Machine Learning Uncovers the Transcriptional Regulatory Network for the Production Host *Streptomyces albidoflavus* J1074

Mathias Jönsson[1], Renata Sigrist[1], Mykhaylo Semenov Petrov[1], Nils Marcussen[1], Tetiana Gren[1], Bernhard O. Palsson[1,2], Lei Yang[1], Emre Özdemir[1]

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800, Kgs. Lyngby, Denmark
[2] Department of Bioengineering, University of California, San Diego, La Jolla, USA

Abstract

Streptomyces albidoflavus J1074 is a popular and genetically tractable platform strain used for drug discovery and production via the expression of heterologous biosynthetic gene clusters (BGCs). However, its transcriptional regulation network (TRN) and its impact on secondary metabolism are poorly understood. Here we characterized its TRN by applying an independent component analysis to a compendium of 142 high quality RNA-Seq transcriptomes, from both in-house and public sources spanning 58 unique growth conditions. We obtained 63 independently modulated sets of genes (iModulons), that quantitatively describe the TRN and its activity state across diverse conditions. Through analyses of condition-dependent TRN activity states, we (i) describe how the TRN adapts to different growth conditions, (ii) conduct a cross-species iModulon comparison, uncovering shared features and unique characteristics of the TRN across lineages, (iii) identify putative regulators of several biosynthetic gene clusters, including Surugamide, Cyclofaulknamycin and Dudomycin, and (iv) infer potential functions of 40% of the uncharacterized genes in the *S. albidoflavus* J1074 genome. Our findings provide a comprehensive and quantitative understanding of the TRN of *S. albidoflavus* J1074, and suggest new strategies for rational strain design to optimize its production capabilities.

Introduction

Streptomyces is a genus of Gram-positive soil-dwelling actinomycetes, which are well-known for their ability to produce valuable secondary (or specialized) metabolites, including herbicides, antibiotics, immunosuppressants, antifungal, and anticancer drugs (Challis, 2014; Nett et al., 2009). *Streptomyces* have distinct life-stages regulated by a complex transcriptional regulatory network (TRN), that includes multicellular differentiation leading to morphological changes and compartmentalization (Chater, 2016; Claessen et al., 2006; de Lima Procópio et al.,

2012; McCormick & Flärdh, 2012). The production of secondary metabolites appear to mainly stem from the activation of biosynthetic gene clusters (BGCs) through a hierarchical network of transcriptional regulators in response to various internal and external signals, such as nutrient availability, stress, and intercellular communication (Barbuto Ferraiuolo et al., 2021; Chen et al., 2010; Wohlleben et al., 2017; Xia et al., 2020; Zhang et al., 2020). However, many BGCs remain silent or poorly expressed under standard laboratory conditions, limiting the discovery and production yield of desired compounds (Dolya et al., 2023). While several strategies have been developed to activate and optimize BGC expression, such as genome editing, heterologous expression, co-cultivation, and elicitor addition, the TRN of *Streptomyces* is poorly understood compared to other model bacteria, which limits the potential and applicability of rational strain design (Barbuto Ferraiuolo et al., 2021; Chater, 2016; Chen et al., 2010).

Streptomyces albidoflavus (previously *Streptomyces albus* J1074) is a well characterized and widely used host for production of complex secondary metabolites (Baltz, 2010; Bilyk et al., 2016; Chen et al., 2008; Gullón et al., 2006; Huang et al., 2011; Myronovskyi et al., 2016, 2018). Its small genome, capacity to produce versatile precursors, and dispersed mycelial growth compared to other *Streptomyces,* make it one of the most convenient *Streptomyces* production host to scale up in industrial bioreactors (Marín et al., 2017; Myronovskyi & Luzhetskyy, 2019; Zaburannyi et al., 2014). However, the production yields are often low, requiring tedious chassis optimization and/or production processes (Dolya et al., 2023; Gummerlich et al., 2021; Li et al., 2023). Genome-wide studies and characterization of the TRN in other *Streptomyces* has identified regulators, promoter regions, and transcription start sites, which may improve rational strain design decisions and production yield (Hwang et al., 2019; Jeong et al., 2016; Luo et al., 2018; Yi et al., 2017). However, despite the significant effort to characterize the TRN of *S. albidoflavus* J1074 specifically, our knowledge remains limited.

Independent Components Analysis (ICA) is a powerful unsupervised machine learning method that can decompose complex gene expression data into sets of independently modulated genes (iModulons), and their corresponding activity profiles across different conditions (Fang et al., 2017; Rychel et al., 2021, Sastry et al., 2019). ICA can capture both linear and nonlinear relationships among genes, and identify both known and novel regulatory signals, without requiring prior knowledge or assumptions (Chauhan et al., 2021; Yoo et al., 2022). ICA has been successfully used to infer genome-scale TRNs and functional modules for several bacterial species, including *Escherichia coli*, *Bacillus subtilis*, and *Mycobacterium tuberculosis* (Lamoureux et al., 2023; Rychel et al., 2020; Sastry et al., 2019; Yoo et al., 2022).

To gain a comprehensive and systematic understanding of the TRN of *S. albidoflavus*, we performed ICA on a compendium of 142 high-quality gene expression profiles, of which 85 were newly generated for this purpose. We identified

63 iModulons which reveal novel insights into the transcriptional responses of *S. albidoflavus* to relevant conditions, such as nutrient limitation, BGC removals, heterologous expression, and over time. We also conduct a large-scale comparison of iModulon structure across six different organisms to validate and extend our findings to gain an evolutionary context of the TRN. Furthermore, we used our iModulons to investigate the regulation and expression of several BGCs, including Antimycin and Surugamide, identifying promising targets for future studies.

Results

Independent component analysis reveals 63 signals in the transcriptomic data

We compiled a dataset of *S. albidoflavus* J1074 transcriptomes consisting of 85 expression profiles generated in-house and all publicly available transcriptome datasets from SRA NCBI (Fig 1A, Supplementary table S1). Our in-house samples doubled the publicly available data. After quality control the final compendium consisted of 142 samples from 58 unique growth conditions, with a high median Pearson R score of 0.99 between replicates (Fig 1B, Supplementary Fig S1). ICA decomposed this compendium into 63 iModulons, with 35.8% of the genes in the genome being assigned to at least one iModulon. The iModulons cover a wide range of functions, including secondary metabolism (20.6%), carbon metabolism (17.5%), and uncharacterized functions (22.2%) (Fig 1C). Moreover, the iModulons explain 80% of the variance in gene expression across our compendium and thus represent the variation derived from transcriptional regulation (Fig 1D).

