clustering analysis revealed four distinct clusters of TFs, one each coloured light green (Cluster 1), light blue (Cluster 2), purple (Cluster 3), and pink (Cluster 4). These clusters

indicate groups of TFs with similar binding properties or regulatory functions.

Within each identified cluster, TFs tend to share conserved motif patterns, indicating potential functional relationships. For instance, it was found that the TFs in cluster 1 were enriched in adenine (A) and thymine (T) bases. Further, cluster 4 was more likely to have an inverted repeat (IR) sequence, such as PA0367 and PA1015. This conservation suggests that TFs within the same cluster co-regulate similar sets of genes. Additionally, the clustering uncovered novel associations between TFs from different DBD families, indicating potential new regulatory interactions or evolutionary links that warrant further investigation.

Genome-wide synergistic co-association of TFs in *P. aeruginosa*.

To further investigate the crosstalk among TFs in *P. aeruginosa*, we computed the co-association scores based on the overlaps between the peaks of all pairs of TFs and ChIP-seq data **(Fig. S4A and Table S6)**. Among a total of 16,175 TF pairs, we found 5,716 significant TF interactions with a co-association score of up to 0.1 based on an elbow statistic **(Fig. S4B)**. The relationships among TFs can be further determined as core clusters if their co-association scores are more than 0.4. The clustering analysis of core networks using Glay clustering[42] resulted in seven associated modules, each containing a diverse number of TFs with the DBD families highlighted in different colours **(Fig. 3A)**.

Next, we summarised the potential functional crosstalk among the middle-scale co-associated five TFs (PA2417, PA2718, PA0756, PmiR, and AgtR) in cluster 3. Among a total of 1,241 target genes, 372 were co-bound by at least three TFs, and 35 were co-bound by all five TFs, indicating that these TFs have a high level of collaboration in regulatory patterns **(Fig. 3B)**. There were smaller intersections among different combinations of TFs, highlighting potential co-regulation or shared binding sites. Binding targets co-bound by more than four TFs are summarised in **Fig. 3C**. The dense network and connectivity indicate a complex regulatory landscape, where these TFs may orchestrate a coordinated regulation of gene expression. For example, peak

visualization of PA2504 revealed distinct binding patterns of these five TFs compared with the control group **(Fig. 3D)**. PA2504 is the sole partner of the PppH RNA hydrolase related to several important cellular factors[43]. In addition, four of the five TFs (barring PA2718) can bind to the promoter region of *pqsH*, which is involved in the *pqs* pathway **(Fig. 3E)**. Notably, PmiR has previously been verified to bind to the *pqsH* promoter region and thereby regulate virulence[44]. PA2504 and *pqsH* are highlighted in blue and green, respectively, in **Fig. 3C**.

Twenty-four master regulators are associated with virulence.

After elucidating the binding targets of 172 TFs, we investigated their potential functions in the virulence-related pathways and pathogenesis of *P. aeruginosa*. For this purpose, we performed an enrichment analysis using a hypergeometric test to define the master regulators (BH-adjusted $P < 0.05$) that might play important roles in regulating specific pathways. We primarily focused on TF-binding profiles involving the promoter regions of genes involved in nine pathways, namely QS, motility, biofilm production, antibiotic resistance, T6SS, T3SS, reactive oxygen species (ROS) resistance, pyocyanin production, and siderophore production. Accordingly, 134 of the 172 TFs were found to bind to at least one target gene associated with the abovementioned virulent pathways, and 24 were identified as master regulators of genes involved in six pathways, namely siderophore production, biofilm production, pyocyanin production, QS, motility, and ROS resistance pathways **(Fig. 4A and Table S7)**.

The intersection of master regulators across these virulent pathways revealed that six TFs (PA0167, PyrR, PA0707, PA0877, PA1504, and PA1484) only played an important role in regulating the motility pathway, and four TFs (PA1380, PA0815, PA5428, and PA3973) jointly regulated the motility and other pathways, including biofilm production, pyocyanin production, and QS pathways **(Fig. 4B)**. Specifically, PA0815 co-regulated motility, biofilm production, and QS pathways. PA1380 jointly regulated motility and pyocyanin production pathways. PA3973 and PA5428 jointly regulated motility and biofilm production pathways. The regulatory network of these

four master regulators and their target genes is summarised in **Fig. S5**, represented by lines connecting the TFs to the targets involved in specific pathways.

Predicted functions of regulators characterised in two metabolomic pathways.

P. aeruginosa uses different global regulators associated with metabolism that ensure its survival and enhance its adaptability under fluctuating environmental conditions. For instance, it uses different regulatory mechanisms to respond to nutrient changes, oxygen limitation, or CF-associated lung infection conditions[45-48].

To define the core TFs that regulate metabolism, we performed master regulator analysis in metabolic pathways (*p.adjust* < 0.05) based on the TFBSs located in the promoter regions revealed by ChIP-seq for all TFs. We identified 14 TFs potentially involved in two pathways, namely the tricarboxylic acid (TCA) cycle and ribosome function pathways **(Fig. S6A-B)**. The radial plot highlights the master regulators within the TCA cycle and ribosome function pathways, and their joint regulation with their targets is presented in the regulatory network below. PA2957 and PA5438 were found to play putative roles in the TCA cycle by potentially binding to the promoter of certain genes, including *icd*, *sucA*, and *sdhC*, indicating that they coordinate the expression of genes crucial for central metabolism and energy production **(Fig. S6B)**. Furthermore, certain TFs, such as PA0611, PA0403 (PyrR), and PA1283, were found to be involved in regulating a broad array of ribosomal genes, such as *rpsB*, *rpsA*, and *rplL*. Notably, TF PyrR was found to play a putative role in regulating motility in *P. aeruginosa*. Overall, these regulators, which are related to the core metabolic pathways that enhance the adaptation of *P. aeruginosa* to different infection conditions and contribute to its pathogenicity, could offer potential drug targets for tackling *P. aeruginosa* infection in the future.

A global summary of transcription regulatory networks and functions in *P. aeruginosa.*

After studying the TFBSs of most of the TFs in *P. aeruginosa* both *in vivo* and *in vitro*, we organised these datasets to yield a comprehensive connection of virulence-related

master regulators. The QS and motility pathways consisting of 18 and 20 TFs, respectively, were the key pathways in the virulent landscape and showed a close collaboration with the other five pathways **(Fig. 5A)**. Four TFs (ExsA, CprR, AlpR, and PA2449) were involved in T3SS, a direct host infection mechanism used by *P. aeruginosa*. Additionally, six TFs (PA4436, PA0929, PA3771, PA2534, PA3699, and PA5218) were involved in siderophore production, as identified using HT-SELEX, and showed correlations with only one TF, PA2534, involved in the ROS resistance pathway. Four TFs (CprR, PA0708, PA1520, and PA2479) were characterised in the stringent response and persister cell pathways and were correlated with other virulence-related pathways, namely QS, ROS resistance, T3SS, and antibiotic resistance pathways.

To investigate the phylogenetic classification of the binding targets of the identified TFs on a genome-wide scale, we derived a global atlas revealing the relationships between the protein functions of target genes and the family categories of TFs **(Fig. 5B)**. The Clusters of Orthologous Groups of proteins (COGs) database was used to classify proteins based on the orthology concept[49]. Based on the annotated COGs of *P. aeruginosa*, we found that most of the TF targets were involved in metabolism, followed by poorly characterised categories. The *P. aeruginosa* PAO1 genome is 6.3 Mbp long, and approximately 40% of it remains uncharacterised but is believed to encode well-conserved hypothetical proteins (HPs), which might have indispensable and similar functions. Given that most of the TF targets are poorly characterised, it is crucial to study these HPs in the future to provide more insights into *P. aeruginosa* pathogenicity. Altogether, our results provide a resource of TFs and their putative target genes in *P. aeruginosa* and, consequently, a basis for understanding the regulatory network of virulence and orthological functions.

A Web-based database of all TFBSs in *P. aeruginosa*.

To enable a quick search of the TFBSs in *P. aeruginosa*, we developed a Web-based database containing our ChIP-seq and HT-SELEX results at https://jiadhuang0417.shinyapps.io/PATF_Net/, which is available to the public. It has

four parts: homepage, data page, upload, and contact. The homepage briefly describes the background information and schematic workflow of ChIP-seq and HT-SELEX. The data page provides two types of data: a network plot of TFs and targets, and detailed TF–target information in a table format. The network plot illustrates the top 50 peaks for ease of visualisation. The code was generated using R, and the figures were created using BioRender (https://www.biorender.com/).

Overall, this platform provides several services, including individual TF–target network visualisation, detailed peak annotations in a table format, TF–target interaction visualisation, data download, and data upload for those who want to share their results on this online platform. The primary purpose of PATF_Net is to store, search, and mine valuable information on *P. aeruginosa* TFs for researchers investigating *P. aeruginosa* infection in the future.

Conservation and evolution of TFs in *P. aeruginosa*.

Differences between species are mainly due to differences in genes, whereas essential genes are usually well conserved, especially in subspecies. To investigate the conservation of TFs in *P. aeruginosa*, we performed a pan-genome analysis of 100 strains containing model strains and clinical strains using the Roary software[50]. A total of 21,432 genes were identified across the 100 *P. aeruginosa* strains, of which only 4,544 genes were classified as core genes (existing in more than 99% of strains), revealing that most of the genes were cloud genes (existing in less than 15% of strains) **(Fig. 6A)**. However, when focusing on the TFs among these 100 strains, almost 89% of TFs (331 of 373) were found to be well-conserved **(Fig. 6B)**, which implied that these recognised TFs from PAO1 play essential roles in biological activities in all of the *P. aeruginosa* subspecies.

The bar plot provides detailed information on the rest of the non-core TFs in PAO1, in addition to several important TFs, such as LasR and ExsA, that are lacking in some strains **(Fig. 7C)**. As is known, the *lasR* mutant of *P. aeruginosa* is associated with CF lung disease progression and is considered a marker of an early CF-adaptive

phenotype[51]. ExsA is considered an important activator of the T3SS pathway in *P. aeruginosa*, which is characterised by an acute infection phenotype. Additionally, VqsM might have a relatively special function in PAO1 among *P. aeruginosa* subspecies in addition to playing an important role in the QS and antibiotic resistance pathways. Taken together, the core- and non-core TFs were both vital and worth studying for understanding the metabolism and virulence-related regulatory mechanisms in *P. aeruginosa*.

