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

Mathias Jönsson1, Renata Sigrist1, Tetiana Gren1, Mykhaylo Semenov Petrov1, Nils Marcussen1, Anna Svetlova1, Pep Charusanti1, Peter Gockel1, Bernhard O. Palsson1,2, Lei Yang1,\*, Emre Özdemir1,\*

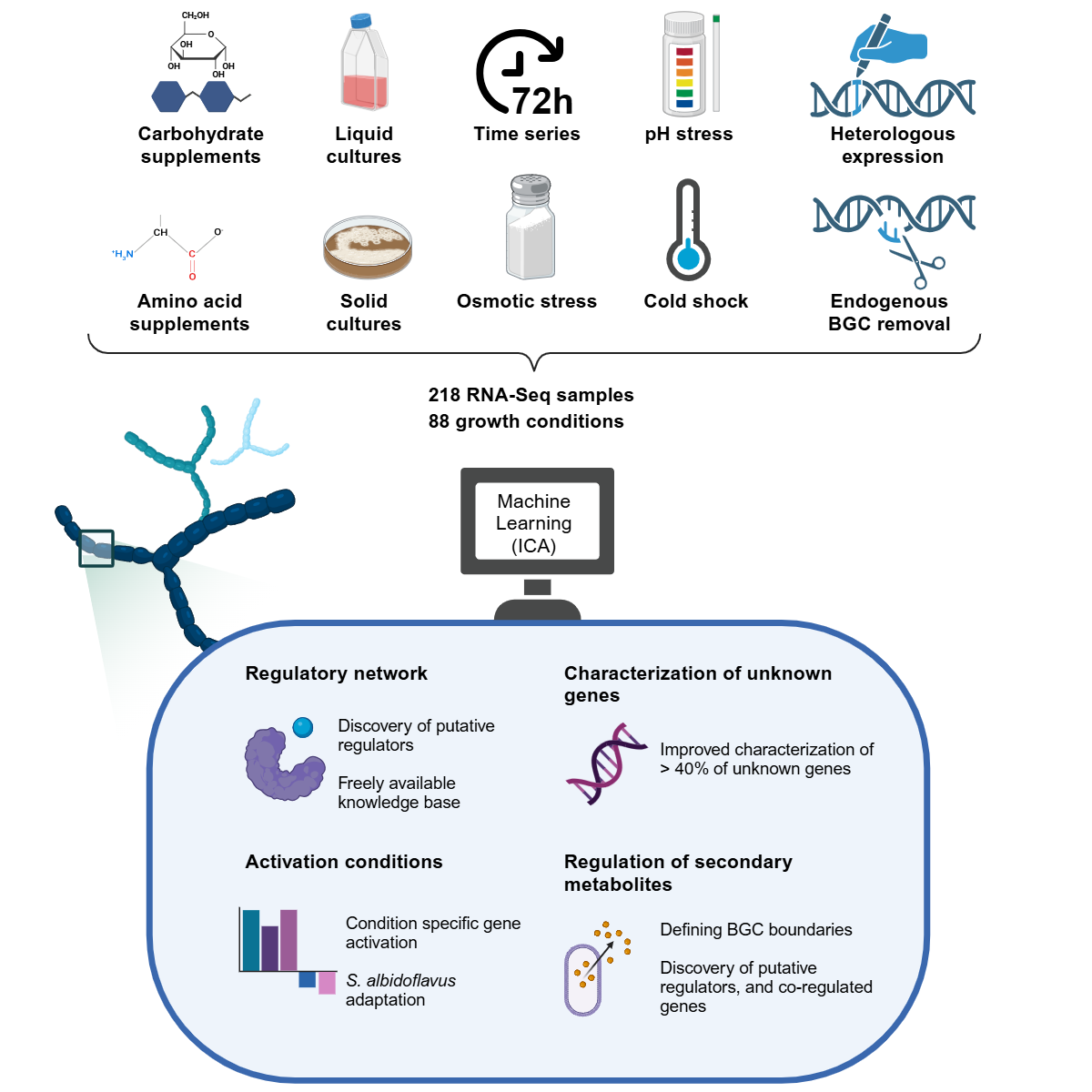
1 Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800, Kgs. Lyngby, Denmark

2 Department of Bioengineering, University of California, San Diego, La Jolla, USA

\* To whom correspondence should be addressed.

Email: [leiya@biosustain.dtu.dk](mailto:leiya@biosustain.dtu.dk), [emoz@biosustain.dtu.dk](mailto:emoz@biosustain.dtu.dk)

## Graphical Abstract



## Summary

*Streptomyces albidoflavus* is a widely used strain for natural product discovery and production through heterologous biosynthetic gene clusters (BGCs). However, the transcriptional regulatory network (TRN) and its impact on secondary metabolism remain poorly understood. Here we characterize the TRN using independent component analysis on 218 RNA-Seq transcriptomes across 88 unique growth conditions. We identify 78 independently modulated sets of genes (iModulons) that quantitatively describe the TRN across diverse conditions. Our analyses reveal (i) TRN adaptation to different growth conditions, (ii) conserved and unique characteristics of the TRN across diverse lineages, (iii) transcriptional activation of several endogenous BGCs, including surugamide, minimycin and paulomycin, and (iv) inferred functions of 40% of uncharacterized genes in the *S. albidoflavus* genome. These findings provide a comprehensive and quantitative understanding of the *S. albidoflavus* TRN, offering a knowledge base for further exploration and experimental validation.

## Introduction

*Streptomyces* is a well-known genus of Gram-positive bacteria capable of producing diverse secondary (specialized) metabolites, including antibiotics, anticancer agents, and antifungals1. These products are encoded in biosynthetic gene clusters (BGCs), with most *Streptomyces* species encoding 20-30 BGCs2–5; Many BGCs, however, are not expressed in laboratory settings. Consequently, a more thorough understanding of the regulation of these BGCs may lead to activation of more BGCs and the discovery of more novel compounds6,7.

The production of secondary metabolites is controlled, in part, by the transcriptional regulatory network (TRN), which controls the expression of genes in response to genetic and environmental changes. *Streptomyces* genomes are often large (6 - 12 Mbp), and bottom-up approaches to characterize the TRN have revealed an intricate hierarchy of regulators — ranging from global to pleiotropic and local — that activate under specific growth conditions and developmental stages, reflecting the complex life cycle of *Streptomyces*8–13. *Streptomyces albidoflavus* (previously *Streptomyces albus* J1074) is a well characterized and widely used heterologous host for studying cryptic Actinomycete BGCs and producing industrially important secondary metabolites14–23. *S. albidoflavus* has one of the smallest *Streptomyces* genomes of 6.8 Mbp, containing around 23 endogenous BGCs24. Moreover, it contains a large proportion of genomic regulators, including over 400 transcription factors and around 30 sigma factors24. The number of potential regulators together with the large genome present significant challenges for bottom-up approaches in characterizing the TRN on a large-scale.

Independent Components Analysis (ICA) is a powerful unsupervised machine learning method that provides a top-down approach to characterizing the TRN25,26. It decomposes gene expression data into sets of independently modulated sets of genes (iModulons) and their corresponding activity profiles across different conditions27–29. ICA has been used to infer genome-scale TRNs and functional modules for several bacterial species, including *Escherichia coli*, *Bacillus subtilis*, and *Mycobacterium tuberculosis,* often with strong overlap with regulons detailed in the literature28–32. Importantly, the modularization of the transcriptome into independent signals has enabled the iModulons to be comparable and transferable across species, enabling new pathways in host species, showcasing great potential for engineering microbial cell factories and biomanufacturing33.

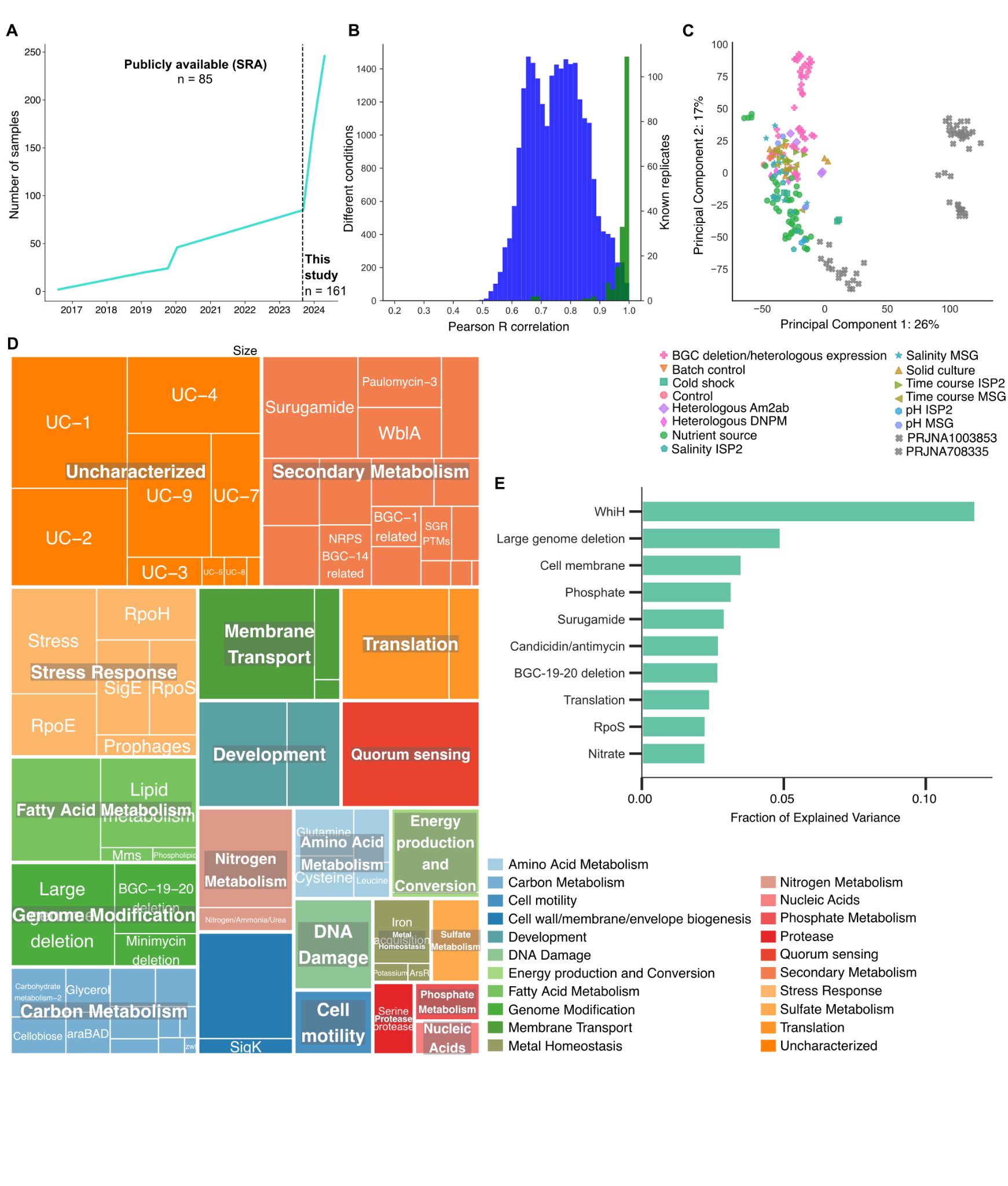
To gain a comprehensive and systematic understanding of the TRN of *S. albidoflavus*, we performed ICA on a compendium of 218 high-quality gene expression profiles, of which 161 were newly generated for this purpose. We identified 78 iModulons which reveal novel insights into the transcriptional responses of *S. albidoflavus* to relevant conditions, such as nutrient limitation, BGC removals, heterologous expression, and to various stressors. We also conduct a large-scale comparison of iModulon structure across seven different organisms to validate our findings and to gain an evolutionary context of the TRN. Our analysis reveals novel insights into BGC regulation, identifying co-regulated genes, defining BGC borders, and putative regulators. To promote the use of this knowledge base, all iModulons can be searched, browsed, and studied through [imodulondb.org](http://imodulondb.org).

