nature metabolism

Article

https://doi.org/10.1038/s42255-024-01024-9

Elucidation of genes enhancing natural product biosynthesis through co-evolution analysis

Received: 24 March 2023

Accepted: 6 March 2024

Published online: 12 April 2024

Check for updates

Xinran Wang[1,11], Ningxin Chen[1,11], Pablo Cruz-Morales[2,11], Biming Zhong[1], Yangming Zhang[1], Jian Wang[3], Yifan Xiao[4], Xinnan Fu[4], Yang Lin[5], Suneil Acharya[6], Zhibo Li[1], Huaxiang Deng[1], Yuhui Sun[3,7], Linquan Bai[4], Xiaoyu Tang[5] ✉, Jay D. Keasling[1,6,8,9] ✉ & Xiaozhou Luo[1,10] ✉

Streptomyces has the largest repertoire of natural product biosynthetic gene clusters (BGCs), yet developing a universal engineering strategy for each *Streptomyces* species is challenging. Given that some *Streptomyces* species have larger BGC repertoires than others, we proposed that a set of genes co-evolved with BGCs to support biosynthetic proficiency must exist in those strains, and that their identification may provide universal strategies to improve the productivity of other strains. We show here that genes co-evolved with natural product BGCs in *Streptomyces* can be identified by phylogenomics analysis. Among the 597 genes that co-evolved with polyketide BGCs, 11 genes in the 'coenzyme' category have been examined, including a gene cluster encoding for the cofactor pyrroloquinoline quinone. When the *pqq* gene cluster was engineered into 11 *Streptomyces* strains, it enhanced production of 16,385 metabolites, including 36 known natural products with up to 40-fold improvement and several activated silent gene clusters. This study provides an innovative engineering strategy for improving polyketide production and finding previously unidentified BGCs.

Microbial-based natural products and their derivatives comprise a notable portion of the pharmaceuticals in clinical use[1], with polyketides and their derivatives constituting 20% of the top-selling drugs, generating over US$20 billion in annual revenue worldwide[2]. In addition to the need to enhance production of these drugs for cost-effectiveness, the rise of antibiotic resistance also calls for discovery of new antibiotics. *Streptomyces* have the largest repertoire of natural product biosynthetic gene clusters (BGCs) known so far. However, due to the distinct physiology and metabolic characteristics among different species, few engineering strategies can be applied to different *Streptomyces* hosts

[1]Shenzhen Key Laboratory for the Intelligent Microbial Manufacturing of Medicines, Key Laboratory of Quantitative Synthetic Biology, Center for Synthetic Biochemistry, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [3]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Ministry of Education), Wuhan University, Wuhan, China. [4]State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [5]Institute of Chemical Biology, Shenzhen Bay Laboratory, Shenzhen, China. [6]Department of Chemical and Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, CA, USA. [7]School of Pharmacy, Huazhong University of Science and Technology, Wuhan, China. [8]Joint BioEnergy Institute, Emeryville, CA, USA. [9]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [10]Shenzhen Infrastructure for Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. [11]These authors contributed equally: Xinran Wang, Ningxin Chen, Pablo Cruz-Morales. ✉e-mail: xtang@szbl.ac.cn; keasling@berkeley.edu; xz.luo@siat.ac.cn

to enhance the titre of natural products or to explore a strain for the discovery of new antibiotics.

The current engineering strategies developed for natural product production are mostly based on two strategies: (1) metabolic engineering to enhance precursor biosynthetic pathways[2] or delete competitor pathways[3] or (2) gene regulation engineering such as modulating pathway regulation or global regulation processes[4]. These strategies came from numerous investigations of natural product biosynthetic processes. However, given the wide diversity of natural product producers and pathways, we are far from being able to apply a universal strategy to improve production of all natural products.

The popularization of sequencing technologies has granted public access to genomes of natural product producers, allowing the performance of bottom-up studies aimed at pinpointing the mechanisms behind complex traits, and paving the way for methods development in synthetic biology for natural products overproduction as well as identification of new molecules. Large-scale genomic studies have shown that some bacterial and fungal taxa are more talented than others at making a diverse range of natural products[5]. This observation indicates the existence of key genes related to high productivity. Identifying these genes is important to develop metabolic engineering and synthetic biology strategies to improve production and accelerate the discovery of these valuable molecules. Pan-genomic analysis has been developed as a tool to explore the genomes of a species for desired traits, and the corresponding 'gene-trait-application' strategy has been successfully applied in many species, including plants[6], fungi[7] and bacteria[8], to look for genes correlated to specific traits and guide further optimization of the species for the desired applications.

Streptomyces is the producer of most polyketide drugs and was shown to encode for more natural product BGCs than other bacteria[9]. *Streptomyces* is a large genus containing around 22,900 known species[10]. Genomic and pan-genomic analysis addressing the abundance and distribution of natural product BGCs inside *Streptomyces* have shown that the potential to produce natural products diverges at the species level[9,11–15], and the relationship between taxonomy and biosynthetic proficiency has been established, indicating that some *Streptomyces* lineages have more BGCs and potentially produce more natural products than others[16,17]. However, the identity and function of the genes associated with biosynthetic proficiency are not known.

