bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Architecture of genome-wide transcriptional regulatory network reveals dynamic functions and evolutionary trajectories in *Pseudomonas syringae*

Yue Sun[#1], Jingwei Li[#1], Jiadai Huang[#1], Shumin Li[2], Youyue Li[1], Xin Deng[*1,3,4,5]

[1] Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, China

[2] The University of Hong Kong, Pokfulam, Hong Kong, China

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

[4] Tung Biomedical Sciences Center, City University of Hong Kong, Hong Kong, China

[5] Chengdu Research Institute, City University of Hong Kong, Chengdu, China

*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.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The model Gram-negative plant pathogen *Pseudomonas syringae* utilises hundreds of transcription factors (TFs) to regulate its functional processes, including virulence and metabolic pathways that control its ability to infect host plants. Although the molecular mechanisms of regulators have been studied for decades, a comprehensive understanding of genome-wide TFs in *P. syringae* remains limited. Here, we investigated the binding characteristics of 170 of 301 annotated TFs through ChIP-seq. Fifty-four TFs, 62 TFs and 147 TFs were identified in top-level, middle-level and bottom-level, reflecting multiple higher-order network structures and direction of information-flow. More than forty thousand TF-pairs were classified into 13 three-node submodules which revealed the regulatory diversity of TFs in *P. syringae* regulatory network. We found that bottom-level TFs performed high co-associated scores to their target genes. Functional categories of TFs at three levels encompassed various regulatory pathways. Three and 25 master TFs were identified to involve in virulence and metabolic regulation, respectively. Evolutionary analysis and topological modularity network revealed functional variability and various conservation of TFs in *P. syringae*. Overall, our findings demonstrated the global transcriptional regulatory network of genome-wide TFs in *P. syringae*. This knowledge can advance the development of effective treatment and prevention strategies for related infectious diseases.

Keywords: transcription factor; *Pseudomonas syringae*; regulatory network

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Transcription is a pivotal process in cellular life events. Transcription factors (TFs) play a crucial role in this process by acting as key regulators that coordinate various biological activities(Lee & Young, 2013; Papavassiliou & Papavassiliou, 2016). TFs control the recruitment of RNA polymerase by identifying and binding to the promoters of downstream genes, thereby either activating or repressing the expression of target genes(Lambert et al., 2018; Wade, 2015). Numerous studies have focused on regulatory events in eukaryotic species, such as humans(Jolma et al., 2013), mice(Badis et al., 2009), and *Saccharomyces cerevisiae*(C. Zhu et al., 2009), and prokaryotic species, such as *Escherichia coli*(Shen-Orr, Milo, Mangan, & Alon, 2002). However, limited comprehensive TF binding datasets for microbial pathogens are available.

Pseudomonas syringae, an important Gram-negative phytopathogen and a model pathogenic bacterium, infects many plants, including economically valuable crops, resulting in substantial annual economic losses globally(Hirano & Upper, 2000). Upon entering host cells, *P. syringae* employs several strategies, such as changing its motility type and secreting phytotoxins, to overcome the plant's immune defences and establish colonies(Bender, Alarcón-Chaidez, & Gross, 1999; Ichinose, Taguchi, & Mukaihara, 2013; Taguchi & Ichinose, 2011). *P. syringae* causes severe disease by secreting various effector proteins through the needle-like type III secretion system (T3SS); this process is regulated by a cluster of TFs(Cunnac, Lindeberg, & Collmer, 2009; Hendrickson, Guevera, & Ausubel, 2000; Huang, Yao, Sun, Ji, & Deng, 2022;

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Jingru Wang et al., 2018). The alternative sigma factor RpoN activates the transcription of another alternative sigma factor, HrpL, which, in turn, binds to the pathogenicity (*hrp*) box in the promoter region of T3SS genes, regulating most of these T3SS genes(Alfano & Collmer, 1997; Lan, Deng, Zhou, & Tang, 2006; Yingxian Xiao & Hutcheson, 1994). HrpS is one of the most important TFs that regulate numerous biological processes(Jingru Wang et al., 2018). Its heterodimeric complex, HrpRS, is modulated by at least six two-component systems (TCSs): RhpRS(Deng et al., 2014), CvsRS(Fishman, Zhang, Bronstein, Stodghill, & Filiatrault, 2018), GacAS(Chatterjee et al., 2003), AauRS(Yan, Rogan, Pang, Davis, & Anderson, 2020), CbrAB2 and EnvZ-OmpR(Shao et al., 2021). In particular, RhpRS serves as a master regulator of T3SS in *P. syringae*. RhpRS senses plant-derived signals, such as polyphenols, through the histidine kinase Pro40 within RhpS and controls the expression of T3SS genes in response to environmental stress(Deng et al., 2014; Yanmei Xiao et al., 2007; Xie et al., 2021). Within the sensor region, the cognate response regulator RhpR undergoes modulation in its phosphorylation state by RhpS, thereby regulating a group of T3SS genes(Deng et al., 2010). Phosphorylated RhpR directly binds to the *hrpRS* promoter, suppressing the *hrpRS* operon and the subsequent *hrpRS-hrpL-hrp* cascade(Deng et al., 2010; Deng et al., 2014; Deng, Xiao, Lan, Zhou, & Tang, 2009; Shao, Xie, Zhang, & Deng, 2019; Yanmei Xiao et al., 2007).

Recently, through a combined analysis of RNA sequencing (RNA-seq) and chromatin immunoprecipitation sequencing (ChIP-seq), we identified seven additional TCSs

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(ErcS, Dcsbis, PhoBR, CzcSR, AlgB/KinB, MerS and CopRS) that regulate the virulence of *P. syringae*(Xie et al., 2022). In addition, we developed an intricate PSTCSome (*P. syringae* TCS regulome) network containing numerous functional genes that respond to changing environmental conditions(Xie et al., 2022). Furthermore, we examined the overall crosstalk between 16 virulence-related regulators under different growth conditions, such as King's B and minimal media. By analysing differentially expressed genes and binding peaks, we constructed a *P. syringae* regulatory network (PSRnet), revealing the involvement of hundreds of functional genes in virulence pathways(Shao et al., 2021). We also elucidated the molecular mechanisms and functions of TFs binding within coding sequences (CDS) and found that CDS-binding TFs interact with cryptic promoters in coding regions, thereby regulating the expression of subgenus and antisense RNAs(Hua et al., 2022). We propose a luminescence reporter system designed to quantitatively measure the translational elongation rates (ERs) of T3SS-related proteins. Our findings demonstrate the key roles of transfer RNAs (tRNAs) and elongation factors in modulating translational ERs and facilitating T3SS protein synthesis(Sun et al., 2022).

Although many key virulence regulators in *P. syringae* have been studied, the global regulatory mechanism and interactions of all 301 annotated TFs across various biological processes remain unclear. To comprehensively explore the DNA-binding features and map the transcriptional regulatory network of all TFs in *P. syringae*, we constructed 170 TF-overexpressing strains and used ChIP-seq, a highly effective and important technology for analysing protein–DNA interactions(Mathelier, Shi, &

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wasserman, 2015). This analysis not only provided insights into the interactions between TFs and their target genes but also revealed the hierarchy (top, middle and bottom) and co-association scores of all these TFs. We found that more than half of 270 TFs in downstream position tended to be regulated by top TFs and bound to the target genes with high co-associated scores. Different TF-pairs were classified into 13 basic three-node submodules, including ringent loops and locked loops. In addition, we mapped the hierarchical binding network of TFs and identified three virulence-related master TFs and 23 metabolic master TFs. Furthermore, we employed ChIP-seq to determine the binding sites of 5 TFs in 4 *P. syringae* lineages, revealing the diversity of TF binding events and the varying functions of TFs among different *P. syringae* strains. Topological modularity classification of the network, including TFs and target genes, revealed the diverse biological functions of TFs in *P. syringae*. This study provides a global and convenient platform for understanding the transcriptional regulatory characteristics and biological functions of TFs in *P. syringae*. In addition, this study provides valuable insights that can inform the development of effective therapies for not only *P. syringae* but also other associated infectious diseases.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

ChIP-seq analysis revealed the binding specificities of 170 previously uncharacterised TFs in *P. syringae*

Based on the current annotations available on '*Pseudomonas* Genome DB' (https://www.pseudomonas.com/)(Winsor et al., 2016), we initially determined the locations of all 301 annotated TFs in the *P. syringae* genome **(Figure S1a)**. To elucidate the binding preferences and functional characteristics of TFs of *P. syringae*, we performed ChIP-seq for the 170 TFs, including three (1.8%) predicted transcriptional regulators, 132 (77.6%) annotated transcriptional regulators and 35 (20.6%) functional proteins with DNA-binding annotations. Based on the DNA-binding domains as annotated in the transcription factor prediction database(Wilson, Charoensawan, Kummerfeld, & Teichmann, 2008), we categorised the 170 analysed TFs into 25 families(Fan et al., 2020). The majority of TFs belonged to the LysR, TetR, AsnC, GntR and AraC families. Among these TFs, PSPPH4700, PSPPH3798, CysB, PSPPH1951, PSPPH4638, PSPPH3504, PSPPH3268 and Irp exhibited over 1,000 binding peaks **(Figure S1b)**. The enriched loci of these binding peaks indicated that these TFs displayed a significant preference for binding to promoters, directly regulating the transcription of downstream targets **(Figure S1c)**. The peak loci of 10 TFs (PSPPH0286, PSPPH0411, PSPPH0711, PSPPH1734, PSPPH2357, PSPPH2407, PSPPH2862, PSPPH3155, PSPPH3431, PSPPH3468, PSPPH4127, PSPPH4622 and PSPPH4768) were completely enriched in the promoter region, with the majority of them belonging to the LysR family. Taken

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

together, the 170 tested TFs in *P. syringae* had over 26,000 DNA-binding peaks distributed across different regions of target genes, suggesting their direct regulatory functions.

Hierarchical TFs reflected multiple higher-order network structures

Transcriptional changes in bacteria are often manipulated by a complex network of TFs. However, bacterial TFs usually have been studied individually or in small clusters with related functions. To comprehensively investigate the associations of all TFs in *P. syringae* at a system level, we constructed a hierarchical network of 270 analysed TFs. The findings revealed 1,757 TF interactions among these 26,000 binding events (**Supplementary Table 1a**). Subsequently, we computed information flow parameters for each TF(Gerstein et al., 2012). In brief, we defined out-degree (O) and in-degree (I) as the number of interactions of a TF in the hierarchical network, representing the regulation of other factors by this TF and the regulation of this TF by other factors, respectively. The difference between O and I indicated the direction of information flow in the network. Hierarchy height (H) was defined as the normalised metric of information circulation, calculated as $H = (O - I)/(O + I)$. When H was close to 1 ($H \approx 1$), it indicated that these TFs tended to regulate other factors and occupy upstream positions in the network. Conversely, when H was close to −1 ($H \approx -1$), it indicated that these TFs were more likely to be regulated than to regulate other TFs, occupying downstream positions in the network. Based on these criteria, we categorised the 270 analysed TFs into three levels: 54 (20%) executive TFs (such as AlgQ, LexA2 and PSPPH0222) at the top level, 62 (23%) communicative TFs (such

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

as MexT, PsrA and PSPPH1100) at the middle level and 147 (54%) foreman TFs (such as PobR, DksA2 and PSPPH0755) at the bottom level **(Figure 1a, Supplementary Table 1b)**. The presence of a larger number of TFs (147) at the bottom level indicated a high degree of information flow, suggesting the maximisation of the number of upward-pointing edges in the network.

The hierarchical network revealed a downward information flow, suggesting the prioritisation of collaboration between different hierarchy levels. Therefore, we defined the direct binding between two TFs as a direct interaction and investigated collaborations within and between hierarchy levels, specifically intra-level (TT, MM and BB) and inter-level (TM, TB and MB) interactions **(Figure 1b, Supplementary Table 1a)**. In terms of the top-level TFs, physical interactions became more enriched as the hierarchy level of their collaborators decreased. Direct interactions between TB pairs constituted the most substantial portion, accounting for nearly half of all interactions. A similar pattern was observed among the bottom-level TFs, where interactions diminished as the hierarchy level of their collaborators decreased. Compared with interactions among the top- and bottom-level TFs, middle-level TFs, serving as information transmission centres, exhibited lower levels of intra-level collaborations. In summary, transcriptional regulation within the TF hierarchy was predominantly manipulated by top-level TFs, which directed the flow of information to downstream TFs.

Multiple three-node submodules revealed the regulatory diversity of TFs in the *P. syringae* regulatory network

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Natural networks, including transcriptional regulation networks, usually show complex characteristics(Newman, 2001; Strogatz, 2001). Among complex networks, some small-scale networks demonstrate numerous connections between individual information nodes and information clusters(Amaral, Scala, Barthelemy, & Stanley, 2000; Jeong, Tombor, Albert, Oltvai, & Barabási, 2000). To investigate the basic structural features of our transcription network, we defined directed edges as direct interactions between two TF nodes and identified global submodules comprising different TF nodes. In this study, we specifically focused on three-node modules, which were considered as 'network motifs'(Milo et al., 2002). Using algorithms designed to detect recurring modules(Shen-Orr et al., 2002), we scanned our hierarchy network and identified 40,307 different pairs across 13 basic three-node submodules **(Figure 1c)**.

In the first six submodules, we observed that two TF nodes established a relationship only through another node. We denoted these submodules (M1 to M6) as 'ringent loops'. These seemingly simpler regulation modules appeared more in the *P. syringae* transcriptional regulatory network, especially the first module (M1, n = 24,479), indicating that *P. syringae* favours the use of simple but efficient modes for modulating transcriptional regulation. For example, PhnF and FruR were directly regulated by CysB (M1), and CapA was coregulated by Anr and PSPPH4700 simultaneously (M3, n = 6,177). In addition, M6 (n = 20) contained pairs of mutually regulating TFs (toggle switches), such as PSPPH3268, PSPPH2315 and PSPPH2733.

The remaining seven submodules, denoted as 'locked loops' (M7 to M13), comprised

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

subordinate three-node regulatory modules within our network. Notably, no instances of a 'fully connected triad' (M12) were observed in our network (n = 0). We found 50 'self-loop' submodules (M7) in our network, including CspD, PSPPH4694 and PSPPH3798, which engaged in mutual regulatory interactions. Among these submodules, M9 (n = 10) was the least common and contained six pairs of toggle switches involving mutually regulating TFs (such as PSPPH1951, PSPPH4612 and CysB), which were similar to those found in the human transcriptional regulatory network(Gerstein et al., 2012). Notably, the most enriched locked loop in our network was M13 (n = 696), denoted as a 'feed-forward loop', which has been extensively studied in other species such as humans and *E. coli*. In this submodule, upstream TFs regulated targeted TFs either by binding directly or manipulating other TFs (**Figure S2**). For example, TF Irp directly controlled PSPPH4899 and also indirectly regulated it by binding to Lon.

Taken together, the simplest and most effective submodule M1 and the coregulatory submodule M13 played crucial roles in the transcriptional regulation of TFs in *P. syringae*. In addition, we found 92 auto-regulators in our hierarchy network. These auto-regulators are important and always act as repressors in scenarios of multi-stability, such as in plant intercellular spaces where bacteria grow **(Figure 1d)**(Alon, 2007). These regulators are regarded as bistable switches that further influence the expression of downstream genes(Burda, Krzywicki, Martin, & Zagorski, 2011). For example, our previous study demonstrated that Lon is a dual-function regulator involved in the regulation of virulence and metabolism in *P. syringae*(Hua et

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

al., 2020). Lon was identified as an auto-regulator in this study. Furthermore, DksA2, which is widely regarded as a protective protein against oxidative stress(Fortuna et al., 2022), was identified as a new bistable switch in this study. In summary, the classification of the TF hierarchy and the identification of enriched network modules not only offer functional predictions for transcriptional regulators but also provide insights into the communication network that governs TF regulation in *P. syringae*.

