scientific data

OPEN

DATA DESCRIPTOR

Tandem mass spectral metabolic profiling of 54 actinobacterial strains and their 459 mutants

Dillon W. P. Tay[1,2]✉, Lee Ling Tan[2,3], Elena Heng[2,3], Nadiah Zulkarnain[3], Elaine Jinfeng Chin[4], Zann Yi Qi Tan[4], Chung Yan Leong[4], Veronica Wee Pin Ng[4], Lay Kien Yang[4], Deborah C. S. Seow[4], Yoganathan Kanagasundaram[4], Siew Bee Ng[4], Yee Hwee Lim[1,2,5] & Fong Tian Wong[1,2,3]✉

Natural products encompass a diverse range of compounds with high impact applications in consumer care, agriculture and most notably, therapeutics. However, despite the expansive chemical repertoire indicated in genomic information of microbes, only a small subset can be obtained under laboratory conditions. To increase accessible chemical space and realize Nature's full chemical potential, a multi-pronged genetic- and cultivation-based strategy has been employed to activate and upregulate natural product biosyntheses in native and heterologous strains. This data descriptor documents a characterized collection of 2,138 liquid chromatography-tandem mass spectrometry (LC/MS-MS) spectra of fermentation extracts from 54 native actinobacterial strains collected from soil and marine environments in Singapore, and their 459 activated mutants in 3 to 5 media. A total of 743 unique metabolites have been identified, with the activated mutants demonstrating an approximately 2-fold expansion in accessible chemical space over wild type strains. Interrogating this expanded chemical diversity with cheminformatic tools can provide direction for the discovery of novel natural products with desirable functional activity.

Background & Summary

Nature's vast chemical diversity[1] has been a rich reservoir for various applications in personal care[2], agriculture[3], and health[4]. Methods to discover these valuable natural products have evolved from trial and error, to high-throughput screening[5], and presently to the artificial intelligence revolution combined with modern bioinformatics and cheminformatics[6,7]. An enduring interest in the exploration and characterisation of natural products has yielded a diverse collection of valuable specialty chemicals exemplified by medicines[8], herbicides[9], and fragrances[10]. Natural products are typically synthesised through the concerted effort of multiple enzymes encoded by gene clusters in the microbial genome, also known as biosynthetic gene clusters (BGCs)[11]. Despite the vast chemical repertoire suggested by genomic information in observed BGCs, only a small subset can be obtained under laboratory conditions[12], with the rest remaining unexpressed (silent) or with its predicted compound unobserved (cryptic)[13].

To interrogate the untapped potential in these silent or cryptic BGCs, we have developed a multi-pronged activation strategy[14] synergising integrase mediated genetic-based activation[15] with the "one strain many compounds" (OSMAC)[16] cultivation-based approach to significantly expand accessible metabolite space by approximately 2-fold. A series of 54 actinobacterial strains isolated from soil and marine environments in Singapore[17] were integrated with 5 different regulators (Table 1) – cyclic AMP receptor protein (Crp)[18], A-factor dependent

[1]Institute of Sustainability for Chemicals, Energy and Environment (ISCE2), Agency for Science, Technology and Research (A*STAR), 8 Biomedical Grove, #07-01 Neuros Building, Singapore, 138665, Republic of Singapore. [2]Singapore Integrative Biosystems and Engineering Research (SIBER), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. [3]Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, #07-06, Proteos, Singapore, 138673, Republic of Singapore. [4]Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, #01-02, Nanos, Singapore, 138669, Republic of Singapore. [5]Synthetic Biology Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Drive, Singapore, 117597, Republic of Singapore. ✉e-mail: dillon_tay@isce2.a-star.edu.sg; wongft@imcb.a-star.edu.sg

Regulator	Description	Accession code
Crp[18]	Cyclic AMP receptor protein	SCO3571
AdpA[19]	AraC/XylS transcriptor family protein	SCO2792
RedD[20]	AfsR/SARP family transcriptor regulator, RedD	SLIV09220
FAS[21]	Fatty acyl CoA synthase	SCO6196
SarA[22]	Putative membrane protein gene	SCO4069

Table 1. List of global regulators examined, their descriptions, and accession codes.

protein A (AdpA)[19], highly conserved *Streptomyces* antibiotic regulatory protein (SARP, RedD)[20], fatty acyl CoA synthase (FAS)[21], and a sporulation and antibiotics related gene A protein (SarA)[22].