## Results

### Independent component analysis reveals 78 iModulons.

To capture a wide range of expression states we cultured *S. albidoflavus* in 68 unique growth conditions. These conditions included different stressors, genome modifications, and nutrient sources, which were chosen based on activity screening using Biolog Phenotype Microarrays (Table S1.1). This in-house generated dataset comprises 161 high quality RNA-Seq transcriptomes, increasing publicly available RNA-Seq datasets from SRA NCBI by 190% (Figure 1A and Table S1.2). We combined our in-house generated dataset with public datasets from the SRA database, and after quality control, the final compendium consisted of 218 samples from 88 unique growth conditions, with a high median Pearson R score of 0.99 between replicates (Figure 1B, C).

ICA decomposed the RNA-Seq compendium into 78 iModulons (Figure 1D), where each iModulon contains a set of genes whose expression levels vary concurrently with each other, but independently of all other genes not in the given iModulon. Overall, 48.4% of all genes were assigned into an iModulon, explaining 80.2% of the variance in the gene expression (Figure 1E, Table S1.3 and Table S1.4). The iModulons cover a wide range of functions, including secondary metabolism (21.8%), carbon metabolism (15.4%), stress response (7.7%), as well as currently uncharacterized sets (11.5%) (Figure 1D). We have provided the information for each iModulon in the form of an interactive dashboard on [imodulondb.org](http://imodulondb.org)28. The dashboard is a user-friendly way for researchers to search for or browse the details of iModulons or genes of interest.

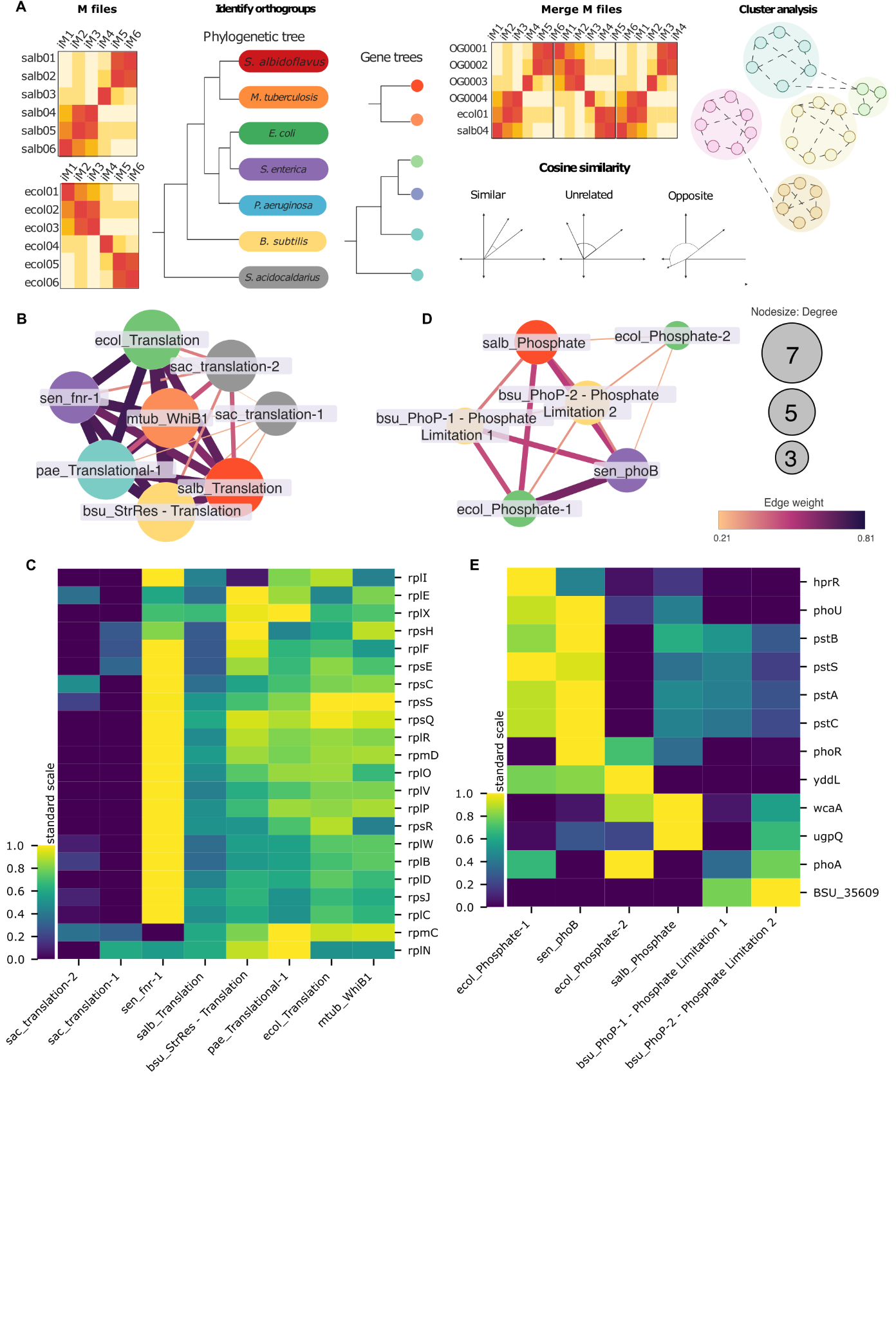


***Figure 1. Overview of the S. albidoflavus iModulons.*** *(****A****) The availability of public RNA-Seq (SRA) data for S. albidoflavus over time. (****B****) Histogram of the correlations between log-tpm profiles of all samples across all conditions (blue) and known replicates (green). The median Pearson R score between replicates is 0.99. Replicates that did not exceed 0.90 in R score were removed from further analysis. (****C****) Principal components analysis of 218 samples used for ICA. The color and symbol depict the specific project where the sample was generated. All in-house generated samples for this study are colored, and the two public datasets included in this study are labeled in grey. (****D****) A treemap depiction of the 78 identified iModulons and their associated functional categories. The dimensions of the individual boxes correlate with the number of genes found in each respective iModulon. (****E****) Highlight of the top 10 iModulons contributing the greatest proportion of explained variance. These iModulons represent key functions such as development, translation, secondary metabolite biosynthesis, and stress.*

### Cross-species iModulon comparison help characterize new iModulons

Given the limited number of regulons described for *S. albidoflavus* in the literature, we developed a method to compare and validate its iModulon structures with those of other species available on iModulonDB. We identified orthologous gene weight similarities across all iModulons from seven species, including model species such as *E. coli* and *B. subtilis*, and visualized the iModulons in a similarity-based network that cluster together based on their function, regardless of species, which we refer to as the iModulome (Figure 2A, Table S2.1–S2.5 and Figure S1). Of the 78 *S. albidoflavus* iModulons, 94.9% clustered together with at least one other iModulon allowing us to compare and characterize the respective gene contents.

For example, the Translation-related cluster of the iModulome comprises eight iModulons from the seven included species, with a high average edge weight of 0.52, indicating significant similarity in the gene weights of orthologous genes (Figure 2B). This high similarity is largely based on the expression patterns of 22 core genes, which all encode for conserved ribosomal subunits (Figure 2C and Figure S2). *S. acidocaldarius* has two iModulons in this cluster, but they are the least connected, with an average edge weight of 0.31 to the other iModulons, indicative of their more distant phylogeny. Moreover, the phosphate-related cluster of the iModulome consists of six iModulons from four species, with an average edge weight of 0.36 (Figure 2D). We identified the core genes as related to the *pstSCAB* and *PhoPR* systems, however there is greater heterogeneity regarding co-expressed genes, suggesting potential lineage-specific adaptations to gene expression under phosphate-related conditions (Figure 2E and Figure S3).



***Figure 2. S. albidoflavus iModulons similarities and differences with other organisms.*** *(****A****) Overview of the construction of the iModulome. M files generated from ICA analysis of seven organisms were collected. Orthogroups were identified using orthofinder (48) and used to merge the M files based on the specific orthogroup. Cosine similarity score and MCL clustering was used to create a clustered network of iModulons based on similarity. (****B****) A cluster from the iModulome of iModulons across seven species (abbreviations and colour: sal, Red, S. albidoflavus; mtub, Orange, M. tuberculosis; ecol, Green, E. coli; sen, Purple, S. enterica; pae, Blue, P. aeruginosa; bsub, Yellow, B. subtilis; sac, Grey, S. acidocaldarius) based on cosine similarity of ICA derived gene weights, containing iModulons related to translation processes. (****C****) Clustermap of core genes in the cluster, the scale bar depicts the standardized weights of orthologous genes, expressed as z-scores. In this example, 22 ribosomal subunits are highly weighted in iModulons related to translation across several different species, suggesting very similar iModulon structures. Notably, the Archaea S. acidocaldarius appears to lack many of the genes that are highly expressed in other organisms, indicating a greater difference in iModulon structure. (****D****) A cluster of phosphate related iModulons based on cosine similarity of ICA derived gene weights, containing iModulons related to phosphate uptake. Coloured nodes depict species as described in (B). (****E****) Clustermap of core orthologous genes in the Phosphate related cluster. The pstSCAB operon has high weights in nearly all iModulons, while for some genes (eg. yddL, phoA) iModulon membership can be species subjective.*

### iModulons give insights into the regulation and activation of BGCs.

Of the 23 predicted BGC regions in *S. albidoflavus*, 12 had core biosynthetic genes as members of 16 iModulons, including paulomycin, cyclofaulknamycin, candicidin, and surugamide (Table 1). The remaining BGCs had evenly distributed transcriptional activity across all conditions and thus were not modularized by ICA. The iModulons provide a comprehensive overview of the expression of the core biosynthetic genes across different growth conditions along with co-regulated genes, both within and outside of the BGC. For example, the BGC regions predicted by ICA were on average 62.7% smaller compared to antiSMASH predictions, with Surugamide F having the lowest coverage (4.55%) and Streptamidine covering 100% of the antiSMASH predicted region (Figure S4). Notable co-regulated genes located outside of the respective BGCs included putative regulators, amino acid transporters, and cytochrome P450 enzymes (Table S1.4). Furthermore, we detected enriched DNA binding motifs that appear upstream of the operons in BGC-related iModulons, further suggesting co-regulation (Table 1 and <https://github.com/biosustain/salb_imodulons>). Below we analyze some BGC-related iModulons in more detail, to provide novel insights into their potential regulatory mechanisms and expression.