High co-association pairs occurred more in bottom-level TFs in *P. syringae*

In addition to direct interactions between TFs, we found a notable preference for co-binding peaks among different TF pairs. Briefly, we counted overlapping regions within the binding peaks of all TF pairs. The ratio of intersection regions to the union set of all peaks between two TFs was identified as the genome-wide co-association of specific TFs. We first computed co-association scores between TF pairs across three hierarchy levels. The majority of TFs tended to cooperate with other TFs and co-bind to specific genomic regions **(Figure S3a, c and e)**. To identify the potential functions of TFs in each level, we performed functional annotations using hypergeometric tests (BH-adjusted $p < 0.05$) based on gene sets derived from the Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) databases **(Figure S3b, d and f, Supplementary Table 2a, c–d)**. The functional categories of TFs at these three levels encompassed various regulatory pathways. For example, the top-level TF PSPPH4700 was involved in siderophore transportation and phosphorelay signal transmembrane transportation. The middle-level TF Lon regulated GTP binding and ribosome transcription. The bottom-level TF PSPPH3486 was involved in amino acid

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transportation, PSPPH0101 in ATP binding and PSPPH1049 in catalytic activity. Notably, we found that TFs at the top level, without cooperating TFs, exhibited a large number of binding peaks **(Figure S3a)**. This finding suggested that these TFs preferred to regulate target genes by recruiting to specific sites of other TFs, facilitating the direct binding between other TFs and their specific targets, a phenomenon defined as tethered binding(Jie Wang et al., 2012). For example, the top-level TF PSPPH4700 (yielding over 1,700 peaks) cooperated with only 24 executive TFs, resulting in an average co-association score of 0.05 **(Supplementary Table 2b)**. PSPPH4700 directly bound to the promoters of nearly half of the tested TFs (66 TFs) and then indirectly regulated numerous downstream genes.

We examined the 125 TFs that were analysed through ChIP-seq and exhibited high co-association scores. We determined the co-association patterns among these regulators **(Figure 2a)** and classified them into four clusters, denoted as C1 to C4. Notably, C1 and C4 contained higher proportions of bottom-level TFs. This finding indicated that the bottom-level TFs, which were more easily regulated, tended to cooperate with downstream genes and other intra-level TFs. When comparing the co-association scores of TF pairs, we observed a stronger tendency toward cooperation among lower-level TFs, especially bottom-level TFs. In particular, 35 bottom-level TFs in C4 (83%) exhibited coregulation at the same peak locations with high co-association scores. For instance, PhnF and PSPPH4692 co-bound to three target genes (PSPPH4117, PSPPH4216 and PSPPHB0021), with co-association scores as high as 0.92. This co-binding relationship between all TF pairs was defined

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

as an indirect interaction. In contrast to the trend observed for direct interactions and cooperativity, we observed stronger correlations between top-level TFs and TFs situated at higher levels (top and middle levels; **Figure 1b)**. For example, the top-level TFs PSPPH3798 and CysB co-bound to the promoter region of *flhB*, which encodes a flagellar biosynthesis protein. We found a similar co-association pattern among bottom-level TFs. However, middle-level TFs displayed the weakest correlations internally, indicating that regulation by higher-level TFs was distributed across different regions in targeted promoters. For instance, we observed that PSPPH3798 bound upstream of *flhB*, whereas the peak of CysB was located in the overlapStart region.

We found that the correlation between bottom-level TFs was weaker in C3 than in the other clusters. To further explore the DNA-binding characteristics of all TFs within the same cluster, we investigated the peak features of the LysR-family TF MexT in C3 as an example. The analysis of the peak locations of MexT demonstrated that MexT showed closer co-association relationships with top-level TFs **(Figure 2b)**. Within the binding peaks of MexT, the target genes PSPPH3643 (LysR family TF), PSPPH2618 (sulphite reductase (NADPH) haemoprotein, CysI) and PSPPH2411 (hypothetical protein) displayed high co-association scores with other TFs. UpSet plots demonstrated that 12 TFs in C3 bound to a highly overlapping region of CysI **(Figure 2c)**. These results suggest that these genes were prioritised for co-regulation by MexT. To further explore the binding features in C3, we determined the motifs of 11 TFs based on their binding sequences obtained via ChIP-seq using MEME(Bailey et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2009), even though the binding motif of MexT was previously investigated(Tian et al., 2009). We identified a 15-bp consensus motif (ATN11AT) throughout the 11 analysed TFs **(Figure 2d)**, demonstrating a high degree of consensus in co-association patterns among TFs in *P. syringae*.

Virulence-associated pathways were primarily regulated by top-level TFs in *P. syringae*

Because the pathogenicity of *P. syringae* mainly depends on T3SS and other virulence-associated pathways(He, 1998), we particularly focused on TFs that bind to numerous virulence-related genes. Seven pathways manipulate the virulence of *P. syringae*(Huang et al., 2022; Xie, Shao, & Deng, 2019), namely T3SS, biofilm production, motility, nucleotide-based secondary messenger function, quorum sensing (QS), phytotoxin production and siderophore production. To comprehensively investigate the transcription regulatory mechanism underlying the virulence of *P. syringae*, we calculated the hierarchical heights of TFs involved in virulence regulation and the virulence genes modulated by them. This analysis provided insights into the organisation of the virulence regulatory network in *P. syringae*, where virulence-involved TFs were categorised into three tiers **(Figure 3a)**.

We found three transcriptional regulatory channels governing virulence regulation in *P. syringae*. The first channel was the direct trigger, which has been extensively studied in previous studies and is referred to as the ‘one-step trigger’ here(Fan et al., 2020). These TFs are recognised as master regulators that directly respond to

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

biological events without additional intermediaries(Chan & Kyba, 2013). In our previous study, we identified TrpI, RhpR, GacA and PSPPH3618 as master regulators in T3SS and 16 master regulators in other virulence pathways(Fan et al., 2020). In line with this definition, we recognised 35 TFs (PSPPH4644, AlgQ, PSPPH1100, CysB, PSPPH2555, PSPPH0239, PSPPH1762, Irp, PSPPH1951, PSPPH2193, PSPH3504, PSPPH1435, PSPPH2983, PSPPH0700, PSPPH5132, PSPPH1776, PSPPH4700, SfsA, PSPPH3522, PSPPH3268, PSPPH4356, PSPPH3798, PSPPH4638, PSPPH4127, Lon, PSPPH4324, PSPPH4673, MexT, PSPPH4844, PSPPH2214, PSPPH0755, PSPPH4012, PSPPH4920, Anr and PSPPH2476) that participate in various virulence pathways. More than 68% of these TFs (24 of 35 TFs) were at the top levels, indicating that the virulence of *P. syringae* is primarily regulated by top-level TFs. Among these TFs in the network, the top-level TFs PSPPH1951, PSPPH2193 and PSPPH3268 were found to have abundant virulence-associated target genes based on ChIP-seq results. The de novo motif analysis of their peak sequences revealed the presence of three head-to-head motifs: a 17-bp motif (AT-N13-AT) for PSPPH1951, a 15-bp motif (ATC-N9-GAT) for PSPPH2193 and a 10-bp motif (AC-N6-GT) for PSPPH3268 **(Figure 3b–d, Figure S4a–c)**.

Three master TFs were identified to participate in virulence

To further verify the biological functions of these three uncharacterised TFs, we first purified the TF proteins and performed an electrophoretic mobility shift assay (EMSA) to confirm their direct interactions with key virulence genes *in vitro*. Next, we

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

generated TF deletion strains to detect the transcription levels of target genes. We found that PSPPH1951 directly regulated multiple T3SS genes, including *hrpRhopAE1* and *hopAH2* **(Figure 3b)**. Among them, the expression of *hrpR* in ΔPSPPH1951 was significantly increased by more than four-fold compared with the WT, suggesting that PSPPH1951 acts as a repressor of *P. syringae* T3SS **(Figure 3b)**. In addition, PSPPH1951 was found to bind to the promoters of type IV pili genes (*pilG*, *pilF* and *pilZ*; **Figure S4d)**.

P. syringae enhances its ability to infect the host by increasing bacterial motility(Buell et al., 2003). ChIP-seq data (*in vivo*) and EMSA (*in vitro*) results showed that PSPPH2193 interacted with the promoters of motility-related genes, such as *fleQ* (encoding a flagellar regulator) and *flhF* (encoding a flagellar biosynthesis regulator; **Figure 3c)**. RT-qPCR results indicated that *fleQ* and *flhF* were downregulated two-fold in ΔPSPPH2193 compared with the WT. As expected, ΔPSPPH2193 exhibited weaker motility than the WT and complementary strain in King's B medium **(Figure 3c)**, indicating that PSPPH2193 serves as an activator of motility in *P. syringae*.

PSPPH3268 was found to influence the pathogenicity of *P. syringae* by regulating multiple virulence-related pathways. During the initial stage of *P. syringae* infection, the bacteria produce biofilm components, including extracellular polysaccharides (EPSs), type IV pili and other highly viscous compounds. These components help the bacteria to establish colonies, providing protection against the host's immune defences and antimicrobial agents(Whitchurch, Tolker-Nielsen, Ragas, & Mattick,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2002). We found that PSPPH3268 strongly interacted with the promoters of key genes involved in biofilm production such as *hrpR*, the alginate biosynthesis gene *alg44* and the type IV pilus assembly gene *pilM* **(Figure 3d)**. Furthermore, the transcription levels of *alg44*, *algX* (encoding the alginate biosynthesis protein) and *pilM* were markedly enhanced in ΔPSPPH3268. This resulted in enhanced biofilm formation and EPS production when the PSPPH3268 gene was deleted and then restored when PSPPH3268 was expressed **(Figure 3d)**. These results demonstrated that PSPPH3268 acts as a master regulator in various virulence-related pathways. In addition, we identified that the TF PSPPH3798 binds to the promoters of flagellar-related genes (*fliK*, *fliE*, *fliD* and *fleQ*; **Figure S4e**).

In addition to the 'one-step trigger' mechanism, we found that TFs also regulate downstream genes through one or two other TFs at different levels, which were regarded as 'one jump-point trigger' and 'two jump-point trigger'. For example, PSPPH2555 indirectly influenced biofilm formation (*algD*), motility (*fleQ*), T3SS (*hrpR*), QS (*ahlR* and *secE*) and phytotoxin production (*aprD*) by directly regulating the bottom-level TF PSPPH4920 **(Supplementary Table 3)**. PSPPH3504 was found to be involved in a PSPPH3504-Lon/PSPPH4324/PSPPH4844-PSPPH0755/PSPPH4920/PSPPH4012-target gene pathway **(Figure S4f, Supplementary Table 3, '/' represents sibling nodes and '-' represents downward regulation)**. Among these TFs, PSPPH0755, PSPPH4920 and PSPPH4012 are considered key performer TFs because they mediate most of the transcription regulatory signals from multiple TFs. We also found reverse

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

regulatory pathways in our network. For example, the middle-level TF PSPPH4673 was found to directly regulate the top-level TF PSPPH4700 and then indirectly control the transcription of many virulence-related genes through regulating the bottom-level TFs PSPPH0755, PSPPH4920, Anr, and PSPPH2476 **(Supplementary Table 3)**. In summary, TFs regulate the pathogenicity of *P. syringae* through diverse pathways, either by directly binding to target genes or indirectly controlling other TFs.

Systematic mapping of TF targets revealed key metabolic regulators in *P. syringae*

In addition to enhancing pathogenicity and resisting host defences, *P. syringae* adjusts its metabolic activities to survive in unpredictable environments(Rico, McCraw, & Preston, 2011). To comprehensively understand metabolic regulation in *P. syringae*, we constructed a hierarchical network that includes key regulators and the genes they trigger, similar to the virulence hierarchical network. We focused on eight metabolic pathways, namely amino acid biosynthesis, DNA replication, ATP-binding cassette (ABC) transportation, oxidative phosphorylation, tricarboxylic acid (TCA) cycle, RNA polymerase, phosphonate metabolism and methyl-accepting chemotaxis **(Figure 4a)**. Compared with the virulence network shown in Figure 3A, we identified more TFs involved in metabolic regulation, many of which exhibited numerous interactions with genes related to oxidative phosphorylation (178 binding peaks) and the TCA cycle (154 binding peaks; **Figure S5a–c**). In a previous study, 12 master regulators were reported to control various metabolic pathways, including LexA1, PSPPH3004

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and PSPPH1960, involved in reactive oxygen species (ROS) resistance(Fan et al., 2020). RhpR participates in several metabolic pathways, such as ABC transporters and oxidoreductase activity(Shao et al., 2019). Lon is involved in glucokinase and oxidoreductase activity(Hua et al., 2020). MgrA, GacA, PilR, PsrA, RpoN, CvsR, OmpR and CbrB2 participate in oxidation resistance, amino acid transportation and other metabolic pathways(Shao et al., 2021). Here, we identified 111 TFs regulating these eight metabolic pathways. Similar to the aforementioned virulence network, three transcriptional regulatory channels were observed for the metabolic pathways. To provide a detailed view of metabolic regulation in *P. syringae*, we counted the number of functionally annotated genes related to each pathway and calculated the proportion of targets for each TF, highlighting key regulators in these eight metabolic pathways using radar plots **(Figures 4b and S6a)**.

We found that the TFs CysB and PSPPH3268 regulate all eight metabolic pathways, whereas the TFs PSPPH1951, PSPPH3798, PSPPH3504 and PSPPH4700 were predicted to regulate seven metabolic pathways. Notably, the TF PSPPH0755 was found to bind to the promoters of PSPPH5210 (encoding ATP synthase F0F1 subunit delta) and PSPPH3109 (encoding the NADH dehydrogenase subunit A NuoA). The monomer motif (CTGAA) of PSPPH0755 was identified through MEME analysis. The interactions in these two metabolic pathways were confirmed through EMSA **(Figure S6b)**. The TF PSPPH3798 was predicted to bind to the promoters of genes in two pathways, including PSPPH3881 (encoding the methyl-accepting chemotaxis protein WspA) and PSPPH5119 (encoding the phosphate transport system regulatory

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

protein PhoU). The 15-bp binding motif of PSPPH3798 was determined to have a head-to head orientation (ATCG-N7-CGAT). EMSA results confirmed these interactions **(Figure S6c)**. In addition to the TF PSPPH3798, the TF PSPPH4638 had a binding site in the promoter region of the PSPPH3881 gene. PSPPH4638 was also predicted to interact with PSPPH0550 that encodes phosphoserine phosphatase SerB. The 8-bp monomer motif (ATTTTCAA) of PSPPH4638 was identified, and the binding interactions were confirmed through EMSA **(Figure S6d)**.

In yeast, TFs in a functional category appear to bind to genes in the same category(Simon et al., 2001), and we observed a similar pattern in *P. syringae*. For example, TF Irp, a key regulator in the methyl-accepting pathway, bound to the promoters of PSPPH3798, PSPPH4638 and PSPPH4700, which were also identified as key regulators in the same pathway **(Figure 4c)**. In addition, we found that TFs from different categories often bound to the promoters of TFs responsible for other cellular processes. For example, key regulators controlling oxidative phosphorylation (highlighted in red and blue; PSPPH1951, PSPPH2193, PSPPH2832, PSPPH3268, PSPPH3504, PSPPH3798, PSPPH4638, PSPPH4673 and PSPPH4700) bound to TFs playing key roles in the methyl-accepting pathway (highlighted in brown and blue; PSPPH1951, PSPPH3268, PSPPH3504, PSPPH3798, PSPPH4638, PSPPH4700 and Irp; **Figure 4d)**. These results demonstrate that many regulatory processes are often achieved through coregulation by a series of multifunctional TFs throughout the network, enabling *P. syringae* to coordinate transcriptional regulation processes across multiple cellular processes.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

TFs indicated large functional variability across different pathovars in *P. syringae*

Although TF functions exhibit both inter- and intra-species variability, most previous studies on TFs have focused on the molecular mechanism of a single strain(Galardini et al., 2015). To investigate the regulatory mechanism of TFs across different strains of *P. syringae*, we selected four model strains: *P. syringae* pv. *syringae* 1448A, *P. syringae* pv. *tomato* DC3000, *P. syringae* pv. *syringae* B728a and *P. syringae* pv. *actinidiae* C48. We used the genome of 1448A as a reference and conducted a homology analysis of 1448A protein sequences with those of the other three strains. We determined a high proportion of homologous proteins in the three strains (4983 in DC3000, 4982 in B728a and 4984 in C48; **Supplementary Table 4**). Across the four strains, all 301 annotated TFs were present. We selected five TFs (Irp, PSPPH2193, PSPPH3122, PSPPH4127 and OmpR) to construct TF-overexpressing strains in DC3000, B728a and C48 before performing ChIP-seq analysis. We identified the binding sites of all these TFs in the four strains and found divergent binding preferences for the same TFs in different strains. Most target genes of each TF in one or two strains were unique. In particular, Irp bound to 19 target genes that were conserved in all four tested strains, including *purB* (encoding adenylosuccinate lyase), *cceA2* (encoding the chemotaxis sensor histidine kinase) and *gidA* (encoding the tRNA uridine 5-carboxymethylaminomethyl modification protein). Four highly conserved target genes were also found to directly interact with PSPPH4127. Evolutionary analysis of the binding peaks of TFs suggested high binding specificity

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and varying levels of conservation of these TFs in the tested strains (**Figure 5a**).