The modifications yielded 459 mutants from 124 unique regulator-strain combinations. These native and engineered strains were then fermented in 3 to 5 media to yield a total of 2,138 fermentation extracts. High-throughput liquid chromatography-tandem mass spectrometry (LC-MS/MS) was employed to separate and characterize the complex chemical composition of these fermentation extracts, and the resulting data analyzed and organized into a curated dataset (Fig. 1).

Here, we report a curated LC-MS/MS dataset[23] describing the metabolic profiling of the 54 actinobacterial strains and their 459 mutants, as well as molecular networking analyses and suggested data applications not found in the original manuscript[14]. By analyzing the tandem mass spectra of 2,138 fermentation extracts using molecular networking (Fig. 2), 743 distinct metabolites grouped into 69 clusters (each containing at least two metabolites), and an additional 126 orphan metabolites were identified. Detailed information on these annotated metabolites and clusters are reported here for the first time[24]. All natural product spectral libraries from GNPS were referenced for comprehensive coverage despite potential risk of false positive matches to natural products outside of the actinobacteria metabolic space. The LC-MS/MS spectral collection of 2,138 fermentation extracts has been deposited on the Global Natural Product Social Networking (GNPS)[25] website and is available as a MassIVE dataset with accession number MSV000092237[23].

Although originally designed to investigate the chemical potential of silent and cryptic BGCs, this substantive collection of metabolite profiles also provides the opportunity to interrogate a diverse pool of potentially novel natural products for starting points toward new therapeutics[26], natural colors[27], or other biomolecules with desirable functional activity.

Methods

Fermentation, extraction, and sample preparation. 54 wild type strains (A1090, A1123, A11345, A1137, A1301, A1532, A1636, A2056, A2278, A2705, A2957, A30639, A33995, A34001, A34053, A40707, A40926, A4217, A44034, A5252, A53961, A58051, A5858, A61715, A6562, A80510, A8274, A8567, ATCC 23862, ATCC 31975, T10, T108, T118, T1195, T12, T1236, T1312, T1415, T1416, T1425, T1628, T168, T175, T265, T271, T298, T302, T343, T354, T36, T39, T467, T4680, T676) and their 459 edited mutants were received from the Agency for Science, Research and Technology (A*STAR)'s Natural Organism Library[17]. They were cultured on ISP2 plates [malt extract 10 g/L, Bacto yeast extract 4 g/L, glucose 4 g/L, Bacto agar 20 g/L] at 28 °C for 5 days. Three agar plugs of 5 mm diameter from the culture plate were then used to inoculate into 250 mL Erlenmeyer flasks each containing 50 mL SV2 seed media [glucose 15 g/L, glycerol 15 g/L, soya peptone 15 g/L, calcium carbonate 1 g/L, pH 7.0] and incubated for 4 days at 28 °C, with shaking at 200 rpm. A volume of 2.5 mL of the homogenized seed cultures were then inoculated into 250 mL Erlenmeyer flasks each containing 50 mL fermentation medium (Table 2). Marine actinomycetes strains were fermented in the same media with addition of 40 g/L sea salt, these media are annotated with the "M" prefix (i.e., MCA02LB instead of CA02LB). All cultures were fermented at 28 °C for 9 days shaking at 200 rpm with 50 mm throw. At the end of the incubation periods, cultures were freeze dried. A total of 2,138 fermentation samples were prepared. The lyophilized samples were extracted overnight (16 h) with methanol (14 mL) with shaking at 150 rpm. The extracted methanolic mixture was passed through cellulose filter paper (Whatman Grade 4, 1004-185) and the filtrate concentrated on a rotary evaporator, 0.1 mg of the dried methanol extract was then submitted for LC-MS/MS analysis.