***Table 1. Biosynthetic gene cluster representation in iModulons.*** *List of antiSMASH predicted BGC regions and corresponding iModulons. The coverage represents the percentage of genes from the antiSMASH predicted BGC region that is present in the iModulon. The highest activity column depicts the growth conditions of the samples with the highest iModulon activation scores. The motif column shows the IUPAC ambiguity codes of the top enriched motif for the operons and genes in the iModulon.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **iModulon** | **antiSMASH BGC** | **Coverage (%)** | | **Highest activity** | **Motif** |
| BGC-1 related | Region 1 | 25.00 | Solid culture | | CCYBGDSCAGVMSCTTGMCCKCCWGGDCGVCGTCSTCMHC |
| SGR PTMs | SGR PTMs | 48.28 | Nutrient source (DNPM, ISP2) | | CSGGYSSGAGMCSRSCHWRTWCRSWTCRCKCMCRWKRA |
| Paulomycin-1 | Paulomycin | 65.71 | Nutrient source (Galactose), Salinity (5% in MSG) | | CTWGGMAADGMADBSCATRDSDAHCAGRAASHSCGAVDK |
| Paulomycin-2 | Paulomycin | 77.14 | Nutrient source (Galactose), Salinity (5% in MSG) | | CSBVMHTTACVKWKGSATWNCWGGCGGSTSYCMWGRAGG |
| Paulomycin-3 | Paulomycin | 42.86 | BGC deletion/heterologous expression, Nutrient source (Starch) | | CSAGVTCGCCGWCVWSGCCGVCGVCSTSGWCGNCCKSVTS |
| Cyclofaulknamycin-1 | Cyclofaulknamycin | 48.65 | BGC deletion/heterologous expression | | CGBCCTSSWSGAVGASCWGWYCSASCTCGBCG |
| Cyclofaulknamycin-2 | Cyclofaulknamycin | 48.65 | Nutrient source (Starch), Heterologous Am2ab, BGC deletion/heterologous expression | | - |
| LAP/Thiopeptide BGC-11 | Region 11 | 51.85 | Nutrient source (Starch, DNPM) | | AGCCGGRGTSGDMGVSCWTCGRGAKCWTGNNGTASYC |
| SAL-2242 | SAL-2242 | 23.53 | Nutrient source (ISP2) | | AGGYCSAGAAKCSSCAKSWCGTMCTCSTYCGTGTCGA |
| NRPS BGC-14 related | Region 14 | 13.79 | PRJNA1003853 | | CGMGGBSACCGSCGCCSTCGYCGTCKMCG |
| Surugamide | Surugamide | 31.82 | Nutrient source (Histidine, Glycerol) | | GGYGVYSRSSRCGRBCGCSGCSRCGRCCGYSR |
| Surugamide F | Surugamide | 4.55 | BGC deletion/heterologous expression | | - |
| Surugamide repressor | Surugamide | 11.36 | Early growth stages | | CASVSCTTCRVHCTWKYTGAYATWCAACTASTAGAAC |
| Minimycin/dudomycin | Dudomycin | 21.05 | Heterologous DNPM | | GTMMKCSWSGNHGBYGHGGVCNTCBMCGWCGACDVSSYC |
| Minimycin/dudomycin | Minimycin | 7.50 | Heterologous DNPM | | GTMMKCSWSGNHGBYGHGGVCNTCBMCGWCGACDVSSYC |
| Isorenieratene/streptamidine | Streptamidine | 100.00 | Heterologous Am2ab, Nutrient source (ISP2, DNPM) | | WATGRAAAYGATTWTCATYWMYAWHHVSGGGRS |
| Candicidin/antimycin | Candicidin/- Antimycin | 30.48 | Nutrient source (DNPM, ISP2), Salinity (2%), pH (9) | | GNCGHCGASYTCGTSGARGYCCTCSVYSMCGNCSDSGASG |

### Novel insights into the regulation of surugamide.

We identified three iModulons that contain core biosynthetic genes for surugamide, a cyclic non-ribosomal peptide synthetase (NRPS) with antibacterial and antifungal properties34,35. These iModulons, named Surugamide, Surugamide Repressor, and Surugamide F, consist of 89, 7, and 2 genes, respectively (Figure 3A, B). Surugamide F appears to be an artifact due to partial BGC deletion in certain conditions (Figure S5). The Surugamide and Surugamide F iModulons comprise 13 and 2 genes, respectively, from the predicted BGC region, including the core biosynthetic genes *surABCD*, transporters, and putative regulators (Figure 3A). Additionally, 22.5% of the Surugamide iModulon genes are related to amino acid transport and metabolism, along with four co-regulated cytochrome P450 genes.

In contrast, the Surugamide Repressor iModulon contains the previously characterized repressor *surR*, which regulates transcription of the *surABCD* operon (Figure 3C)36. In addition, we identified an undescribed *surR*-like gene located approximately 2 Mbp downstream of the surugamide BGC (Figure 3C). Both *surR* and the *surR*-like factor exhibit similar expression patterns across various conditions, and we identified an enriched DNA binding motif within this iModulon, suggesting co-regulation (Figure 3D and 3E). Surprisingly, neither *surR* nor *surR*-like are significantly correlated with any genes from the *surABCD* operon, suggesting that there are likely other factors involved in the regulation of the surugamide BGC (Figure 3F).

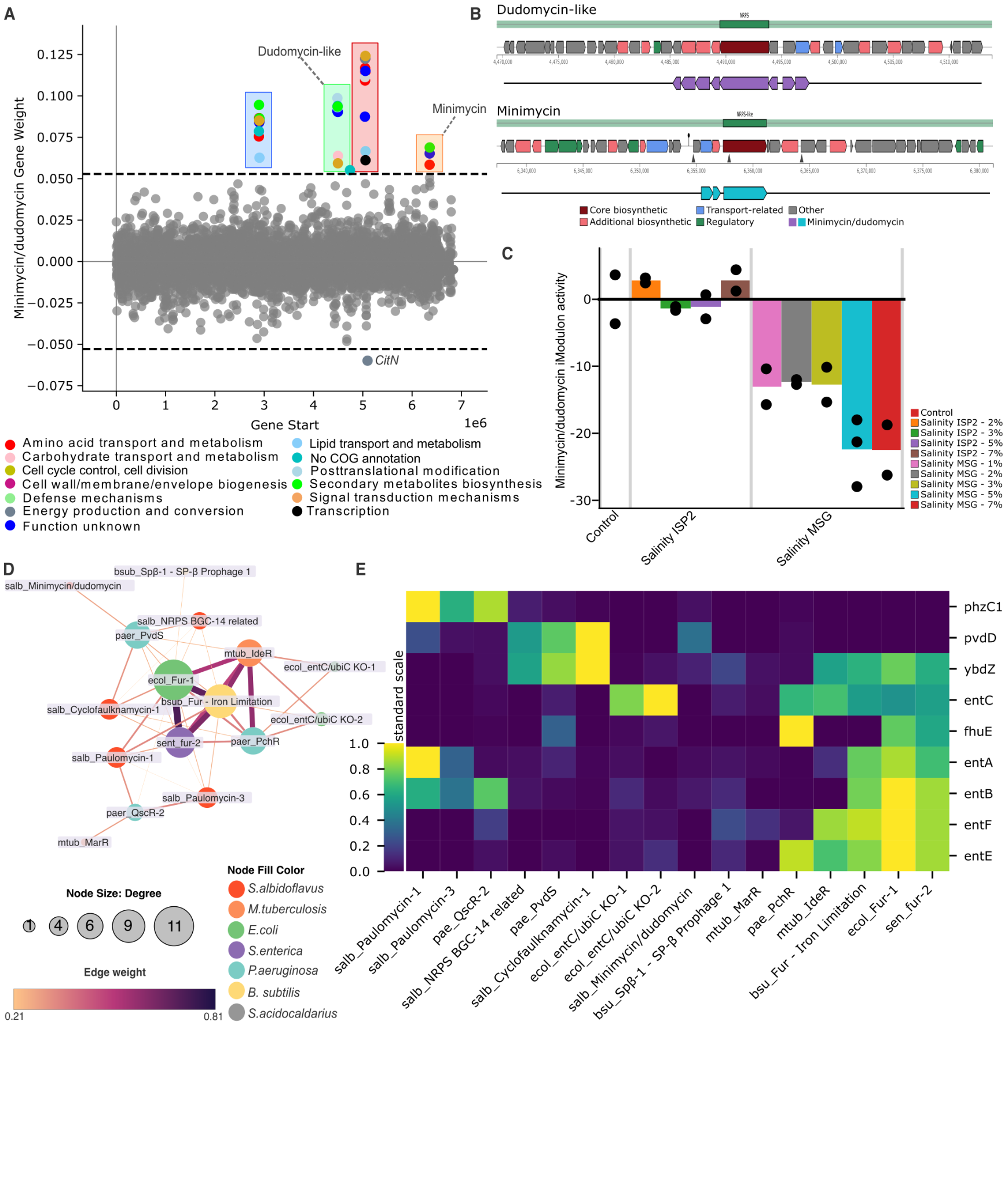


***Figure 3. Surugamide activation and repression.*** *(****A****) Gene locations of the surugamide BGC region as depicted by antiSMASH v7.10.0 and the gene memberships in corresponding iModulons. (****B****) Gene weights of the Surugamide iModulon. The y-axis depicts the individual gene weightings with positive values depicting activation within the iModulon, whereas genes with negative gene weights are repressed in that iModulon. The x-axis shows the location of the gene in the genome. Genes within the surugamide BGC are located within the purplebox. Colour depicts COG category. 22.5% of the genes in this iModulon are related to amino acid transport and metabolism. (****C****) Gene weights for the Surugamide-repressor iModulon, which contains the regulator surR and a surR-like located 2Mbp downstream of the surugamide BGC. (****D****) Sequence logo of a significantly enriched DNA binding motif in the Surugamide-repressor iModulon. (****E****) Scatterplot depicting the correlation in gene expression across all samples for surR and the surR-like, indicating similar expression patterns across most growth conditions. (****F****) Scatterplot illustrating the correlation between the iModulon activity of Surugamide repressor and the gene expression of surA.*

### Identification of co-regulated gene clusters with two NRPS BGCs.

The Minimycin/dudomycin iModulon contains core biosynthetic genes for two NRPS BGCs predicted by antiSMASH, minimycin (also known as oxazinomycin) and dudomycin-like (Figure 4A and 4B). This iModulon is downregulated under salinity stress in minimal media (Figure 4C). In addition to the two BGCs, this iModulon contains two additional gene clusters, one that appears to be involved in amino acid transport, with genes such as *ilvE*, aminotransferase, and tryptophan 2,3-dioxygenases, and one that contains genes such as *ectB*, SidA/IucD/PvdA family monooxygenase, and an AMP-binding enzyme, suggesting these gene clusters may be important for the biosynthesis and transport of necessary components of the biosynthesis of minimycin and the dudomycin-like.