In addition to the intersection between 1448A and C48, we observed differences between the target genes of all five TFs (**Figure 5b**) and the peak locations (TF target interactions; **Figure 5c**) in these four pathovars. The inconsistency between the number of targets and peaks suggested that some target genes were regulated by at least one different TF in these four strains, which is similar to regulation in *P. aeruginosa*(Trouillon et al., 2021). To confirm the presence of target genes regulated by the same TF or different TFs in various strains, we compared the peak locations of *gidA* and *rpoD* (RNA polymerase sigma factor) as two examples. TF Irp was found to bind to the promoter of *gidA* in four strains (**Figure 5d**) and had 19 conserved target genes in these strains (**Supplementary Table 4**). In addition, PSPPH4127 had four conserved target genes (*rpoD*, PSPPHPPH_1001, PSPPHPPH_1998 and PSPPHPPH_5016) in all four strains (**Figure 5e**). These results showed that Irp and PSPPH4127 exhibited higher functional conservation than the other three TFs. We also found that PSPPH2193 in 1448A and Irp in DC3000 bound to the promoter of *rpoD* (**Figure 5e**). Differences in the regulation of the same targets by different TFs were also observed in more than 1,500 target genes, suggesting the potential diversity of the transcriptional regulation of TFs in our network.

Topological modularity of the transcription regulatory network exhibited various functions in biological processes in *P. syringae*

Complex networks in nature often exhibit topological and/or functional

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

modularity(Dittrich, Klau, Rosenwald, Dandekar, & Müller, 2008; Olesen, Bascompte, Dupont, & Jordano, 2007). We used a partitioning algorithm (with a resolution of 0.9) to classify network elements into different subsets using Gephi. Our analysed TFs and their target genes were divided into 16 modules, each represented by a different colour **(Figure 6a)**. Each module contained 2.7% to 12.1% of the total elements and exhibited correlations with each other. We found that almost all nodes in the network had connections both within and between modules, indicating that the 16 modules were not isolated and contributed to extensive information flow throughout the network to regulate transcription in *P. syringae* **(Figure 6b)**. Module 12 appeared to play a central role in facilitating large information flows with other modules. Module 15 also exhibited transcriptional information exchanges both between modules and within the same module.

Among the 16 modules, Module 2 was involved in most nodes (443 elements, including 41 TFs and 402 target genes). Module 3 contained the least TFs (181 elements, including two TFs and 179 target genes). To investigate the potential correlation between topological modularity and biological functions, we performed GO term and KEGG pathway enrichment analysis for each module. As expected, 15 modules were enriched in specific biological functions **(Figure 6c and d)**. In some modules, hundreds of elements were assessed. For example, genes in Module 2 (443 nodes), Module 12 (327 nodes) and Module 14 (432 nodes) were mainly enriched in the regulation of transcription and DNA binding **(Figure 6c)**. DksA, a TF in Module 2, played a key role in regulating transcription-coupled DNA repair(Meddows, Savory,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Grove, Moore, & Lloyd, 2005) and also participated in oxidative phosphorylation, amino sugar and nucleotide sugar metabolism and RNA degradation in *P. syringae*. Twenty TFs (such as CapA, CapB, CysB, FruR and MarR) classified in Module 12 were identified to be involved in transcriptional regulation. The TF PSPPH3798 located in Module 14 was observed to be involved in flagellar assembly, and these interactions were confirmed by EMSA (**Figure S4b**). The genes in Module 4 were enriched in oxidoreductase activity, and our analysis revealed that the TF PSPPH0755 played a role in regulating the oxidative phosphorylation pathway **(Figure 6d and S6b)**. MexT in Module 5, which was previously associated with motility in *P. syringae*(Kawakita et al., 2012), was found to participate in in biofilm formation and QS pathways in our study. In addition, we not only identified the crucial roles of the TFs PSPPH1951 (Module 6), PSPPH2193 (Module 11) and PSPPH3268 (Module 15) in T3SS pathways, bacterial motility and biofilm formation, respectively (**Figure 3b–d**), but also reported their potential biological functions in aminoacyl-tRNA biosynthesis (PSPPH1951), RNA degradation (PSPPH2193) and catalytic activity (PSPPH3268) (**Figure 6d**). Our results allowed us to identify the potential regulators of specific pathways and perform functional predictions for hypothetical proteins in *P. syringae*. For instance, PSPPH1503 in Module 15, which encodes a hypothetical protein, was possibly correlated with glycerophospholipid metabolism.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Most microbial studies on genome-wide transcriptional regulatory network focus on *S. cerevisiae* and *E. coli*, which reveal the principles of architecture and interactions of their regulatory networks. The analysis of the transcription regulatory associations in *S. cerevisiae* mainly rely on the databases such as YEASTRACT (YEAst Search for Transcriptional Regulators And Consensus Tracking)(Teixeira et al., 2018). In *E.coli*, relative complete transcriptional regulatory network has been generated through integrating three different data sources (RegulonDB, Ecocyc and TRN-SO)(Ma et al., 2004). However, few study has yet comprehensively evaluated TFs in other prokaryotic species throughout a genome(Ishihama et al., 2016). In this study, we successfully generated the most complete transcriptional regulatory network and data source, which profiled the transcriptional regulatory features of both the aforementioned 100 TFs and an additional 170 TFs in *P. syringae* through ChIP-seq. By mapping the TF-target hierarchical regulatory networks, we identified several novel master regulators involved in significant biological processes. Furthermore, our evolutionary analysis and assessment of the topological functional modularity of TFs and their respective targets revealed the evolutional conservation and functional diversity of TFs in *P. syringae*.

Although transcriptional regulatory networks are considered conserved(Perez & Groisman, 2009), many studies reveal highly functional variability of TFs in inter- and intra-species(Galardini et al., 2015). These observed diversities between different strains of the same species mainly result from the expression levels of TFs, contents

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of target genes, and differences of binding sequences(Trouillon et al., 2021). In our study, we observed large differences of DNA-binding characteristic of TF Irp between the C48 strain and the DC3000 strain. The functional diversity of TFs may arise from the large difference in the contents of target genes and TFs, which are regarded as the main determinant of transcriptional regulatory(van Duin, Krautz, Rennie, & Andersson, 2023), although Irp display high homology in these pathovars.

Collaborations between TFs at higher levels (top and middle) were enriched, a pattern similar to the tendencies observed in human TFs(Gerstein et al., 2012). In particular, TT TF pairs exhibited a greater degree of cooperative gene regulation, whereas TB TF pairs accounted for nearly half of direct interactions within all communications. Furthermore, we observed that both direct physical regulation and cooperative interactions were the least common among MM TF pairs. By contrast, in humans, direct regulation tends to occur between TT or TM TF pairs. Furthermore, interactions between TFs, in any form, within human and yeast transcriptional regulatory networks are more likely to appear between middle TFs, which act as information transfer centres(Bhardwaj, Yan, & Gerstein, 2010). This may be attributed to the higher abundance of bottom-level TFs than higher-level TFs observed in prokaryotic microorganisms, a pattern also found in *E. coli*(Bhardwaj et al., 2010). This finding indicates that the bottom-level TFs that are more likely to be regulated are evolutionarily preferred in multicellular eukaryotic organisms. However, in *E. coli*, the deficient bottom-level TFs are more commonly co-associated with other same-level TFs. When comparing co-associated TF pairs and the cooperativity of TFs,

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

we observed a distinct and inverse relationship in *P. syringae* compared with yeast or *E. coli* (**Figure 1B**). The enriched cooperativity of bottom-level TFs with high co-associated scores indicated that these bottom-level TFs preferred to coregulate target genes by binding to the same peak locations. Notably, seven TFs without correlations with other TFs appeared to independently participate in biological processes. These findings not only shed light on the inherent properties of direct regulation and co-association across various species but also indicate the unique characteristics of *P. syringae* to response to dynamic environmental variations.

The fundamental units of a transcriptional regulatory network are positive and negative loops. For a more comprehensive description, these regulatory units can be classified into six submodules, namely autoregulation, multicomponent loops, feedforward loops, single-input, multi-input and regulator chain(Lee et al., 2002). In yeast, only 10 TFs were found to autoregulate themselves, whereas the majority of regulatory units among 116 TFs in *E. coli* exhibit autoregulation(Thieffry, Huerta, Pérez-Rueda, & Collado-Vides, 1998). Similarly, we identified 92 autoregulators in our transcriptional regulatory network, which may function as auto-inhibitors in *E. coli*. Feedforward loops are highly prevalent in prokaryotic transcriptional regulatory networks, such as human and yeast(Gerstein et al., 2012; Lee et al., 2002). The 696 feedforward loops (M13) identified in our study highlighted this cooperative regulation of TFs in response to small signals in prokaryotic species. This enriched submodule is regarded as a temporal switch that provides constant feedback to respond rapidly to signal impacts(Shen-Orr et al., 2002). Multistep regulation assists

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

master regulators in enhancing the initial information flow(Goldbeter & Koshland Jr, 1984). Unlike M3, the most abundant module (868) in the human TF regulatory network, we observed that M1 was the most prevalent module (24,479) in the *P. syringae* transcriptional regulatory network. This finding indicates that *P. syringae* prefers achieving transcriptional regulation through few TFs to ensure rapid transmission and response to environmental signals.

Despite more than 20 years of research, our understanding of the global regulatory network in the plant pathogen *P. syringae* remains inadequate. Our previous studies reported 16 key virulence-related regulators(Shao et al., 2021), 25 master virulence-related regulators(Fan et al., 2020) and seven global regulators acting as TCSs(Xie et al., 2022). However, the regulatory relationships and functional crosstalk among all TFs in *P. syringae* remain unknown. In this study, we integrated all available interaction information concerning almost all TFs in *P. syringae* and mapped the first comprehensive transcriptional regulatory network in this plant pathogen (**Figure 7**). This network offers a global view of the multiple functions of TFs in *P. syringae*. We also identified 35 vital TFs that participate in virulence pathways and 111 key TFs involved in metabolic pathways across the global transcriptional regulatory network in *P. syringae*. This analysis uncovered new functions of previously characterised TFs, such as MexT. In addition to the previously reported *mexEF-oprN* operon,(Sawada et al., 2018) this study identified *fleQ* (a flagellar regulator) and *shcF* (a type III chaperone protein) as the targets of MexT. In addition, we investigated the functional evolution and potential intra-species variability of TFs

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in *P. syringae*, demonstrating the functional diversity of TFs among *P. syringae* species during their long course of evolution. Based on the aforementioned results, we established the *P. syringae* transcriptional regulatory network (PSTRnet) database, which contains detailed binding peak information for all TFs in *P. syringae* (https://jiadhuang0417.shinyapps.io/PSTF_NET/). This database serves as a valuable platform for presenting, searching and downloading regulatory information of transcription in *P. syringae*.

We observed crosstalk between not only TFs but also various pathways. For example, the TF PSPPH4700 directly regulated *fleQ* and *hrpR* and indirectly regulated *fleQ* and *hrpR* through the PSPPH4700/PSPPH4324/PSPPH0755 cascade. Feedforward loop modules are usually coherent, meaning that the direct effect of downstream TFs has the same regulatory direction (positive or negative) as the indirect effect of upstream TFs(Shen-Orr et al., 2002). This observation enhanced our understanding of the influence of TFs in the network on their target genes based on the identified effects of other TFs. Furthermore, the TF PSPPH4700 was identified to bind to the promoters of genes involved in seven metabolic pathways, namely the TCA cycle, oxidative phosphorylation, methyl-accepting, amino acid biosynthesis, RNA polymerase, ABC transportation and DNA replication. Our results proved the possibility of crosstalk between different pathways in *P. syringae*.

Taken together, our study provides comprehensive insights into the DNA-binding characteristics and potential regulatory pathways of almost all annotated TFs in *P. syringae*. The global transcriptional regulatory network can not only contribute to the

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

development of novel drugs to combat *P. syringae* infections but also advance research on the molecular mechanisms of TFs in other pathogens.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

Bacterial strains, culture media, plasmids, and primers

The bacterial strains, plasmids and primers used in this study are listed in **Supplementary Table 5**. *P. syringae* 1448A, DC3000, B728a and C48 strains were grown at 28 °C in King' B (KB) medium shaking at 220 rpm or KB agar plates(King, Ward, & Raney, 1954). *E. coli* BL21(DE3) or DH5α strains were grown at 37 °C in Luria-Bertani medium (LB) shaking at 220 rpm or on LB agar plates. Antibiotics used for *P. syringae* 1448A, DC3000 and B728a wide type (WT) strains and mutants were rifampicin at 50 μg/ml; *P. syringae* 1448A, DC3000 and B728a overexpression strains were rifampicin at 50 μg/ml and spectinomycin at 50 μg/ml, C48 overexpression strains were spectinomycin at 50 μg/ml; *P. syringae* 1448A strains with pK18mobsacB plasmids for mutant construction were rifampicin at 50 μg/ml and kanamycin at 50 μg/ml. Antibiotics used for *E. coli* with pET28a plasmids for protein purification, and pK18*mobsacB* plasmids were kanamycin at 50 μg/ml; *E. coli* with pHM1 plasmids for overexpression strain construction were spectinomycin at 50 μg/ml.

Construction of overexpression strains and mutants

Overexpression strains and mutants of *P. syringae* were constructed as previously described(Kvitko & Collmer, 2011; Shao et al., 2021). In brief, for overexpression strain, the open reading frame (Caillet, Baron, Boni, Caillet-Saguy, & Hajnsdorf) of each TF-coding gene was amplified by PCR from *P. syringae* genome and cloned into

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pHM1 plasmid. The ligated fragments were inserted into the digested pHM1 plasmids (*Hind*III) using ClonExpress MultiS One Step Cloning Kit (Vazyme). The recombinant plasmids were transformed into the *P. syringae* wild-type strain. The single colonies were confirmed by western blot. For mutants, the upstream (~1500-bp) and downstream (~1000-bp) fragments of TF ORF were amplified by PCR from the *P. syringae* genome and digested with *Xba*I respectively, and then ligated by T4 DNA ligase (Jumper et al.). The ligated fragments were inserted into the digested pK18*mobsacB* plasmids (*Xba*I) using ClonExpress MultiS One Step Cloning Kit (Vazyme). The recombinant plasmids were transformed into the *P. syringae* wild-type strain. The single colonies were selected to the sucrose plates, and further screened in two kinds of KB plates (with kanamycin/rifampin and with rifampin) concurrently. The single colonies losing kanamycin resistance were further verified by real-time quantitative PCR (RT-qPCR) to detect the mRNA level of corresponding TF genes.

ChIP-seq analyses

As previous description(Blasco et al., 2012), the overexpression strains of corresponding TFs and *P. syringae* WT with empty pHM1 plasmid was cultured in 30 mL KB medium to OD_{600}=0.6. Bacterial cultures were cross-linked with 1% formaldehyde for 10 min at 28°C and then the reaction was stopped by the addition of 125 mM glycine for 5 min. The centrifugated bacteria were washed twice with Tris Buffer (20 mM Tris-HCl [pH 7.5] and 150 mM NaCl) and washed again with 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).

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The centrifugated bacteria were preserved at 80℃ or continued for the next experiments. The bacteria were subjected with IP Buffer and then sonicated to pull down the DNA fragments (150-300-bp). The supernatant was the DNA-TF-HA tag complex and used as IP samples. IP experiments and control sample were incubated with agarose conjugated anti-HA antibodies (Sigma) in IP Buffer. The complex of DNA-TF-anti-HA agarose was applied to washing, crosslink reversal, and purification(Blasco et al., 2012). The 150-250-bp DNA fragments were selected for library construction. The libraries were sequenced using the HiSeq 2000 system (Illumina). Two biological replications have been performed for all ChIP-seq experiments. ChIP-seq reads were mapped to the *P. syringae* 1448A genome (NC_005773.3), DC3000 (NC_004578.1), B728a (NC_007005.1) and C48 (NZ_CP032631.1) using Bowtie2 (version 2.3.4.3)(Zhang et al., 2008). Subsequently, binding peaks ($q < 0.01$) were identified using MACS2 software (version 2.1.0). The enriched loci for each TF were annotated using the R package ChIPpeakAnno (version 3.18.2)(L. J. Zhu et al., 2010).