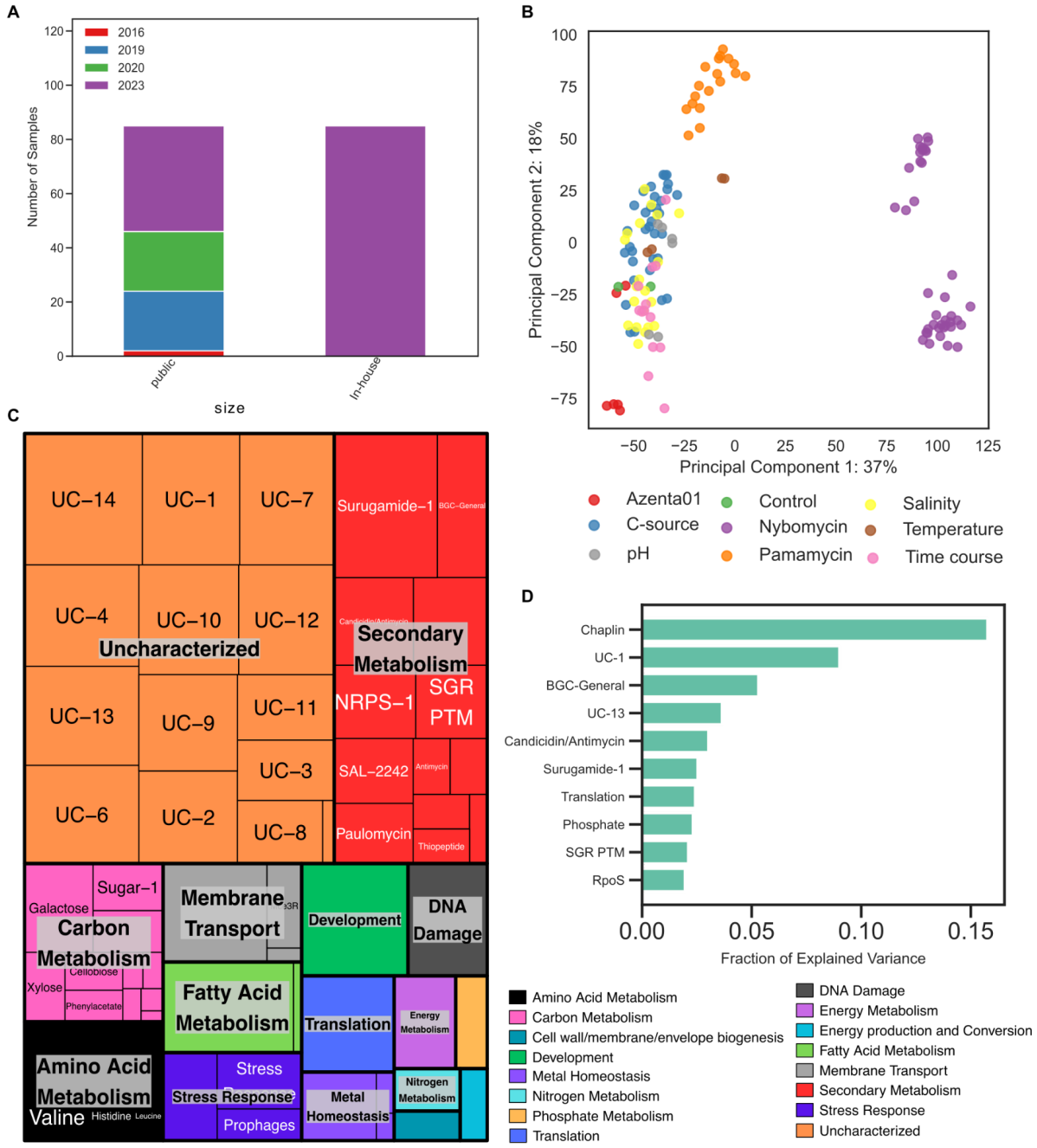

Figure 1. Overview of the S. albidoflavus J1074 iModulons. *(**A**) Bar plot showing the number of RNA-Seq transcriptomes available for S. albidoflavus J1074 on SRA NCBI. The color indicates the year of submission. Our newly generated in-house transcriptomes will increase the publicly available datasets by 100%. (**B**) Principal components analyses of 142 samples used for ICA. The color depicts the specific project where the sample was generated. (**C**) A treemap depiction of the 63 identified iModulons and their associated functional categories. The dimensions of the individual boxes correlate with the number of genes found in each respective iModulon. (**D**) Highlight of the top 10 iModulons contributing the greatest proportion of explained variance. These iModulons encapsulate key functions such as development, translation, secondary metabolite biosynthesis, stress, and unknown functions.*

Comparison of iModulon activity across the phylogenetic tree reveal conserved and lineage-specific transcriptional regulation in *S. albidoflavus* J1074

To further explore the TRN of *S. albidoflavus* J1074, we compared its iModulons with those of other species with more annotated TRNs. We used imodulondb.org and orthofinder to identify orthologous gene weight similarities in different iModulons (Fig 2A) (Emms & Kelly, 2019; Rychel et al., 2021). We visualized the iModulons in a similarity based network that cluster together based on their function, regardless of species, which we call the iModulome (Fig 2B-D, Supplementary Fig S2). This analysis helped us assign functions to some of the iModulons that were difficult to characterize. Moreover, the iModulome revealed conserved and lineage-specific mechanisms of gene regulation in *S. albidoflavus* J1074. For many iModulons, there were highly conserved genes that play a key role in the activation of the iModulon across all the species studied here, such as the ribosomal subunits genes in the Translation related iModulons (Fig 2C). Similarly, the Leucine/Lysine cluster contained iModulons related to Leucine metabolism processes across all 6 species, including the *leuLABCD* operon, however, this analysis also highlights some lineage-specific transcriptional patterns, such as a co-regulation of two uncharacterized genes in *S. albidoflavus* J1074, and *ilvH* and *gabP* which are an acetolactate synthase and an amino acid transporter, respectively. Moreover, the phosphate related iModulons were also highly similar between *E. coli*, *S. enterica*, and *S. albidoflavus* J1074, where genes in the pstSCAB operon appear to play a conserved role in phosphate uptake (Fig 2D). Additionally, the *S. albidoflavus* J1074 phosphate iModulon contained genes such as *waaH*, *kdsC*, and *bacA*, which were also co-regulated, suggesting potential lineage-specific adaptations to phosphate uptake.

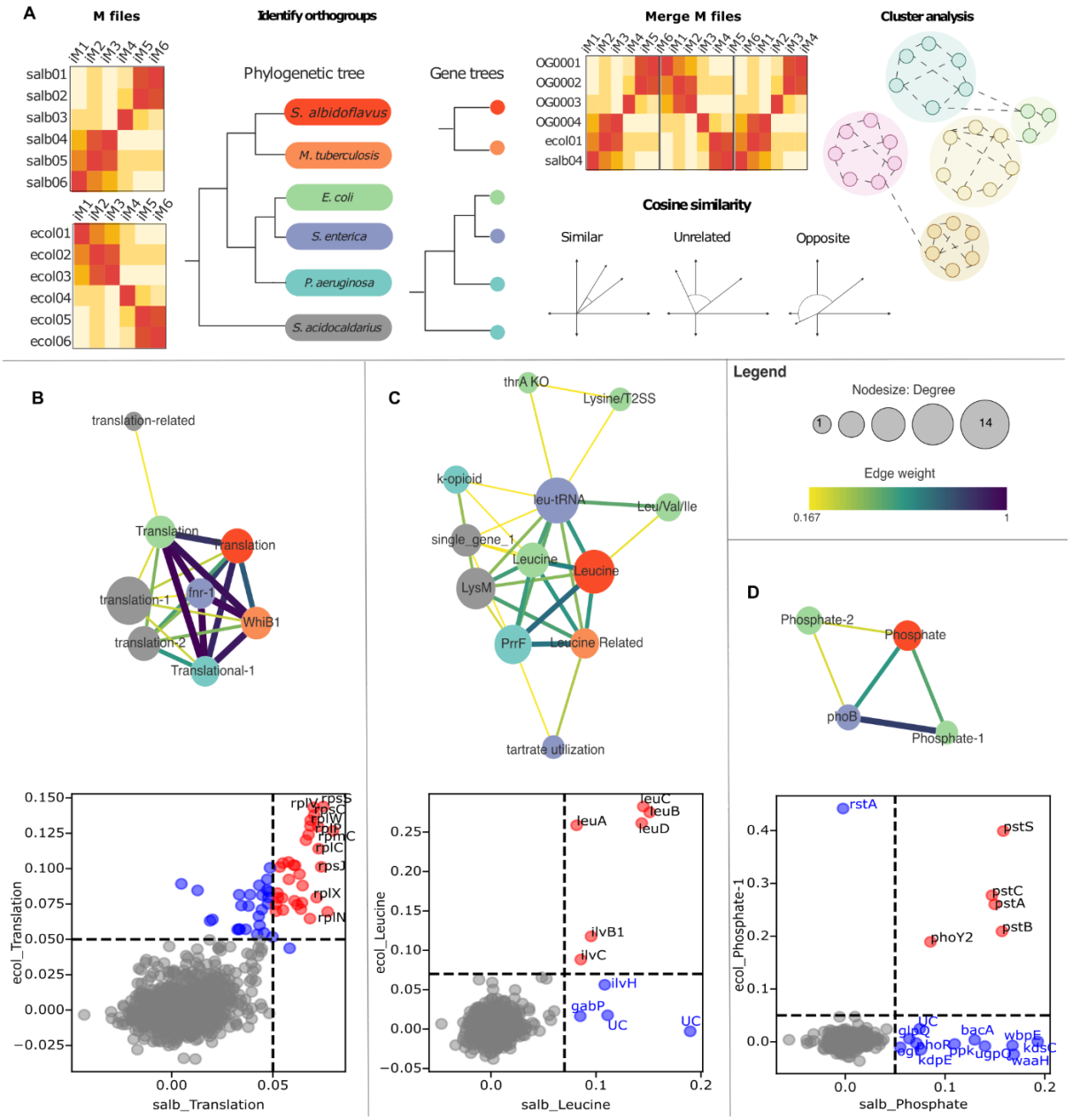

Figure 2. S. albidoflavus J1074 iModulons similarities and differences with other organisms. *(A) Overview of the construction of the iModulome. M files generated from ICA analysis of six organisms were collected. Orthogroups were identified using orthofinder (Emms & Kelly, 2019), and used to merge the M files based on the specific orthogroup. Cosine similarity score was used to create a clustered network of iModulons based on similarity. (B) A subcluster from the iModulome (top) of iModulons across six species (coloured nodes; Red, S. albidoflavus; Orange, M. tuberculosis; Green, E. coli; Purple, S. enterica; Turquoise, P. aeruginosa; S. acidocaldarius) based on cosine similarity of ICA derived gene weights, containing iModulons related to Translation processes. (Bottom) a scatterplot of gene weights between the E. coli Translation iModulon (y-axis) and the J1074 Translation iModulon. Red orthologous genes are highly weighted in both species, while blue genes are highly weighted in just one species. (C) Subcluster (top) and scatterplot (bottom) of Leucine related iModulons. (D) Subcluster (top) and scatterplot (bottom) of Phosphate related iModulons.*

iModulons give insights into the regulation of the surugamide BGC

We detected 11 different BGCs across 13 iModulons, including Paulomycin, Antimycin, Candicidin, and Surugamide (Table 1). The coverage of the antiSMASH predicted BGC ranged from 10.8% in the NRPS-1 related iModulon to 80% in the Paulomycin related iModulon (Table 1). Some of these iModulons contain co-regulated genes located outside the correspondent BGC, giving insights into how these secondary metabolites may be regulated.