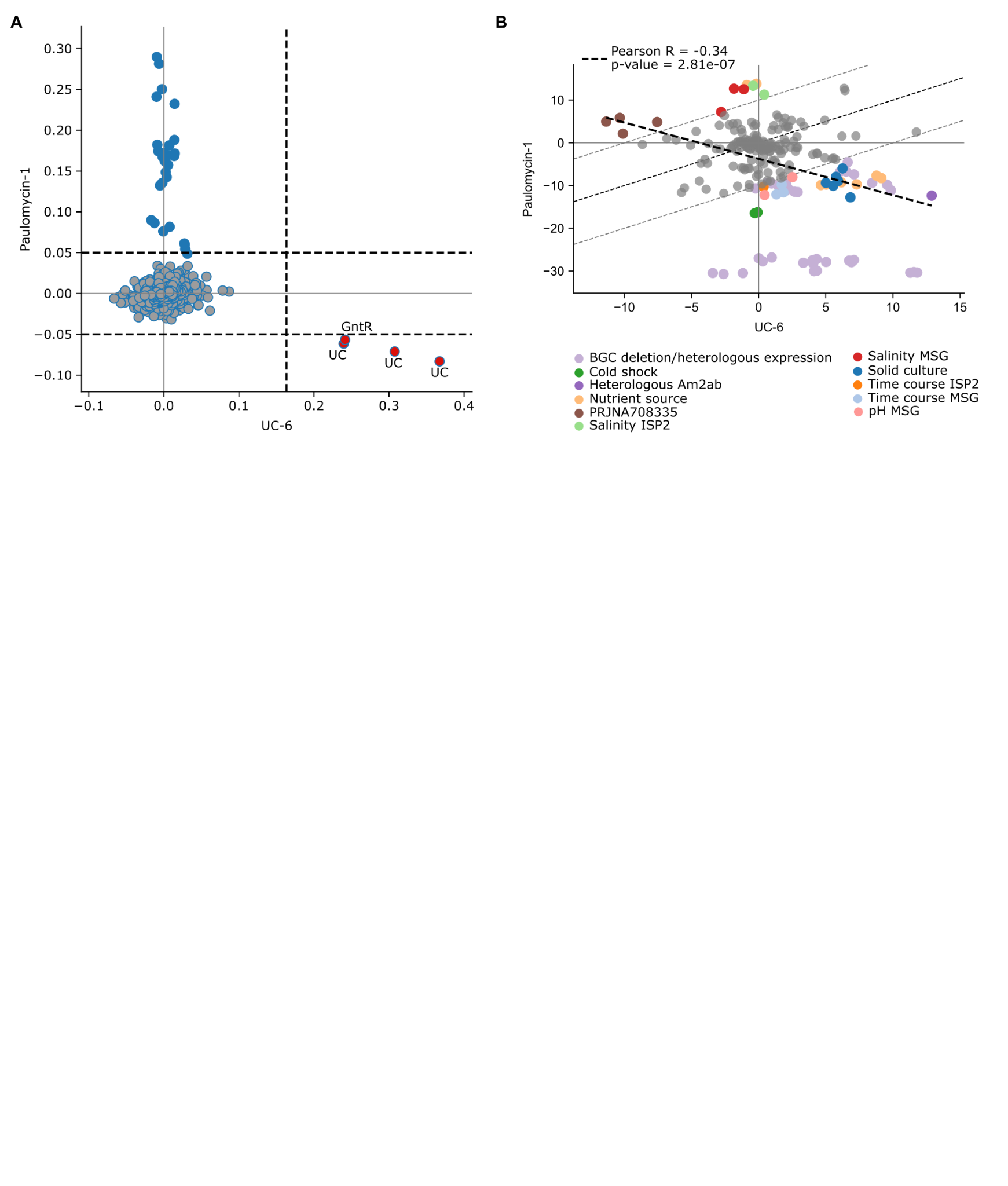
iModulons involving the NRPS BGCs minimycin/dudomycin, cyclofaulknamycin-1 and NRPS BGC-14 and two iModulons related to the transcription of paulomycin, a polyketide synthase (PKS)-like BGC, cluster with iron limitation-related iModulons of other species in the iModulome (Figure 4D). Core genes of this cluster of iModulons are orthologous of *Fur* regulated and iron dependent genes in *E. coli* involved in the enterobactin biosynthesis pathway*,* such as *entA*, *entB*, *entC*, *entE*, *entF*, *ybdZ*, and *fhuE*. Additional core genes include orthologs to *phzC1* and *pvdD,* which encode for the phenazine biosynthesis protein *phzC*, and pyoverdine synthetase D, respectively, in *P. aeruginosa* (Figure 4E).



***Figure 4. NRPS activation in iron limited conditions.*** *(****A****) Gene weights of the Minimycin/dudomycin iModulon, where two distinct gene clusters, in addition to the two BGCs, are highlighted. Genes are plotted against their position in the genome. The color of the dots depicts the COG function according to the legend. (****B****) Genome location and orientation of the minimycin and dudomycn-like BGCs predicted by antiSMASH, and their coverage within the iModulon. (****C****) Activity plot of the Minimycin/dudomycin iModulon across control and salinity stress conditions. (****D****) A cluster from the iModulome with iModulons related to iron acquisition across many organisms including Five iModulons from S. albidoflavus related to NRPS or PKS-like BGCs. (****E****) Clustermap of core genes in the cluster, which are mainly orthologous genes of the enterobactin synthesis pathway in E. coli.*

### Uncharacterized iModulons provide a road map to discover novel functions.

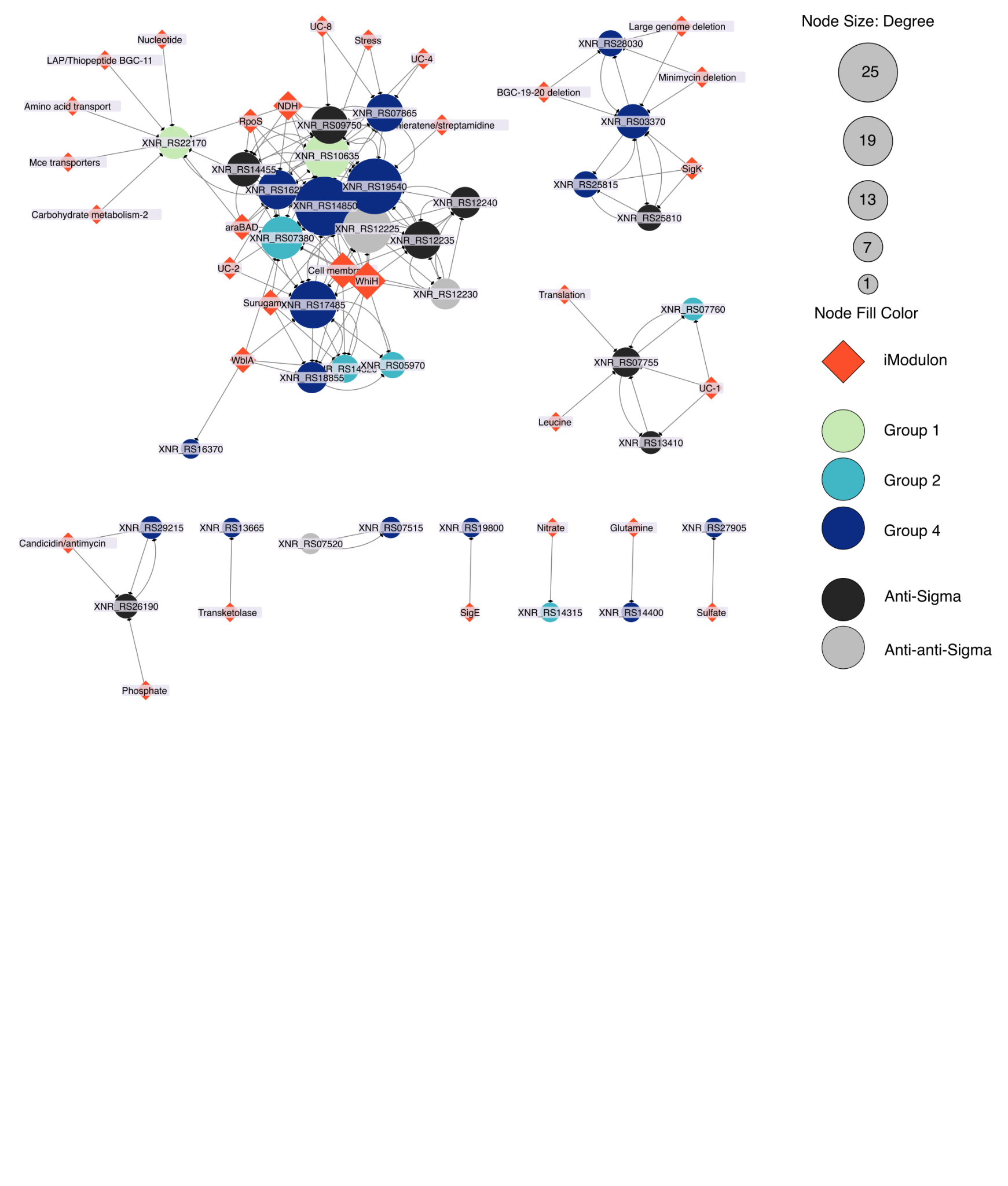
When analyzing an iModulon covering the BGC for paulomycin, we identified a cluster of four genes located 1.8 Mbp downstream of the paulomycin BGC which is down-regulated in the Paulomycin-1 iModulon (Figure 5A). Interestingly, this cluster consists of one putative GntR family transcription factor and three uncharacterized genes, which together make up a separate uncharacterized iModulon called UC-6 (Figure 5A). Upregulation of UC-6 appears to lower the expression of the Paulomycin BGC in a condition-dependent manner. For instance, cultures grown on solid media appear to upregulate UC-6 while down-regulating Paulomycin-1 (Figure 5B). Characterizing the function of UC-6 and fully understanding its effect on the paulomycin BGC require experimental verification of the uncharacterized genes. Around 30% of the genes in the *S. albidoflavus* genome haveno COG annotation, and many more lack experimental validation24. 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 (Table S1.4).



***Figure 5. Activities of uncharacterized iModulons.*** *(****A****) Scatterplot of gene weights between the UC-6 and Paulomycin-1 iModulons. Dashed lines indicate the upper and lower thresholds of the respective iModulon. Red genes are members of both iModulons, with high activity in UC-6 and low activity in Paulomycin-1. (****B****) Activity of UC-6 and Paulomycin-1 iModulons across all samples shows conditions which activate the iModulons. Paulomycin-1 is for example down-regulated in solid cultures while UC-6 is upregulated.*

### Uncovering the Sigma factor regulatory network of *S. albidoflavus.*

Sigma factors are important regulators of gene expression; however, their regulatory roles in *S. albidoflavus* are not well understood, with many of their full regulons not yet being defined37. We identified 31 sigma, 11 anti-sigma, and three anti-anti-sigma factors annotated in the *S. albidoflavus* genome, with 17 being members of least one iModulon, providing potential regulatory targets (Table S3.1). To expand on this, we performed a network analysis by comparing sigma factor activities with iModulon activities across our RNA-Seq compendium28,38. This similarity-based network revealed 34 sigma-related factors with high expression similarities between themselves or iModulons (Figure 6 and Table S3.2). Notably, the expression levels of Group 1 primary Sigma factors *hrdD* (XNR\_RS10635) and *hrdA* (XNR\_RS22170) exhibited significant correlations with multiple iModulons, including Cell membrane, RpoS, Amino acid transport, and Carbohydrate metabolism-2 (Figure 6). Furthermore, some Group 4 sigma factors such as XNR\_RS17485, XNR\_RS14850 (*sigT*), and XNR\_RS19540, are significantly correlated with several large iModulons, including Surugamide, WblA, WhiH, and RpoS, suggesting potential global regulatory roles for these factors, meriting further investigation.