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) data acquisition. Fermentation extract samples were analysed on an Agilent 1290 Infinity LC System coupled to an Agilent 6540 accurate-mass quadrupole time-of-flight (QTOF) mass spectrometer. 5 µL of extract was injected onto a Waters Acquity UPLC BEH C18 column, 2.1 × 50 mm, 1.7 µm. Mobile phases were water (A) and acetonitrile (B), both with 0.1% formic acid. The analysis was performed at flow rate of 0.5 mL/min, under gradient elution of 2% B to 100% B in 8 min. LC-MS/MS data was acquired in positive electrospray ionization (ESI) mode MS1 was acquired between *m/z* 100–2500 at a scan rate of 3 spectra/sec while MS/MS was acquired between *m/z* 100–2000 at a scan rate of 4 spectra/sec. For MS/MS fragmentation, a ramped collision energy method was employed, whereby the collision energy was determined according to the following formula:

$$collision\ energy\,(eV) = \frac{(precursor\ mz \times 5)}{100} + 2.5$$

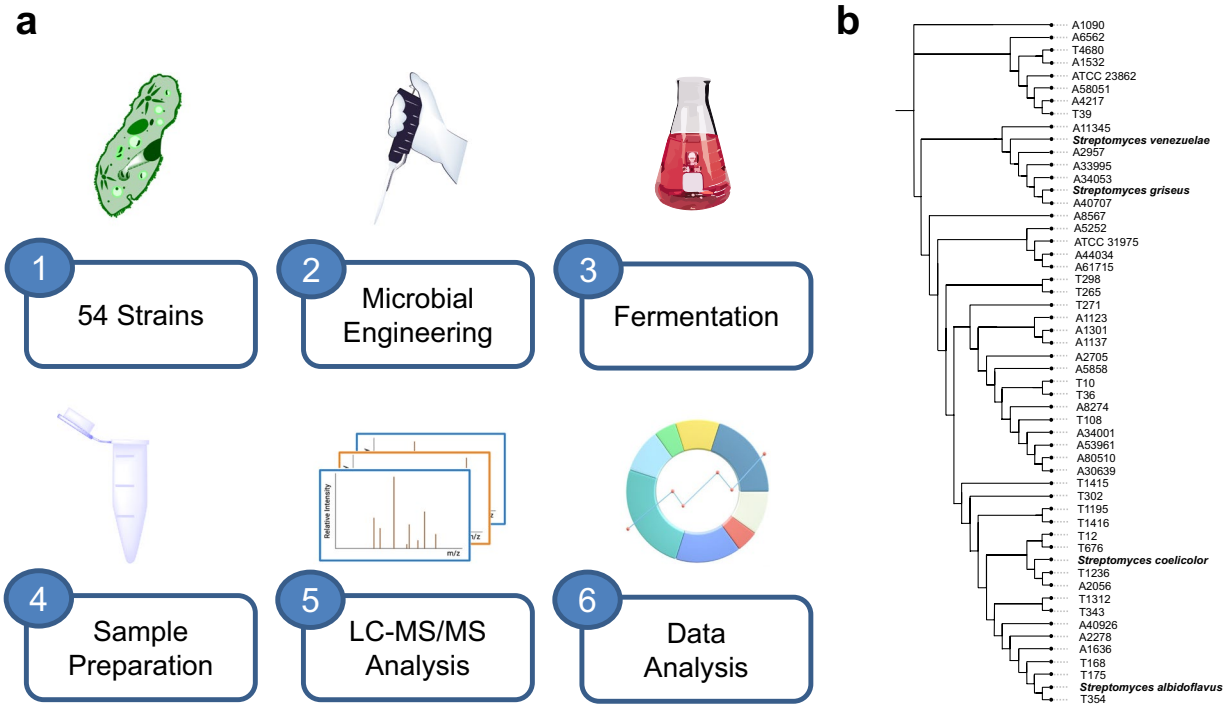

Fig. 1 Overview of LC-MS/MS dataset describing the metabolic profiling of the 54 actinobacterial strains and their 459 mutants. (**a**) The experimental workflow involves the following steps: we start from (1) an actinobacterial collection of 54 wild type strains, which are firstly (2) genetically activated by integrating overexpression cassettes of 5 global regulators (Crp, AdpA, RedD, FAS, SarA), (3) followed by OSMAC fermentation of these strains and their mutants in 3–5 different media to generate 2,138 samples. (4) Subsequently, sample preparation is carried out through lyophilization and solvent extraction, (5) followed by LC-MS/MS analysis of the fermentation extracts and (6) overall molecular networking data analysis. (**b**) Phylogenetic tree of 16S rRNA sequences of 50 strains[14] and 4 model *Streptomyces* (*Streptomyces venezuelae*, *Streptomyces griseus*, *Streptomyces coelicolor* and *Streptomyces albidoflavus*).