To further investigate the conservation of TFs in *P. aeruginosa*, we re-analysed the ChIP-seq data of PhoB and RpoN in the PA14 strain[52,53] and compared them with those in the PAO1 strain. **Fig. S7A** shows genome-wide binding peaks of PhoB and RpoN throughout the genomes of PAO1 and PA14. We found that in PA14 and PAO1, PhoB had 725 and 63 peaks, while RpoN yielded 363 and 1,238 peaks, respectively. Next, we used MEME[54] to compare the motifs of PhoB and RpoN in PAO1 and PA14 based on their binding sequences from ChIP-seq **(Fig. S7B)**. PhoB and RpoN shared a well-conserved motif in the two strains even though they showed different peak distributions in the genomes. They had different numbers of peaks, and of the 690 and 345 respective annotation targets of PhoB and RpoN in PA14, 114 and 66 targets, respectively, were unique to PA14. The rest of the targets were homologous between PAO1 and PA14. We next performed target annotation and observed that the genes regulated by these two TFs were significantly (Fisher's test) overrepresented in the two tested strains **(Fig. S7C)**. For example, RpoN has been found to play an important role in regulating T6SS in both PAO1 and PA14[53,55]. Overall, the conservation of the representative TFs PhoB and RpoN in PAO1 and PA14 indicates that conserved TFs may share similar functions in different species. This knowledge is expected to guide drug development for different infectious strains.

TFs play a pivotal role in regulating gene expression, and studying their conservation and evolution could provide insights into how gene expression patterns are conserved or modified across different species and help to predict gene regulatory networks in different species. For example, the conservation of a TF across multiple species

suggests that its targets and regulatory interactions are likely to be conserved as well. Therefore, to study the conservation and evolution of *P. aeruginosa* TFs across other species, we conducted phylogenetic analyses of the 373 TFs of *P. aeruginosa* across other bacteria, fungi, archaea, plants, and animals. We found that three *P. aeruginosa* species showed the most conserved TFs, followed by *P. savastanoi*, a phytopathogen **(Fig. 6D)**. Compared with bacteria, the conservation values of the TFs were lower in fungi, archaea, plants, and animals. However, several orthologues of TFs were still noted among these organisms. For instance, PA2032 had more than 25% and approximately 30% identity with aminoadipate aminotransferase in humans and mice, respectively. Altogether, the high-level evolutionary conservation of TFs in *P. aeruginosa* strongly suggests that the regulatory mechanisms are common to a wide range of bacteria.

Discussion

Bacterial pathogens use many mechanisms to establish infection and cause diseases in human hosts, making them a major public health concern worldwide. These bacteria can secrete a wide range of molecular particles that recognise and bind to host cell targets, thus damaging or preventing host responses[56]. Pathogenic bacteria use strategies to adapt to environmental stressors such as nutrient limitation or antibiotic treatment[57]. The underlying mechanism of these strategies involves sensing external signals and then responding to different environmental conditions. TFs are key molecules involved in responding to host or environmental signals by precisely modulating transcription levels; hence, their pivotal roles in infection conditions cannot be overstated. Nevertheless, our comprehension of bacterial TFs remains considerably limited, even in extensively examined model organisms such as *P. aeruginosa*. To address this knowledge gap and shed more light on TFs in *P. aeruginosa*, we conducted ChIP-seq experiments aimed at elucidating the TFBSs for 172 TFs with relatively unexplored biological functions.

The basic function of TFs is to recognise specific DNA sequences (i.e., motifs) at the promoter region and then upregulate or downregulate the transcription of the target gene[58]. Apart from promoter regions, TFs have also been identified to directly regulate the transcription of coding regions in bacteria[59]. ChIP-seq is a valuable technique to obtain high-resolution DNA–protein interaction mapping on a genome-wide scale to study the potential binding targets of TFs. The technique is highly sensitive and can detect low-abundance DNA-binding events, allowing the detection of weak TF–DNA interactions. *In vitro* assays, such as DAP-seq or HT-SELEX, offer superior scalability to ChIP-seq but present their own challenges[29,60]. It is difficult to investigate the crosstalk and interactions between TFs by studying proteins individually. Thus, we devised a strategy involving TF-overexpressing plasmids. This strategy can enhance the visualisation of more subtle binding peaks, particularly when investigating co-associations and TFs lacking well-defined signals. Notably, the literature underscores the collaborative nature of TFs, which often exhibit interactions

with other auxiliary proteins at multiple promoter sites[61]. As shown in **Figures 3 and S3**, we present a TF-binding motifs clustering tree and a TF co-association regulatory network based on ChIP-seq data and core clusters of TFs with high co-association scores. Overall, our above-described analysis has illustrated a regulatory atlas to study the biological functions of *P. aeruginosa* TFs in specific contexts.

P. aeruginosa uses TFs to control biological functions, including metabolism and virulence. Given the public health burden of *P. aeruginosa* infection, many studies have investigated the roles of TFs in this pathogen. These studies have characterised the regulatory mechanisms of individual TFs, such as FleQ, AlgR, LasR, VqsM, and PvrA[62-66]. Additionally, a TRN comprising 690 genes and 1,020 regulatory interactions with six biological modules and main motifs was established in 2011 using published data[67]. Furthermore, we developed a PAGnet containing 20 key virulence-associated TFs based on ChIP-seq and RNA-seq data[68]. A previous study used machine learning to identify modules of a group of genes related to known TFs from published data[69]. Another study used DAP-seq to determine a regulatory network of 55 RRs in *P. aeruginosa*[29]. To further identify and characterise unknown TFs, we used HT-SELEX to describe the binding specificities of 182 TFs[28]. Compared with the abovementioned studies, the present work profiled a comprehensive regulatory network of *P. aeruginosa*, combining existing and new data with experimental verification.

We confirmed the functions and provided potential molecular mechanisms of a number of previously identified TFs to explain their phenotypes. PA0797 is known to regulate the *pqs* system and pyocyanin production[70]. In the present study, it was also found to bind to the *pqsH* promoter region and its motif was visualised. PA5428 was found to bind to the promoter regions of *aceA* and *glcB* genes[71], which was also demonstrated in our ChIP-seq results. PA4381 (CloR) was found to be associated with polymyxin resistance in a previous study[72] and to be possibly related to ROS resistance in the present study. Furthermore, PA5032 plays a putative role in biofilm regulation and also forms an operon with PA5033, an HP associated with biofilm

formation[73].

Our study is the first to illustrate the hierarchical regulatory network and ternary motifs of and genome-wide co-association relationships among TFs in *P. aeruginosa*, contributing to a better understanding of the fundamental traits of pathogenicity of this species. Traditional bacterial eradication strategies involving antibiotics often focus on pathogen extermination, yet this approach can inadvertently foster selective growth advantages and precipitate the emergence of drug-resistant strains. Therefore, prioritising TF-targeted drugs could potentially alleviate selective growth pressure and enhance future preventative measures against *P. aeruginosa* infection. Overall, we gathered the binding site data of most of the uncharacterised TFs in *P. aeruginosa* and established a Web-based database for storing and searching this information. The TF regulatory mechanisms clarified in this database can help to decode the pathogenic mechanisms and guide the development of new drugs targeting virulence-related TFs to combat *P. aeruginosa* infection. The conservation and evolutionary patterns of TFs also suggest that these TFs perform similar functions in diverse species, thus underscoring the broader applicability of our research to studies investigating TFs in other microbes.

Methods

Strains, plasmids, and primers

The bacterial strains, plasmids, and primers used in the present study are listed in **Table S8**. The *P. aeruginosa* PAO1 strain, and its derivatives were grown at 37 °C in LB (Luria-Bertani) broth with shaking at 220 rpm or on LB agar plates. Antibiotics were used for *Escherichia coli* at the following concentrations: kanamycin at 50 μg/ml and carbenicillin at 60 μg/ml.

ChIP-seq and analysis

The chromatin immunoprecipitation (ChIP) procedures were modified from a

previous study[74]. Briefly, for the VSVG-tagged, the ORF was amplified by PCR from the PAO1 genome and cloned into pAK1900 plasmid by HindIII/BamHI for the overexpressed TFs through HindIII site by using ClonEx- press MultiS One Step Cloning Kit (Vazyme, China). Wild-type *P. aeruginosa* containing empty pAK1900 or pAK1900-TF-VSV-G was cultured in LB medium supplemented at 37°C with shaking until mid-log phase (OD600 = 0.6), then we cross-linked the samples with 1% formaldehyde for 10 min. Subsequently, cross-linking was stopped by the addition of 125 mM glycine. Samples were centrifuged and washed thrice with a Tris buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl). The resulting pellets were resuspended in 500 μl IP buffer (50 mM HEPES–KOH [pH 7.5], 150 mM NaCl, 1 mM EDTA, 1 % Triton X- 100, 0.1% sodium deoxycholate, 0.1% SDS, and mini-protease inhibitor cocktail (Roche), and then s The DNA was broken into pieces (100 to 300 bp) with an ultrasonic processor. Insoluble cellular debris was removed by centrifugation at 4 °C, and the supernatant was used as the input sample in IP experiments. We next added 25 μl agarose-conjugated anti-VSV antibodies (Sigma) in the IP buffer. Washing, reverse cross-linked, and purification of the ChIP DNA were conducted. We used the NEXTflex ChIP-Seq kit (Bio Scientific) to construct the DNA fragment library, and agarose gel was used to cut DNA fragments between 150 and 250 bp. After sequencing, the raw reads were trimmed by Trim Galore[75] with default parameters. Then, the filtered reads were aligned to the PAO1 genome (NC_002516) using bowtie2. Only the unique reads after alignment will be used to perform peakcalling with MACS2 ($P < 0.001$). MEME-ChIP was used to identify consensus motifs with all peaks. TF target genes were then annotated by the R package ChIPpeakAnno.

Identification of virulence-related master regulators

We first generated gene lists associated with nine pathways: QS, motility, biofilm production, antibiotic resistance, T6SS, T3SS, ROS resistance, pyocyanin, and siderophores. Then, we calculated the number of overlap genes between target genes of TF and genes involved in the above eight pathways. The statistical significance of the master regulator was identified using the Hypergeometric test (BH-adjusted

$P < 0.05$). R package ggplot2[76] was used to visualize the result.