***Figure 6. Sigma and related factors activity with iModulons.*** *Gene/iModulon Network analysis of Sigma, anti-sigma, anti-anti-sigma factors (circles), and iModulon (diamonds) activities across the transcriptional compendium. Node size depicts the number of connections in the network. Nodes in the clusters are significantly correlated.*

## Discussion

By applying a machine learning algorithm, ICA, to a large compendium of both in-house generated and publicly available RNA-Seq datasets, we have identified 78 iModulons which provide a systems-level view of the transcriptome and a window into the complex TRN of *S. albidoflavus*3,24,39,40. Like previous studies, this modularization of the TRN resulted in new discoveries, including co-regulation of peripheral gene clusters with BGCs, putative regulators, numerous uncharacterized genes which are promising targets for future research, and provided a detailed compendium of growth conditions and their effect on the TRN28–32,38,41–43. iModulons are becoming valuable tools in strain design and bioproduction, suggesting our analysis will further improve the engineering efforts of *S. albidoflavus*33. Importantly, the iModulon knowledge base can continuously be updated as new RNA-Seq data becomes available, ensuring more detailed reconstructions of the TRN over time29,30,44. The iModulons are freely available at [imodulondb.org](http://imodulondb.org).

**The iModulome places the iModulons in an evolutionary context.**

While many of the *S. albidoflavus* iModulons appear to have significant overlap with known regulons, there remains a comparatively low number of studies detailing regulators in *S. albidoflavus*, hindering characterization and annotation efforts10–12,40. To assist with characterization and to validate the iModulons, we introduced the ‘iModulome’ — a comprehensive network of iModulons from seven distinct species, which clusters iModulons together based on orthologous gene weights — allowing a cross-species comparison of gene contents across iModulons. This approach not only facilitated the annotation of iModulons in *S. albidoflavus* but also placed the iModulons in an evolutionary context, opening a potentially new “pan-modulome” area of research akin to pan-genomics and pan-transcriptomics, which may reveal conserved and lineage-specific aspects of the TRN45,46. While this level of analysis was beyond the scope of this study, the similarity between the iModulons of *S. albidoflavus* and those of other species suggests that our results will be broadly applicable to other *Streptomyces* and *Actinomycetes* more generally. Moreover, these methodologies hold promise for application to other non-model organisms to increase our understanding of microbial gene regulation42,47.

**iModulons give insights into the regulation of secondary metabolites.**

The iModulons captured over half of the predicted endogenous BGCs in the *S. albidoflavus* genome, including polyketide synthases, NRPSs, lanthipeptides, and ribosomally synthesized and post-translationally modified peptide products (RiPPs), detailing growth conditions that transcriptionally up- or down-regulate these clusters. In addition to improving BGC border definitions, the iModulons revealed several co-regulated genes and gene clusters outside the predicted BGC region. Our findings include a *surR*-like transcription factor that is co-expressed with a major regulator of surugamide biosynthesis, *surR,* across most growth conditions35. In addition, the core biosynthetic genes of surugamide (*surABCD*) are members of an iModulon that include several genes with amino acid transport functions, cytochrome P450 enzymes, and an acyl-CoA dehydrogenase, which is crucial for avenolide activity, a key secondary metabolite signaling molecule48,49. Taken together, this iModulon may contain several genes necessary for surugamide biosynthesis, providing a roadmap for experimental validation. Moreover, we identified two off-site gene clusters related to the transcription of the minimycin and dudomycin-like BGCs in *S. albidoflavus*, one which appears involved in amino acid transport and one which contains genes such as *ectB*, SidA/IucD/PvdA family monooxygenase, and an AMP-binding enzyme, suggesting these gene clusters may be important for biosynthesis and transport of necessary components50,51.

Moreover, through the iModulome we identified several BGC-related iModulons that clustered together with *Fur* regulated and iron scavenging-related iModulons of other species52–54. The *S. albidoflavus* BGCs included cyclofaulknamycin, minimycin and dudomycin-like, an unknown NRPS-like, and the PKS-like paulomycin. While the biosynthesis of minimycin appears to involve a non-heme iron dependent monooxygenase, there is very little known about the involvement of iron in the expression of these BGCs, or potential iron scavenging properties55–59. This underscores the potential of iModulons as a framework for unraveling the intricate regulatory mechanisms that govern BGC expression, particularly in relation to metal ion homeostasis. The association of these iModulons with iron-dependent pathways not only hints at a broader regulatory landscape but also paves the way for targeted investigations into the iron-mediated control of secondary metabolism.

Our findings of co-regulated peripheral gene sets with BGCs, along with potential improvements in BGC border definitions, may support future strain design efforts for *S. albidoflavus* and facilitate BGC engraftments into other organisms33. Additionally, many BGC related iModulons were enriched in DNA binding motifs within 500 bp upstream of the BGC, presenting promising targets for identifying new promoters to increase production yields60. These findings and the information available on iModulonDB may provide a valuable foundation for experimental efforts seeking to understand BGC regulation.

**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* genome consists of 30% genes with uncharacterized or unknown functions, and many more that lack experimental validation24. iModulons have been used to characterize several uncharacterized genes in *E. coli* and have identified many lucrative targets for other species33,61. Our analyses enable us to infer potential functions for 40% of all genes that lack a COG annotation, providing promising experimental targets going forward. For instance, several of the iModulons related to BGCs contain co-regulated genes with unknown functions, many of which are located outside of the main BGC and therefore may be necessary components for biosynthesis. The iModulons generated in this study provides a roadmap to experimentally validate these uncharacterized genes.

**Sigma factors may regulate several iModulons.**

By analyzing the correlation of activity between iModulons and Sigma factors we generated a network of sigma factors that are potentially orchestrating the regulation of specific iModulons within the *S. albidoflavus* genome. Out of 31 sigma, 11 anti-sigma, and three anti-anti-sigma factors, 17 are integral components of at least one iModulon, and 23 appear significantly correlated with iModulon activities, highlighting the intricate regulatory mechanisms governing gene expression in this species. In addition to the two primary Sigma factors *hrdA* and *hrdD* which appear highly connected and correlated with many iModulons, several extracytoplasmic function sigma factors stood out, including a paralog to *sigT* in *S. coelicolor*, which regulates development and actinorhodin production in response to nitrogen stress62,63, and XNR\_RS17485 and XNR\_RS19540 which are significantly correlated with iModulons related to BGCs, development, and uncharacterized functions. Furthermore, the discovery of enriched DNA binding motifs upstream of each operon within the iModulons provides a valuable blueprint for future research. Studies employing methodologies such as ChIP-, RIViT-, or DAP-seq could leverage our findings to elucidate the direct interactions between these sigma factors and their target genes64–66. Such investigations would not only validate the regulatory roles suggested by our network analysis but also enhance our understanding of the dynamic control of BGCs, potentially unlocking new avenues for the optimization of secondary metabolite production in *Streptomyces*.

**Limitations of the study**

Successful modularization of the transcriptome through ICA requires activation of gene sets through specific growth conditions. Consequently, ICA cannot modularize genes that are lowly expressed or equally expressed across all conditions29. In this study, 48.4% of *S. albidoflavus* genes were assigned to at least one iModulon, indicating many more growth conditions remain to be tested, including for example oxidative stress, heat stress, and additional gene knockouts. Moreover, the lack of detailed information about regulons limits our ability to fully characterize *S. albidoflavus* iModulons. While our iModulome analyses aided in categorizing these iModulons, further experimental studies on regulons are essential to improve our understanding about the TRN. Despite these limitations, ICA remains a powerful tool for uncovering gene regulatory networks. Importantly, our capture of several BGCs within iModulons is promising for secondary metabolism research. However, while we demonstrate transcriptional activation of these BGCs, complementary metabolic studies are needed to confirm the production and secretion of the corresponding metabolites.

**Conclusion**

In this study we compiled a large compendium of in-house generated and publicly available transcriptomes to identify iModulons in *S. albidoflavus*, providing a comprehensive and quantitative understanding of its TRN. We showed how the iModulons can be used to (i) compare the TRN of *S. albidoflavus* with other organisms on imodulondb, revealing both conserved and lineage-specific features, (ii) elucidate the regulation of BGCs, such as surugamide, minimycin, and paulomycin by identifying putative regulators, co-regulated genes, 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 and process design to optimize the production of valuable natural products by *S. albidoflavus* and other *Streptomyces* species.

## Resource Availability

### Lead Contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Emre Özdemir ([emoz@biosustain.dtu.dk](mailto:emoz@biosustain.dtu.dk)) or Lei Yang ([leiya@biosustain.dtu.dk](mailto:leiya@biosustain.dtu.dk)).

### Materials Availability

This study did not generate new unique reagents.

### Data and Code Availability

* RNA-seq raw files data have been deposited at NCBI as Bioproject: PRJNA1062162 and are publicly available as of the date of publication.
* All original code has been deposited at GitHub (<https://github.com/biosustain/salb_imodulons>) and Zenodo at (ZENODO DOI LINK) and is publicly available as of the date of publication.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## Acknowledgements

We are grateful to Tilmann Weber, Kai Blin, Thomas Booth and Simon Shaw for the informative discussions and input when characterizing the iModulons, and to Edward Catoiu for integrating the dataset to imodulondb.

## Author Contributions

MJ performed computational analysis, generated presented results, and wrote the manuscript. RS, TG, MSP, NM, AS, PC, PG developed protocols and performed experiments. BP, LY, EO conceptualized the study. All authors contributed to interpretation of results, reviewed and/or edited the manuscript.

## Declaration of Interests

The authors declare no competing interests.