The typical QTOF operating parameters were as follows: sheath gas nitrogen, 12 L/min at 325 °C; drying gas nitrogen flow, 12 L/min at 350 °C; nebulizer pressure, 50 psi; nozzle voltage, 1.5 kV; capillary voltage, 4 kV. Lock masses in positive ion mode: purine ion at m/z 121.0509 and HP-0921 ion at m/z 922.0098.

Molecular networking. MSConvert v3.0.22198-0867718 from Proteowizard[28] was used for initial processing of raw liquid chromatography-tandem mass spectrometry (LC-MS/MS) data into an open-source file format (.mzML). All tandem mass spectra (MS/MS) signals with intensity values below 1000 signal intensity were removed as background correction. Classical molecular networking was performed on resulting MS/MS spectra using the online workflow from the GNPS website (http://gnps.ucsd.edu). All peaks in a +/−17 Da around the precursor ion mass were deleted to remove residual precursor ions, and peaks not in the top 6 most intense peaks in a +/−50 Da window were filtered out. The precursor ion mass tolerance was set to 0.02 Da and the MS/MS fragment ion tolerance was set to 0.02 Da. Nearly identical MS/MS spectra with precursor ion m/z within the mass tolerance are combined into a single representative spectrum via the MS-Cluster algorithm[29] and annotated as individual metabolites. Representative spectra created from a minimum number of 2 MS/MS spectra were considered for molecular networking. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to unlimited. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks.

Data Records

The dataset comprising of (1) unprocessed raw Agilent LC-MS/MS data (.d) as well as (2) converted open source file format (.mzML) copies of the 2,138 fermentation extracts from 54 actinobacterial strains and their 459 mutants in 3–5 media, has been deposited and is publicly accessible via MassIVE with the accession number MSV000092237 (https://doi.org/10.25345/C53X83W53)[23]. Detailed information on the 2,138 fermentation extracts analyzed, as well as the 743 individual metabolites and 69 clusters identified are available on figshare (https://doi.org/10.6084/m9.figshare.26144116)[24].