GO analysis

GO enrichment analysis of TFs target genes was identified using R package clusterProfile[77]. The enriched GO terms with BH-adjusted $P < 0.05$ were defined as significantly enriched.

Pan-genome analysis

The information on the model and clinical *P. aeruginosa* strains[78,79] can be found in **Table S9**. Briefly, the genome sequence of all strains were annotated by prokka[80] and the outputs were then used as input files for roary.

Statistical analysis

Student's t-test was performed to analyze the RT-qPCR and lux production results using Microsoft Office Excel 2019. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$, and results represent means ± SD. All experiments were repeated at least twice.

Data availability

The ChIP-seq data was uploaded to National Center for Biotechnology Information Gene Expression Omnibus database under accession GSE241603 and GSE271817. Analysis codes have been deposited at https://github.com/dengxinb2315/PS-PATRnet-code.

Acknowledgment

This study was funded by the Health and Medical Research Fund (20190942), Shenzhen Science and Technology Fund (JCYJ20210324134000002), National Natural Science Foundation of China grants (32172358), the General Research Funds of Hong Kong (11103221, 11101722, and 11102223), and the Theme-based Research Scheme (T11-104/22-R). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

Contributions

X.D. designed the project and obtained funding, material support and study supervision. X.D., J.H. wrote the manuscript and generated the figures and tables. J.H. performed bioinformatic analysis. J.H., Y.S., F.C., and J.L performed the experiments. S.L. helped to write the code. Y.S. and L.H. helped to edit the manuscript. All authors reviewed the manuscript.

Declare of interest

The authors declare no competing interests.

Figure legend

Figure 1. Overview of ChIP-seq results. (**A).** Density of all TFs (green) and ChIPed TFs (orange) in this study throughout the *P. aeruginosa* genome. (**B).** Annotation heatmap of all peak distribution with 6 locations: Upstream, OverlapStart, Inside, OverlapEnd, Downstream, and IncludeFeature. **(C).** Peak distance to the translational start site (TSS) of each DBD family. **(D).** Treemap of the 172 TFs peak numbers based on DBD family. Each box's size represents the family's size (number of peaks), and the explained variance of each DBD type means the color shades of each box. DBD families of ChIPed TFs are classified into 20 different categories: LuxR, LysR, Two DBDs (Two DNA binding domains), AraC, TetR, ArsR, CRP, OmpR, GntR, MarR, AsnC, Cro/CI, TyrR, Rrf2, MerR, IclR, Fis, RpiR, DeoR, and undetectable. **(E).** The dot plot shows the top 10 gene ontology (GO) terms from the PseudoCAP annotation of ChIPed TFs. The size of the dots indicates the significance of each functional category, quantified by −log10 (p. adjust).

Figure 2. Hierarchical networks and network motifs. (A). Overview of hierarchy regulatory network after removing TFs with degree O+I <= 10. Nodes indicate TFs, and the color represents the hierarchical level. The top level was highlighted in red, the middle in yellow, and the bottom highlighted in blue. The edges with arrows indicate the regulatory direction. Gray means downward-pointing, and red means upward-pointing. **(B).** The auto-regulator motif of 9 TFs. **(C).** Three transcription-factor motifs occurrences with five basic triangular motifs and eight toggle switch motifs. The circle presents TF, and the arrow indicates the regulatory direction**. (D).** Alluvial diagram reveals basic triangle motif 3 (n = 1956) depending on DBD types. The color of splines is highlighted in different DBD families, and the name of DBD families are labeled.

Figure 3. Core co-association regulatory networks. (A). Core clusters of significant co-binding patterns of TFs. Each TF is highlighted in different colors based on DBD types. The co-association score by pair of TFs was calculated by Jaccard statistics, which measures the ratio of the number of base pairs in overlapped binding peaks on

both TFs to the number of base pairs in their union. **(B).** The histogram's overlapped target genes of TFs in cluster 3 represent the number of target peaks in the individual/overlapped set. **(C).** The network of co-regulation of TFs in cluster 3 with co-bound targets of more than 4. **(D-E).** Genome browser view of TFs in cluster 3 binding intensities at the PA2504 and *pqsH* locus.

Figure 4. Newly identified virulence-related master regulators. (A). Overview of all identified master regulators related to 6 pathways, including QS, motility, biofilm, siderophore, pyocyanin, and ROS. Each circle represents one TF, with the height indicating the significance, quantified by -log10 (p.adjust), and the size indicates the number of targets associated with the virulence pathway. **(B).** Intersection of master regulators in 6 virulent pathways. The bar chart on the top shows the number of intersections, while the matrix below indicates which TFs are involved in specific biological processes such as siderophore production, pyocyanin production, biofilm formation, ROS response, QS, and motility.

Figure 5. Overview of transcriptional network of TFs in *P. aeruginosa*. (A). The interaction network of virulence-related master regulators in *P. aeruginosa*. The ten virulence pathways are highlighted in different colors. The size of the circle represents the degree, and the width of the edges indicates the overlapped number of TFs between two pathways. **(B).** The target annotated using the COG database shows four orthology classification profiling from different DBD types TFs. The size of rectangle indicates the number of targets.

Figure 6. Conservation and variability of TFs in PAO1. (A). The pie chart shows the proportions for all genes. **(B).** The pie chart of proportion for all TFs identified from PAO1. **(C).** The bar plot of proportion for non-core TFs. Core genes indicate genes existed in 99% ~ 100% strains; Soft core genes indicate genes existed in 95% ~ 99% strains; Shell genes indicate genes existed in 15% ~ 95% strains; Cloud genes indicate genes existed in 0% ~ 15% strains. **(D).** The conservation and evolutionary relationship of all 373 TFs in PAO1 among bacteria, archaebacteria, fungi, plants, and

animals. The conservation value was normalized after blastp alignment[82]. The phylogenetic trees were constructed using MEGA11[83] and plotted via R package ggtree[84].

Supplementary information

Fig. S1. RNA-seq for all TFs and ChIPed TFs. Previous RNA-seq data shows the expression level of all TFs and ChIPed TFs. The circle size indicates the log2-transformation raw counts, and the color highlighted for ChIPed TFs represented different DBD families.

Fig. S2. Distribution of Hierarchy Height h. $h = (O-I)/(O+I)$, O presents the extent they regulate other TFs, and I indicates the level other TFs regulate them. **(A).** TFs can be divided into almost equal three levels depending on index h: Bottom level ($-1.0 < h < 0.75$), Middle level ($-0.75 < h < 0.75$), and Top level ($0.75 < h < 1.0$). **(B).** The number of TFs resided in three levels after removing TFs with $O+I < 10$.

Fig. S3 Circular phylogenetic tree of TF binding motifs. Circular phylogenetic tree representing the clustering of TFs based on their motif similarity. The tree was constructed using hierarchical clustering with the Ward.D2 method applied to the pairwise similarity matrix of motifs. The tree were visualized using ggtree[81]. Each node represents a TF, color-coded by its DBD family classification as indicated in the legend on the right. The motif logos are aligned with the respective TFs, with their orientations adjusted to maintain legibility around the circular layout.

Fig. S4. Co-association network of TFs. (A). Heatmap reveals a full co-association pattern of all TFs. **(B).** The density of co-association score for all TF pairs. Determining co-association score as 0.1 (dashed in red) of significant TF co-associations based on an elbow statistic.

Fig. S5. Co-regulation of virulence-related master regulators. Graph diagram of interactions involving target genes of 4 TFs, including PA1380, PA0815, PA5428, and PA3973.

Fig. S6. Regulators involved in TCA and Ribosome pathways. (A). Radar plots show the putative regulator in TCA and ribosome pathways. **(B).** Graph diagram of interactions involving target genes of 14 TFs, including PA0611, PA5218, PA5431, PA1759, PA0403 (PyrR), PA 1283, PA4381, PA5255 (AlgQ), PA4784, PA0475, PA0893, PA0448, PA2957 and PA5438.

Fig. S7. Conservation comparison of PhoB and RpoN. (A). The coverage of PhoB and RpoN peak regions over PAO1 and PA14 chromosomes. Each line shows the location and log2 Fold Enrichment of peaks signal in the chromosome. **(B).** The binding motifs of PhoB and RpoN were analyzed via MEME-ChIP. All peaks were used to define the binding motif. **(C).** Comparison of overlapped targets enriched in PAO1 and PA14 of PhoB and RpoN. The Fisher test made the significance of the overlap ratio.

Table S1. List of 172 ChIPed TFs.

Table S2. Enriched GO terms of 172 ChIPed TFs.

Table S3. Hierarchical regulatory network.

Table S4. Ternary regulatory motifs.

Table S5. PWM pairwise similarity scores.

Table S6. Co-association network.

Table S7. Virulence-related master regulators.

Table S8. Strains and primers used in this study.

Table S9. List of strains for pan-genome analysis.