## Main figure titles and legends

***Figure 1. Overview of the S. albidoflavus iModulons.*** *(****A****) The availability of public RNA-Seq (SRA) data for S. albidoflavus over time. (****B****) Histogram of the correlations between log-tpm profiles of all samples across all conditions (blue) and known replicates (green). The median Pearson R score between replicates is 0.99. Replicates that did not exceed 0.90 in R score were removed from further analysis. (****C****) Principal components analysis of 218 samples used for ICA. The color and symbol depict the specific project where the sample was generated. All in-house generated samples for this study are colored, and the two public datasets included in this study are labeled in grey. (****D****) A treemap depiction of the 78 identified iModulons and their associated functional categories. The dimensions of the individual boxes correlate with the number of genes found in each respective iModulon. (****E****) Highlight of the top 10 iModulons contributing the greatest proportion of explained variance. These iModulons represent key functions such as development, translation, secondary metabolite biosynthesis, and stress.*

***Figure 2. S. albidoflavus iModulons similarities and differences with other organisms.*** *(****A****) Overview of the construction of the iModulome. M files generated from ICA analysis of seven organisms were collected. Orthogroups were identified using orthofinder (48) and used to merge the M files based on the specific orthogroup. Cosine similarity score and MCL clustering was used to create a clustered network of iModulons based on similarity. (****B****) A cluster from the iModulome of iModulons across seven species (abbreviations and colour: sal, Red, S. albidoflavus; mtub, Orange, M. tuberculosis; ecol, Green, E. coli; sen, Purple, S. enterica; pae, Blue, P. aeruginosa; bsub, Yellow, B. subtilis; sac, Grey, S. acidocaldarius) based on cosine similarity of ICA derived gene weights, containing iModulons related to translation processes. (****C****) Clustermap of core genes in the cluster, the scale bar depicts the standardized weights of orthologous genes, expressed as z-scores. In this example, 22 ribosomal subunits are highly weighted in iModulons related to translation across several different species, suggesting very similar iModulon structures. Notably, the Archaea S. acidocaldarius appears to lack many of the genes that are highly expressed in other organisms, indicating a greater difference in iModulon structure. (****D****) A cluster of phosphate related iModulons based on cosine similarity of ICA derived gene weights, containing iModulons related to phosphate uptake. Coloured nodes depict species as described in (B). (****E****) Clustermap of core orthologous genes in the Phosphate related cluster. The pstSCAB operon has high weights in nearly all iModulons, while for some genes (eg. yddL, phoA) iModulon membership can be species subjective.*

***Figure 3. Surugamide activation and repression.*** *(****A****) Gene locations of the surugamide BGC region as depicted by antiSMASH v7.10.0 and the gene memberships in corresponding iModulons. (****B****) Gene weights of the Surugamide iModulon. The y-axis depicts the individual gene weightings with positive values depicting activation within the iModulon, whereas genes with negative gene weights are repressed in that iModulon. The x-axis shows the location of the gene in the genome. Genes within the surugamide BGC are located within the purplebox. Colour depicts COG category. 22.5% of the genes in this iModulon are related to amino acid transport and metabolism. (****C****) Gene weights for the Surugamide-repressor iModulon, which contains the regulator surR and a surR-like located 2Mbp downstream of the surugamide BGC. (****D****) Sequence logo of a significantly enriched DNA binding motif in the Surugamide-repressor iModulon. (****E****) Scatterplot depicting the correlation in gene expression across all samples for surR and the surR-like, indicating similar expression patterns across most growth conditions. (****F****) Scatterplot illustrating the correlation between the iModulon activity of Surugamide repressor and the gene expression of surA.*

***Figure 4. NRPS activation in iron limited conditions.*** *(****A****) Gene weights of the Minimycin/dudomycin iModulon, where two distinct gene clusters, in addition to the two BGCs, are highlighted. Genes are plotted against their position in the genome. The color of the dots depicts the COG function according to the legend. (****B****) Genome location and orientation of the minimycin and dudomycn-like BGCs predicted by antiSMASH, and their coverage within the iModulon. (****C****) Activity plot of the Minimycin/dudomycin iModulon across control and salinity stress conditions. (****D****) A cluster from the iModulome with iModulons related to iron acquisition across many organisms including Five iModulons from S. albidoflavus related to NRPS or PKS-like BGCs. (****E****) Clustermap of core genes in the cluster, which are mainly orthologous genes of the enterobactin synthesis pathway in E. coli.*

***Figure 5. Activities of uncharacterized iModulons.*** *(****A****) Scatterplot of gene weights between the UC-6 and Paulomycin-1 iModulons. Dashed lines indicate the upper and lower thresholds of the respective iModulon. Red genes are members of both iModulons, with high activity in UC-6 and low activity in Paulomycin-1. (****B****) Activity of UC-6 and Paulomycin-1 iModulons across all samples shows conditions which activate the iModulons. Paulomycin-1 is for example down-regulated in solid cultures while UC-6 is upregulated.*

***Figure 6. Sigma and related factors activity with iModulons.*** *Gene/iModulon Network analysis of Sigma, anti-sigma, anti-anti-sigma factors (circles), and iModulon (diamonds) activities across the transcriptional compendium. Node size depicts the number of connections in the network. Nodes in the clusters are significantly correlated.*

## Main tables and legends

***Table 1. Biosynthetic gene cluster representation in iModulons.*** *List of antiSMASH predicted BGC regions and corresponding iModulons. The coverage represents the percentage of genes from the antiSMASH predicted BGC region that is present in the iModulon. The highest activity column depicts the growth conditions of the samples with the highest iModulon activation scores. The motif column shows the IUPAC ambiguity codes of the top enriched motif for the operons and genes in the iModulon.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **iModulon** | **antiSMASH BGC** | **Coverage (%)** | | **Highest activity** | **Motif** |
| BGC-1 related | Region 1 | 25.00 | Solid culture | | CCYBGDSCAGVMSCTTGMCCKCCWGGDCGVCGTCSTCMHC |
| SGR PTMs | SGR PTMs | 48.28 | Nutrient source (DNPM, ISP2) | | CSGGYSSGAGMCSRSCHWRTWCRSWTCRCKCMCRWKRA |
| Paulomycin-1 | Paulomycin | 65.71 | Nutrient source (Galactose), Salinity (5% in MSG) | | CTWGGMAADGMADBSCATRDSDAHCAGRAASHSCGAVDK |
| Paulomycin-2 | Paulomycin | 77.14 | Nutrient source (Galactose), Salinity (5% in MSG) | | CSBVMHTTACVKWKGSATWNCWGGCGGSTSYCMWGRAGG |
| Paulomycin-3 | Paulomycin | 42.86 | BGC deletion/heterologous expression, Nutrient source (Starch) | | CSAGVTCGCCGWCVWSGCCGVCGVCSTSGWCGNCCKSVTS |
| Cyclofaulknamycin-1 | Cyclofaulknamycin | 48.65 | BGC deletion/heterologous expression | | CGBCCTSSWSGAVGASCWGWYCSASCTCGBCG |
| Cyclofaulknamycin-2 | Cyclofaulknamycin | 48.65 | Nutrient source (Starch), Heterologous Am2ab, BGC deletion/heterologous expression | | - |
| LAP/Thiopeptide BGC-11 | Region 11 | 51.85 | Nutrient source (Starch, DNPM) | | AGCCGGRGTSGDMGVSCWTCGRGAKCWTGNNGTASYC |
| SAL-2242 | SAL-2242 | 23.53 | Nutrient source (ISP2) | | AGGYCSAGAAKCSSCAKSWCGTMCTCSTYCGTGTCGA |
| NRPS BGC-14 related | Region 14 | 13.79 | PRJNA1003853 | | CGMGGBSACCGSCGCCSTCGYCGTCKMCG |
| Surugamide | Surugamide | 31.82 | Nutrient source (Histidine, Glycerol) | | GGYGVYSRSSRCGRBCGCSGCSRCGRCCGYSR |
| Surugamide F | Surugamide | 4.55 | BGC deletion/heterologous expression | | - |
| Surugamide repressor | Surugamide | 11.36 | Early growth stages | | CASVSCTTCRVHCTWKYTGAYATWCAACTASTAGAAC |
| Minimycin/dudomycin | Dudomycin | 21.05 | Heterologous DNPM | | GTMMKCSWSGNHGBYGHGGVCNTCBMCGWCGACDVSSYC |
| Minimycin/dudomycin | Minimycin | 7.50 | Heterologous DNPM | | GTMMKCSWSGNHGBYGHGGVCNTCBMCGWCGACDVSSYC |
| Isorenieratene/streptamidine | Streptamidine | 100.00 | Heterologous Am2ab, Nutrient source (ISP2, DNPM) | | WATGRAAAYGATTWTCATYWMYAWHHVSGGGRS |
| Candicidin/antimycin | Candicidin/- Antimycin | 30.48 | Nutrient source (DNPM, ISP2), Salinity (2%), pH (9) | | GNCGHCGASYTCGTSGARGYCCTCSVYSMCGNCSDSGASG |

### STAR★Methods

### Key resources table

|  |  |  |
| --- | --- | --- |
| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
| Antibodies | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Bacterial and virus strains | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Biological samples |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Chemicals, peptides, and recombinant proteins | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Critical commercial assays | | |
| RNeasy Mini QIAcube kit | Qiagen | cat. No./ ID: 74116 |
| Phenotype MicroArrayTM analysis |  |  |
| NEBNext rRNA Depletion kit (Bacteria) | New England Biolabs | cat. No. E7850X |
|  |  |  |
|  |  |  |
| Deposited data | | |
| RNA-Seq *Streptomyces albidoflavus* | This study. | SRA: PRJNA1062162 |
| RNA-Seq on *Streptomyces albus* engineered to produce nybomycin. | Stegmüller et al. 67 | SRA: PRJNA1003853 |
| RNA-Seq on *Streptomyces albus* J1074/R2 engineered to produce pamamycin derivatives. | Gläser et al. 68 | SRA: PRJNA708335 |
| Original code for this study. | This study. | Doi to zenodo link |
| Experimental models: Cell lines | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Experimental models: Organisms/strains | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Oligonucleotides | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Recombinant DNA | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| Software and algorithms | | |
| DuctApe 0.18.2 | Galardini et al. 69 | [DuctApe · PyPI](https://pypi.org/project/DuctApe/) |
| iModulonMiner | Sastry et al. 70 | [GitHub - SBRG/iModulonMiner: Workflow to download, process, and explore microbial RNA-seq data from NCBI SRA](https://github.com/SBRG/iModulonMiner?tab=readme-ov-file) |
| Orthofinder 2.5.5 | Emms and Kelly71 | [GitHub - davidemms/OrthoFinder: Phylogenetic orthology inference for comparative genomics](https://github.com/davidemms/OrthoFinder) |
| Cytoscape 3.10.2 | Smoot et al. 72 | [Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization](https://cytoscape.org/) |
| clusterMaker | Morris et al. 73 | [Cytoscape App Store - clusterMaker](https://apps.cytoscape.org/apps/clustermaker) |
| PyModulon 0.2.1 | Sastry et al. 74 | [PyModulon: Analyzing iModulons in Python — pymodulon 0.2.1 documentation](https://pymodulon.readthedocs.io/en/latest/) |
| antiSMASH 7.10.0 | Blin et al. 75 | [antiSMASH bacterial version](https://antismash.secondarymetabolites.org/" \l "!/start) |
| Other | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Method details

### RNA extraction and library preparation

We assessed *S. albidoflavus* activity across various environmental and nutritional factors using Biolog microplates. Based on the metabolic activities observed, we identified the most promising conditions that supported robust growth and activated diverse metabolic pathways (Table S1.1). Next, we imitated these favorable conditions during our *S. albidoflavus* culturing process for subsequent RNA extraction. We grew *S. albidoflavus* and extracted RNA from 211 samples, each with at least one biological replicate, under 83 unique growth conditions. These conditions included supplemented minimal media with different carbon sources, salinity, pH, and temperature stresses, genome-reduced strain (up to 10 endogenous BGCs deleted), and heterologous expression of actinorhodin BGC.