Electrophoretic mobility-shift assay

DNA probes were amplified from *P. syringae* 1448A genome by PCR using primers listed in Supplementary Table 5. The probes (20 ng) were mixed with various concentrations of proteins in 20 μL of gel shift buffer (10 mM Tris-HCl, pH 7.4, 50 mM KCl, 5 mM $MgCl_2$, 10% glycerol). After incubation at room temperature for 20 min, the samples were analyzed by 6% native polyacrylamide gel electrophoresis (90 V for 90 min for sample separation). The gels were subjected to Gel Red dye (Tiangen

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Biotech) for 5 min and photographed by using a gel imaging system (Bio-Rad). The assay was repeated at least three times with similar results.

RT-qPCR

The RT-qPCR primers used are shown Supplementary Table 5 in the Supplemental Information. The cultured bacteria were grown to OD_{600}=0.6 and the total RNA were purified with Bacteria Total RNA Isolation Kit (Sangon Biotech). The RNA concentrations were measured using a NanoDrop 2000 spectrophotometer (ThermoFisher). cDNA synthesis was performed using HiScript III RT SuperMix (Vazyme, China). RT-qPCR was performed with a SuperReal Premix Plus (SYBR Green) kit (Vazyme, China) according to the manufacturer's instructions. The reactions used 100 nM primers and were run for 40 cycles at 95°C for 30 s and 95°C for 10 s, and at 60°C for 30 s. The fold change represents relative expression level of mRNA, which can be estimated by the values of $2^{-(\Delta\Delta Ct)}$. The relative expression of target genes in WT was set to 1. All the reactions were conducted with three repeats.

Motility assay

The motility assay was performed based on our previous study(Shao et al., 2021). Swimming plates were KB agar plates containing 0.3% agar (MP Biomedicals, UK) and rifampicin at 25 μg/ml. Overnight bacterial cultures were inoculated on swimming plates as 2 μL aliquots and incubated at room temperature for 3-5 days. Finally, the diameter of motility trace represented the swimming motility of *P. syringae* strains. Photographs were taken by using the Bio-Rad imaging system. The

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

assay was repeated at least three times with similar results.

Congo red assay

Congo red assay was performed as previous study(Shao et al., 2021). Congo red plates were KB agar plates containing 1.0% agar (MP Biomedicals, UK) and rifampicin at 25 μg/ml. Overnight bacterial cultures were inoculated on Congo red plates as 2 μL aliquots and incubated at 28℃. The colony staining was photographed after 5-7 days. The assay was repeated at least three times with similar results.

Biofilm formation assay

Biofilm formation assay was performed as previously describe(Shao et al., 2019). Overnight bacterial cultures were transferred to a sterile 10 mL borosilicate tube containing 2 ml KB medium with rifampicin (25 μg/ml) with the original concentration OD_{600} = 0.1. The cultures grew at room temperature for 36 hr without shaking. 0.1% crystal violet was used to stain the biofilm adhered to the tube for 20 min without shaking. Tubes were washed for more than three times with distilled deionized water gently and other components on the tube loosely was washed off. The tubes were dry and photographed. The remaining crystal violet was fully dissolved in 1 ml 95% ethanol with constantly shaking and measured its optical density at 590 nm (Biotek microplate reader). The assay was repeated at least three times with similar results.

Network and functional enrichment analysis

Network analyses were performed on Gephi 0.10. Functional annotations were

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

retrieved from the *Pseudomonas* database and GO functional enrichment analyses and KEGG analysis were performed using DAVID v6.8.

Statistical analysis

Two-tailed Student's *t* tests were performed using Microsoft Office Excel 2010. $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$ and results represent means ± SD. All experiments were repeated at least twice.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data Availability

Sequencing data have been deposited and publicly available in Gene Expression Omnibus (GEO) under accession number GSE247395. Source data contain the numerical data used to generate the figures of EMSA. Hierarchical information and functional categories of TFs are available in **Supplementary Table 1-3**. Evolutionary details are provided in **Supplementary Table 4**. Primers and strains used in this paper are provided in **Supplementary Table 5**.

Funding

This study was supported by grants from the National Natural Science Foundation of China (31670127 and 31870116 to X.D.), General Research Funds of Hong Kong (21103018, 11101619, and 11102720 to X.D.). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication.

Acknowledgments

X.D., Y.S., J.W. and J. H. conceived the project. Y.S., J.W., J.H., S.L. and Y.L. carried out experiments. J.H, S.L. and Y.L. performed data analysis. X.D., Y.S, J.W., and J.H. wrote the paper.

Ethics declarations

Competing interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Alfano, J. R., & Collmer, A. (1997). The type III (Hrp) secretion pathway of plant pathogenic bacteria: trafficking harpins, Avr proteins, and death. *Journal of bacteriology*, *179*(18), 5655-5662.

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450-461.

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the national academy of sciences*, *97*(21), 11149-11152.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., . . . Chen, X. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science*, *324*(5935), 1720-1723.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(suppl_2), W202-W208.

Bender, C. L., Alarcón-Chaidez, F., & Gross, D. C. (1999). *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and molecular biology reviews*, *63*(2), 266-292.

Bhardwaj, N., Yan, K.-K., & Gerstein, M. B. (2010). Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proceedings of the national academy of*

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sciences, 107(15), 6841-6846.

Blasco, B., Chen, J. M., Hartkoorn, R., Sala, C., Uplekar, S., Rougemont, J., . . . Cole, S. T. (2012). Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. *PLoS Pathogens, 8*(3), e1002621.

Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., . . . Kolonay, J. F. (2003). The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proceedings of the national academy of sciences, 100*(18), 10181-10186.

Burda, Z., Krzywicki, A., Martin, O. C., & Zagorski, M. (2011). Motifs emerge from function in model gene regulatory networks. *Proceedings of the national academy of sciences, 108*(42), 17263-17268.

Caillet, J., Baron, B., Boni, I. V., Caillet-Saguy, C., & Hajnsdorf, E. (2019). Identification of protein-protein and ribonucleoprotein complexes containing Hfq. *Sci Rep, 9*(1), 14054. doi:10.1038/s41598-019-50562-w

Chan, S. S.-K., & Kyba, M. (2013). What is a master regulator? *Journal of stem cell research & therapy, 3*.

Chatterjee, A., Cui, Y., Yang, H., Collmer, A., Alfano, J. R., & Chatterjee, A. K. (2003). GacA, the response regulator of a two-component system, acts as a master regulator in *Pseudomonas syringae* pv. *tomato* DC3000 by controlling regulatory RNA, transcriptional activators, and alternate sigma factors. *Molecular Plant-Microbe Interactions, 16*(12), 1106-1117.

Cunnac, S., Lindeberg, M., & Collmer, A. (2009). *Pseudomonas syringae* type III

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

secretion system effectors: repertoires in search of functions. *Current opinion in microbiology, 12*(1), 53-60.

Deng, X., Lan, L., Xiao, Y., Kennelly, M., Zhou, J.-M., & Tang, X. (2010). *Pseudomonas syringae* two-component response regulator RhpR regulates promoters carrying an inverted repeat element. *Molecular Plant-Microbe Interactions, 23*(7), 927-939.

Deng, X., Liang, H., Chen, K., He, C., Lan, L., & Tang, X. (2014). Molecular mechanisms of two-component system RhpRS regulating type III secretion system in *Pseudomonas syringae*. *Nucleic Acids Research, 42*(18), 11472-11486.

Deng, X., Xiao, Y., Lan, L., Zhou, J.-M., & Tang, X. (2009). *Pseudomonas syringae* pv. *phaseolicola* mutants compromised for type III secretion system gene induction. *Molecular Plant-Microbe Interactions, 22*(8), 964-976.

Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T., & Müller, T. (2008). Identifying functional modules in protein–protein interaction networks: an integrated exact approach. *Bioinformatics, 24*(13), i223-i231.

Fan, L., Wang, T., Hua, C., Sun, W., Li, X., Grunwald, L., . . . Yin, Y. (2020). A compendium of DNA-binding specificities of transcription factors in *Pseudomonas syringae*. *Nature Communications, 11*(1), 4947.

Fishman, M. R., Zhang, J., Bronstein, P. A., Stodghill, P., & Filiatrault, M. J. (2018). Ca2+-induced two-component system CvsSR regulates the type III secretion system and the extracytoplasmic function sigma factor AlgU in *Pseudomonas*

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

syringae pv. *tomato* DC3000. *Journal of bacteriology, 200*(5), 10.1128/jb.00538-00517.

Fortuna, A., Collalto, D., Schiaffi, V., Pastore, V., Visca, P., Ascenzioni, F., . . . Leoni, L. (2022). The *Pseudomonas aeruginosa* DksA1 protein is involved in H2O2 tolerance and within-macrophages survival and can be replaced by DksA2. *Scientific Reports, 12*(1), 10404.

Galardini, M., Brilli, M., Spini, G., Rossi, M., Roncaglia, B., Bani, A., . . . Bacci, G. (2015). Evolution of intra-specific regulatory networks in a multipartite bacterial genome. *PLoS Computational Biology, 11*(9), e1004478.

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., . . . Alexander, R. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature, 489*(7414), 91-100.

Goldbeter, A., & Koshland Jr, D. E. (1984). Ultrasensitivity in biochemical systems controlled by covalent modification. Interplay between zero-order and multistep effects. *Journal of Biological Chemistry, 259*(23), 14441-14447.

He, S. Y. (1998). Type III protein secretion systems in plant and animal pathogenic bacteria. *Annual review of phytopathology, 36*(1), 363-392.

Hendrickson, E. L., Guevera, P., & Ausubel, F. M. (2000). The alternative sigma factor RpoN is required for *hrp* activity in *Pseudomonas syringae* pv. *maculicola* and acts at the level of *hrpL* transcription. *Journal of bacteriology, 182*(12), 3508-3516.

Hirano, S. S., & Upper, C. D. (2000). Bacteria in the leaf ecosystem with emphasis on

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pseudomonas syringae—a pathogen, ice nucleus, and epiphyte. *Microbiology and molecular biology reviews, 64*(3), 624-653.

Hua, C., Huang, J., Wang, T., Sun, Y., Liu, J., Huang, L., & Deng, X. (2022). Bacterial transcription factors bind to coding regions and regulate internal cryptic promoters. *MBio, 13*(5), e01643-01622.

Hua, C., Wang, T., Shao, X., Xie, Y., Huang, H., Liu, J., . . . Jiang, L. (2020). *Pseudomonas syringae* dual-function protein Lon switches between virulence and metabolism by acting as both DNA-binding transcriptional regulator and protease in different environments. *Environmental Microbiology, 22*(7), 2968-2988.

Huang, J., Yao, C., Sun, Y., Ji, Q., & Deng, X. (2022). Virulence-related regulatory network of *Pseudomonas syringae*. *Computational and Structural Biotechnology Journal*.

Ichinose, Y., Taguchi, F., & Mukaihara, T. (2013). Pathogenicity and virulence factors of *Pseudomonas syringae*. *Journal of general plant pathology, 79*, 285-296.

Ishihama, A., Shimada, T., & Yamazaki, Y. (2016). Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Research, 44*(5), 2058-2074.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature, 407*(6804), 651-654.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., . . . Wei, G. (2013). DNA-binding specificities of human transcription factors. *Cell, 152*(1),

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

327-339.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *nature*, *596*(7873), 583-589. doi:10.1038/s41586-021-03819-2

Kawakita, Y., Taguchi, F., Inagaki, Y., Toyoda, K., Shiraishi, T., & Ichinose, Y. (2012). Characterization of each *aefR* and *mexT* mutant in *Pseudomonas syringae* pv. *tabaci* 6605. *Molecular genetics and genomics*, *287*, 473-484.

King, E. O., Ward, M. K., & Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescin. *The Journal of laboratory and clinical medicine*, *44*(2), 301-307.

Kvitko, B. H., & Collmer, A. (2011). Construction of *Pseudomonas syringae* pv. *tomato* DC3000 mutant and polymutant strains. *Plant immunity: methods and protocols*, 109-128.

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., . . . Weirauch, M. T. (2018). The human transcription factors. *Cell*, *172*(4), 650-665.

Lan, L., Deng, X., Zhou, J., & Tang, X. (2006). Genome-wide gene expression analysis of *Pseudomonas syringae* pv. *tomato* DC3000 reveals overlapping and distinct pathways regulated by *hrpL* and *hrpRS*. *Molecular Plant-Microbe Interactions*, *19*(9), 976-987.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., . . . Simon, I. (2002). Transcriptional regulatory networks in *Saccharomyces*

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cerevisiae. *Science*, *298*(5594), 799-804.

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell*, *152*(6), 1237-1251.

Ma, H.-W., Kumar, B., Ditges, U., Gunzer, F., Buer, J., & Zeng, A.-P. (2004). An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Research*, *32*(22), 6643-6649.

Mathelier, A., Shi, W., & Wasserman, W. W. (2015). Identification of altered cis-regulatory elements in human disease. *Trends in Genetics*, *31*(2), 67-76.

Meddows, T. R., Savory, A. P., Grove, J. I., Moore, T., & Lloyd, R. G. (2005). RecN protein and transcription factor DksA combine to promote faithful recombinational repair of DNA double-strand breaks. *Molecular microbiology*, *57*(1), 97-110.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, *298*(5594), 824-827.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, *98*(2), 404-409.

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the national academy of sciences*, *104*(50), 19891-19896.

Papavassiliou, K. A., & Papavassiliou, A. G. (2016). Transcription factor drug targets.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Journal of cellular biochemistry, 117(12), 2693-2696.

Perez, J. C., & Groisman, E. A. (2009). Evolution of transcriptional regulatory circuits in bacteria. *Cell, 138*(2), 233-244.

Rico, A., McCraw, S. L., & Preston, G. M. (2011). The metabolic interface between *Pseudomonas syringae* and plant cells. *Current opinion in microbiology, 14*(1), 31-38.

Sawada, T., Eguchi, M., Asaki, S., Kashiwagi, R., Shimomura, K., Taguchi, F., . . . Toyoda, K. (2018). MexEF-OprN multidrug efflux pump transporter negatively controls N-acyl-homoserine lactone accumulation in *Pseudomonas syringae* pv. *tabaci* 6605. *Molecular genetics and genomics, 293*, 907-917.

Shao, X., Tan, M., Xie, Y., Yao, C., Wang, T., Huang, H., . . . Han, L. (2021). Integrated regulatory network in *Pseudomonas syringae* reveals dynamics of virulence. *Cell Reports, 34*(13).

Shao, X., Xie, Y., Zhang, Y., & Deng, X. (2019). Biofilm formation assay in *Pseudomonas syringae*. *Bio-protocol, 9*(10), e3237-e3237.

Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature genetics, 31*(1), 64-68.

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N. J., Volkert, T. L., . . . Jaakkola, T. S. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell, 106*(6), 697-708.

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410*(6825), 268-276.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Sun, Y., Shao, X., Zhang, Y., Han, L., Huang, J., Xie, Y., . . . Deng, X. (2022). Maintenance of tRNA and elongation factors supports T3SS proteins translational elongations in pathogenic bacteria during nutrient starvation. *Cell & Bioscience*, *12*(1), 1-17.

Taguchi, F., & Ichinose, Y. (2011). Role of type IV pili in virulence of *Pseudomonas syringae* pv. *tabaci* 6605: correlation of motility, multidrug resistance, and HR-inducing activity on a nonhost plant. *Molecular Plant-Microbe Interactions*, *24*(9), 1001-1011.

Teixeira, M. C., Monteiro, P. T., Palma, M., Costa, C., Godinho, C. P., Pais, P., . . . Pedreira, T. (2018). YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *46*(D1), D348-D353.

Thieffry, D., Huerta, A. M., Pérez-Rueda, E., & Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays*, *20*(5), 433-440.