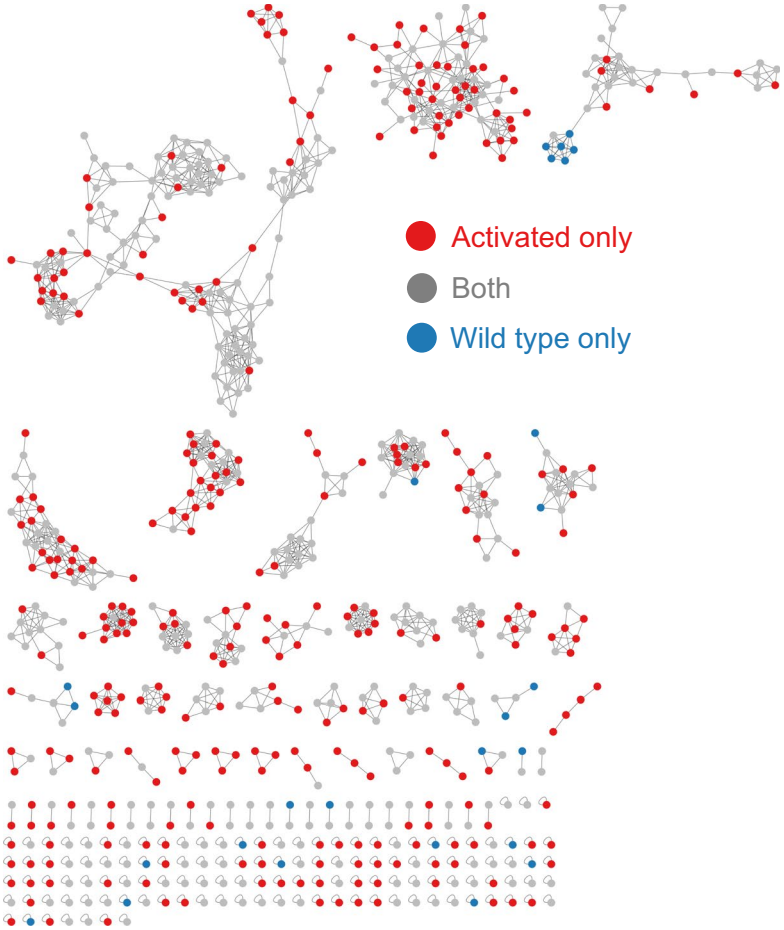

Fig. 2 Molecular networking analysis performed on tandem mass spectra of 2,138 fermentation extracts from 54 actinobacterial strains and their 459 mutants in 3 to 5 media[14]. 743 unique metabolites arranged in 69 clusters (with 2 or more metabolites) and 126 orphan metabolites are visualized with their connecting edges specifying a cosine score of more than 0.7. Red = metabolites present only in mutant strains. Blue = metabolites present only in wild type strains. Grey = metabolites present in both wild type and mutant strains.

CA02LB	CA07LB	CA08LB	CA09LB	CA10LB
Mannitol 20 g/L	Glycerol 15 g/L	Glucose 15 g/L	Beef Extract 10 g/L	Soluble Starch 20 g/L
Soybean Meal 20 g/L	Oatmeal 30 g/L	Cane Molasses 20 g/L	Yeast Extract 4 g/L	Soybean Meal 15 g/L
pH 7.5	Yeast Extract 5 g/L	Soluble Starch 40 g/L	Glucose 20 g/L	KH_2PO_4 3 g/L
	KH_2PO_4 5 g/L	$CaCO_3$ 8 g/L	Glycerol 3 g/L	$Na_2HPO_4.12H_2O$ 2 g/L
	$Na_2HPO_4.12H_2O$ 5 g/L	Cotton seed flour 25 g/L	pH 7.0	$Mg/LSO_4.7H_2O$
	$MgCl_2.6H_2O$ 1 g/L	pH 7.2		*Trace Salts Solution 1 mL
	pH natural			pH 7.2

Table 2. Media compositions employed for the fermentation of 54 wild type and 459 mutant strains. *Trace Salts Solution: solution containing 0.2% each of the following: $FeSO_4.7H_2O$, $MnCl_2.4H_2O$, $ZnSO_4.7H_2O$, $CuSO_4.5H_2O$, $CoCl_2.2H_2O$.

Technical Validation

LC-MS/MS retention time consistency. A combination of 4 compounds (Table 3) was used as quality control for retention time stability between different samples run over the 17-month period of data acquisition. The quality control samples were analyzed using the same experimental methodology as for fermentation extract analysis, identified via their unique precursor m/z, and their retention times recorded. Low coefficient of variation (%CV $\leq$ 1.3) indicates stable elution times between sample runs.

Intra-study quality control samples. For each 96-well plate of samples analyzed via LC-MS/MS, a minimum of ten quality control samples, and ten methanol blanks were run to ensure consistency in retention

Compound	m/z	Average Retention Time* (min)	%CV
Phenylpropylamine	152.1075	1.81 ± 0.024	1.3
Caffeine	195.0882	2.28 ± 0.017	0.7
Boldine	328.1549	2.49 ± 0.019	0.7
Reserpine	609.2812	4.14 ± 0.026	0.6

Table 3. Performance of quality control sample consisting of four reference compounds collected over seventeen months. *1 standard deviation given

time and background noise across the samples analyzed in this study. However, no additional intra-study quality controls such as pooled or representative fermentation extract samples were run, which is a limitation in experimental design.