Table 1. Biosynthetic gene cluster representation in iModulons. *List of BGCs and corresponding iModulons. The coverage represents the percentage of genes from the BGC is present within the iModulon. The Co-regulated genes contains information about genes of interest present in the iModulon, outside of the BGC range.*

BGC	iModulon	Coverage (%)	Co-regulated genes
Antimycin and Uncharacterized	Antimycin	20.4	Pyruvate dehydrogenase and a FUSC family protein
Candicidin and Antimycin	Candicidin/Antimycin	43.2	smoEFG and ctaD.
Cyclofaulknamycin	Cyclofaulknamycin-1	25.0	Sensor domain-containing protein and an alpha/beta hydrolase.
Cyclofaulknamycin	Cyclofaulknamycin-2	11.5	Ten co-regulated genes, 50% of them have COG signal transduction mechanisms.
Dudomycin and Minimycin	Dudomycin/Minimycin	17.0	Two co-regulated gene clusters primarily involved with secondary metabolite biosynthesis and amino acid transport
Uncharacterized	NRPS-1	10.8	82% of the genes in this iModulon are outside of the BGC, including 44% with unknown functions that are downregulated.
Paulomycin	Paulomycin	80.0	-
SAL-2242	SAL-2242	27.8	Alpha/beta hydrolase and a downregulated sigma-70 factor.
SGR PTM	SGR PTM	20.7	37% of the genes have unknown functions.
Surugamide	Surugamide-1	36.4	33% genes involved with amino acid transport and metabolism.
Surugamide	Surugamide-2	9.1	surR-like.
Thiopeptide	Thiopeptide	42.4	Two members of the antimycin BGC.

Furthermore, we identified two iModulons related to the production of surugamide, which is a cyclic NRPS with antibacterial and antifungal properties (Ninomiya et al., 2016; Takada et al., 2013). Surugamide-1 and 2 contain 122 and 10 genes, respectively (Fig 3A, B). Surugamide-1 contains 20 genes from the BGC, including the four core NRPS biosynthetic genes (Fig 3A). In addition, 33.6% of the genes in the iModulon are related to amino acid transport and metabolism, which may provide precursors for surugamide synthesis, suggesting that this iModulon may represent the activation of the surugamide BGC. In contrast, Surugamide-2 iModulon contains the previously described repressor *surR,* which binds to the promoter region of the core NRPS genes in the BGC and inhibits their transcription (Fig 3B) (Xu et al., 2017). In addition, Surugamide-2 contains another GntR-like transcription factor, which is closely related to *surR* and located around 2 Mbp downstream of the surugamide BGC (Fig 3C). This transcription factor may act as a secondary repressor or a regulator of other genes in the BGC. Furthermore, Surugamide-2 contains several uncharacterized genes within the BGC and located next to the GntR-like gene, suggesting these could have important functions for surugamide regulation. Moreover, Surugamide-1 and 2 are activated in distinct conditions that appear to be related to salinity levels, growth phase, and carbon sources (Fig 3D). Surugamide-1 genes are active in high salinity and late-stage growth phases, while Surugamide-2 genes are active in low salinity conditions and in rich media at early growth phases. In combination, these iModulons appear to describe key genes and conditions which activates or represses the biosynthesis of surugamide, providing targets for future study.

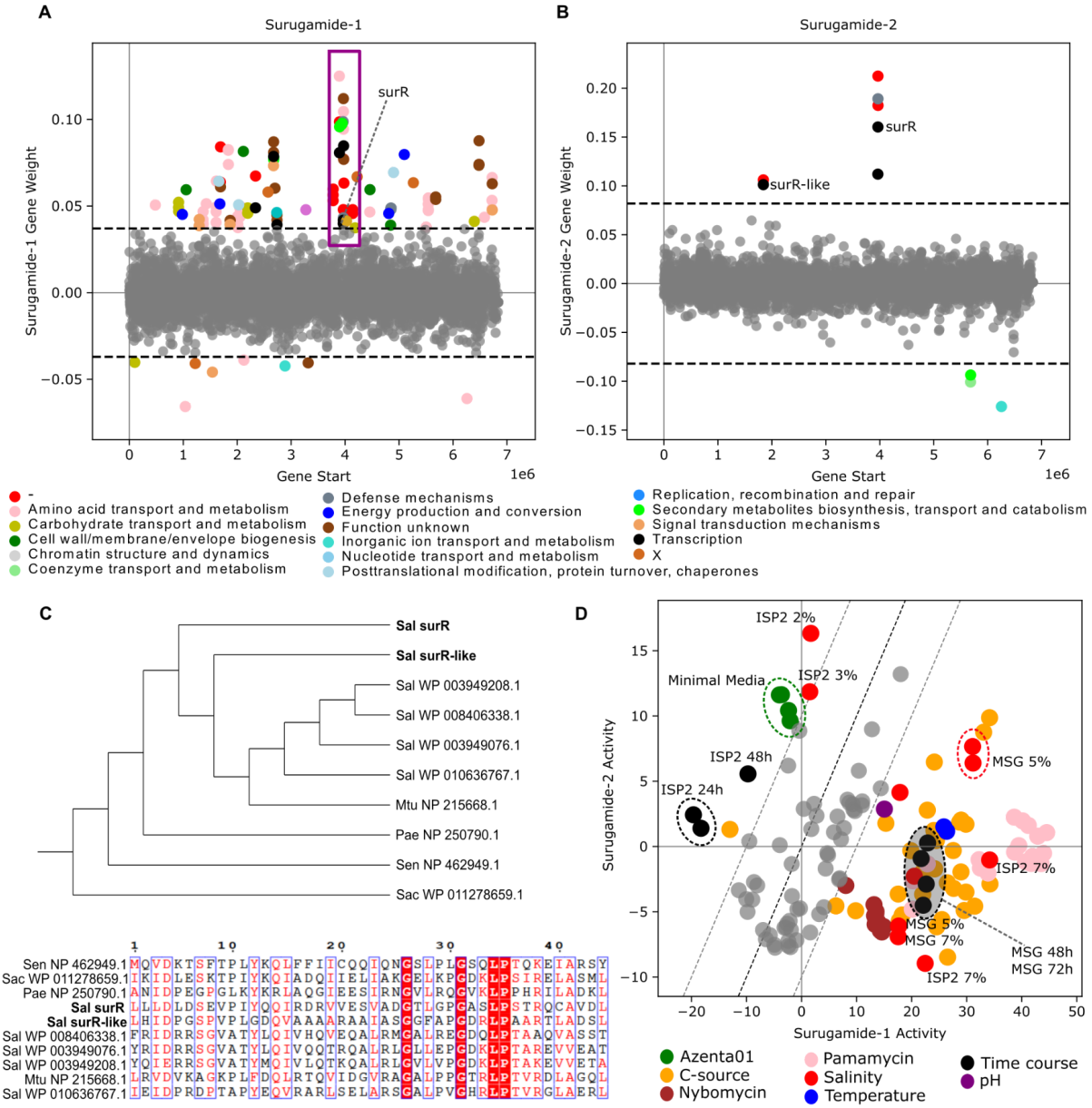

Figure 3. Surugamide activation and repression. *(A) Weightings of all the genes in the component associated with the Surugamide-1 iModulon. y axis depicts the individual gene weightings, and x-axis shows the location of the gene in the genome. Genes within the surugamide BGC are located within the purple square box. Over one third of the genes enriched in this iModulon are related to amino acid transport and metabolism. Colour depicts COG category. (B) Gene weights for the Surugamide-2 iModulon, which contains the repressor surR and another related GntR-like located 2Mbp downstream of the surugamide BGC. (C) Phylogenetic tree and sequence alignment of genes within the same orthogroup as surR. The GntR-like gene present in Surugamide-2 appears to be closely related to surR. (D) Activity of Surugamide-1 and Surugamide-2 iModulons across all samples shows conditions which activate the iModulons, respectively. Low salinity, and rich media in early growth phases appear to activate Surugamide-2, while high salinity, and MSG at late growth phases appear to activate Surugamide-1.*

Activation of several non ribosomal peptide BGCs in *S. albidoflavus* J1074 is correlated with iron limitation.