Tian, Z.-X., Fargier, E., Mac Aogain, M., Adams, C., Wang, Y.-P., & O'Gara, F. (2009). Transcriptome profiling defines a novel regulon modulated by the LysR-type transcriptional regulator MexT in *Pseudomonas aeruginosa*. *Nucleic Acids Research*, *37*(22), 7546-7559.

Trouillon, J., Imbert, L., Villard, A.-M., Vernet, T., Attrée, I., & Elsen, S. (2021). Determination of the two-component systems regulatory network reveals core and accessory regulations across *Pseudomonas aeruginosa* lineages. *Nucleic*

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acids Research, *49*(20), 11476-11490.

van Duin, L., Krautz, R., Rennie, S., & Andersson, R. (2023). Transcription factor expression is the main determinant of variability in gene co-activity. *Molecular Systems Biology*, e11392.

Wade, J. T. (2015). Mapping transcription regulatory networks with ChIP-seq and RNA-seq. *Prokaryotic systems biology*, 119-134.

Wang, J., Shao, X., Zhang, Y., Zhu, Y., Yang, P., Yuan, J., . . . Chen, S. (2018). HrpS is a global regulator on type III secretion system (T3SS) and non-T3SS genes in *Pseudomonas savastanoi* pv. *phaseolicola*. *Molecular Plant-Microbe Interactions*, *31*(12), 1232-1243.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., . . . Cheng, Y. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome research*, *22*(9), 1798-1812.

Whitchurch, C. B., Tolker-Nielsen, T., Ragas, P. C., & Mattick, J. S. (2002). Extracellular DNA required for bacterial biofilm formation. *Science*, *295*(5559), 1487-1487.

Wilson, D., Charoensawan, V., Kummerfeld, S. K., & Teichmann, S. A. (2008). DBD-taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Research*, *36*(suppl_1), D88-D92.

Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., & Brinkman, F. S. (2016). Enhanced annotations and features for comparing thousands of

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pseudomonas genomes in the *Pseudomonas* genome database. *Nucleic Acids Research*, *44*(D1), D646-D653.

Xiao, Y., & Hutcheson, S. W. (1994). A single promoter sequence recognized by a newly identified alternate sigma factor directs expression of pathogenicity and host range determinants in *Pseudomonas syringae*. *Journal of bacteriology*, *176*(10), 3089-3091.

Xiao, Y., Lan, L., Yin, C., Deng, X., Baker, D., Zhou, J.-M., & Tang, X. (2007). Two-component sensor RhpS promotes induction of *Pseudomonas syringae* type III secretion system by repressing negative regulator RhpR. *Molecular Plant-Microbe Interactions*, *20*(3), 223-234.

Xie, Y., Ding, Y., Shao, X., Yao, C., Li, J., Liu, J., & Deng, X. (2021). *Pseudomonas syringae* senses polyphenols via phosphorelay crosstalk to inhibit virulence. *EMBO reports*, *22*(12), e52805.

Xie, Y., Li, J., Ding, Y., Shao, X., Sun, Y., Xie, F., . . . Deng, X. (2022). An atlas of bacterial two-component systems reveals function and plasticity in signal transduction. *Cell Reports*, *41*(3).

Xie, Y., Shao, X., & Deng, X. (2019). Regulation of type III secretion system in *Pseudomonas syringae*. *Environmental Microbiology*, *21*(12), 4465-4477.

Yan, Q., Rogan, C. J., Pang, Y.-Y., Davis, E. W., & Anderson, J. C. (2020). Ancient co-option of an amino acid ABC transporter locus in *Pseudomonas syringae* for host signal-dependent virulence gene regulation. *PLoS Pathogens*, *16*(7), e1008680.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Li, W. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome biology, 9*(9), 1-9.

Zhu, C., Byers, K. J., McCord, R. P., Shi, Z., Berger, M. F., Newburger, D. E., . . . Radhakrishnan, M. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome research, 19*(4), 556-566.

Zhu, L. J., Gazin, C., Lawson, N. D., Pagès, H., Lin, S. M., Lapointe, D. S., & Green, M. R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics, 11*, 1-10.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure legends

Figure 1. Hierarchical height and collaboration of TFs reveal the multiple regulatory patterns in *P. syringae*.

a, Close-up representation of 262 TFs hierarchy in *P. syringae*. Nodes depict TFs. Colors of edges represent source-bases. **b,** Enrichment of different collaborating (direct interaction, indirect interaction and cooperativity) TF-pairs at top (T), middle (M) and bottom (B) levels. Gray nodes below the graph represent TFs. The dashed orange line indicates the averaged level of collaboration. **c,** Thirteen Three-node sub-modules with the number of occurrences and an example. Spire loop is the most enriched sub-module. Edges represent the regulatory direction. **d,** Autoregulations are accompanied by the number of occurrences and 13 auto-regulators as examples.

Figure 2. Bottom TFs share the same binding sequences to coregulate in *P. syringae*.

a, The co-association map for 170 TFs in *P. syringae* shows the co-associated scores of binding peaks of these TFs (rows) that overlap each TF peak (columns). The three colored rectangles represent three different TF levels. C1-C4 represent four clusters of TFs according to the co-associated scores. The TFs in corresponding cluster are shown in the circle diagram. Orange nodes represent top TFs. Green nodes represent middle TFs. Purple nodes represent bottom TFs. The colors of edges are the mixture of two source TF colors. **b,** The heat-map of MexT indicates the associated scores of binding peaks of TFs in C3 (columns) that overlap the binding peaks (rows) of MexT.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PSPPH2411, PSPPH3643 and CysI are the top 3 TFs with high associated scores. **c,** UpSet plot shows the number of genes uniquely targeted TFs or co-targeted by multiple TFs in C3. The vertical black lines represent shared TF-binding sites. The y axis represents the number of overlapped binding sites across the linked TFs. The x axis represents the number of binding sites for each TF. Orange line represents the most enriched gene *cysI* that is co-targeted by 12 TFs in C3. **d,** Motifs of MexT and other 10 TFs in C3 which show similar binding sequences.

Figure 3. Virulence hierarchical regulatory network reveals 35 TFs involved in virulence.

a, Virulence hierarchical regulatory network shows the TF hierarchy and the large pool of target genes of multi-TF. Target genes are related with seven key virulence pathways, including biofilm formation, secondary messengers, motility, T3SS, QS, phytotoxin production and siderophore transporter. Orange nodes represent top TFs. Green nodes represent middle TFs. Purple nodes represent bottom TFs. Blue nodes represent target genes. Yellow edges represent downward point. Red edges represent upward point. **b,** The head-to-head binding motif of PSPPH1951, the validation of the binding sites of PSPPH1951 by EMSA, and the detection of expression of target gene *hopAE1* in WT, ΔPSPPH1951 and complementary strain by RT-qPCR. The validated binding sites are from the promoters of the *hrpR*, *hopAE1* and *hopAH2*. **c,** The binding motif of PSPPH2193 is head-to-head. EMSA confirms the direct binding of PSPPH2193 to the promoters of *fleQ* and *flhF*. RT-qPCR confirms the positive regulation of PSPPH2193 on the expression of *fleQ* and *flhF*. Motility assay validates

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the weaker motility of ΔPSPPH2193 than WT and complementary strain. **d,** The binding motif of PSPPH3268 is head-to-head. EMSA confirms the direct binding of PSPPH3268 to the promoters of *hrpR*, *alg44* and *pilM*. RT-qPCR confirms the negative regulation of PSPPH3268 on the expression of *alg44*, *algX* and *pilM*. Crystal violate staining assay and the quantification of biofilm formation validate the negative regulation of PSPPH3268 on the biofilm formation. Congo red assay confirms the negative regulation of PSPPH3268 on colony morphologies and EPS production. Student's *t* test. n.s., not significant, $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$.

Figure 4. Hundreds of TFs are identified to participate in metabolic pathways.

a, Metabolic hierarchical regulatory network shows the TF hierarchy and the large pool of target genes of multi-TF. Target genes are related with eight key metabolic pathways, including biosynthesis of amino acids, DNA replication, ABC transporter, oxidative phosphorylation, TCA cycle, RNA polymerase, phosphonate metabolism and methyl-accepting chemotaxis. Orange nodes represent top TFs. Green nodes represent middle TFs. Purple nodes represent bottom TFs. Blue nodes represent target genes. Yellow edges represent the direct interaction. **b,** Radar plots show the putative key regulators identified in six different metabolic pathways, including TCA cycle, oxidative phosphorylation, methyl-accepting, biosynthesis of amino acids, RNA polymerase and ABC transporter. Each radiation line represents a key regulator, and the radial length of the thick colored line is the rate of target genes to the associated genes, representing the significance of the enrichment of the TF target genes within each pathway. **c,** TFs involved in the methyl-accepting pathway bound to the

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

promoters of TFs in the same category. Brown nodes represent the TFs that are responsible for methyl-accepting pathway. The brown arrows point to the targeted TFs. **d,** TFs involved in the oxidative phosphorylation pathway bind to the promoters of TFs in the methyl-accepting pathway. Red nodes represent the TFs that are responsible for oxidative phosphorylation pathway. Brown nodes represent the TFs that are responsible for methyl-accepting pathway. Blue nodes represent the TFs involved in these two pathways. The arrows point to the targeted TFs and the arrow colors are source-based.

Figure 5. Various conservations are observed in TFs between different *P. syringae* pathovars.

a, Proportion of the TF target genes detected in one, two, three or four tested genomes. c1-c4 represent the conservation of targets in one, two, three and four strains. **b-c**, Repartition of the total pool of target genes (**b**) or TF-target interactions (**c**) in four tested strains. **d,** Enrichment coverage tracks of ChIP-seq against negative controls for TF Irp with binding sites on the promoter of *gidA* in all four genomes. **e,** Enrichment coverage tracks of ChIP-seq against negative controls (gray tracks) for the three TFs (Irp, PSPPH2193 and PSPPH4127) with binding sites on the promoter of *rpoD* in four tested strains.

Figure 6. Functional modularized regulatory network in *P. syringae* exhibits the specific functions of both TFs and their target genes.

a, The functional categorical regulation network in *P. syringae* analyzed by Gephi

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(resolution 0.9). The 16 modules (both TFs and their target genes) are labeled in different colors. TF nodes are shown as corresponding sized circles representing their expression level. Their target genes are shown as corresponding-colored dots in the background. TF-target edges are shown as corresponding-colored lines between nodes. **b,** Graph diagram indicates the connections between TF and their targets in modules. Module nodes are shown as corresponding-colored circles with size proportional to the number of nodes within. Edge colors are source-based, and edge thicknesses represent the connected quantity between modules. **c-d,** Functional category enrichment analysis of genes in each module.

Figure 7. Global transcriptional regulatory network in *P. syringae*.

The integrated transcriptional regulatory network reflects the interactions between all TFs classified into 39 families from different DNA binding domain types and target genes annotated from pathway annotations. The targets are shown in 12 pathways with various colors. TF nodes are gray as corresponding sizes representing TF number in the family. Edge colors are target-based.

Figure S1. Summary of ChIP-seq results in *P. syringae*.

a, Locations of all the 301 annotated TFs in *P. syringae* genome. Blue lines represent the TF loci. **b,** ChIPed TFs are classified as respective family with different colors. Square size represents the targeted enrichment of each TF. **c,** The preferential enrichment at different genome loci of each ChIPed TF, including upstream, overlapStart, inside, overlapEnd, downstream and includeFeature regions.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S2. Graph diagram of feed-forward loop in *P. syringae*.

TF columns in feed-forward loop (M13, n=696) are classified as families. TF-TF edges are distinguished with separate colors.

Figure S3. Co-association and virulence-related functional category of TFs at three different levels.

a-b, The co-associated map and functional category of top TFs. The peak number of top TFs are shown at upper corresponding to the TFs in the lower map. **c-d,** The co-associated map and functional category of middle TFs. **e-f,** The co-associated map and functional category of bottom TFs.

Figure S4. The validation of the binding sites of virulence-related TFs in *P. syringae*.

a-c, The promoter of *rpoD* is the negative control of PSPPH1951 and PSPPH2193. The promoter of PSPPH3658 is the negative control of PSPPH3268. **d,** The validation of binding sites of PSPPH1951. The validated binding sites are from the promoters of *pilZ*, *pilF* and *pilG*. **e,** The validation of binding sites of PSPPH3798. The validated binding sites are from the promoters of *fliK*, *fliE*, *fliD* and *fleQ*. The promoter of PSPPH2263 is the negative control. **f**, A regulatory cascade of PSPPH3504. / means sibling nodes. – means downward regulation.

Figure S5. Metabolic functional category of TFs at three different levels.

a-c, Functional category according to different metabolic pathways of top (**a**), middle

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

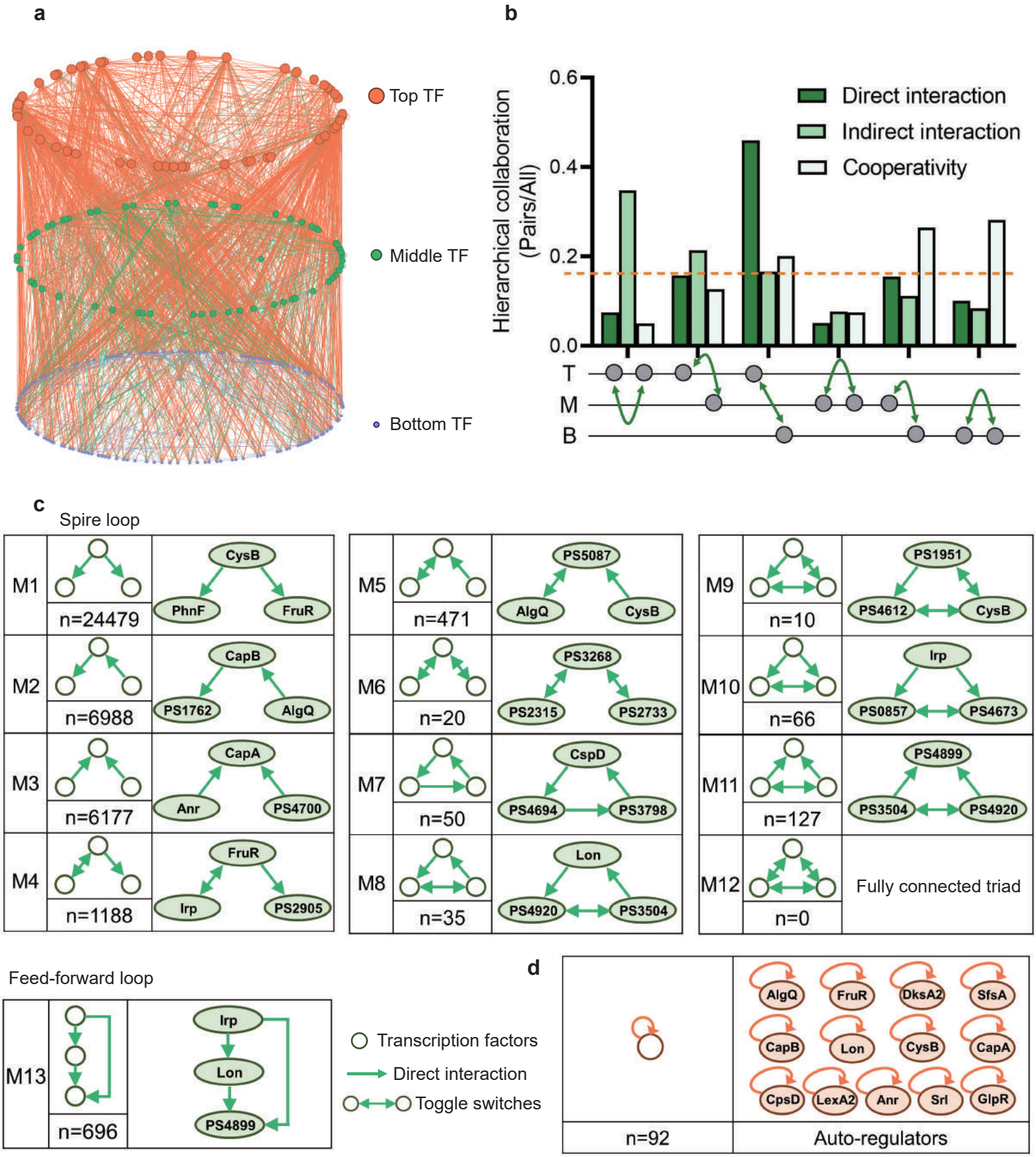
(**b**) and bottom (**c**) TFs.