Reference

1 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432 (2019). https://doi.org:10.1093/nar/gky995

2 Salgado, H., Martinez-Antonio, A. & Janga, S. C. Conservation of transcriptional sensing systems in prokaryotes: a perspective from *Escherichia coli*. *FEBS Lett* **581**, 3499-3506 (2007). https://doi.org:10.1016/j.febslet.2007.06.059

3 Seshasayee, A. S., Sivaraman, K. & Luscombe, N. M. An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. *Subcell Biochem* **52**, 7-23 (2011). https://doi.org:10.1007/978-90-481-9069-0_2

4 Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502 (2007). https://doi.org:10.1126/science.1141319

5 Rhee, H. S. & Pugh, B. F. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell* **147**, 1408-1419 (2011). https://doi.org:10.1016/j.cell.2011.11.013

6 Bartlett, A. *et al*. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* **12**, 1659-1672 (2017). https://doi.org:10.1038/nprot.2017.055

7 Jolma, A. *et al.* DNA-Binding Specificities of Human Transcription Factors. *Cell* **152**, 327-339 (2013). https://doi.org:10.1016/j.cell.2012.12.009

8 Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet* **13**, 840-852 (2012). https://doi.org:10.1038/nrg3306

9 Fu, L. Y. *et al*. ChIP-Hub provides an integrative platform for exploring plant regulome. *Nat Commun* **13** (2022). https://doi.org:ARTN 3413 10.1038/s41467-022-30770-1

10 Minch, K. J. *et al*. The DNA-binding network of *Mycobacterium tuberculosis*. *Nat Commun* **6**, 5829 (2015). https://doi.org:10.1038/ncomms6829

11 Ayala, J. C., Silva, A. J. & Benitez, J. A. H-NS: an overarching regulator of the *Vibrio cholerae* life cycle. *Res Microbiol* **168**, 16-25 (2017). https://doi.org:10.1016/j.resmic.2016.07.007

12 Garcia-Pastor, L., Sanchez-Romero, M. A., Jakomin, M., Puerta-Fernandez, E. & Casadesus, J. Regulation of bistability in the std fimbrial operon of *Salmonella enterica* by DNA adenine methylation and transcription factors HdfR, StdE and StdF. *Nucleic Acids Res* **47**, 7929-7941 (2019). https://doi.org:10.1093/nar/gkz530

13 Seo, S. W. *et al*. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. *Nat Commun* **5**, 4910 (2014). https://doi.org:10.1038/ncomms5910

14 Ishihama, A., Shimada, T. & Yamazaki, Y. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res* **44**, 2058-2074 (2016). https://doi.org:10.1093/nar/gkw051

15 Hebdon, S. D., Gerritsen, A. T., Chen, Y. P., Marcano, J. G. & Chou, K. J. Genome-Wide Transcription Factor DNA Binding Sites and Gene Regulatory Networks in *Clostridium thermocellum*. *Front Microbiol* **12**, 695517 (2021). https://doi.org:10.3389/fmicb.2021.695517

16 Kilic, S., White, E. R., Sagitova, D. M., Cornish, J. P. & Erill, I. CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res* **42**, D156-160 (2014). https://doi.org:10.1093/nar/gkt1123

17 Oliveira Monteiro, L. M. *et al.* PredicTF: prediction of bacterial transcription factors in complex microbial communities using deep learning. *Environ Microbiome* **17**, 7 (2022). https://doi.org:10.1186/s40793-021-00394-x

18 Dudek, C. A. & Jahn, D. PRODORIC: state-of-the-art database of prokaryotic gene regulation.

Nucleic Acids Res **50**, D295-D302 (2022). https://doi.org:10.1093/nar/gkab1110

19 Makita, Y., Nakao, M., Ogasawara, N. & Nakai, K. DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res* **32**, D75-77 (2004). https://doi.org:10.1093/nar/gkh074

20 Galan-Vasquez, E., Luna-Olivera, B. C., Ramirez-Ibanez, M. & Martinez-Antonio, A. RegulomePA: a database of transcriptional regulatory interactions in *Pseudomonas aeruginosa* PAO1. *Database-Oxford* (2020). https://doi.org:ARTN baaa106
10.1093/database/baaa106

21 Bodey, G. P., Bolivar, R., Fainstein, V. & Jadeja, L. Infections caused by *Pseudomonas aeruginosa*. *Rev Infect Dis* **5**, 279-313 (1983). https://doi.org:10.1093/clinids/5.2.279

22 Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959-964 (2000). https://doi.org:10.1038/35023079

23 Lee, J. & Zhang, L. H. The hierarchy quorum sensing network in *Pseudomonas aeruginosa*. *Protein Cell* **6**, 26-41 (2015). https://doi.org:10.1007/s13238-014-0100-x

24 Girard, G. & Bloemberg, G. V. Central role of quorum sensing in regulating the production of pathogenicity factors in *Pseudomonas aeruginosa*. *Future Microbiol* **3**, 97-106 (2008). https://doi.org:10.2217/17460913.3.1.97

25 Schuster, M. & Greenberg, E. P. Early activation of quorum sensing in *Pseudomonas aeruginosa* reveals the architecture of a complex regulon. *Bmc Genomics* **8** (2007). https://doi.org:Artn 287
10.1186/1471-2164-8-287

26 Kazmierczak, B. I., Schniederberend, M. & Jain, R. Cross-regulation of *Pseudomonas* motility systems: the intimate relationship between flagella, pili and virulence. *Curr Opin Microbiol* **28**, 78-82 (2015). https://doi.org:10.1016/j.mib.2015.07.017

27 Kearns, D. B. A field guide to bacterial swarming motility. *Nat Rev Microbiol* **8**, 634-644 (2010). https://doi.org:10.1038/nrmicro2405

28 Wang, T. T. *et al.* An atlas of the binding specificities of transcription factors in *Pseudomonas aeruginosa* directs prediction of novel regulators in virulence. *Elife* **10** (2021). https://doi.org:ARTN e61885
10.7554/eLife.61885

29 Trouillon, J. *et al.* Determination of the two-component systems regulatory network reveals core and accessory regulations across *Pseudomonas aeruginosa* lineages. *Nucleic Acids Research* **49**, 11476-11490 (2021). https://doi.org:10.1093/nar/gkab928

30 Winsor, G. L. *et al.* Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic Acids Res* **44**, D646-653 (2016). https://doi.org:10.1093/nar/gkv1227

31 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009). https://doi.org:10.1186/gb-2009-10-3-r25

32 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137

33 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010). https://doi.org:10.1186/1471-2105-11-237

34 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE

data. *Nature* **489**, 91-100 (2012). https://doi.org:10.1038/nature11245

35 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

36 Cheng, C. *et al.* Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. *PLoS Comput Biol* **7**, e1002190 (2011). https://doi.org:10.1371/journal.pcbi.1002190

37 Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-461 (2007). https://doi.org:10.1038/nrg2102

38 Burda, Z., Krzywicki, A., Martin, O. C. & Zagorski, M. Motifs emerge from function in model gene regulatory networks. *Proceedings of the National Academy of Sciences* **108**, 17263-17268 (2011).

39 Huang, J., Yao, C., Sun, Y., Ji, Q. & Deng, X. Virulence-related regulatory network of *Pseudomonas syringae*. *Comput Struct Biotechnol J* **20**, 6259-6270 (2022). https://doi.org:10.1016/j.csbj.2022.11.011

40 Shen, D. K., Filopon, D., Kuhn, L., Polack, B. & Toussaint, B. PsrA is a positive transcriptional regulator of the type III secretion system in *Pseudomonas aeruginosa*. *Infect Immun* **74**, 1121-1129 (2006). https://doi.org:10.1128/Iai.74.2.1121-1129.2006

41 Kojic, M., Jovcic, B., Vindigni, A., Odreman, F. & Venturi, V. Novel target genes of PsrA transcriptional regulator of *Pseudomonas aeruginosa*. *Fems Microbiol Lett* **246**, 175-181 (2005). https://doi.org:10.1016/j.femsle.2005.04.003

42 Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135-3137 (2010). https://doi.org:10.1093/bioinformatics/btq596

43 Drabinska, J., Steczkiewicz, K., Kujawa, M. & Kraszewska, E. Searching for biological function of the mysterious PA2504 protein from *Pseudomonas aeruginosa*. *International Journal of Molecular Sciences* **22**, 9833 (2021).

44 Cui, G. *et al.* PmiR senses 2-methylisocitrate levels to regulate bacterial virulence in *Pseudomonas aeruginosa*. *Science Advances* **8**, eadd4220 (2022).

45 Moreno, R. & Rojo, F. The importance of understanding the regulation of bacterial metabolism. *Environ Microbiol* **25**, 54-58 (2023). https://doi.org:10.1111/1462-2920.16123

46 Dolan, S. K. *et al.* Contextual Flexibility in *Pseudomonas aeruginosa* Central Carbon Metabolism during Growth in Single Carbon Sources. *Mbio* **11** (2020). https://doi.org:ARTN e02684-19
10.1128/mBio.02684-19

47 Oberhardt, M. A., Goldberg, J. B., Hogardt, M. & Papin, J. A. Metabolic Network Analysis of *Pseudomonas aeruginosa* during Chronic Cystic Fibrosis Lung Infection. *J Bacteriol* **192**, 5534-5548 (2010). https://doi.org:10.1128/Jb.00900-10

48 Korgaonkar, A. K. & Whiteley, M. *Pseudomonas aeruginosa* Enhances Production of an Antimicrobial in Response to N-Acetylglucosamine and Peptidoglycan. *J Bacteriol* **193**, 909-917 (2011). https://doi.org:10.1128/Jb.01175-10

49 Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**, 33-36 (2000). https://doi.org:10.1093/nar/28.1.33

50 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**,

3691-3693 (2015).

51 Hoffman, L. R. *et al.* *Pseudomonas aeruginosa lasR* mutants are associated with cystic fibrosis lung disease progression. *Journal of Cystic Fibrosis* **8**, 66-70 (2009).

52 Bielecki, P. *et al.* Cross talk between the response regulators PhoB and TctD allows for the integration of diverse environmental signals in *Pseudomonas aeruginosa*. *Nucleic Acids Research* **43**, 6413-6425 (2015).

53 Allsopp, L. P., Collins, A. C. Z., Hawkins, E., Wood, T. E. & Filloux, A. RpoN/Sfa2-dependent activation of the *Pseudomonas aeruginosa* H2-T6SS and its cognate arsenal of antibacterial toxins. *Nucleic Acids Res* **50**, 227-243 (2022). https://doi.org:10.1093/nar/gkab1254

54 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**, W202-208 (2009). https://doi.org:10.1093/nar/gkp335

55 Shao, X. *et al.* RpoN-Dependent Direct Regulation of Quorum Sensing and the Type VI Secretion System in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **200** (2018). https://doi.org:10.1128/JB.00205-18

56 Wilson, J. W. *et al.* Mechanisms of bacterial pathogenicity. *Postgrad Med J* **78**, 216-224 (2002). https://doi.org:10.1136/pmj.78.918.216

57 Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol* **13**, 42-51 (2015). https://doi.org:10.1038/nrmicro3380

58 Perez-Rueda, E. *et al.* Abundance, diversity and domain architecture variability in prokaryotic DNA-binding transcription factors. *PLoS One* **13**, e0195332 (2018). https://doi.org:10.1371/journal.pone.0195332

59 Hua, C. *et al.* Bacterial Transcription Factors Bind to Coding Regions and Regulate Internal Cryptic Promoters. *Mbio* **13**, e0164322 (2022). https://doi.org:10.1128/mbio.01643-22

60 Baumgart, L. A. *et al.* Persistence and plasticity in bacterial gene regulation. *Nat Methods* **18**, 1499-1505 (2021). https://doi.org:10.1038/s41592-021-01312-2

61 Browning, D. F., Butala, M. & Busby, S. J. W. Bacterial Transcription Factors: Regulation by Pick "N" Mix. *J Mol Biol* **431**, 4067-4077 (2019). https://doi.org:10.1016/j.jmb.2019.04.011

62 Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S. & Ramphal, R. A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in *Pseudomonas aeruginosa* in a cascade manner. *J Bacteriol* **179**, 5574-5581 (1997). https://doi.org:10.1128/jb.179.17.5574-5581.1997

63 Kong, W. *et al.* ChIP-seq reveals the global regulator AlgR mediating cyclic di-GMP synthesis in *Pseudomonas aeruginosa*. *Nucleic acids research* **43**, 8268-8282 (2015).