We further analysed the NRPS clusters of *S. albidoflavus* J1074, several of which are well represented in the iModulons, including Cyclofaulknamycin, Dudomycin and Minimycin (Fig 4A, B). When comparing the iModulon activities of these with other species, they grouped in a large cluster of iron limitation-related iModulons, including *Fur* regulated iModulons in *E. coli,* and siderophore related iModulons such as pyoverdines from *P. aeruginosa* and enterobactin from *E. coli.* Interestingly, there were four iModulons from *S. albidoflavus* J1074 in this cluster, which were all related to the production of NRPS: Cyclofaulknamycin-1, NRPS-1, BGC-General, and Dudomycin/Minimycin (Fig 4C). Moreover, we identified 9 putative core genes of this iModulon cluster that are orthologous to key genes in the enterobactin biosynthesis pathway in *E. coli*: '*entA*', '*entB*', '*entC*', '*entE*', '*entF*', '*fpvB*', '*pfeA*', '*ybdB*', and '*ybdZ*'. A clustered heatmap of the filtered weights, further elucidated the relationships and patterns among these core genes within the iModulons (Fig 4D). In addition, the four NRPS related iModulons were mainly active in stress related conditions, such as low salinity, cold temperature, and low pH (Supplementary Fig S3). These results suggest that *S. albidoflavus* J1074 may have evolved a strategy to cope with iron limitation by producing NRPS.

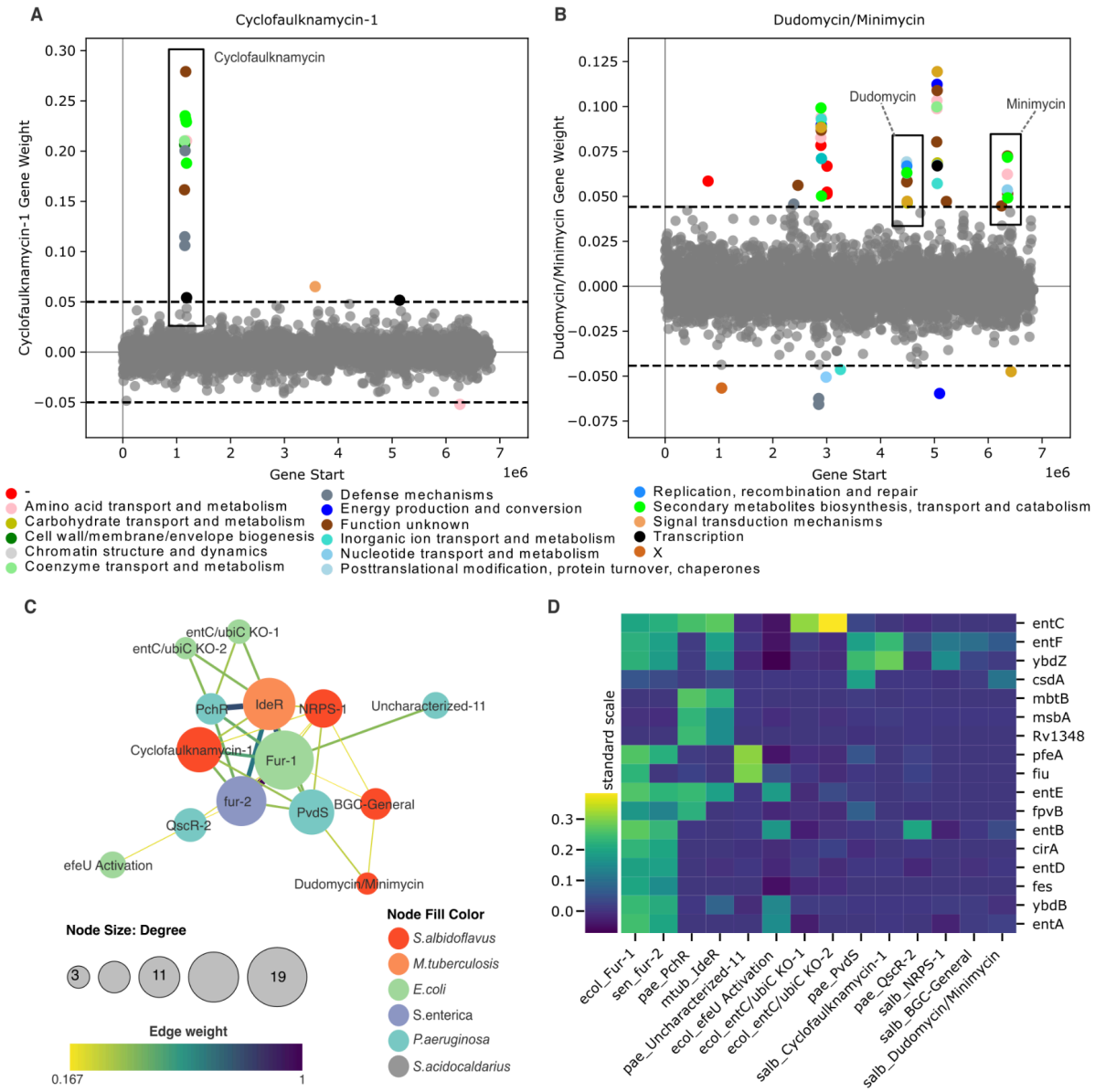

Figure 4. NRPS activation in iron limited conditions. *(A) Gene weights of the Cyclofaulknamycin-1 iModulon. Genes are plotted against their position in the genome. The color depicts the COG function according to the legend. (B) Gene weights of the Dudumycin/Minimycin iModulon. Parts of these BGCs are co-regulated with two other clusters which appear to be mainly involved with amino acid transport and metabolism. (C) A subcluster from the iModulome with iModulons related to iron acquisition across many organisms. Four iModulons from S. albidoflavus J1074 are present, which are all related to NRPS biosynthesis. (D) Clustermap of core genes in the subcluster, which are mainly orthologous genes involved with siderophore production.*

Uncharacterized iModulons provide a road map to discover novel functions

Streptomyces contain around 30% genes annotated as uncharacterized or with unknown function, and many more with no experimental validation. Thus, there were

several difficult to characterize iModulons, however, the iModulome analysis allowed us to infer putative functions of the more difficult to characterize iModulons. For instance, UC-3 contains 48 genes, 18.8% of which are uncharacterized. Orthologous genes of this iModulon are active in RpoS regulated iModulons of other species, suggesting UC-3 is involved in stress response (Fig 5A). This cluster of iModulons contains 15 putative core genes, including *osmY*, *otsA*, and *otsB* (Fig 5B). Moreover, UC-3 shows an increase in gene expression at 48h and 72h post-induction, suggesting it might be a stress response involved in development and metabolic shift from primary to secondary metabolism (Fig 5C).

Another two iModulons, UC-5 and UC-10, contain six and 90 genes, respectively, with 66.7% uncharacterized genes in UC-5 and 34.4% in UC-10. These iModulons cluster together with iModulons related to iron acquisition and transport in other organisms (Fig 5D). This cluster contains 20 core genes mainly involved in iron acquisition and transport (Fig 5E). The genes that are most similar between UC-5 and UC-10 with others in the cluster are *entS*, *bfd*, *fepD*, *nfeF*, *aztA*, and *fepG* orthologs, suggesting these iModulons are involved in iron homeostasis (Fig 5E). Interestingly, these two iModulons are significantly downregulated in all samples which have modified the strain to remove endogenous BGCs, and heterologously express Nybomycin and Pamamycin (Fig 5F). Combined, 40% of uncharacterized genes in *S. albidoflavus* were present in at least one iModulon, providing a pathway to validate gene function in future studies.