Usage Notes

This dataset provides the opportunity to interrogate the chemical potential of a collection of 54 actinobacterial strains and their 459 activated mutants for novel natural products with desirable functional activity (e.g., anti-microbials, colorants). Some specific examples of such usage include 1) identification of known molecules with desired bioactivity such as valinomycin for antibiotic activity, then investigating networked metabolites or spectrally similar metabolites for novel antibiotic analogues, or 2) leveraging structural information captured in metabolite MS/MS data to perform spectral matching with known functional molecules to search for potentially novel natural products with similar structural characteristics that could demonstrate the desired functional activity. Additionally, the carefully curated mass spectral dataset presented here can also serve as a foundation for computational modelling applications, including artificial intelligence (AI) and machine learning. This study includes spectral data for various strains as well as their corresponding "activated" mutants. This dataset can be used to identify patterns in the production of different classes of molecules affected by genetic- and cultivation-based activation. This dataset also reveals the metabolic diversity in actinobacteria and the impact of genetic- and cultivation-based activation on metabolite production, this comparative data could facilitate bioinformatics studies aimed at metabolite annotation and pathway reconstruction. Additional metabolite characterisation such unsupervised substructure discovery (e.g. MS2LDA[30]), natural product classification (e.g. MolNetEnhancer[31]), and network annotation propagation (e.g. NAP[32]) may also be explored to provide richer insights.

Code availability

LC-MS/MS data conversion software (MSConvert v3.0.22198-0867718) employed is part of the open-source tool ProteoWizard (https://proteowizard.sourceforge.io/). LC-MS/MS data processing software (MestReNova v12.0.2-20910) is commercially available from Mestrelab Research S.L. Molecular networking tools are available on the Global Natural Products Social Molecular Networking (GNPS) website at https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp.

Received: 24 April 2024; Accepted: 27 August 2024;
Published online: 07 September 2024

References

1. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
2. Heath, R. S., Ruscoe, R. E. & Turner, N. J. The beauty of biocatalysis: sustainable synthesis of ingredients in cosmetics. *Nat. Prod. Rep.* **39**, 335–388 (2022).
3. Ortiz, A. & Sansinenea, E. Recent advancements for microorganisms and their natural compounds useful in agriculture. *Appl. Microbiol. Biotechnol.* **105**, 891–897 (2021).
4. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661 (2016).
5. Mishra, K. P., Ganju, L., Sairam, M., Banerjee, P. K. & Sawhney, R. C. A review of high throughput technology for the screening of natural products. *Biomed. Pharmacother.* **62**, 94–98 (2008).
6. Stone, S., Newman, D. J., Colletti, S. L. & Tan, D. S. Cheminformatic analysis of natural product-based drugs and chemical probes. *Nat. Prod. Rep.* **39**, 20–32 (2022).
7. Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L. & Plisson, F. Natural product drug discovery in the artificial intelligence era. *Chem. Sci.* **13**, 1526–1546 (2022).
8. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **163**, 1297–1300 (2015).
9. Dayan, F. E., Owens, D. K. & Duke, S. O. Rationale for a natural products approach to herbicide discovery. *Pest Manag. Sci.* **68**, 519–528 (2012).
10. Burger, P., Plainfossé, H., Brochet, X., Chemat, F. & Fernandez, X. Extraction of Natural Fragrance Ingredients: History Overview and Future Trends. *Chem. Biodivers.* **16**, e1900424 (2019).
11. Skinnider, M. A. *et al.* Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.* **11**, 6058 (2020).
12. Covington, B. C., Xu, F. & Seyedsayamdost, M. R. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu. Rev. Biochem* **90**, 763–788 (2021).
13. Hoskisson, P. A. & Seipke, R. F. Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. *mBio* **11**, e02642–02620 (2020).
14. Tay, D. W. P. *et al.* Exploring a general multi-pronged activation strategy for natural product discovery in Actinomycetes. *Commun. Biol.* **7**, 50 (2024).
15. Bierman, M. *et al.* Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene* **116**, 43–49 (1992).