Genome-reduced strains were obtained by deleting 10 BGCs from *S. albidoflavus* parental strain using CRISPR-Cas976 or CASCADE-Cas3 systems77. Additionally, actinorhodin BGC was integrated via conjugation into the chromosome of parental and genome reduced hosts.

For cultivation of *S. albidoflavus* strains, spore suspension was obtained from 5-7 days of sporulated MS plates, as described in Tong et al. 76. Precultures were prepared by inoculating 300 µL from cryostocks or spore suspension in 30 mL minimal media. 2 mL preculture was aliquoted, washed twice with sterile MilliQ water and resuspended in 600 µL sterile MilliQ water. 100 µL of washed and resuspended cells were used to inoculate 25 mL of specified media for RNA extraction. The strains were cultivated aerobically in different media: 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 (NH4)2SO4, 1.86 g/L citric acid, 0.21 g/L FeCl3, 3.33 mL/L 2M MgSO4, 250 g/L of monosodium glutamate, 0.1 M phosphate buffer pH 7.0 and 1.6 mL/L trace metal solution78 (20.4 g/L H2SO4 96%, 50 g/L citrate·H2O, 16.75 g/L ZnSO4·7H2O, 2.5 g/L CuSO4·5H2O, 1.5 g/L MnCl2·4H2O, 2 g/L H3BO3, and 2 g/L NaMoO4·2H2O). The recipe for DNPM typically contains soytone but was replaced with bacteriological peptone from meat in our formulation15. In conditions where cultures were grown on solid media, plates containing Mannitol Soya Flour (MS) medium (20g/L Mannitol, 20g/L Soya Flour, 20g/L Agar, and tap water, pH 7) were used. Solid culture samples were recovered using 10 µL inoculation loops. 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. In temperature heat stress conditions, cultures were submerged in a ScanVac HeatSafe 18 water bath set to 42Cº. 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.

### Biolog Phenotype Microarrays

*S. albidoflavus* activity across various nutritional sources was evaluated by Phenotype MicroArrayTM analysis. Uniform spore suspension (80% T, transmittance) was obtained from 5-7 days sporulated MS plates and used to prepare inoculating fluids according to Biolog protocol for *Streptomyces.* IF-0a base inoculating fluid and Redox Dye Mix D were used for this experiment. PM 1-2 (carbon sources), 3 (nitrogen sources) and 4 (phosphorus and sulfur sources) Biolog microplates were selected and inoculated with 100 µL of prepared inoculation fluid. PM microplates were incubated in OmniLog® PM System, at 30°C, 72h. Data analysis was performed with DuctApe (0.18.2)69 using dphenome module as described on manual pages. In brief, phenotype microarray data for *S. albidoflavus* was loaded, background signals based on control wells were subtracted, signals were trimmed, data was analyzed based on default 10 clusters and inconsistent replicates were purged.

### Data acquisition and processing

Data processing and quality control is detailed in Sastry et al. 70 and available at <https://github.com/SBRG/iModulonMiner>. 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 Bowtie79. The read direction was inferred using RSeQC80 before generating read counts using featureCounts81. Lastly, all quality control metrics were compiled using MultiQC82.

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 sample outliers. Only samples with at least two replicates and a Pearson R correlation of > 0.90 between replicates were analyzed. Furthermore, to minimize possible batch effects, each individual experiment was normalized to a reference condition prior to calculating the iModulons (Table S1.2)70. After quality control and normalization, the final data compendium comprised 218 high-quality expression profiles: 161 generated from this study and 57 collected from NCBI-SRA. The two public datasets that met our quality control criteria included PRJNA1003853, which tested the heterologous expression of Nybomycin in a *S. albidoflavus* strain with 14 endogenous BGCs removed67, and PRJNA708335, which heterologously expressed pamamycin in the wild type *S. albidoflavus*68.

### Performing Independent component analysis (ICA)

Independent component analysis was implemented as per McConn et al. 83 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 84,85. 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 ICs with opposite signs, the following distance metric was used for computing the distance matrix:

where 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 of ICA can affect the modularization, we applied the above procedure to the dataset multiple times, ranging the number of dimensions from 10 to 190 (i.e. the approximate size of the dataset) with a step size of 10. 190 dimensions generated the most robust iModulons and were used for this analysis.

### Cross-species comparison of iModulons

To enable comparisons of iModulon structures between organisms we first identified orthologous genes among seven species represented on imodulondb.org, (*Escherichia coli, Salmonella enterica, Pseudomonas aeruginosa, Bacillus subtilis, Mycobacterium tuberculosis, S. albidoflavus,* and *Sulfolobus acidocaldarius*) using Orthofinder with default settings71. These species were chosen to represent a diverse range of phyla, spanning both closely related and distantly related species. All iModulon data files and objects were downloaded from imodulondb.org28. The gene weight files of all species describing the iModulon memberships were merged based on orthogroups 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 pairwisme manner for all iModulons and visualized the similarities using Cytoscape72. Edge weights were filtered (0.648) to only show the top 1500 edge weights, and we used the MCL cluster algorithm from the clusterMaker app to identify clusters of iModulons73.

### Characterization of iModulons

To facilitate iModulon characterization, we used the PyModulon Python package74. *k*-means clustering was used to identify component-specific thresholds, as per PyModulon recommendation, with manual editing of 16 iModulon thresholds to improve characterization efforts. Specifically, in large iModulons with lowly expressed genes, increasing the threshold improved interpretability of the gene contents. Similarly, we found some instances like Paulomycin-1, where lowering the threshold enabled us to identify potentially co-regulated genes. The iModulons were characterized by information from public databases, literature and based on similarities to other iModulons from other species in our iModulome analysis. Gene annotations and operon information was downloaded from [Biocyc.org](http://biocyc.org)86. BGC regions were annotated using antiSMASH with default settings75. Code for our analysis pipeline, manual thresholding, and characterization of iModulons is maintained on github (<https://github.com/biosustain/salb_imodulons>).

## Supplemental information index

**Document S1. Figures S1–S5.**

**Table S1. Excel file containing additional data too large to fit in a PDF, related to Figure 1.**

**Table S2. Excel file containing additional data too large to fit in a PDF, related to Figure 2.**

**Table S3. Excel file containing additional data too large to fit in a PDF, related to Figure 6.**

## References

1. Hopwood, D.A. (2007). Streptomyces in Nature and Medicine: The Antibiotic Makers (Oxford University Press, USA).

2. Katz, L., and Baltz, R.H. (2016). Natural product discovery: past, present, and future. J. Ind. Microbiol. Biotechnol. *43*, 155–176. https://doi.org/10.1007/s10295-015-1723-5.

3. Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature *417*, 141–147. https://doi.org/10.1038/417141a.

4. Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M., and Horinouchi, S. (2008). Genome Sequence of the Streptomycin-Producing Microorganism *Streptomyces griseus* IFO 13350. J. Bacteriol. *190*, 4050–4060. https://doi.org/10.1128/JB.00204-08.

5. Mohite, O.S., Jørgensen, T.S., Booth, T., Charusanti, P., Phaneuf, P.V., Weber, T., and Palsson, B.O. (2024). Pangenome mining of the *Streptomyces* genus redefines their biosynthetic potential. Preprint, https://doi.org/10.1101/2024.02.20.581055 https://doi.org/10.1101/2024.02.20.581055.

6. Scherlach, K., and Hertweck, C. (2009). Triggering cryptic natural product biosynthesis in microorganisms. Org. Biomol. Chem. *7*, 1753. https://doi.org/10.1039/b821578b.

7. Hoskisson, P.A., and Seipke, R.F. (2020). Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. mBio *11*, e02642-20. https://doi.org/10.1128/mBio.02642-20.

8. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Ōmura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat. Biotechnol. *21*, 526–531. https://doi.org/10.1038/nbt820.

9. Ward, A., and Allenby, N. (2018). Genome mining for the search and discovery of bioactive compounds: The Streptomyces paradigm. FEMS Microbiol. Lett. https://doi.org/10.1093/femsle/fny240.

10. Romero-Rodríguez, A., Maldonado-Carmona, N., Ruiz-Villafán, B., Koirala, N., Rocha, D., and Sánchez, S. (2018). Interplay between carbon, nitrogen and phosphate utilization in the control of secondary metabolite production in Streptomyces. Antonie Van Leeuwenhoek *111*, 761–781. https://doi.org/10.1007/s10482-018-1073-1.

11. Van Der Heul, H.U., Bilyk, B.L., McDowall, K.J., Seipke, R.F., and Van Wezel, G.P. (2018). Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. Nat. Prod. Rep. *35*, 575–604. https://doi.org/10.1039/C8NP00012C.

12. Wei, J., He, L., and Niu, G. (2018). Regulation of antibiotic biosynthesis in actinomycetes: Perspectives and challenges. Synth. Syst. Biotechnol. *3*, 229–235. https://doi.org/10.1016/j.synbio.2018.10.005.

13. Palazzotto, E., Tong, Y., Lee, S.Y., and Weber, T. (2019). Synthetic biology and metabolic engineering of actinomycetes for natural product discovery. Biotechnol. Adv. *37*, 107366. https://doi.org/10.1016/j.biotechadv.2019.03.005.

14. Baltz, R.H. (2010). Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. J. Ind. Microbiol. Biotechnol. *37*, 759–772. https://doi.org/10.1007/s10295-010-0730-9.

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

16. Chen, Y., Wendt-Pienkowski, E., and 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. J. Bacteriol. *190*, 5587–5596. https://doi.org/10.1128/JB.00592-08.

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

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

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

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

21. Dolya, B., Hryhorieva, O., Sorochynska, K., Lopatniuk, M., Ostash, I., Tseduliak, V.-M., Sterndorff, E.B., Jørgensen, T.S., Gren, T., Dacyuk, Y., et al. (2023). Properties of Multidrug-Resistant Mutants Derived from Heterologous Expression Chassis Strain Streptomyces albidoflavus J1074. Microorganisms *11*, 1176. https://doi.org/10.3390/microorganisms11051176.

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

23. Li, X., Guo, R., Luan, J., Fu, J., Zhang, Y., and 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. Microb. Cell Factories *22*, 15. https://doi.org/10.1186/s12934-023-02023-3.

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

25. Kong, W., Vanderburg, C.R., Gunshin, H., Rogers, J.T., and Huang, X. (2008). A Review of Independent Component Analysis Application to Microarray Gene Expression Data. BioTechniques *45*, 501–520. https://doi.org/10.2144/000112950.

26. Saelens, W., Cannoodt, R., and Saeys, Y. (2018). A comprehensive evaluation of module detection methods for gene expression data. Nat. Commun. *9*, 1090. https://doi.org/10.1038/s41467-018-03424-4.

27. Fang, X., Sastry, A., Mih, N., Kim, D., Tan, J., Yurkovich, J.T., Lloyd, C.J., Gao, Y., Yang, L., and Palsson, B.O. (2017). Global transcriptional regulatory network for *Escherichia coli* robustly connects gene expression to transcription factor activities. Proc. Natl. Acad. Sci. *114*, 10286–10291. https://doi.org/10.1073/pnas.1702581114.