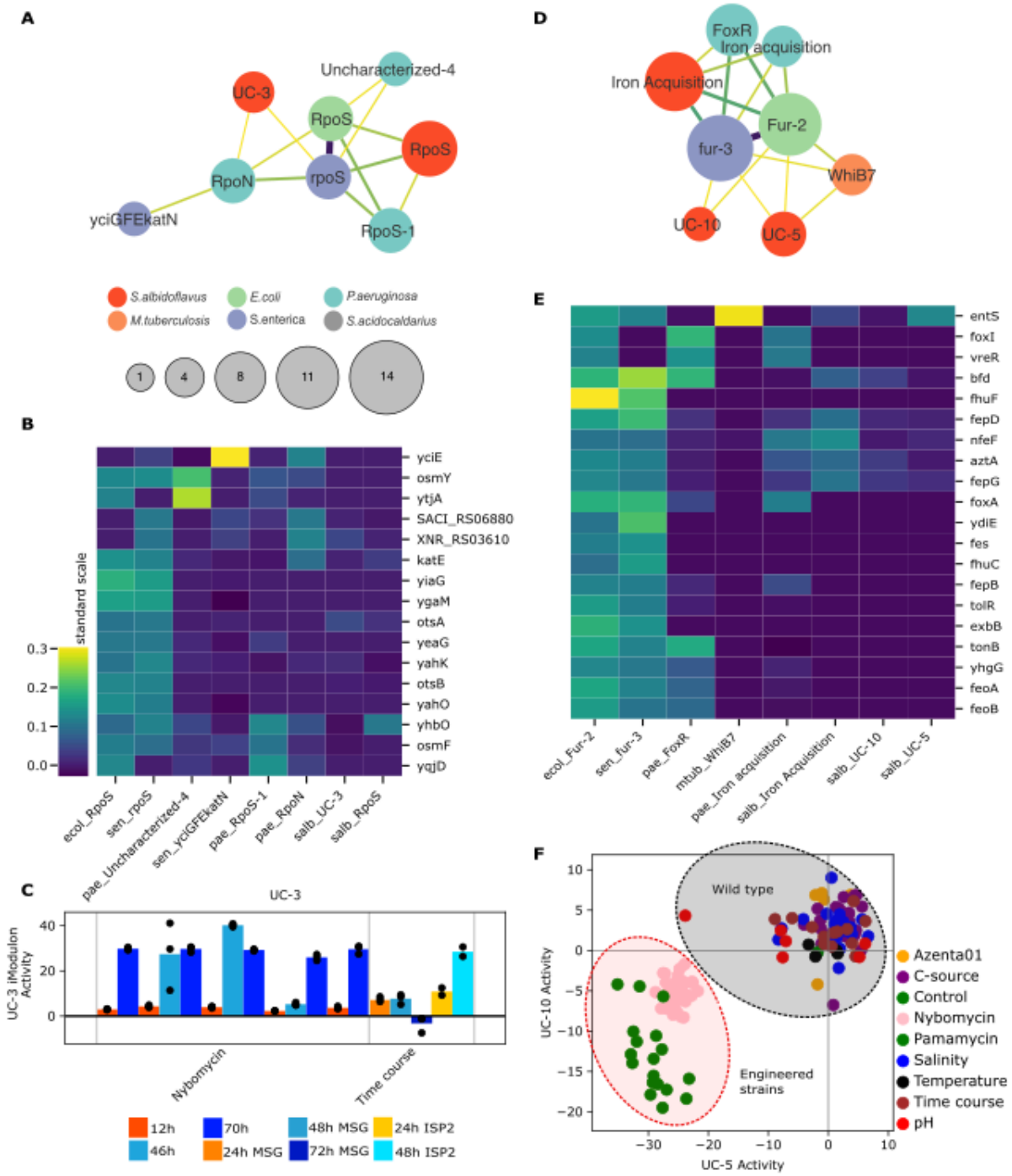

Figure 5. Discerning functions of uncharacterized iModulons. *(A) The UC-3 iModulon cluster together with stress related iModulons of other species in the iModulome. (B) The core genes of the cluster make up a general stress response. (C) UC-3 is mainly activated at later growth phases. (D) UC-5 and UC-10 cluster together with iModulons related to iron acquisition in other species in the iModulome. (E) Core genes of the subcluster are involved with iron homeostasis. (F) UC-5 and UC-10 are deactivated in modified strains that aim to heterologously produce new compounds.*

Discussion

S. albidoflavus J1074 holds great potential as a platform for heterologous expression of valuable compounds. While the current production yields remain a challenge (Gummerlich et al., 2021; Li et al., 2023), our holistic effort using ICA on a large compendium of in-house generated and publicly available RNA-Seq data to generate iModulons has established a knowledge base that spans primary and secondary metabolism, stress responses, and other important biological processes. Notably, this approach captured over half of the predicted BGCs in the *S. albidoflavus* J1074 genome, shedding light on the conditions that activate and suppress these clusters and identifying putative regulators. Moreover, This study represents the first comprehensive, systems-level perspective of global changes in *Streptomyces* gene expression profiles, and provides a detailed comparison of iModulon structure between organisms.

Enriched knowledge base for *S. albidoflavus* J1074

Despite the wide use and extensive studies conducted on *S. albidoflavus* J1074 there remains a significant gap in curated data and knowledge available to the scientific community. Our large compendium of in-house generated RNA-Seq transcriptomes, encompassing a wide range of growth conditions, including growth on different carbon sources, and stresses such as salinity, pH and temperature, significantly increases the number of publicly available datasets on SRA NCBI available to researchers. This compendium, coupled with other high-quality RNA-Seq datasets forms the foundations for our top-down exploration of the transcriptional regulatory network in *S. albidoflavus* J1074.

The iModulome provides insight into conserved and lineage-specific gene regulation

Using imodulondb.org and orthofinder (Emms & Kelly, 2019), we examined orthologous gene weight similarities within iModulons from diverse organisms, constructing a similarity-based network that clusters iModulons based on function rather than species affiliation, which we call the iModulome. This approach not only facilitated the annotation of iModulons in *S. albidoflavus* J1074 but also placed the iModulons in an evolutionary context, revealing conserved and lineage-specific aspects of the TRN. Notably, certain iModulons showed high conservation, with key genes playing pivotal roles across species, exemplified by the ribosomal subunit genes in Translation-related iModulons, many of which are found in the pangenome of the last common ancestor (Hyun & Palsson, 2023). Similarly, the phosphate-related iModulons demonstrated high similarity across different species, emphasizing conserved genes, such as those in the pstSCAB operon, contributing to phosphate uptake, however also revealing coregulated genes in a lineage-specific

manner. The iModulome analysis can help us understand how the TRN adapts to various stimuli and how it reflects the evolutionary history and functional diversity of bacteria. This suggests that our results will be broadly applicable to other non-model organisms and *Streptomyces* species (Dalldorf et al., 2023; Yuan et al., 2022).

iModulons give insights into the regulation of secondary metabolites

Activation of BGCs is crucial for producing valuable compounds in *Streptomycetes*, but their regulation is often poorly understood (Chen et al., 2010; Kallifidas et al., 2018; Myronovskyi et al., 2018; Robertsen et al., 2017). We identified 13 iModulons that capture 12 endogenous BGCs, showing the conditions and co-regulated genes that affect their expression. The high number of iModulons related to secondary metabolism reflects its significance and diversity in *Streptomyces*, often encoding over 20 BGCs (Baltz, 2010; Robertsen et al., 2017; Wohlleben et al., 2017). We found that the expression of BGCs is influenced by various factors, such as carbon source, stress, growth phase, and phosphate availability. Interestingly, some BGC-related iModulons clustered with primary metabolism iModulons, indicating a coordination of metabolic pathways in *Streptomyces* (Barreiro & Martínez-Castro, 2019; Hesketh et al., 2002; Martín et al., 2011; Sola-Landa et al., 2003).

We further investigated the regulation of surugamide, a cyclic NRPS with antimicrobial properties (Ninomiya et al., 2016; Takada et al., 2013) and identified two iModulons with putative opposite effects on surugamide production: one that activates production under stress conditions, and one that represses it under favorable conditions. The activating iModulon contains co-regulated genes involved with amino acid transport and metabolism, NRPS production, and an acyl-CoA dehydrogenase (XNR_RS11615), which is required for avenolide activity, a hormone-like signaling molecule that appears important for the production of secondary metabolites in *S. albidoflavus* J1074 (Ahmed et al., 2017; Kong et al., 2019). The repressing iModulon contains the main repressor of surugamide, *surR* and a *surR*-like located around 2 Mbp downstream of the surugamide BGC (Takada et al., 2013). These iModulons reveal the complex and dynamic regulation of surugamide, which may have ecological and evolutionary significance.