16. Romano, S., Jackson, S. A., Patry, S. & Dobson, A. D. W. Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Mar. Drugs* **16**, 244 (2018).
17. Ng, S. B. *et al*. The 160K Natural Organism Library, a unique resource for natural products research. *Nat. Biotechnol.* **36**, 570–573 (2018).
18. Gao, C., Hindra, Mulder, D., Yin, C. & Elliot, M. A. Crp Is a Global Regulator of Antibiotic Production in *Streptomyces*. *mBio* **3**, e00407–00412 (2012).
19. Lee, H.-N., Kim, J.-S., Kim, P., Lee, H.-S. & Kim, E.-S. Repression of Antibiotic Downregulator WblA by AdpA in Streptomyces coelicolor. *Appl. Environ. Microbiol.* **79**, 4159–4163 (2013).
20. Krause, J., Handayani, I., Blin, K., Kulik, A. & Mast, Y. Disclosing the potential of the SARP-type regulator PapR2 for the activation of antibiotic gene clusters in Streptomycetes. *Front. Microbiol.* **11**, 225 (2020).
21. Wang, W. *et al*. Harnessing the intracellular triacylglycerols for titer improvement of polyketides in Streptomyces. *Nat. Biotechnol.* **38**, 76–83 (2020).
22. Ou, X. *et al*. SarA influences the sporulation and secondary metabolism in Streptomyces coelicolor M145. *Acta Biochim. Biophys. Sin.* **40**, 877–882 (2008).
23. Wong, F. T. A general multipronged activation approach for natural product discovery in Actinomycetes 54 actinobacterial strains with genetic and cultivation based activation. *MassIVE* https://doi.org/10.25345/C53X83W53 (2023).
24. Tay, D. W. P. *et al*. Tandem mass spectral metabolic profiling of 54 actinobacterial strains and their 459 mutants. *figshare* https://doi.org/10.6084/m9.figshare.26144116 (2024).
25. Wang, M. *et al*. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
26. Ghirga, F. *et al*. A unique high-diversity natural product collection as a reservoir of new therapeutic leads. *Org. Chem. Front.* **8**, 996–1025 (2021).
27. Newsome, A. G., Culver, C. A. & van Breemen, R. B. Nature's Palette: The Search for Natural Blue Colorants. *J. Agric. Food. Chem.* **62**, 6498–6511 (2014).
28. Chambers, M. C. *et al*. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
29. Frank, A. M. *et al*. Clustering Millions of Tandem Mass Spectra. *J. Proteome Res.* **7**, 113–122 (2008).
30. van der Hooft, J. J. J., Wandy, J., Barrett, M. P., Burgess, K. E. V. & Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *PNAS* **113**, 13738–13743 (2016).
31. Ernst, M. *et al*. MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* **9**, 144 (2019).
32. da Silva, R. R. *et al*. Propagating annotations of molecular networks using in silico fragmentation. *PLoS Comput. Biol.* **14**, e1006089 (2018).

Acknowledgements

The authors gratefully acknowledge financial support from the National Research Foundation, Singapore (NRF-CRP19-2017-05-00). This work is also supported by the Agency for Science, Technology and Research (A*STAR, Singapore) under C211917003, C211917006, C233017006, and Singapore Integrative Biosystems and Engineering Research Strategic Research & Translational Thrust (SIBER SRTT).

Author contributions

F.T.W. and Y.H.L. conceptualized, designed, and coordinated the study. The following authors conducted the experiments and acquired the data: L.L.T., E.H. and N.Z. performed the molecular biology (integration and screening mutants) work; L.K.Y. and D.C.S.S. acquired LC-MS/MS data; E.J.C., Z.Y.Q.T., C.Y.L. and V.W.P.N. performed fermentation. D.W.P.T. performed data analysis. F.T.W. supervised the molecular biology design and experiments. S.B.N. supervised fermentation. Y.K. supervised chemical analysis. F.T.W. and Y.H.L. supervised the overall data analysis, interpretation, and presentation. D.W.P.T., F.T.W. and Y.H.L. wrote the manuscript with inputs from all the authors.

Competing interests

Patent applications have been filed by some of the authors wherein some of the data are disclosed in this manuscript.

Additional information

Correspondence and requests for materials should be addressed to D.W.P.T. or F.T.W.