64 Gilbert, K. B., Kim, T. H., Gupta, R., Greenberg, E. P. & Schuster, M. Global position analysis of the *Pseudomonas aeruginosa* quorum-sensing transcription factor LasR. *Mol Microbiol* **73**, 1072-1085 (2009).

65 Liang, H., Deng, X., Li, X., Ye, Y. & Wu, M. Molecular mechanisms of master regulator VqsM mediating quorum-sensing and antibiotic resistance in *Pseudomonas aeruginosa*. *Nucleic acids research* **42**, 10307-10320 (2014).

66 Pan, X. *et al.* PvrA is a novel regulator that contributes to *Pseudomonas aeruginosa* pathogenesis by controlling bacterial utilization of long chain fatty acids. *Nucleic Acids Research* **48**, 5967-5985 (2020).

67 Galan-Vasquez, E., Luna, B. & Martinez-Antonio, A. The Regulatory Network of *Pseudomonas aeruginosa*. *Microb Inform Exp* **1**, 3 (2011). https://doi.org:10.1186/2042-5783-1-3

68 Huang, H. *et al.* An integrated genomic regulatory network of virulence-related transcriptional factors in *Pseudomonas aeruginosa*. *Nat Commun* **10** (2019). https://doi.org:ARTN 2931 10.1038/s41467-019-10778-w

69 Rajput, A. *et al.* Machine learning from *Pseudomonas aeruginosa* transcriptomes identifies independently modulated sets of genes associated with known transcriptional regulators. *Nucleic Acids Research* **50**, 3658-3672 (2022).

70 Cui, G. Y. *et al.* PmiR senses 2-methylisocitrate levels to regulate bacterial virulence in *Pseudomonas aeruginosa*. *Sci Adv* **8** (2022). https://doi.org:ARTN eadd4220 10.1126/sciadv.add4220

71 Hwang, W. *et al.* Genome-wide association study of signature genetic alterations among *pseudomonas aeruginosa* cystic fibrosis isolates. *PLoS Pathogens* **17**, e1009681 (2021).

72 Gutu, A. D. *et al.* Polymyxin resistance of *Pseudomonas aeruginosa phoQ* mutants is dependent on additional two-component regulatory systems. *Antimicrob Agents Chemother* **57**, 2204-2215 (2013). https://doi.org:10.1128/AAC.02353-12

73 Zhang, L. *et al.* Identification of genes involved in Pseudomonas aeruginosa biofilm-specific resistance to antibiotics. *PLoS One* **8**, e61625 (2013). https://doi.org:10.1371/journal.pone.0061625

74 Blasco, B. *et al.* Virulence Regulator EspR of *Mycobacterium tuberculosis* Is a Nucleoid-Associated Protein. *Plos Pathog* **8** (2012). https://doi.org:ARTN e1002621 10.1371/journal.ppat.1002621

75 Felix Krueger, F. J., Phil Ewels, Ebrahim Afyounian, Michael Weinstein, Benjamin Schuster-Boeckler, Gert Hulselmans, & sclamons. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path (0.6.10). (2023).

76 Wickham, H. & Wickham, H. Data transformation. In: *ggplot2*: Elegant graphics for data analysis. 203-220 (2016). https://doi.org:10.1007/978-3-319-24277-4_10

77 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* **16**, 284-287 (2012).

78 Hu, Y. Y., Wang, Q., Sun, Q. L., Chen, G. X. & Zhang, R. A novel plasmid carrying carbapenem-resistant gene blaKPC-2 in *Pseudomonas aeruginosa*. *Infect Drug Resist* **12**, 1285-1288 (2019). https://doi.org:10.2147/Idr.S196390

79 Hu, Y. *et al.* Emergence and expansion of a carbapenem-resistant *Pseudomonas aeruginosa* clone are associated with plasmid-borne bla KPC-2 and virulence-related genes. *Msystems* **6**, 10.1128/msystems. 00154-00121 (2021).

80 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014). https://doi.org:10.1093/bioinformatics/btu153

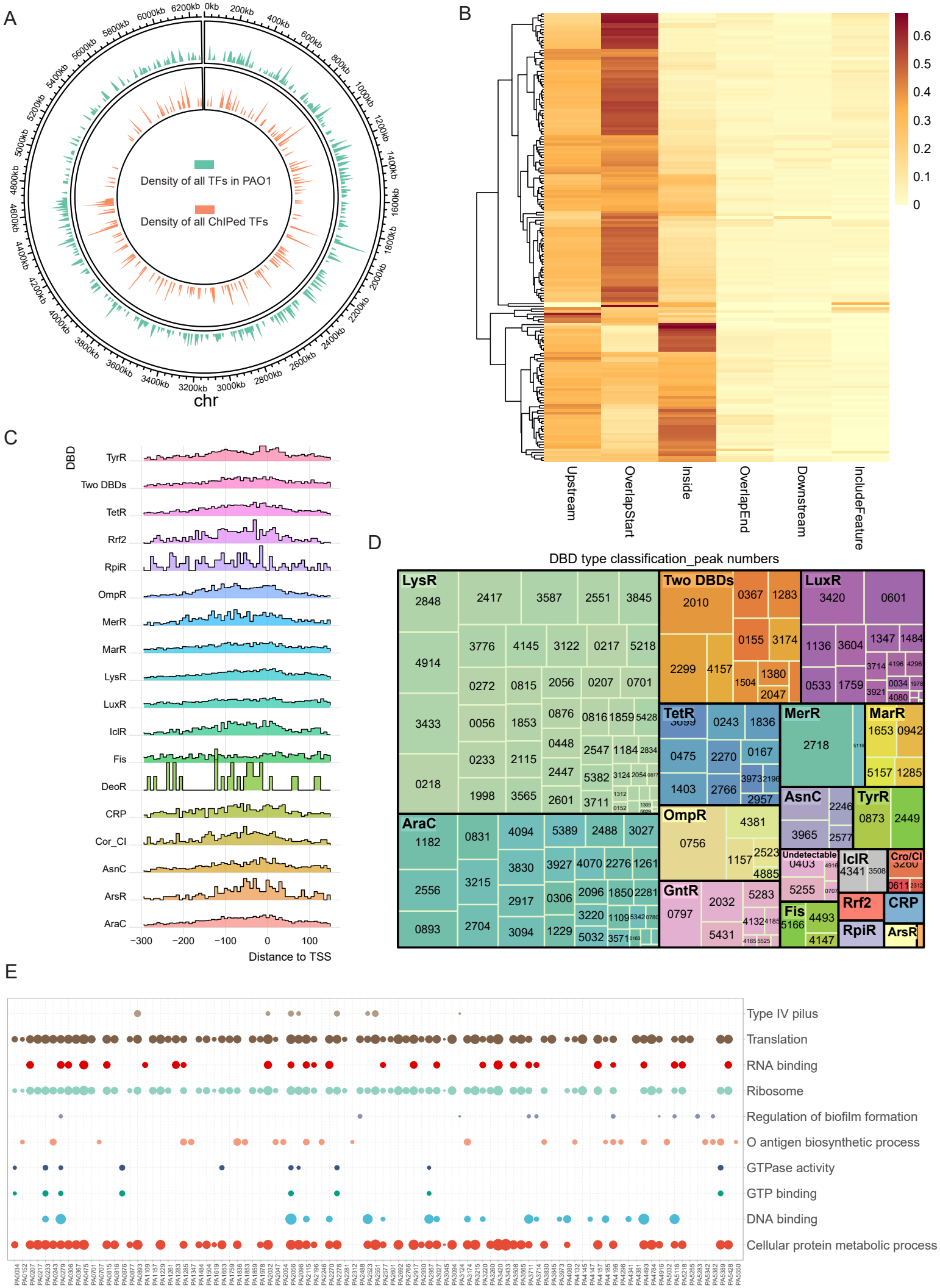
81 Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics* **69**, e96 (2020). https://doi.org:10.1002/cpbi.96