Figure S6. Key TFs in different metabolic pathways.

a, Radar plots show the putative key regulators identified in two different metabolic pathways, including DNA replication, and phosphonate and phosphonate metabolism. Each radiation line represents a key regulator, and the radial length of the thick colored line is the rate of target genes to the associated genes, representing the significance of the enrichment of the TF target genes within each pathway. **b,** The monomer motif of PSPPH0755 and the validation of the binding sites of PSPPH0755 by EMSA. The validated binding sites are from promoters of PSPPH5210 and PSPPH3109. The promoter of PSPPH4598 is negative control. **c,** The head-to-head motif of PSPPH3798 and the validation of the binding sites of PSPPH3798 by EMSA. The validated binding sites are from promoters of PSPPH3881 and PSPPH5119. **d,** The monomer motif of PSPPH4638 and the validation of the binding sites of PSPPH4638 by EMSA. The validated binding sites are from promoters of PSPPH3881 and PSPPH0550. The promoter of PSPPH4598 is the negative control.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1.



bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2.

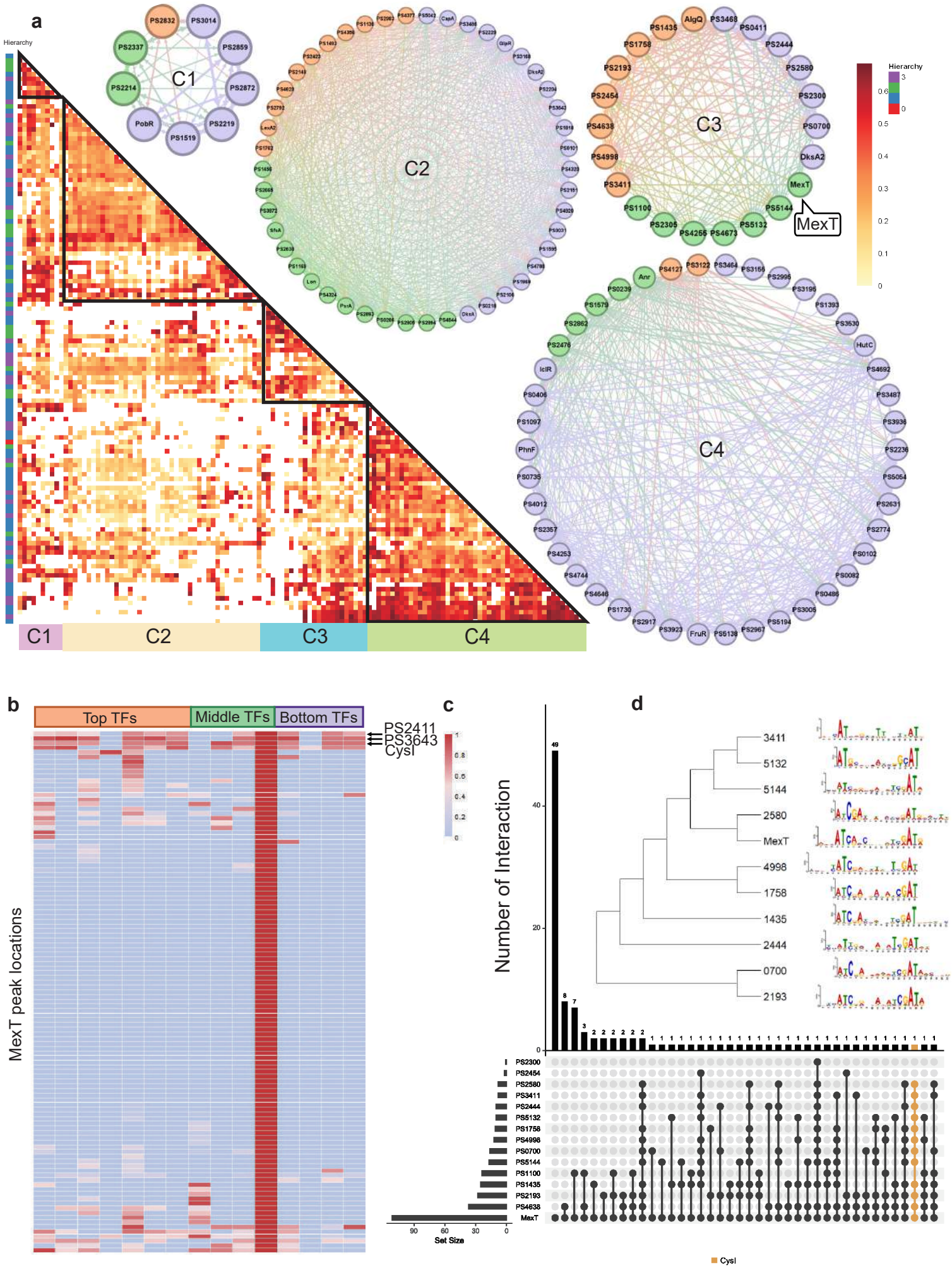


Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

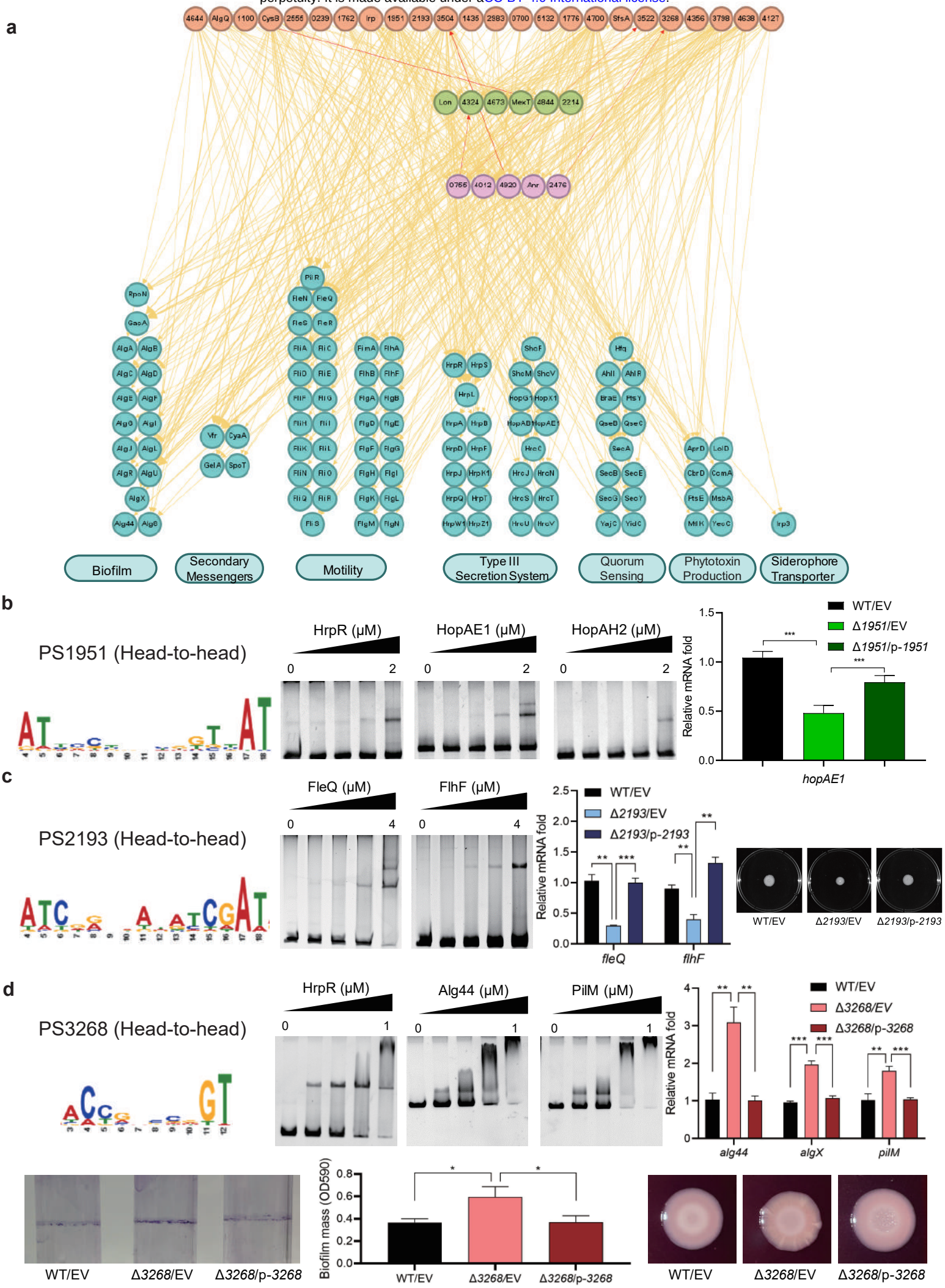
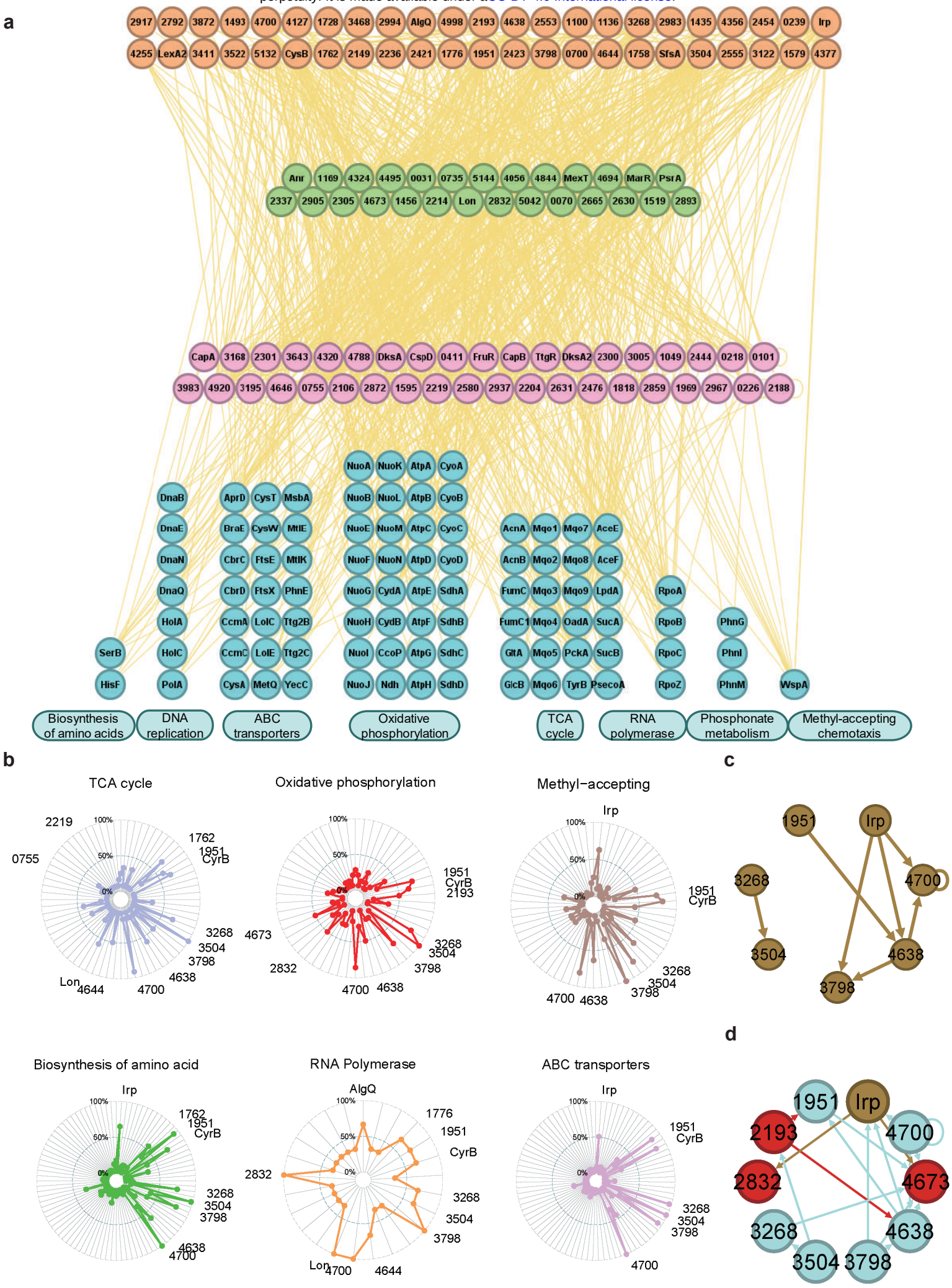


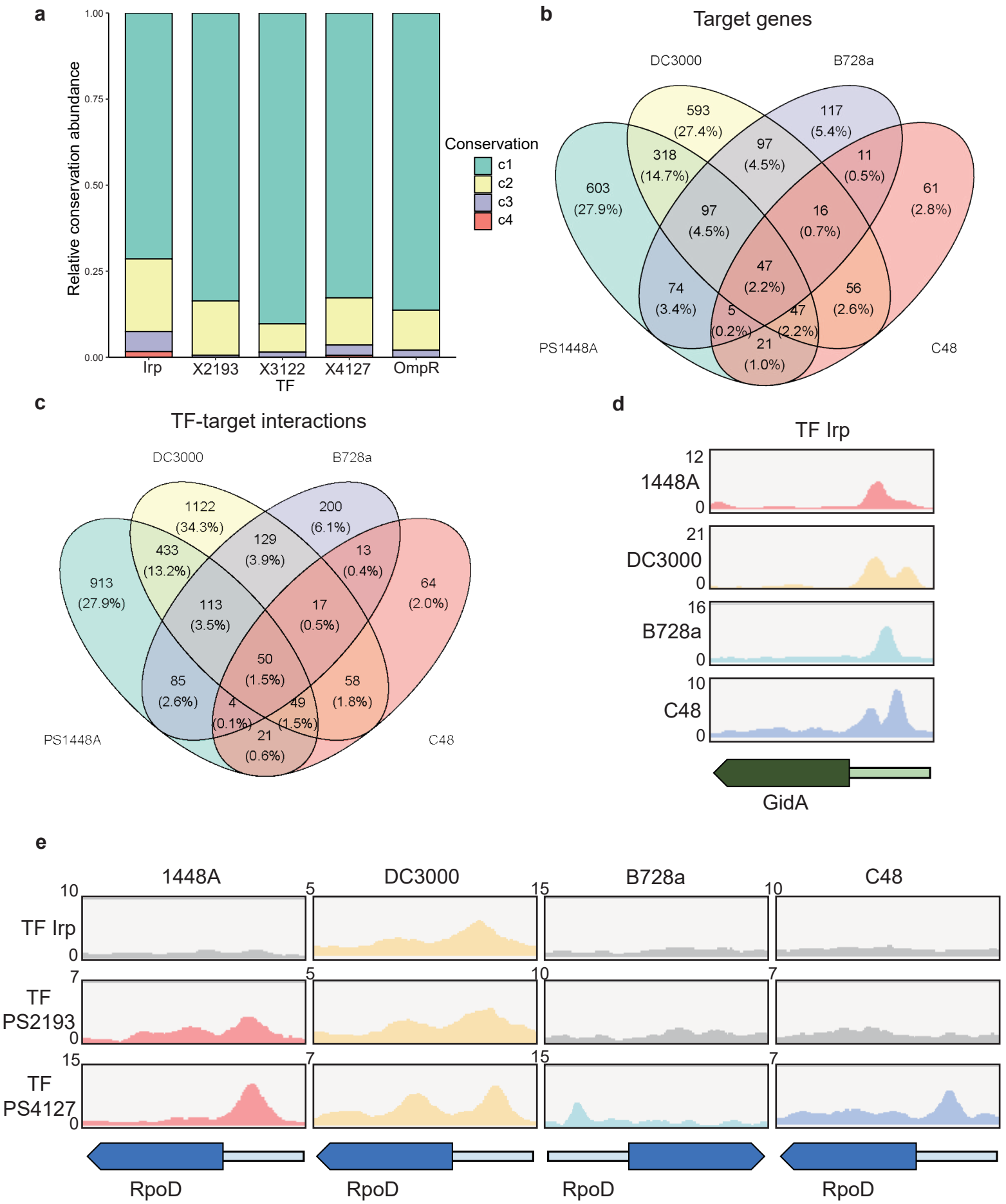
Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