Finally, we observed a relationship between several NRPS compounds, including cyclofaulknamycin, dudomycin, and minimycin, and iron acquisition in other organisms. The genes that drove this clustering appear homologous to genes involved in the synthesis of the siderophore enterobactin in *E. coli*, suggesting that these NRPSs may also function as iron scavengers in nutrient limiting conditions (Crosa & Walsh, 2002; Ehmann et al., 2000; Gehring et al., 1998). The iModulons thus provide valuable insights into the intricate regulatory networks governing secondary metabolism in *Streptomyces*, opening avenues for future research and applications.

iModulons help infer potential functions of uncharacterized genes

Several iModulons were difficult to characterize due to a high number of uncharacterized genes and scarcely available information about regulons. The *S. albidoflavus* J1074 genome consists of 29.5% genes with uncharacterized or unknown functions, and many more that lack experimental validation. However, the iModulon analysis allows us to infer potential functions for 40% of these genes, providing valuable experimental targets going forward. For instance, several of the iModulons related to secondary metabolites contain co-regulated genes with unknown functions, many of which are located outside of the main BGC, and therefore may be necessary components for the successful production of secondary metabolites. Furthermore, two iModulons UC-5 and UC-10 appear to be stress-related iModulons due to genome modifications such as endogenous BGC removal and heterologous expression. Compared with other species, they commonly activate siderophore related genes which can be a part of stress response (Peralta et al., 2016). The iModulons generated in this study provides a roadmap to experimentally validate these uncharacterized iModulons and the uncharacterized genes within iModulons.

Conclusion

This is the first study to compile a large compendium of in-house generated and publicly available transcriptomes to identify iModulons in *Streptomyces*, providing a comprehensive and quantitative understanding of the TRN of *S. albidoflavus* J1074. We showed how the iModulons can be used to (i) compare the TRN of *S. albidoflavus* J1074 with other organisms on imodulondb, revealing both conserved and lineage-specific features, (ii) elucidate the regulation of BGCs, such as surugamide, cyclofaulknamycin, and dudomycin, by identifying putative regulators and condition-dependent activation patterns, and (iii) infer the functions of 40% of the uncharacterized genes in the genome, by associating them with iModulons. Our findings pave the way for future studies to extend and improve the iModulon framework by incorporating new conditions and targeted mutants for transcription factors. Moreover, our results suggest new strategies for rational strain design to optimize the production of valuable natural products by *S. albidoflavus* J1074 and other *Streptomyces* species.

Methods

The methods followed in this study for data acquisition and processing, performing ICA, and iModulon characterization have been adapted from Sastry et al. (Sastry et al., 2019).

RNA extraction and library preparation

We grew *S. albidoflavus* J1074 and extracted RNA from 93 samples, with at least one biological replicate, for 39 unique growth conditions, including supplemented

minimal media with different carbon sources, salinity, pH and temperature stresses. In brief, *S. albidoflavus* J1074 was cultivated aerobically in ISP2 (4 g/L yeast extract, 10 g/L malt extract, 4 g/L dextrose and distilled water, pH 7), DNPM (40 g/L dextrin, 7.5 g/L bacteriological peptone from meat, 5 g/L yeast extract, 21 g/L MOPS, 750 mL miliQ water, 250 mL tap water, pH 7) or minimal medium (6 g/L $(NH_4)_2SO_4$, 1.86 g/L citric acid, 0.21 g/L $FeCl_3$, 3.33 mL/L 2M $MgSO_4$, 250 g/L of monosodium glutamate, 0.1 M phosphate buffer pH 7.0 and 1.6 mL/L trace metal solution (Roubos et al., 2001) (20.4 g/L H_2SO_4 96%, 50 g/L citrate·H_2O, 16.75 g/L $ZnSO_4 \cdot 7H_2O$, 2.5 g/L $CuSO_4 \cdot 5H_2O$, 1.5 g/L $MnCl_2 \cdot 4H_2O$, 2 g/L H_3BO_3, and 2 g/L $NaMoO_4 \cdot 2H_2O$). In culture conditions where different carbon sources were tested, monosodium glutamate was replaced by selected carbon sources at 66 mM. In salinity stress conditions, NaCl solutions were used to supplement the correspondent medium. For routine cultivation, the temperature was set up at 30Cº.

Total RNA was extracted using the RNeasy Mini QIAcube kit (Qiagen; cat. No./ ID: 74116) and treated with DNase I according to the manufacturer's instructions. RNA concentration and integrity were verified using Qubit 2.0 (Invitrogen) and Fragment analyzer (Agilent 530V). rRNA purification, library preparation and RNA-Seq were performed by Azenta Life Sciences. NEBNext rRNA Depletion kit (Bacteria) (New England Biolabs; cat. No. E7850X) was used for rRNA depletion. Barcoded libraries were prepared for each condition in duplicate with NEBNext Ultra II Directional RNA Library Prep Kit for Illumina using default protocols. Libraries for all samples were sequenced using Illumina NovaSeq 6000.

Data acquisition and processing

Data processing and quality control is detailed in Sastry et al. (Sastry et al., 2021) and available at https://github.com/avsastry/modulome-workflow/. Briefly, read trimming was performed using Trim Galore with the default options (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), followed by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, reads were aligned to the *S. albidoflavus* reference genome from NCBI (NC_020990.1) using Bowtie (Langmead & Salzberg, 2012). The read direction was inferred using RSEQC (Wang et al., 2012) before generating read counts using featureCounts (Liao et al., 2014). Lastly, all quality control metrics were compiled using MultiQC (Ewels et al., 2016).

Samples that failed any of the four FASTQC metrics were discarded: per base sequence quality, per sequence quality scores, per base n content, and adapter content. Furthermore, samples with less than 500 000 reads mapped to coding sequences were removed, and hierarchical clustering was used to identify any outlier samples. We only analyzed samples with at least two replicates and if the Pearson R correlation between the replicates was > 0.90. To facilitate cross-dataset comparison, we normalized all expression profiles against the two in-house

generated control samples. After quality control and normalization, the final data compendium comprised 142 high-quality expression profiles: 85 generated from this study and 57 downloaded from NCBI-SRA. The two public datasets that met our quality control criteria were named the Nybomycin project, which tested the heterologous expression of Nybomycin in the J1074 Del14 strain (Bioproject: PRJNA1003853), and the Pamamycin project, which heterologously expressed pamamycin in the wild type J1074 strain in a published study (Gläser et al., 2021).

Performing Independent component analysis (ICA)

Independent components analysis was implemented as per McConn et al. (McConn et al., 2021) on the RNA-Seq compendium. Specifically, the scikit-learn implementation of FastICA was executed 100 times with random seeds and a convergence tolerance of 10–7 (Hyvarinen, 1999; Pedregosa et al., 2011). The resulting independent components (ICs) were clustered using DBSCAN to identify robust ICs, using an epsilon of 0.1 and a minimum cluster seed size of 50. To account for identical components with opposite signs, the following distance metric was used for computing the distance matrix:

$$dx, y = 1 - ||||\rho x, y||||$$

where $\rho x, y$ is the Pearson correlation between components x and y. The final robust ICs were defined as the centroids of the cluster.

Since the number of dimensions selected in ICA can alter the results, we applied the above procedure to the dataset multiple times, ranging the number of dimensions from 10 to 140 (i.e. the approximate size of the dataset) with a step size of 10. 140 dimensions generated the most robust iModulons, and was used for this analysis.

Characterization of iModulons

To facilitate iModulon characterization, we used the PyModulon Python package (Sastry et al., 2021). As *S. albidoflavus* has a poorly documented TRN, k-means clustering was used to identify component-specific thresholds, with manual editing, if necessary, to improve characterization efforts. The iModulons were characterized by information from public databases and literature. Code for our analysis pipeline and description of imodulons is maintained on github (https://github.com/biosustain/salb_imodulons).

Cross-species comparison of iModulons

To enable comparisons of iModulon structures between organisms we first identified orthologous genes among five other species, which currently have iModulon datasets on imodulondb.org using Orthofinder with default settings (Emms & Kelly, 2019). All iModulon data files and objects were downloaded from imodulondb.org (Rychel et al., 2021). The gene weight files of all species describing the iModulon

memberships were merged based on the orthogroup identified by orthofinder. If more than one gene per species were present in a given orthogroup only the highest gene weight was kept per iModulon. Next, we conducted cosine similarity in a pairwise manner for all iModulons, and visualized the similarities using Cytoscape (Smoot et al., 2011). Edge weights were filtered to only show the top 1500 edge weights, and we used the MCL cluster algorithm from the clusterMaker app to identify clusters of iModulons (Morris et al., 2011). Lastly, phylogenetic trees from resolved gene trees generated by orthofinder were visualized using iTol (Letunic & Bork, 2021).Multiple sequence alignments were aligned using MUSCLE (Edgar, 2004), and visualized using ESPript 3 (Robert & Gouet, 2014).

Data/code availability

All RNA-seq raw files generated from this study are available from NCBI at bioproject PRJNA1062162. Source codes for iModulon analysis and figures are available at https://github.com/biosustain/salb_imodulons.

Funding

This work was supported by The Novo Nordisk Foundation (NNF) Center for Biosustainability (CfB) at the Technical University of Denmark (NNF20CC0035580)

Acknowledgements

We are grateful to Pep Charusanti, Anna Svetlova, Tilmann Weber, Kai Blin and Thomas Booth for the informative discussions and input when characterizing the iModulons.