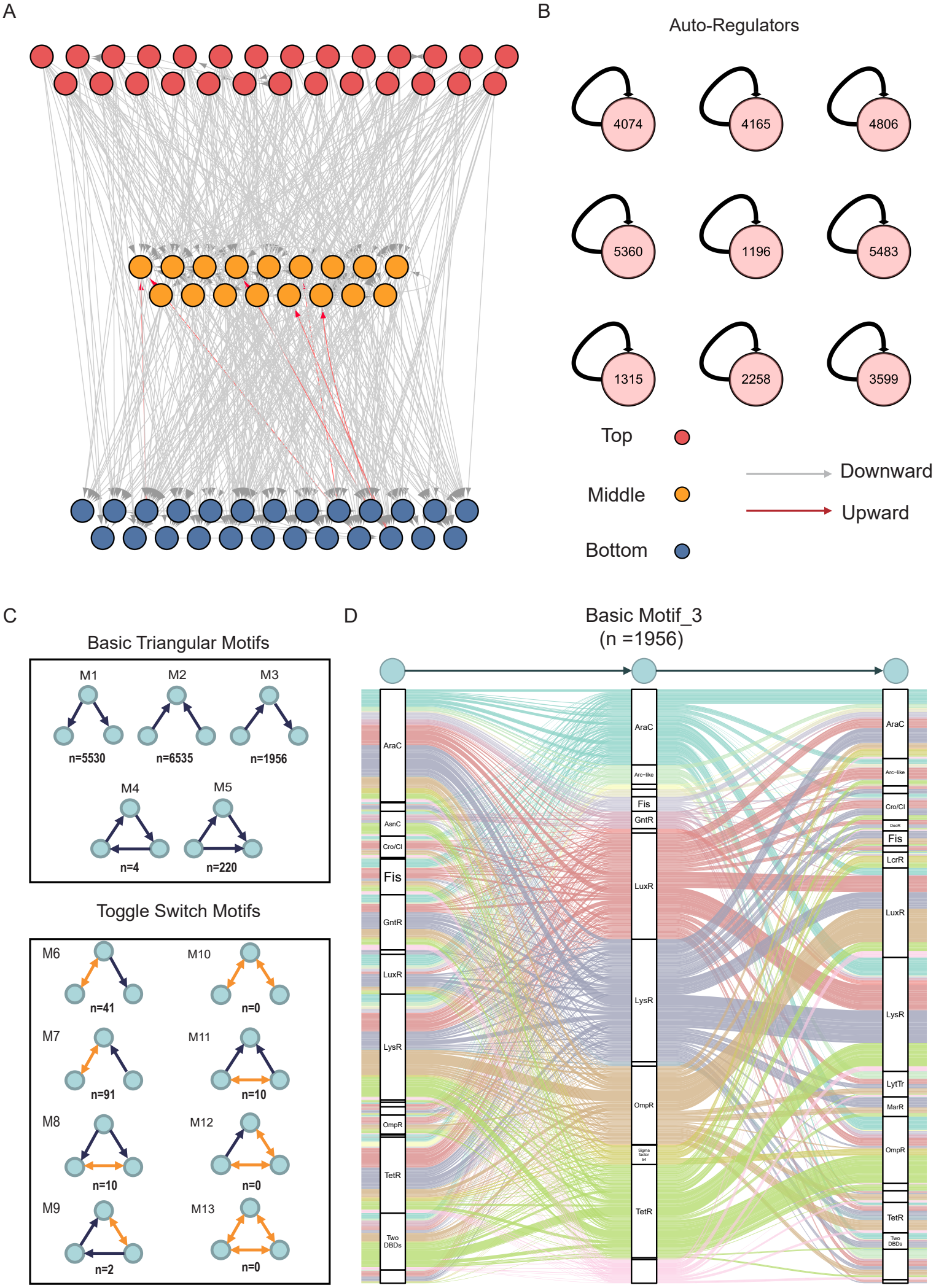
82 Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **41**, D8-D20 (2013). https://doi.org:10.1093/nar/gks1189

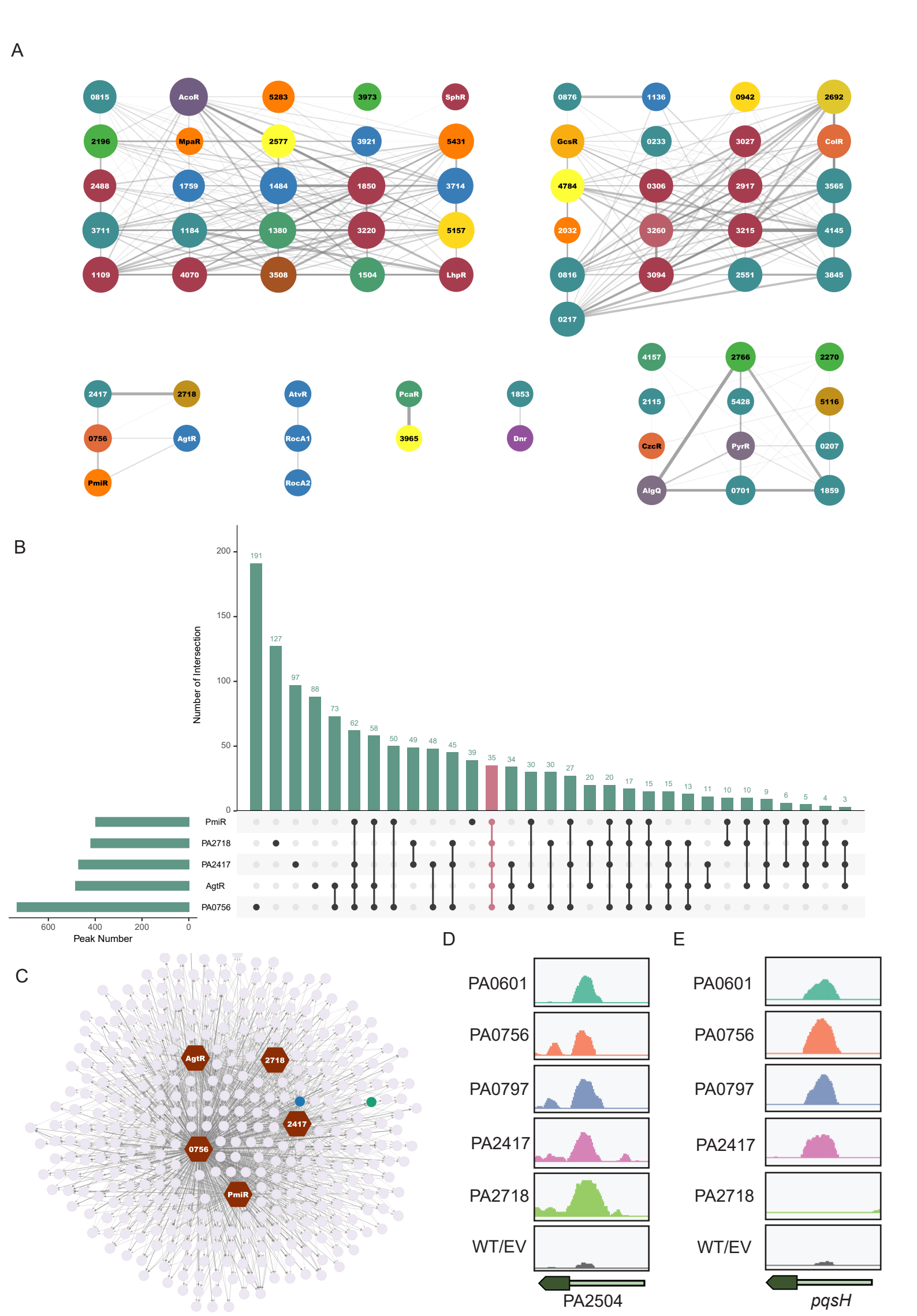
83 Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis

Version 11. *Mol Biol Evol* **38**, 3022-3027 (2021). https://doi.org:10.1093/molbev/msab120

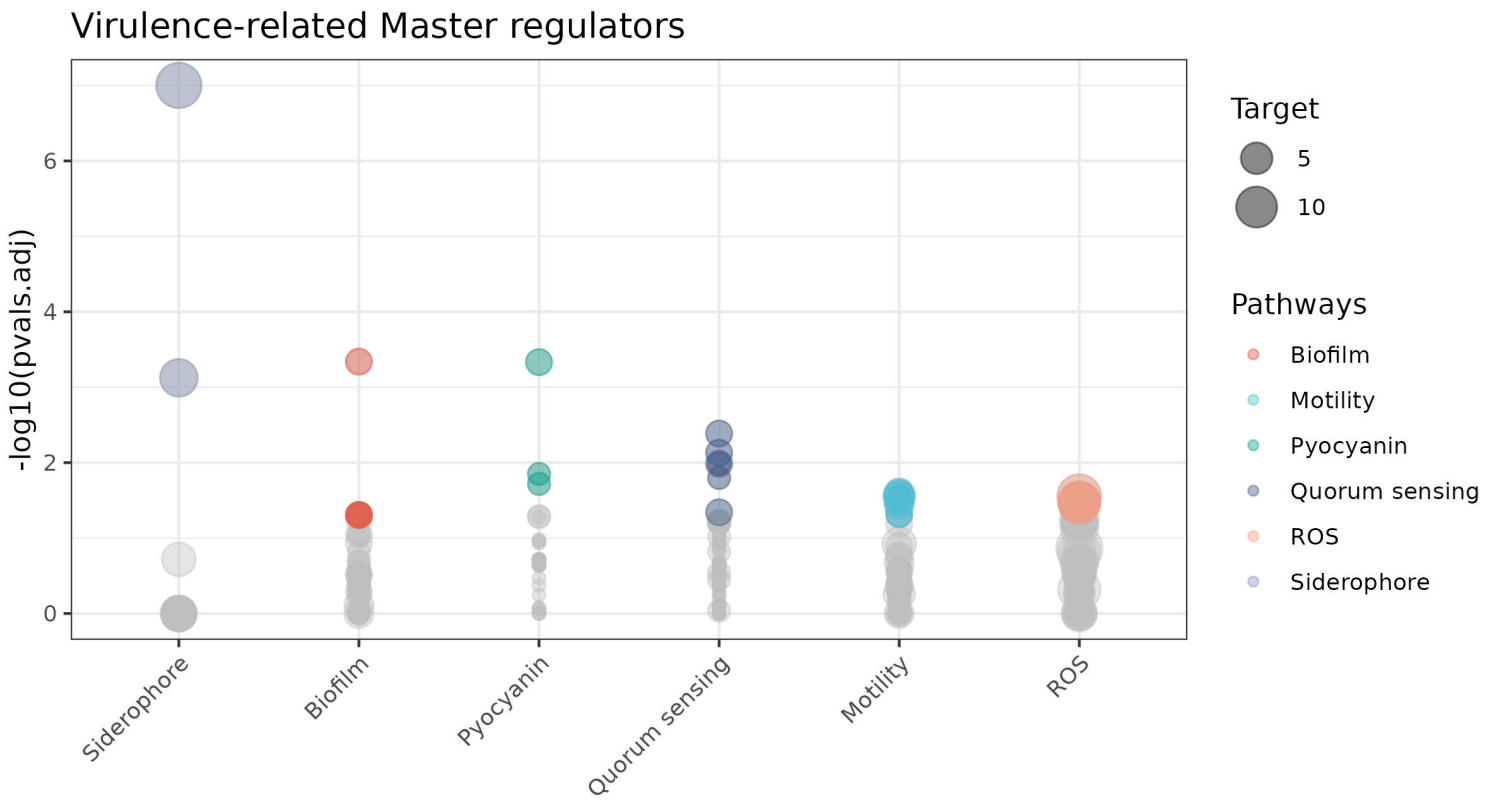
84 Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28-36 (2017).