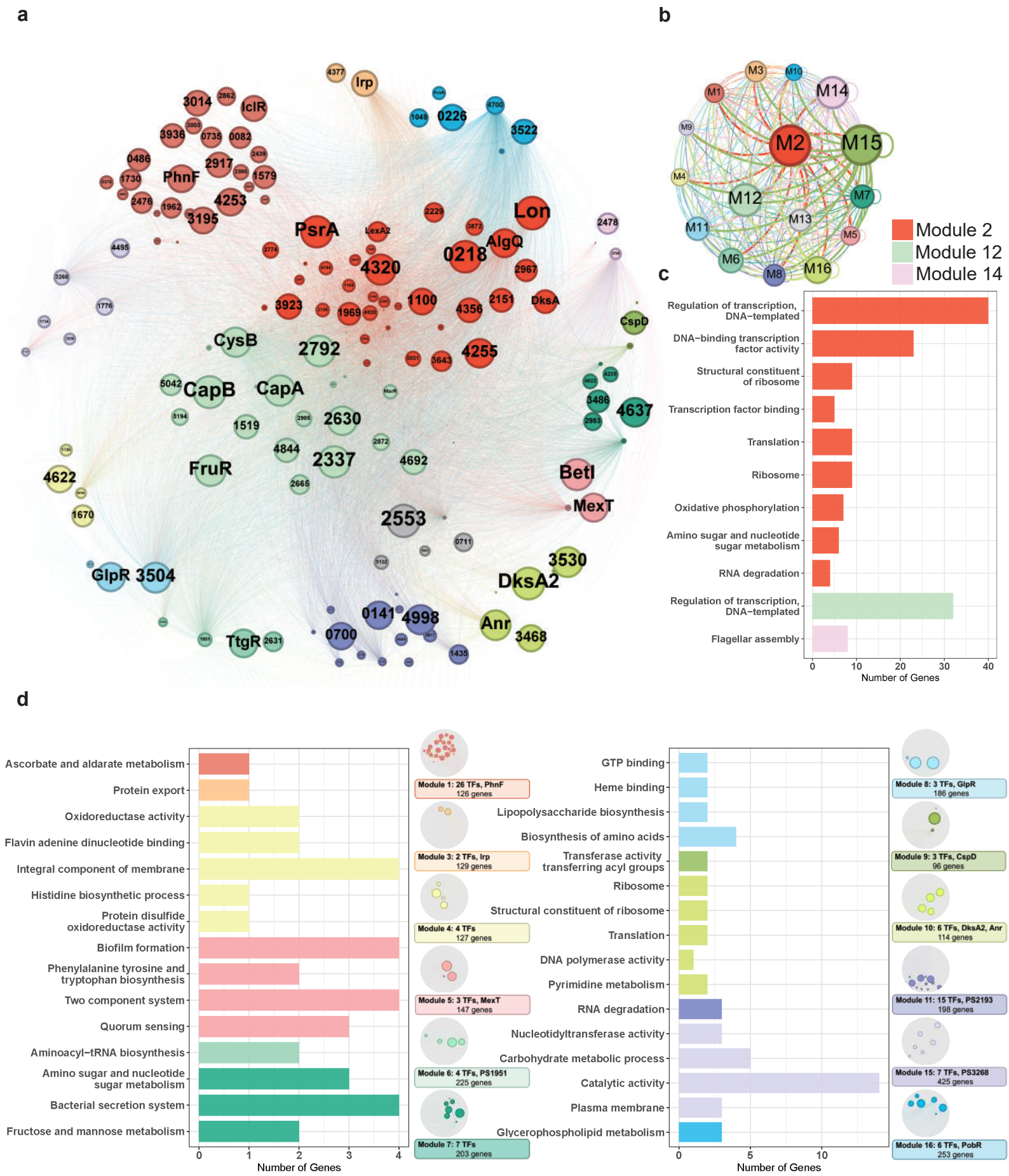
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 5.



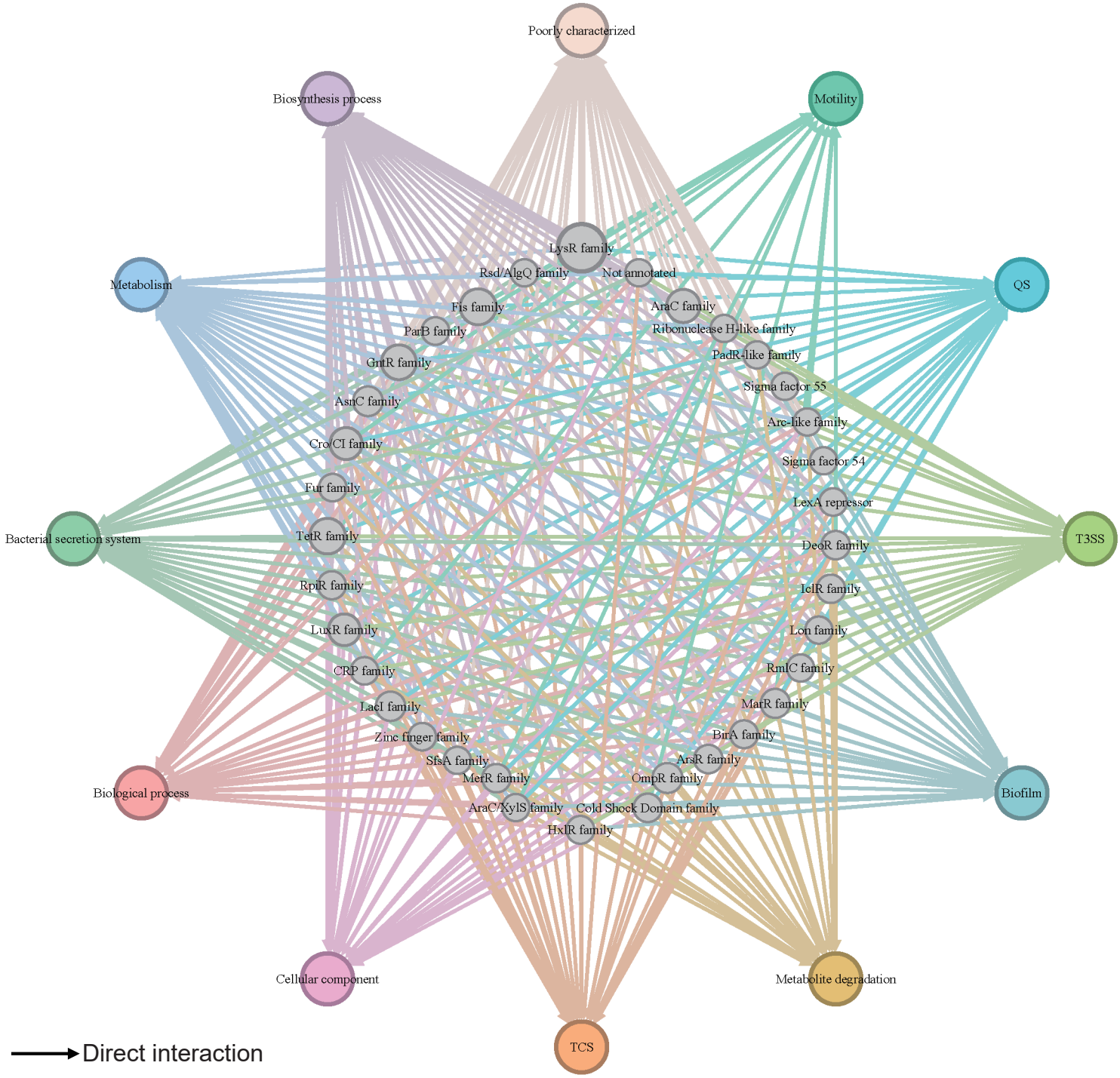
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 6.



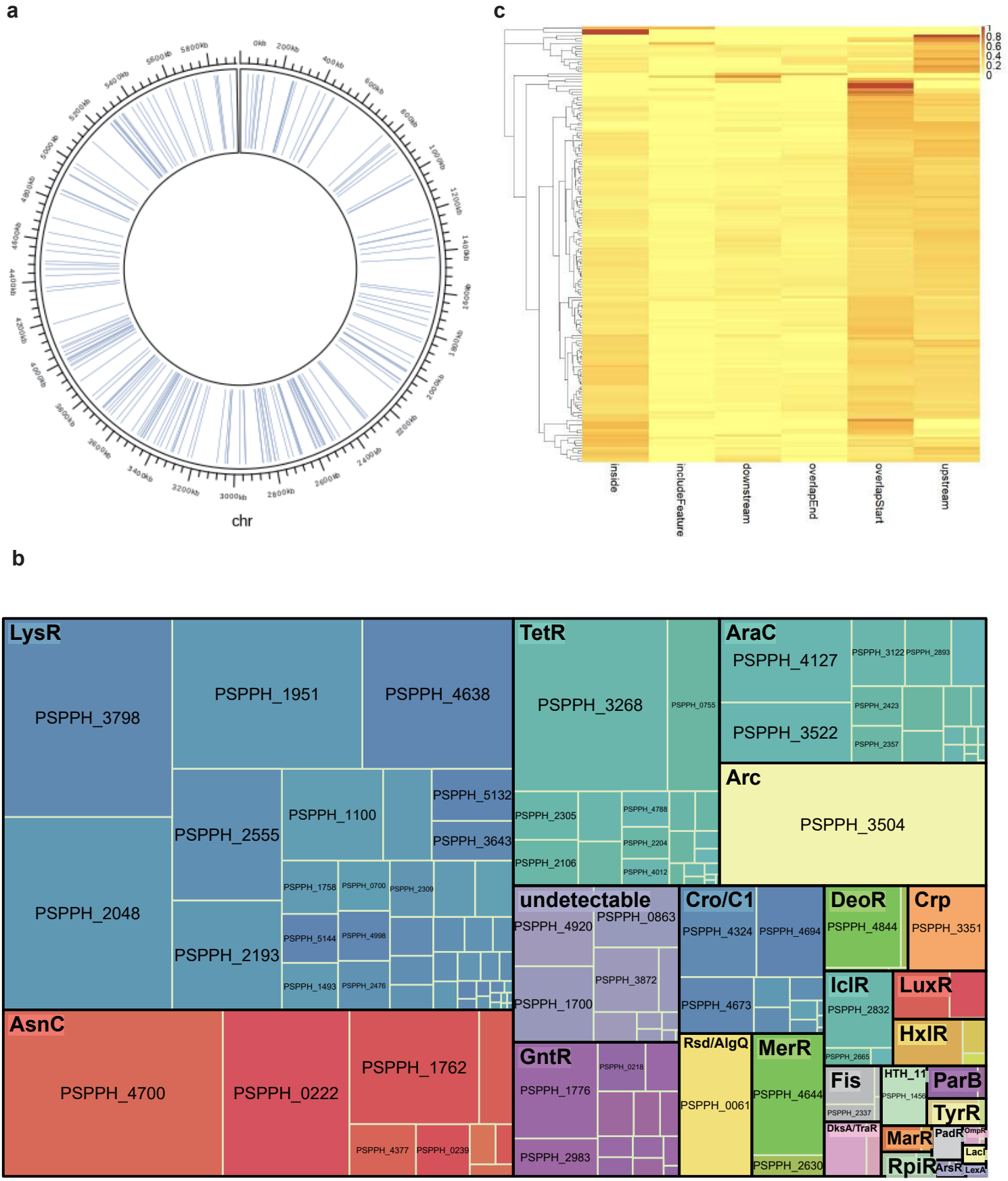
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 7.



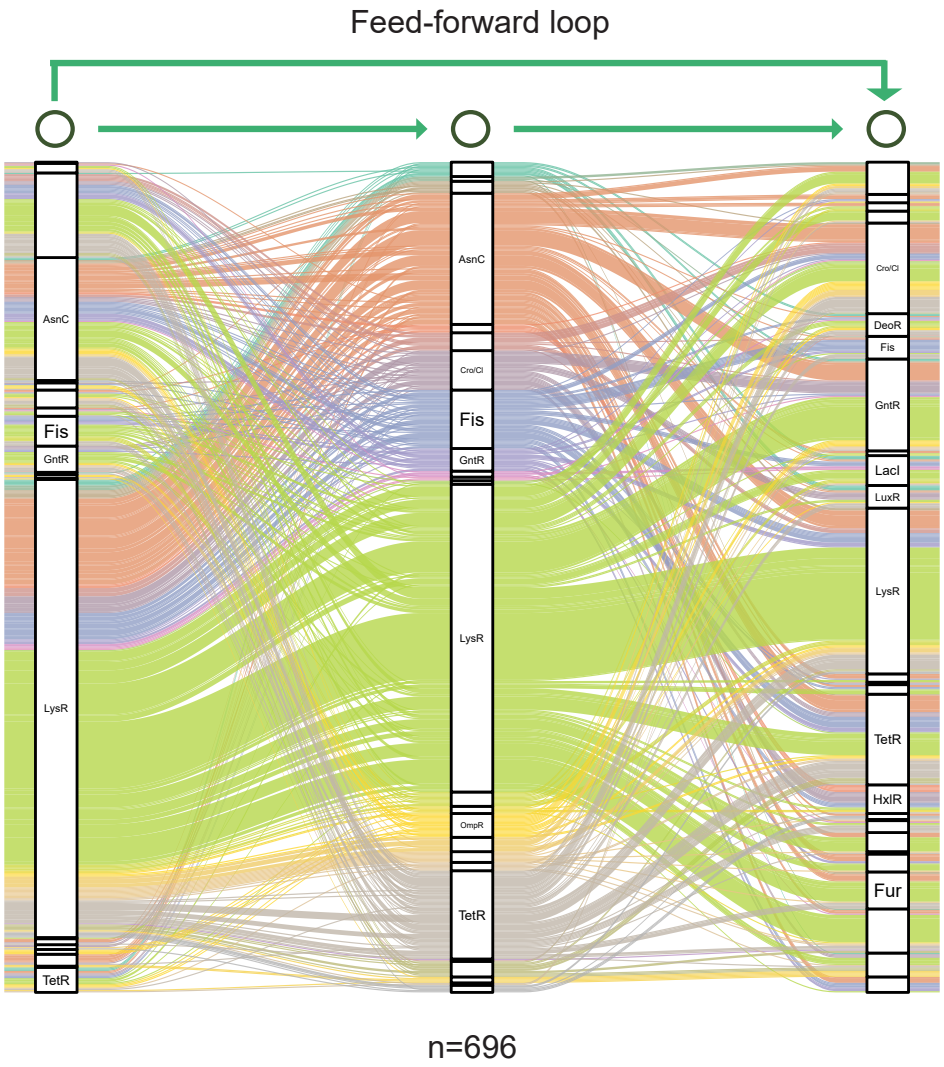
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S1