References

Ahmed, Y., Rebets, Y., Tokovenko, B., Brötz, E., & Luzhetskyy, A. (2017). Identification of butenolide regulatory system controlling secondary metabolism in *Streptomyces albus* J1074. *Scientific Reports*, *7*(1), 9784. https://doi.org/10.1038/s41598-017-10316-y

Baltz, R. H. (2010). *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *Journal of Industrial Microbiology and Biotechnology*, *37*(8), 759–772. https://doi.org/10.1007/s10295-010-0730-9

Barbuto Ferraiuolo, S., Cammarota, M., Schiraldi, C., & Restaino, O. F. (2021). *Streptomycetes* as platform for biotechnological production processes of drugs. *Applied Microbiology and Biotechnology*, *105*(2), 551–568. https://doi.org/10.1007/s00253-020-11064-2

Barreiro, C., & Martínez-Castro, M. (2019). Regulation of the phosphate metabolism in *Streptomyces* genus: Impact on the secondary metabolites. *Applied Microbiology and Biotechnology*, *103*(4), 1643–1658. https://doi.org/10.1007/s00253-018-09600-2

Bilyk, O., Sekurova, O. N., Zotchev, S. B., & Luzhetskyy, A. (2016). Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. *PLOS ONE*, *11*(7), e0158682. https://doi.org/10.1371/journal.pone.0158682

Challis, G. L. (2014). Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *Journal of Industrial Microbiology and Biotechnology*, *41*(2), 219–232. https://doi.org/10.1007/s10295-013-1383-2

Chater, K. F. (2016). Recent advances in understanding *Streptomyces* [version 1; peer review: 4 approved]. *F1000Research*, *5*(F1000 Faculty Rev), 2795. https://doi.org/10.12688/f1000research.9534.1

Chauhan, S. M., Poudel, S., Rychel, K., Lamoureux, C., Yoo, R., Al Bulushi, T., Yuan, Y., Palsson, B. O., & Sastry, A. V. (2021). Machine Learning Uncovers a Data-Driven Transcriptional Regulatory Network for the Crenarchaeal Thermoacidophile *Sulfolobus* acidocaldarius. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.753521

Chen, Y., Smanski, M. J., & Shen, B. (2010). Improvement of secondary metabolite production in *Streptomyces* by manipulating pathway regulation. *Applied Microbiology and Biotechnology*, *86*(1), 19–25. https://doi.org/10.1007/s00253-009-2428-3

Chen, Y., Wendt-Pienkowski, E., & Shen, B. (2008). Identification and Utility of FdmR1 as a *Streptomyces* Antibiotic Regulatory Protein Activator for Fredericamycin Production in *Streptomyces griseus* ATCC 49344 and Heterologous Hosts. *Journal of Bacteriology*, *190*(16), 5587–5596. https://doi.org/10.1128/jb.00592-08

Claessen, D., de Jong, W., Dijkhuizen, L., & Wösten, H. A. B. (2006). Regulation of *Streptomyces* development: Reach for the sky! *Trends in Microbiology*, *14*(7), 313–319. https://doi.org/10.1016/j.tim.2006.05.008

Crosa, J. H., & Walsh, C. T. (2002). Genetics and Assembly Line Enzymology of Siderophore Biosynthesis in Bacteria. *Microbiology and Molecular Biology Reviews*, *66*(2), 223–249. https://doi.org/10.1128/mmbr.66.2.223-249.2002

Dalldorf, C., Rychel, K., Szubin, R., Hefner, Y., Patel, A., Zielinski, D. C., & Palsson, B. O. (2023). The hallmarks of a tradeoff in transcriptomes that balances stress and growth functions. *Research Square*, rs.3.rs-2729651. https://doi.org/10.21203/rs.3.rs-2729651/v1

de Lima Procópio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L., & de Araújo, J. M. (2012). Antibiotics produced by *Streptomyces*. *The Brazilian Journal of Infectious Diseases*, *16*(5), 466–471. https://doi.org/10.1016/j.bjid.2012.08.014

Dolya, B., Hryhorieva, O., Sorochynska, K., Lopatniuk, M., Ostash, I., Tseduliak, V.-M., Sterndorff, E. B., Jørgensen, T. S., Gren, T., Dacyuk, Y., Weber, T.,

Luzhetskyy, A., Fedorenko, V., & Ostash, B. (2023). Properties of Multidrug-Resistant Mutants Derived from Heterologous Expression Chassis Strain *Streptomyces albidoflavus* J1074. *Microorganisms*, *11*(5), 1176. https://doi.org/10.3390/microorganisms11051176

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C., & Walsh, C. T. (2000). The *EntF* and *EntE* adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences*, *97*(6), 2509–2514. https://doi.org/10.1073/pnas.040572897

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fang, X., Sastry, A., Mih, N., Kim, D., Tan, J., Yurkovich, J. T., Lloyd, C. J., Gao, Y., Yang, L., & Palsson, B. O. (2017). Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. *Proceedings of the National Academy of Sciences*, *114*(38), 10286–10291. https://doi.org/10.1073/pnas.1702581114

Gehring, A. M., Mori, I., & Walsh, C. T. (1998). Reconstitution and Characterization of the *Escherichia coli* Enterobactin Synthetase from *EntB, EntE*, and *EntF*. *Biochemistry*, *37*(8), 2648–2659. https://doi.org/10.1021/bi9726584

Gläser, L., Kuhl, M., Stegmüller, J., Rückert, C., Myronovskyi, M., Kalinowski, J., Luzhetskyy, A., & Wittmann, C. (2021). Superior production of heavy pamamycin derivatives using a *bkdR* deletion mutant of *Streptomyces albus* J1074/R2. *Microbial Cell Factories*, *20*(1), 111. https://doi.org/10.1186/s12934-021-01602-6

Gullón, S., Olano, C., Abdelfattah, M. S., Braña, A. F., Rohr, J., Méndez, C., & Salas, J. A. (2006). Isolation, Characterization, and Heterologous Expression of the Biosynthesis Gene Cluster for the Antitumor Anthracycline Steffimycin. *Applied and Environmental Microbiology*, *72*(6), 4172–4183. https://doi.org/10.1128/AEM.00734-06

Gummerlich, N., Manderscheid, N., Rebets, Y., Myronovskyi, M., Gläser, L., Kuhl, M., Wittmann, C., & Luzhetskyy, A. (2021). Engineering the precursor pool to modulate the production of pamamycins in the heterologous host *S. albus* J1074. *Metabolic Engineering*, *67*, 11–18. https://doi.org/10.1016/j.ymben.2021.05.004

Hesketh, A. R., Chandra, G., Shaw, A. D., Rowland, J. J., Kell, D. B., Bibb, M. J., & Chater, K. F. (2002). Primary and secondary metabolism, and

post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. *Molecular Microbiology*, *46*(4), 917–932. https://doi.org/10.1046/j.1365-2958.2002.03219.x

Huang, S., Zhao, Y., Qin, Z., Wang, X., Onega, M., Chen, L., He, J., Yu, Y., & Deng, H. (2011). Identification and heterologous expression of the biosynthetic gene cluster for holomycin produced by *Streptomyces clavuligerus*. *Process Biochemistry*, *46*(3), 811–816. https://doi.org/10.1016/j.procbio.2010.11.024

Hwang, S., Lee, N., Jeong, Y., Lee, Y., Kim, W., Cho, S., Palsson, B. O., & Cho, B.-K. (2019). Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Nucleic Acids Research*, *47*(12), 6114–6129. https://doi.org/10.1093/nar/gkz471

Hyun, J. C., & Palsson, B. O. (2023). Reconstruction of the last bacterial common ancestor from 183 pangenomes reveals a versatile ancient core genome. *Genome Biology*, *24*(1), 183. https://doi.org/10.1186/s13059-023-03028-2

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. https://doi.org/10.1109/72.761722

Jeong, Y., Kim, J.-N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., Kim, B.-G., Roe, J.-H., Kim, S. C., Smith, C. P., & Cho, B.-K. (2016). The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). *Nature Communications*, *7*(1), 11605. https://doi.org/10.1038/ncomms11605

Kallifidas, D., Jiang, G., Ding, Y., & Luesch, H. (2018). Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. *Microbial Cell Factories*, *17*, 25. https://doi.org/10.1186/s12934-018-0874-2

Kong, D., Wang, X., Nie, J., & Niu, G. (2019). Regulation of Antibiotic Production by Signaling Molecules in *Streptomyces*. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02927