A
Density of all TFs in PAO1
Density of all ChIPed TFs
chr
B
Upstream
OverlapStart
Inside
OverlapEnd
Downstream
IncludeFeature
C
DBD
TyrR
Two DBDs
TetR
Rrf2
RpiR
OmpR
MerR
MarR
LysR
LuxR
IclR
Fis
DeoR
CRP
Cor_CI
AsnC
ArsR
AraC
Distance to TSS
D
DBD type classification_peak numbers
LysR
Two DBDs
LuxR
TetR
MerR
MarR
AsnC
TyrR
AraC
OmpR
Undetectable
IclR
Cro/CI
GntR
Fis
Rrf2
CRP
RpiR
ArsR
E
Type IV pilus
Translation
RNA binding
Ribosome
Regulation of biofilm formation
O antigen biosynthetic process
GTPase activity
GTP binding
DNA binding
Cellular protein metabolic process

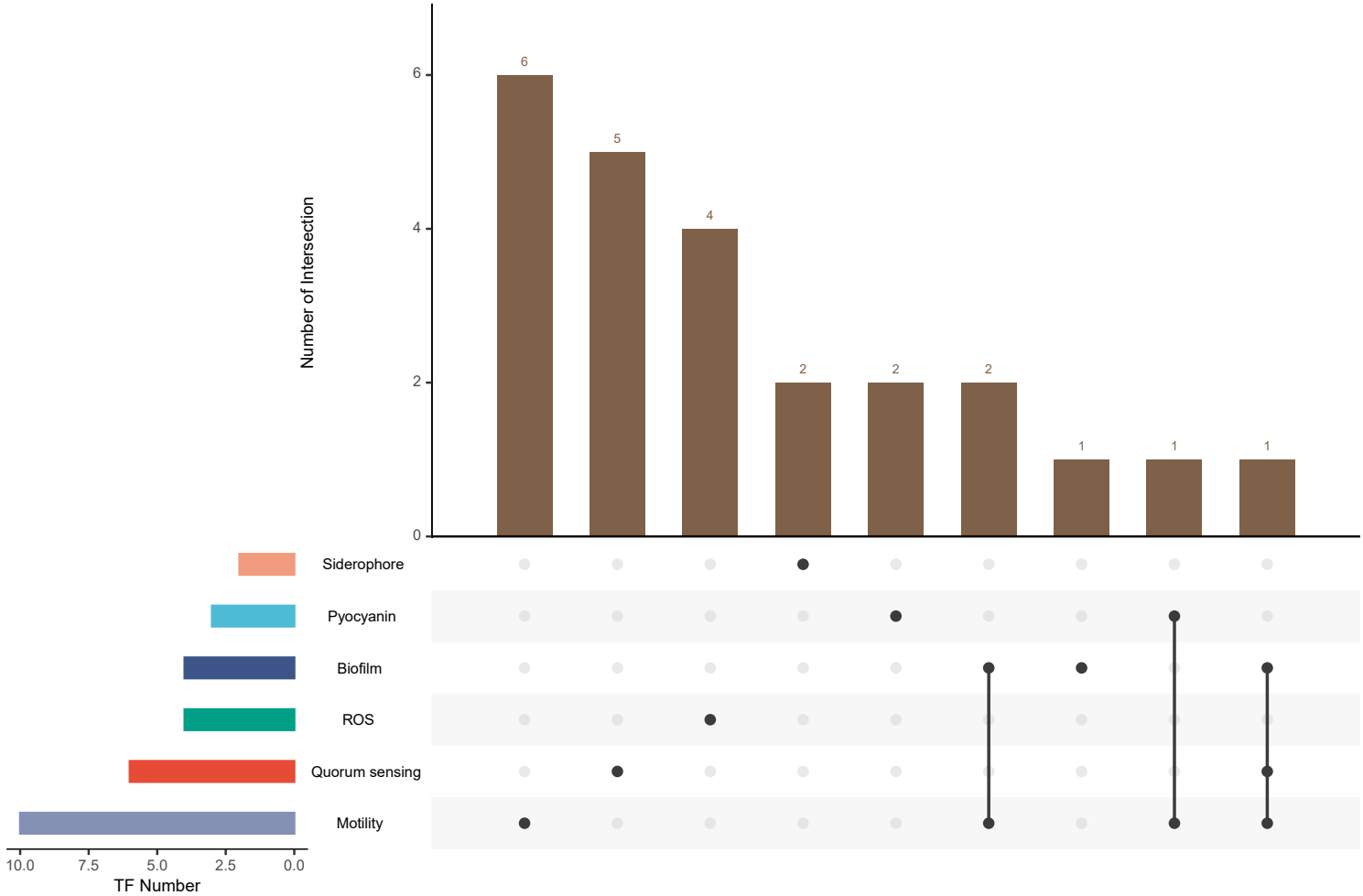

A
B
Auto-Regulators
4074
4165
4806
5360
1196
5483
1315
2258
3599
Top
Middle
Bottom
Downward
Upward
C
Basic Triangular Motifs
M1
M2
M3
n=5530
n=6535
n=1956
M4
M5
n=4
n=220
Toggle Switch Motifs
M6
n=41
M10
n=0
M7
n=91
M11
n=10
M8
n=10
M12
n=0
M9
n=2
M13
n=0
D
Basic Motif_3
(n =1956)
AraC
AsnC
Cro/CI
Fis
GntR
LuxR
LysR
OmpR
TetR
Two DBDs
Arc-like
Sigma factor 54
DeoR
LcrR
LytTr
MarR


A
0815
AcoR
5283
3973
SphR
2196
MpaR
2577
3921
5431
2488
1759
1484
1850
3714
3711
1184
1380
3220
5157
1109
4070
3508
1504
LhpR
0876
1136
0942
2692
GcsR
0233
3027
ColR
4784
0306
2917
3565
2032
3260
3215
4145
0816
3094
2551
3845
0217
2417
2718
0756
AgtR
PmiR
AtvR
RocA1
RocA2
PcaR
3965
1853
Dnr
4157
2766
2270
2115
5428
5116
CzcR
PyrR
0207
AlgQ
0701
1859
B
Number of Intersection
191
127
97
88
73
62
58
50
49
48
45
39
35
34
30
30
27
20
20
17
15
15
13
11
10
10
9
6
5
4
3
PmiR
PA2718
PA2417
AgtR
PA0756
Peak Number
C
AgtR
2718
2417
0756
PmiR
D
PA0601
PA0756
PA0797
PA2417
PA2718
WT/EV
PA2504
E
PA0601
PA0756
PA0797
PA2417
PA2718
WT/EV
pqsH

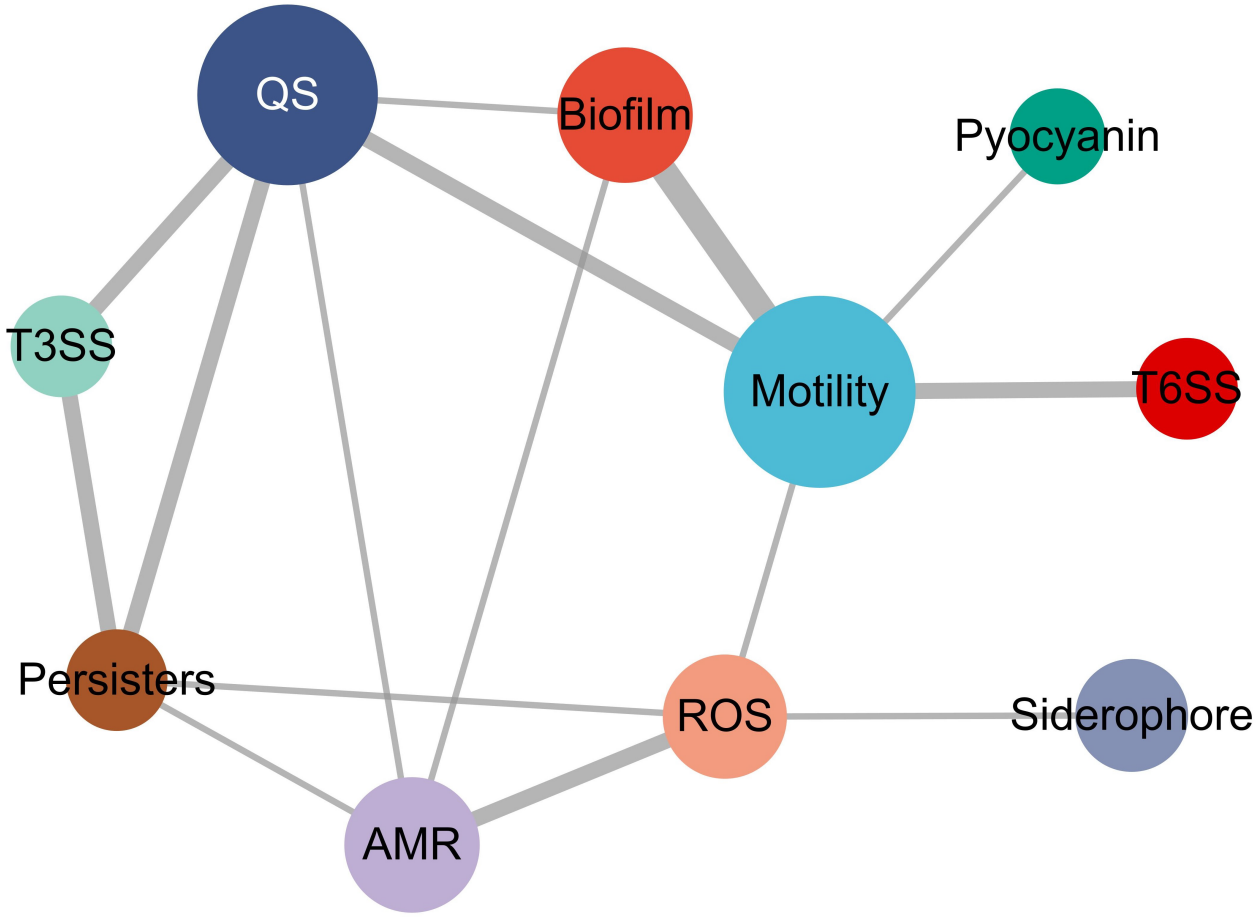
A


Virulence-related Master regulators
-log10(pvals.adj)
0
2
4
6
Siderophore
Biofilm
Pyocyanin
Quorum sensing
Motility
ROS
Target
5
10
Pathways
Biofilm
Motility
Pyocyanin
Quorum sensing
ROS
Siderophore