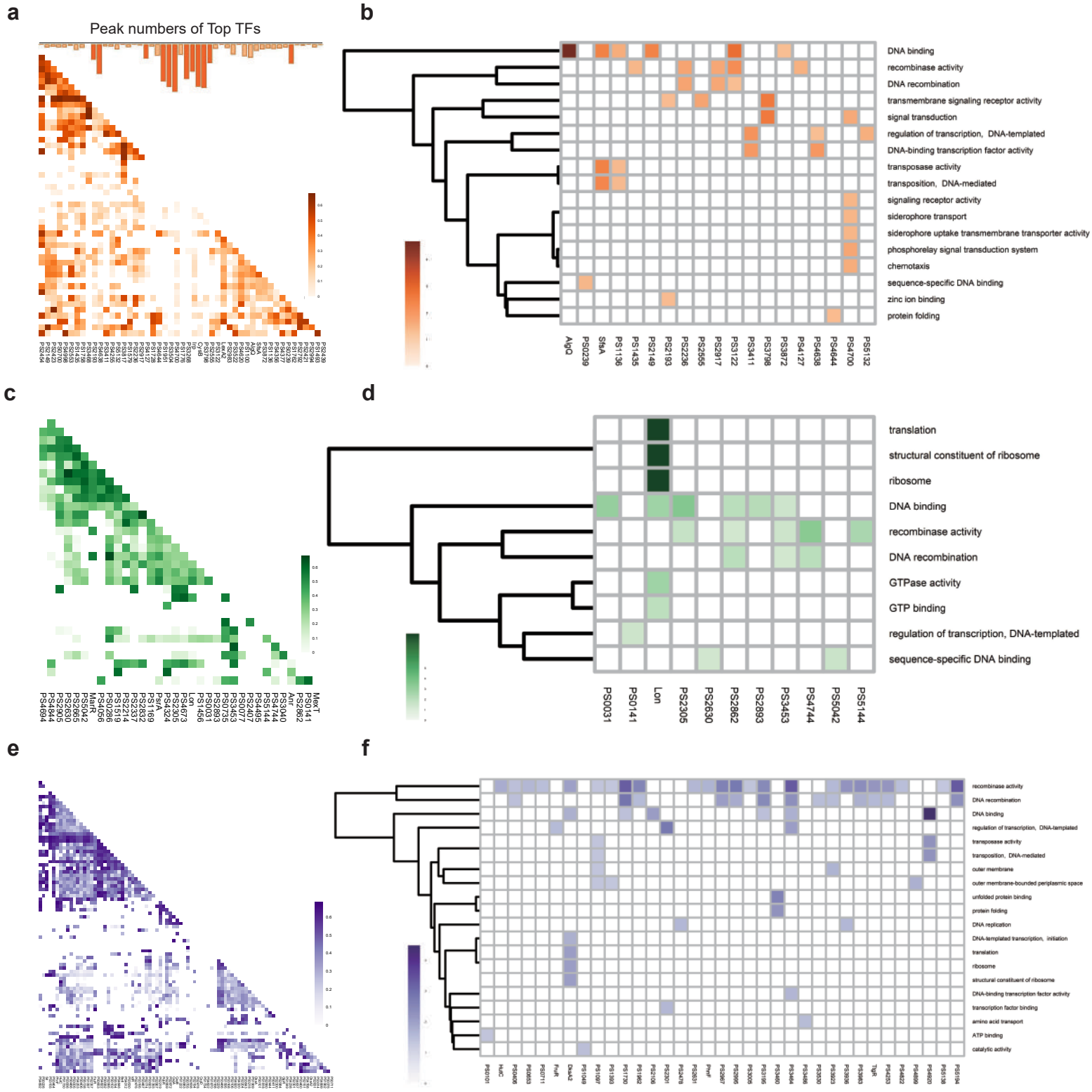
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S2



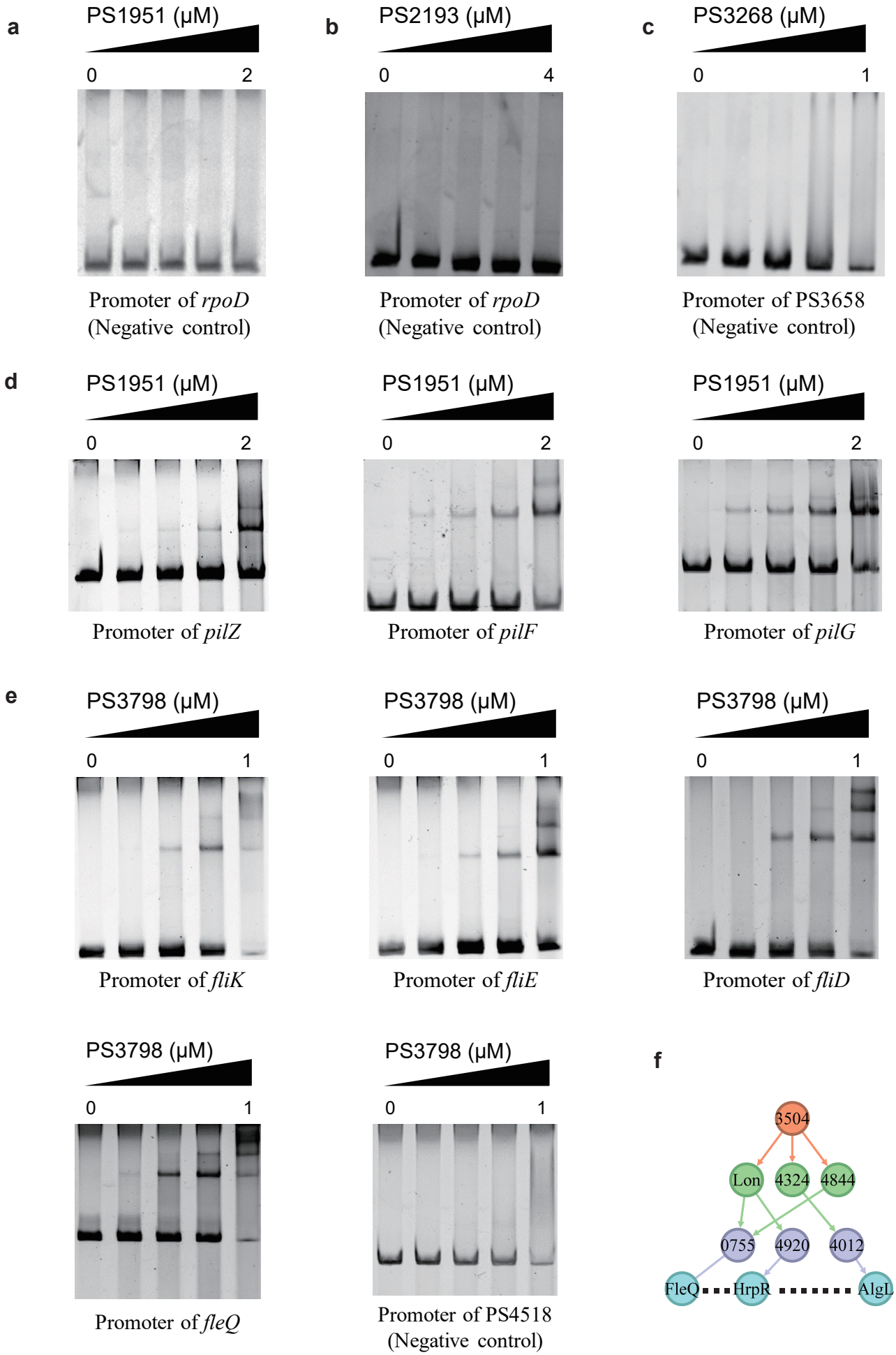
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S3.



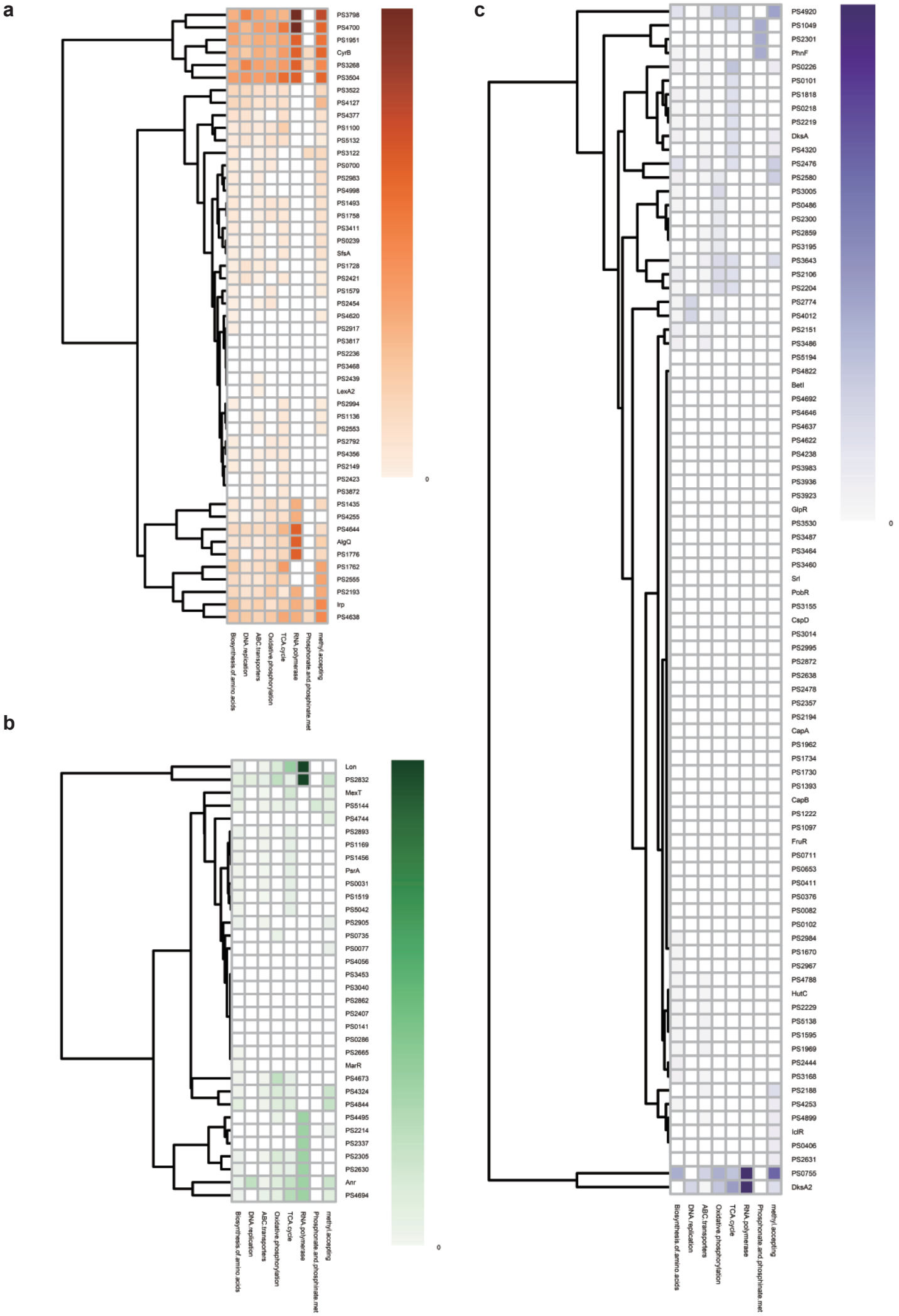
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S4



bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

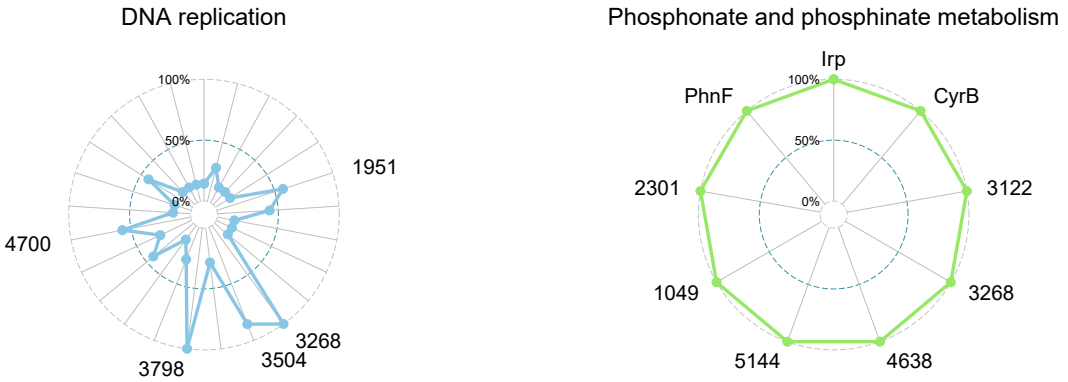
Figure S5.



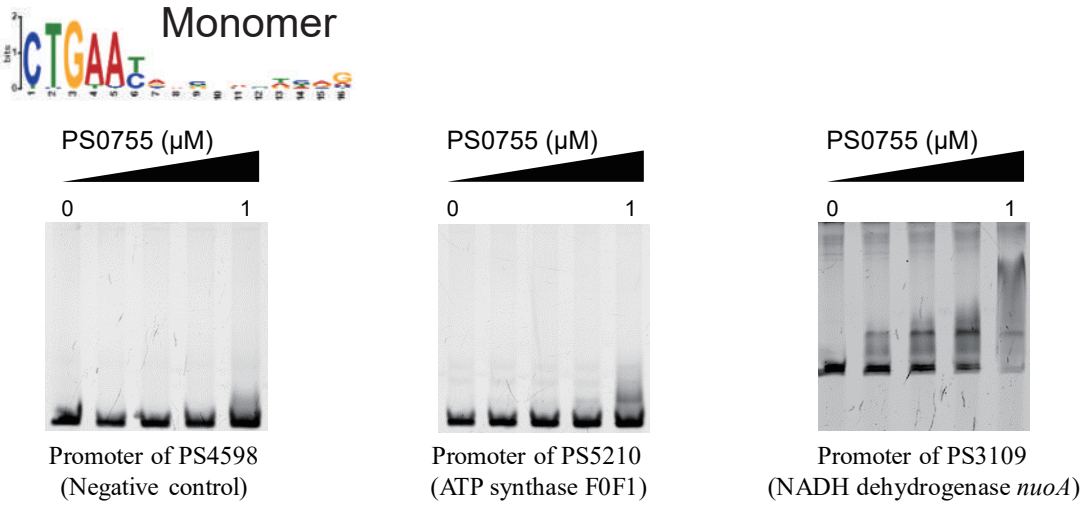
bioRxiv preprint doi: https://doi.org/10.1101/2024.01.18.576191; this version posted January 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S6

a



b

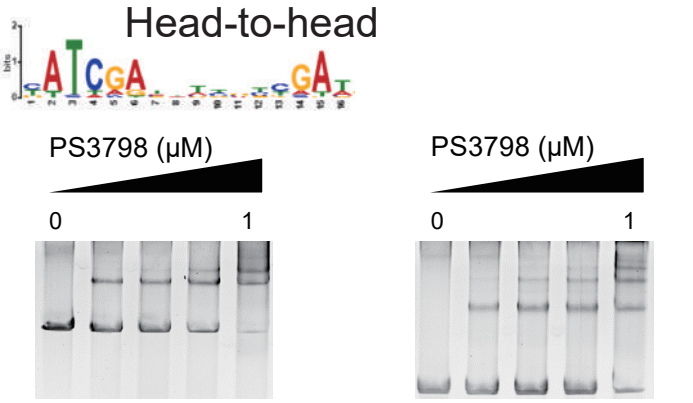


Promoter of PS4598 (Negative control)

Promoter of PS5210 (ATP synthase F0F1)

Promoter of PS3109 (NADH dehydrogenase *nuoA*)

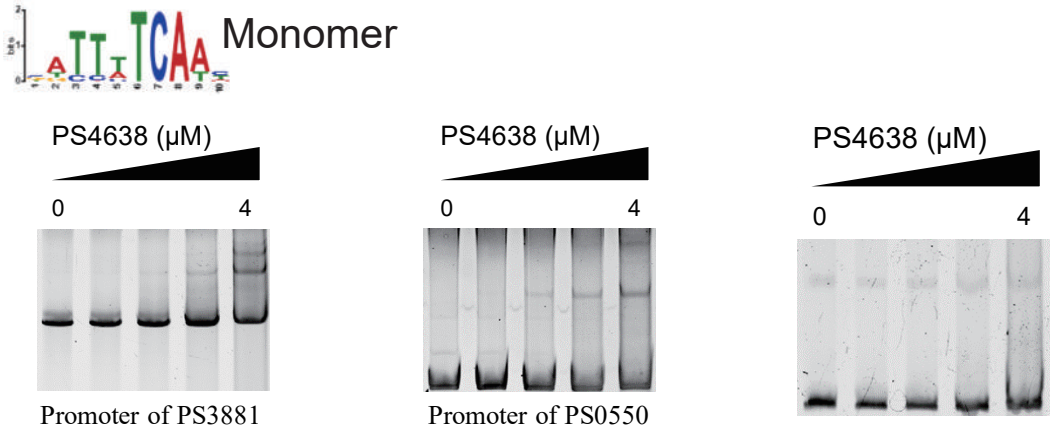
c



Promoter of PS3881 (methyl-accepting chemotaxis gene *wspA*)

Promoter of PS5119 (phosphate transport system regulatory gene *phoU*)

d



Promoter of PS3881 (methyl-accepting chemotaxis gene *wspA*)

Promoter of PS0550 (phosphoserine phosphatase *serB*)

Promoter of PS4117 (Negative control)