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

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

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

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

32. Yoo, R., Rychel, K., Poudel, S., Al-bulushi, T., Yuan, Y., Chauhan, S., Lamoureux, C., Palsson, B.O., and 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*, e00033-22. https://doi.org/10.1128/msphere.00033-22.

33. Choe, D., Olson, C.A., Szubin, R., Yang, H., Sung, J., Feist, A.M., and Palsson, B.O. (2024). Advancing the scale of synthetic biology via cross-species transfer of cellular functions enabled by iModulon engraftment. Nat. Commun. *15*, 2356. https://doi.org/10.1038/s41467-024-46486-3.

34. Ninomiya, A., Katsuyama, Y., Kuranaga, T., Miyazaki, M., Nogi, Y., Okada, S., Wakimoto, T., Ohnishi, Y., Matsunaga, S., and Takada, K. (2016). Biosynthetic Gene Cluster for Surugamide A Encompasses an Unrelated Decapeptide, Surugamide F. ChemBioChem *17*, 1709–1712. https://doi.org/10.1002/cbic.201600350.

35. Takada, K., Ninomiya, A., Naruse, M., Sun, Y., Miyazaki, M., Nogi, Y., Okada, S., and 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. J. Org. Chem. *78*, 6746–6750. https://doi.org/10.1021/jo400708u.

36. Xu, F., Nazari, B., Moon, K., Bushin, L.B., and Seyedsayamdost, M.R. (2017). Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens. J. Am. Chem. Soc. *139*, 9203–9212. https://doi.org/10.1021/jacs.7b02716.

37. Paget, M. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules *5*, 1245–1265. https://doi.org/10.3390/biom5031245.

38. Poudel, S., Tsunemoto, H., Seif, Y., Sastry, A.V., Szubin, R., Xu, S., Machado, H., Olson, C.A., Anand, A., Pogliano, J., et al. (2020). Revealing 29 sets of independently modulated genes in *Staphylococcus aureus* , their regulators, and role in key physiological response. Proc. Natl. Acad. Sci. *117*, 17228–17239. https://doi.org/10.1073/pnas.2008413117.

39. Castro-Melchor, M., Charaniya, S., Karypis, G., Takano, E., and Hu, W.-S. (2010). Genome-wide inference of regulatory networks in Streptomyces coelicolor. BMC Genomics *11*, 578. https://doi.org/10.1186/1471-2164-11-578.

40. Romero-Rodríguez, A., Robledo-Casados, I., and Sánchez, S. (2015). An overview on transcriptional regulators in Streptomyces. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1849*, 1017–1039. https://doi.org/10.1016/j.bbagrm.2015.06.007.

41. Lim, H.G., Rychel, K., Sastry, A.V., Bentley, G.J., Mueller, J., Schindel, H.S., Larsen, P.E., Laible, P.D., Guss, A.M., Niu, W., et al. (2022). Machine-learning from Pseudomonas putida KT2440 transcriptomes reveals its transcriptional regulatory network. Metab. Eng. *72*, 297–310. https://doi.org/10.1016/j.ymben.2022.04.004.

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

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

44. Rajput, A., Tsunemoto, H., Sastry, A.V., Szubin, R., Rychel, K., Chauhan, S.M., Pogliano, J., and Palsson, B.O. (2022). Advanced transcriptomic analysis reveals the role of efflux pumps and media composition in antibiotic responses of *Pseudomonas aeruginosa*. Nucleic Acids Res. *50*, 9675–9688. https://doi.org/10.1093/nar/gkac743.

45. Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. Curr. Opin. Genet. Dev. *15*, 589–594. https://doi.org/10.1016/j.gde.2005.09.006.

46. Vernikos, G., Medini, D., Riley, D.R., and Tettelin, H. (2015). Ten years of pan-genome analyses. Curr. Opin. Microbiol. *23*, 148–154. https://doi.org/10.1016/j.mib.2014.11.016.

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

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

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

50. Louis, P., and Galinski, E.A. (1997). Characterization of genes for the biosynthesis of the compatible solute ectoine from Marinococcus halophilus and osmoregulated expression in Escherichia coli. Microbiology *143*, 1141–1149. https://doi.org/10.1099/00221287-143-4-1141.

51. Kuhlmann, A.U., and Bremer, E. (2002). Osmotically Regulated Synthesis of the Compatible Solute Ectoine in *Bacillus pasteurii* and Related *Bacillus* spp. Appl. Environ. Microbiol. *68*, 772–783. https://doi.org/10.1128/AEM.68.2.772-783.2002.

52. Crosa, J.H., and Walsh, C.T. (2002). Genetics and Assembly Line Enzymology of Siderophore Biosynthesis in Bacteria. Microbiol. Mol. Biol. Rev. *66*, 223–249. https://doi.org/10.1128/MMBR.66.2.223-249.2002.

53. Ehmann, D.E., Shaw-Reid, C.A., Losey, H.C., and 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. Proc. Natl. Acad. Sci. *97*, 2509–2514. https://doi.org/10.1073/pnas.040572897.

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

55. Horbal, L., Stierhof, M., Palusczak, A., Eckert, N., Zapp, J., and Luzhetskyy, A. (2021). Cyclofaulknamycin with the Rare Amino Acid D-capreomycidine Isolated from a Well-Characterized Streptomyces albus Strain. Microorganisms *9*, 1609. https://doi.org/10.3390/microorganisms9081609.

56. Ren, D., Lee, Y.-H., Wang, S.-A., and Liu, H. (2022). Characterization of the Oxazinomycin Biosynthetic Pathway Revealing the Key Role of a Nonheme Iron-Dependent Mono-oxygenase. J. Am. Chem. Soc. *144*, 10968–10977. https://doi.org/10.1021/jacs.2c04080.

57. Hoz, J.F.-D.L., Méndez, C., Salas, J.A., and Olano, C. (2017). Novel Bioactive Paulomycin Derivatives Produced by Streptomyces albus J1074. Molecules *22*, 1758. https://doi.org/10.3390/molecules22101758.

58. González, A., Rodríguez, M., Braña, A.F., Méndez, C., Salas, J.A., and Olano, C. (2016). New insights into paulomycin biosynthesis pathway in Streptomyces albus J1074 and generation of novel derivatives by combinatorial biosynthesis. Microb. Cell Factories *15*, 56. https://doi.org/10.1186/s12934-016-0452-4.

59. Sarmiento-Vizcaíno, A., Braña, A., Pérez-Victoria, I., Martín, J., De Pedro, N., Cruz, M., Díaz, C., Vicente, F., Acuña, J., Reyes, F., et al. (2017). Paulomycin G, a New Natural Product with Cytotoxic Activity against Tumor Cell Lines Produced by Deep-Sea Sediment Derived Micromonospora matsumotoense M-412 from the Avilés Canyon in the Cantabrian Sea. Mar. Drugs *15*, 271. https://doi.org/10.3390/md15090271.

60. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. *102*, 12678–12683. https://doi.org/10.1073/pnas.0504604102.

61. Rodionova, I.A., Gao, Y., Sastry, A., Hefner, Y., Lim, H.G., Rodionov, D.A., Saier, M.H., and Palsson, B.O. (2021). Identification of a transcription factor, PunR, that regulates the purine and purine nucleoside transporter punC in E. coli. Commun. Biol. *4*, 991. https://doi.org/10.1038/s42003-021-02516-0.

62. Feng, W.-H., Mao, X.-M., Liu, Z.-H., and Li, Y.-Q. (2011). The ECF sigma factor SigT regulates actinorhodin production in response to nitrogen stress in Streptomyces coelicolor. Appl. Microbiol. Biotechnol. *92*, 1009–1021. https://doi.org/10.1007/s00253-011-3619-2.

63. Mao, X.-M., Zhou, Z., Cheng, L.-Y., Hou, X.-P., Guan, W.-J., and Li, Y.-Q. (2009). Involvement of SigT and RstA in the differentiation of *Streptomyces coelicolor*. FEBS Lett. *583*, 3145–3150. https://doi.org/10.1016/j.febslet.2009.09.025.

64. Park, P.J. (2009). ChIP–seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. *10*, 669–680. https://doi.org/10.1038/nrg2641.

65. Bartlett, A., O’Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. *12*, 1659–1672. https://doi.org/10.1038/nprot.2017.055.

66. Otani, H., and Mouncey, N.J. (2022). RIViT-seq enables systematic identification of regulons of transcriptional machineries. Nat. Commun. *13*, 3502. https://doi.org/10.1038/s41467-022-31191-w.

67. Stegmüller, J., Estévez, M.R., Shu, W., Gläser, L., Myronovskyi, M., Rückert-Reed, C., Kalinowski, J., Luzhetskyy, A., and Wittmann, C. (2024). Systems metabolic engineering of the primary and secondary metabolism of Streptomyces albidoflavus enhances production of the reverse antibiotic nybomycin against multi-resistant Staphylococcus aureus. Metab. Eng. *81*, 123–143. https://doi.org/10.1016/j.ymben.2023.12.004.

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

69. Galardini, M., Mengoni, A., Biondi, E.G., Semeraro, R., Florio, A., Bazzicalupo, M., Benedetti, A., and Mocali, S. (2014). DuctApe: A suite for the analysis and correlation of genomic and OmniLogTM Phenotype Microarray data. Genomics *103*, 1–10. https://doi.org/10.1016/j.ygeno.2013.11.005.

70. Sastry, A.V., Yuan, Y., Poudel, S., Rychel, K., Yoo, R., Lamoureux, C.R., Li, G., Burrows, J.T., Chauhan, S., Haiman, Z.B., et al. (2024). iModulonMiner and PyModulon: Software for unsupervised mining of gene expression compendia. PLOS Comput. Biol. *20*, e1012546. https://doi.org/10.1371/journal.pcbi.1012546.

71. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

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

73. Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., and 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.

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

75. Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J.N., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. *51*, W46–W50. https://doi.org/10.1093/nar/gkad344.

76. Tong, Y., Whitford, C.M., Blin, K., Jørgensen, T.S., Weber, T., and Lee, S.Y. (2020). CRISPR–Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in streptomycetes. Nat. Protoc. *15*, 2470–2502. https://doi.org/10.1038/s41596-020-0339-z.

77. Whitford, C.M., Gockel, P., Faurdal, D., Gren, T., Sigrist, R., and Weber, T. (2023). CASCADE-Cas3 Enables Highly Efficient Genome Engineering in *Streptomyces* Species. Preprint, https://doi.org/10.1101/2023.05.09.539971 https://doi.org/10.1101/2023.05.09.539971.

78. Roubos, J.A., Krabben, P., Luiten, R.G.M., Verbruggen, H.B., and Heijnen, Joseph.J. (2001). A Quantitative Approach to Characterizing Cell Lysis Caused by Mechanical Agitation of *Streptomyces clavuligerus*. Biotechnol. Prog. *17*, 336–347. https://doi.org/10.1021/bp0001617.

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

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

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

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

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

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

85. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine Learning in Python. Mach. Learn. PYTHON.

86. Karp, P.D., Billington, R., Caspi, R., Fulcher, C.A., Latendresse, M., Kothari, A., Keseler, I.M., Krummenacker, M., Midford, P.E., Ong, Q., et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. Brief. Bioinform. *20*, 1085–1093. https://doi.org/10.1093/bib/bbx085.