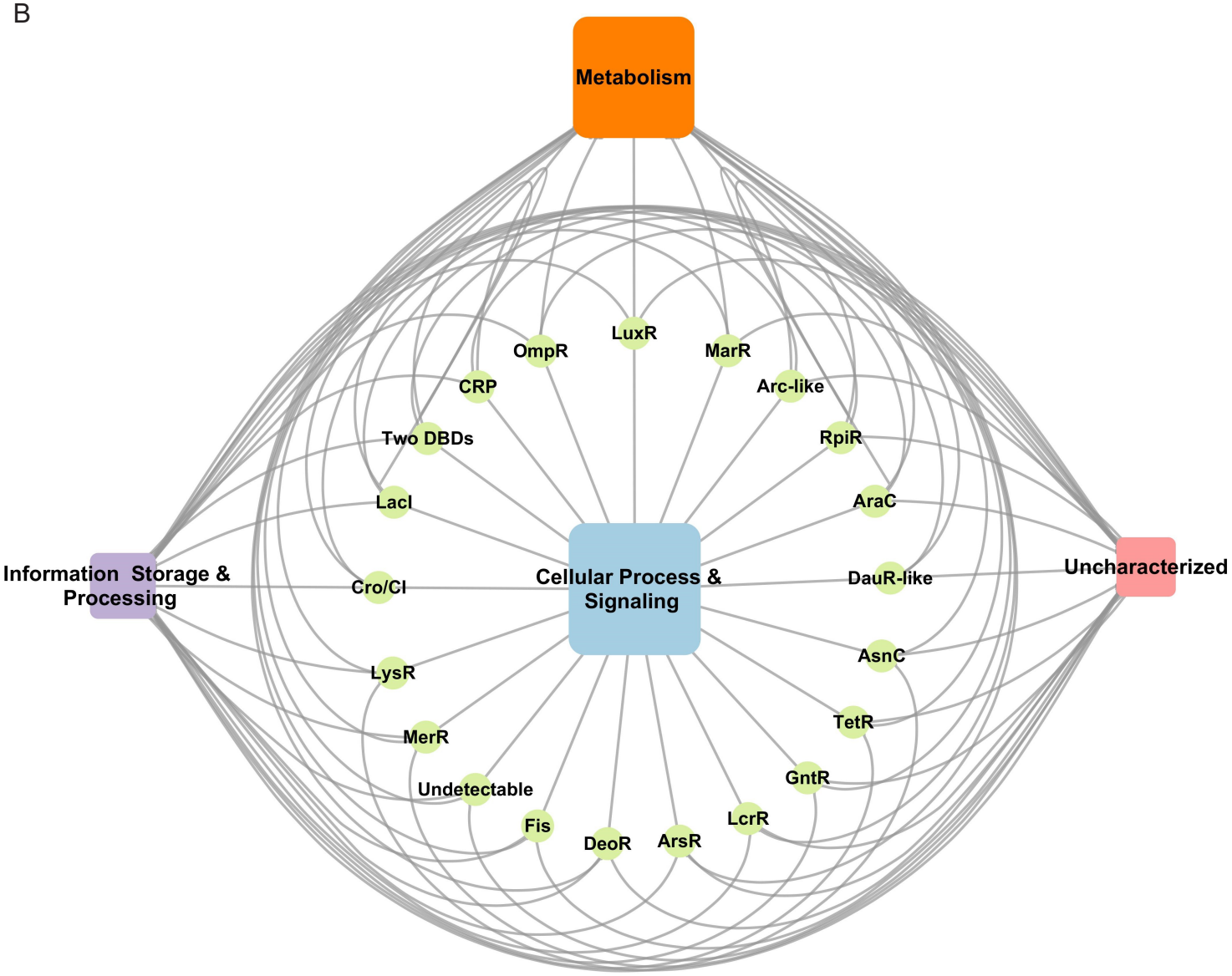
B

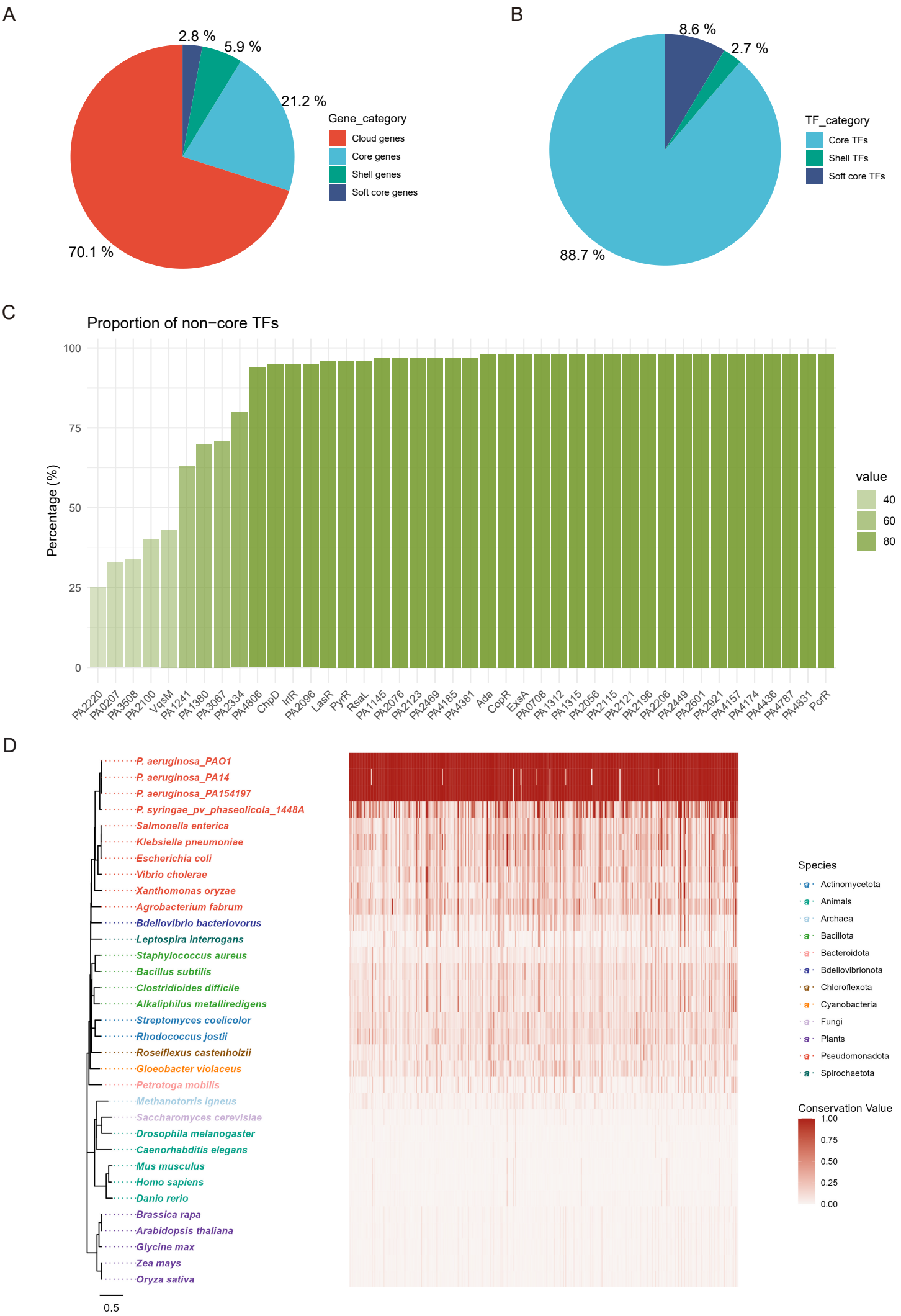

Number of Intersection
0
2
4
6
6
5
4
2
2
2
1
1
1
Siderophore
Pyocyanin
Biofilm
ROS
Quorum sensing
Motility
10.0
7.5
5.0
2.5
0.0
TF Number

A


QS
Biofilm
Pyocyanin
T3SS
Motility
T6SS
Persisters
ROS
Siderophore
AMR

B


Metabolism
LuxR
OmpR
MarR
CRP
Arc-like
Two DBDs
RpiR
LacI
AraC
Information Storage & Processing
Cro/CI
Cellular Process & Signaling
DauR-like
Uncharacterized
AsnC
LysR
TetR
MerR
GntR
Undetectable
LcrR
Fis
DeoR
ArsR


A
2.8 %
5.9 %
21.2 %
70.1 %
Gene_category
Cloud genes
Core genes
Shell genes
Soft core genes
B
8.6 %
2.7 %
88.7 %
TF_category
Core TFs
Shell TFs
Soft core TFs
C
Proportion of non−core TFs
Percentage (%)
100
75
50
25
0
PA2220
PA0207
PA3508
PA2100
VqsM
PA1241
PA1380
PA3067
PA2334
PA4806
ChpD
IrlR
PA2096
LasR
PvrR
RsaL
PA1145
PA2076
PA2123
PA2469
PA4185
PA4381
Ada
CopR
ExsA
PA0708
PA1312
PA1315
PA2056
PA2115
PA2121
PA2196
PA2206
PA2449
PA2601
PA2921
PA4157
PA4174
PA4436
PA4787
PA4831
PcrR
value
40
60
80
D
P. aeruginosa_PAO1
P. aeruginosa_PA14
P. aeruginosa_PA154197
P. syringae_pv_phaseolicola_1448A
Salmonella enterica
Klebsiella pneumoniae
Escherichia coli
Vibrio cholerae
Xanthomonas oryzae
Agrobacterium fabrum
Bdellovibrio bacteriovorus
Leptospira interrogans
Staphylococcus aureus
Bacillus subtilis
Clostridioides difficile
Alkaliphilus metalliredigens
Streptomyces coelicolor
Rhodococcus jostii
Roseiflexus castenholzii
Gloeobacter violaceus
Petrotoga mobilis
Methanotorris igneus
Saccharomyces cerevisiae
Drosophila melanogaster
Caenorhabditis elegans
Mus musculus
Homo sapiens
Danio rerio
Brassica rapa
Arabidopsis thaliana
Glycine max
Zea mays
Oryza sativa
0.5
Species
Actinomycetota
Animals
Archaea
Bacillota
Bacteroidota
Bdellovibrionota
Chloroflexota
Cyanobacteria
Fungi
Plants
Pseudomonadota
Spirochaetota
Conservation Value
1.00
0.75
0.50
0.25
0.00