Lamoureux, C. R., Decker, K. T., Sastry, A. V., Rychel, K., Gao, Y., McConn, J. L., Zielinski, D. C., & Palsson, B. O. (2023). A multi-scale expression and regulation knowledge base for *Escherichia coli*. *Nucleic Acids Research*, *51*(19), 10176–10193. https://doi.org/10.1093/nar/gkad750

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, X., Guo, R., Luan, J., Fu, J., Zhang, Y., & Wang, H. (2023). Improving spinosad production by tuning expressions of the forosamine methyltransferase and the forosaminyl transferase to reduce undesired less active byproducts in the heterologous host *Streptomyces albus* J1074. *Microbial Cell Factories*, *22*(1), 15. https://doi.org/10.1186/s12934-023-02023-3

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Luo, S., Chen, X.-A., Mao, X.-M., & Li, Y.-Q. (2018). Transposon-based identification of a negative regulator for the antibiotic hyper-production in *Streptomyces*. *Applied Microbiology and Biotechnology*, *102*(15), 6581–6592. https://doi.org/10.1007/s00253-018-9103-5

Marín, L., Gutiérrez-del-Río, I., Yagüe, P., Manteca, Á., Villar, C. J., & Lombó, F. (2017). De Novo Biosynthesis of Apigenin, Luteolin, and Eriodictyol in the Actinomycete *Streptomyces albus* and Production Improvement by Feeding and Spore Conditioning. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00921

Martín, J. F., Sola-Landa, A., Santos-Beneit, F., Fernández-Martínez, L. T., Prieto, C., & Rodríguez-García, A. (2011). Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *Microbial Biotechnology*, *4*(2), 165–174. https://doi.org/10.1111/j.1751-7915.2010.00235.x

McConn, J. L., Lamoureux, C. R., Poudel, S., Palsson, B. O., & Sastry, A. V. (2021). Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinformatics*, *22*(1), 584. https://doi.org/10.1186/s12859-021-04497-7

McCormick, J. R., & Flärdh, K. (2012). Signals and regulators that govern *Streptomyces* development. *FEMS Microbiology Reviews*, *36*(1), 206–231. https://doi.org/10.1111/j.1574-6976.2011.00317.x

Morris, J. H., Apeltsin, L., Newman, A. M., Baumbach, J., Wittkop, T., Su, G., Bader, G. D., & Ferrin, T. E. (2011). clusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, *12*, 436. https://doi.org/10.1186/1471-2105-12-436

Myronovskyi, M., Brötz, E., Rosenkränzer, B., Manderscheid, N., Tokovenko, B., Rebets, Y., & Luzhetskyy, A. (2016). Generation of new compounds through unbalanced transcription of landomycin A cluster. *Applied Microbiology and Biotechnology*, *100*, 9175–9186. https://doi.org/10.1007/s00253-016-7721-3

Myronovskyi, M., & Luzhetskyy, A. (2019). Heterologous production of small molecules in the optimized *Streptomyces* hosts. *Natural Product Reports*, *36*(9), 1281–1294. https://doi.org/10.1039/C9NP00023B

Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., & Luzhetskyy, A. (2018). Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters. *Metabolic Engineering, 49,* 316-324. https://doi.org/10.1016/j.ymben.2018.09.004

Nett, M., Ikeda, H., & Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports*, *26*(11), 1362–1384. https://doi.org/10.1039/b817069j

Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S.,

Wakimoto, T., Ohnishi, Y., Matsunaga, S., & Takada, K. (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an Unrelated Decapeptide, Surugamide F. *ChemBioChem*, *17*(18), 1709–1712. https://doi.org/10.1002/cbic.201600350

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12,* 2825–2830.

Peralta, D. R., Adler, C., Corbalán, N. S., García, E. C. P., Pomares, M. F., & Vincent, P. A. (2016). Enterobactin as Part of the Oxidative Stress Response Repertoire. *PLOS ONE*, *11*(6), e0157799. https://doi.org/10.1371/journal.pone.0157799

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), W320–W324. https://doi.org/10.1093/nar/gku316

Robertsen, H. L., Weber, T., Kim, H. U., & Lee, S. Y. (2017). Toward Systems Metabolic Engineering of Streptomycetes for Secondary Metabolites Production. *Biotechnology Journal, 13*(1), 1700465. https://doi.org/10.1002/biot.201700465

Rychel, K., Sastry, A. V., & Palsson, B. O. (2020). Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. *Nature Communications, 11*, 6338. https://doi.org/10.1038/s41467-020-20153-9

Rychel, K., Decker, K., Sastry, A. V., Phaneuf, P. V., Poudel, S., & Palsson, B. O. (2021). iModulonDB: a knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Research*, *49*(D1), D112–D120. https://doi.org/10.1093/nar/gkaa810

Sastry, A. V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., Choudhary, K. S., Yang, L., King, Z. A., & Palsson, B. O. (2019). The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nature Communications, 10,* 5536. https://doi.org/10.1038/s41467-019-13483-w

Sastry, A. V., Poudel, S., Rychel, K., Yoo, R., Lamoureux, C. R., Chauhan, S., Haiman, Z. B., Bulushi, T. A., Seif, Y., & Palsson, B. O. (2021). Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. *bioRxiv*, 2021.07.01.450581. https://doi.org/10.1101/2021.07.01.450581

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., & Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, *27*(3), 431–432. https://doi.org/10.1093/bioinformatics/btq675

Sola-Landa, A., Moura, R. S., & Martín, J. F. (2003). The two-component *PhoR-PhoP* system controls both primary metabolism and secondary metabolite biosynthesis in *Streptomyces lividans*. *Proceedings of the National Academy of Sciences*, *100*(10), 6133–6138. https://doi.org/10.1073/pnas.0931429100

Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., Okada, S., &

Matsunaga, S. (2013). Surugamides A–E, Cyclic Octapeptides with Four d-Amino Acid Residues, from a Marine *Streptomyces* sp.: LC–MS-Aided Inspection of Partial Hydrolysates for the Distinction of d- and l-Amino Acid Residues in the Sequence. *The Journal of Organic Chemistry*, *78*(13), 6746–6750. https://doi.org/10.1021/jo400708u

Wang, L., Wang, S., & Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184–2185. https://doi.org/10.1093/bioinformatics/bts356

Wohlleben, W., Bera, A., Mast, Y., & Stegmann, E. (2017). Regulation of Secondary Metabolites of Actinobacteria. In J. Wink, F. Mohammadipanah, & J. Hamedi (Eds.), *Biology and Biotechnology of Actinobacteria* (pp. 181–232). Springer International Publishing. https://doi.org/10.1007/978-3-319-60339-1_8

Xia, H., Li, X., Li, Z., Zhan, X., Mao, X., & Li, Y. (2020). The Application of Regulatory Cascades in *Streptomyces*: Yield Enhancement and Metabolite Mining. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00406

Xu, F., Nazari, B., Moon, K., Bushin, L. B., & Seyedsayamdost, M. R. (2017). Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens. *Journal of the American Chemical Society*, *139*(27), 9203–9212. https://doi.org/10.1021/jacs.7b02716

Yi, J. S., Kim, M. W., Kim, M., Jeong, Y., Kim, E.-J., Cho, B.-K., & Kim, B.-G. (2017). A Novel Approach for Gene Expression Optimization through Native Promoter and 5′ UTR Combinations Based on RNA-seq, Ribo-seq, and TSS-seq of *Streptomyces coelicolor*. *ACS Synthetic Biology*, *6*(3), 555–565. https://doi.org/10.1021/acssynbio.6b00263

Yoo, R., Rychel, K., Poudel, S., Al-bulushi, T., Yuan, Y., Chauhan, S., Lamoureux, C., Palsson, B. O., & Sastry, A. (2022). Machine Learning of All *Mycobacterium tuberculosis* H37Rv RNA-seq Data Reveals a Structured Interplay between Metabolism, Stress Response, and Infection. *mSphere*, *7*(2), e00033-22. https://doi.org/10.1128/msphere.00033-22

Yuan, Y., Seif, Y., Rychel, K., Yoo, R., Chauhan, S., Poudel, S., Al-bulushi, T., Palsson, B. O., & Sastry, A. V. (2022). Pan-Genome Analysis of Transcriptional Regulation in Six *Salmonella enterica* Serovar Typhimurium Strains Reveals Their Different Regulatory Structures. *mSystems*, *7*(6), e00467-22. https://doi.org/10.1128/msystems.00467-22

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., & Luzhetskyy, A. (2014). Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics, 15,* 97. https://doi.org/10.1186/1471-2164-15-97

Zhang, Z., Du, C., Barsy, F. D., Liem, M., Liakopoulos, A., Wezel, G. P. V., Choi, Y. H., Claessen, D., & Rozen, D. E. (2020). Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Science Advances, 6*(3), eaay5781. https://doi.org/10.1126/sciadv.aay5781