Pyrroloquinoline quinone (PQQ) is a redox cofactor involved in alcohol or glucose dehydrogenase reactions in many Gram-negative bacteria[18]. PQQ participates in a variety of intracellular roles, including the use of carbon sources[19] and the respiratory metabolism related to oxidative stress resistance[18]. The biosynthesis of PQQ requires specific precursors and enzymes encoded by the *pqq* operon, which contains five (*pqqABCDE*) or six biosynthetic genes (*pqqABCDEF*) depending on the source organism[20,21]. In *Streptomyces*, it has been reported that PQQ-dependent dehydrogenases participated in lankacidin biosynthesis in *Streptomyces rochei* 7434AN4 (ref. 22). The linkage of PQQ with other natural products in other *Streptomyces* species has not been previously reported.

In this study, we applied pan-genomic analysis to *Streptomyces* species to look for genes that co-evolved with polyketide BGCs. We found that 597 genes were conserved among a closely related group of *Streptomyces* that encode for an outstandingly large number of polyketide synthases (PKSs), including the biosynthesis of the cofactor PQQ. We confirmed the association of PQQ with biosynthetic proficiency by introduction or overexpression of the PQQ biosynthetic pathway in 11 *Streptomyces* strains and two other industrial actinobacterial strains. PQQ could effectively enhance natural product production and activation of silent gene clusters in *Streptomyces*. Besides previous studies describing PQQ-dependent dehydrogenases involved in lankacidin biosynthesis[22], this is, to our knowledge, the only report of the ubiquitous enhancement of natural product production through the metabolic effect of the PQQ cofactor. This study provides innovative engineering strategies and targets for development of platforms for polyketide biosynthesis production and drug discovery.

Results

Phylogenomic insights on *Streptomyces* PKS-enriched family

For this analysis, we integrated a database with high-quality, publicly available genomes of strains that were accurately classified as members of the *Streptomyces* clade. To avoid variations in quality or errors in taxonomic classification, we filtered a set of 7,762 bacterial genomes from a previously integrated database[23]; from it, we selected 720 actinobacterial genomes assembled in no more than ten contigs and with a completeness of 85% or more (Supplementary Data 1). This cut off allowed for comprehensive retrieval of actinobacterial genomes including those with several plasmids or those that may be minimally fragmented and for which enough markers for phylogenetic reconstruction can be found. We extracted 18 conserved protein markers from the 720 genomes (Supplementary Table 1 and Supplementary Fig. 1) and used them to obtain a species tree. Using this tree as reference, we selected 201 genomes found within the Streptomycetaceae clade including 17 non-*Streptomyces* genomes and 184 members of the genus *Streptomyces*. These genomes have an average completeness of 98% and share 604 orthologous proteins (Supplementary Data 2). This set of orthologues was used to construct a *Streptomyces* species tree that was rooted using the 17 non-*Streptomyces* strains. This tree showed excellent branch support for most nodes. Therefore, we used it as the basis for further analyses (Fig. 1).

To search for potential PKS-associated genes, we established 14 clades and recorded the G+C content, genome length and the number and type of BGC for each strain. We observed that the variations in these traits correlate with the topology of the tree and with our clade classification (Fig. 1). This analysis made evident that clade 12 includes all the strains with the largest genomes among in our set and it also encodes the largest number of PKS-related domains and BGCs, this result is consistent with previous reports[17]. We proposed that, although these BGCs may be expressed under different conditions, this outstanding abundance may be associated with other metabolic traits that support PKS production.

Conserved *pqq* gene cluster in polyketide-rich *Streptomyces*

To explore this hypothesis, we calculated the core genomes (genes exist in all strains of a clade) for each of the 14 clades. Then, to identify clade-specific conserved genes, we compared the clade-level core genomes and looked at the genes that are uniquely conserved in each clade.

Following this approach, we found 597 genes uniquely conserved in clade 12 (Supplementary Data 3). To establish whether correlations between these 597 genes and PKS proficiency exist, we recorded the presence or absence of each of them in the 201 genomes (Supplementary Data 4). Then, the correlation between the 597 genes and KS domain abundance was evaluated as point biserial correlations (Supplementary Data 4). There was a positive correlation ($P < 0.005$, correlation >0) in most unique genes (545 out of 597, 91.2%) (Supplementary Data 4 and Supplementary Fig. 2). These results are consistent with our hypothesis that unique genes identified in the PKS-proficient clade by phylogenomic analysis are correlated with PKS abundance.

To identify traits and genes that may explain biosynthetic proficiency, the unique genes were sorted according to their functions into different families. Of them, 57% could be assigned a function, and classified in one of 12 families, namely regulators, fatty acids, lipids and polyketide related functions, transporters, carbohydrate-active enzymes (cazymes), sigma factors, coenzymes, cell wall biosynthesis, metal related functions, cell division, peptidase, amino acid metabolism

https://doi.org/10.1038/s42255-024-01024-9

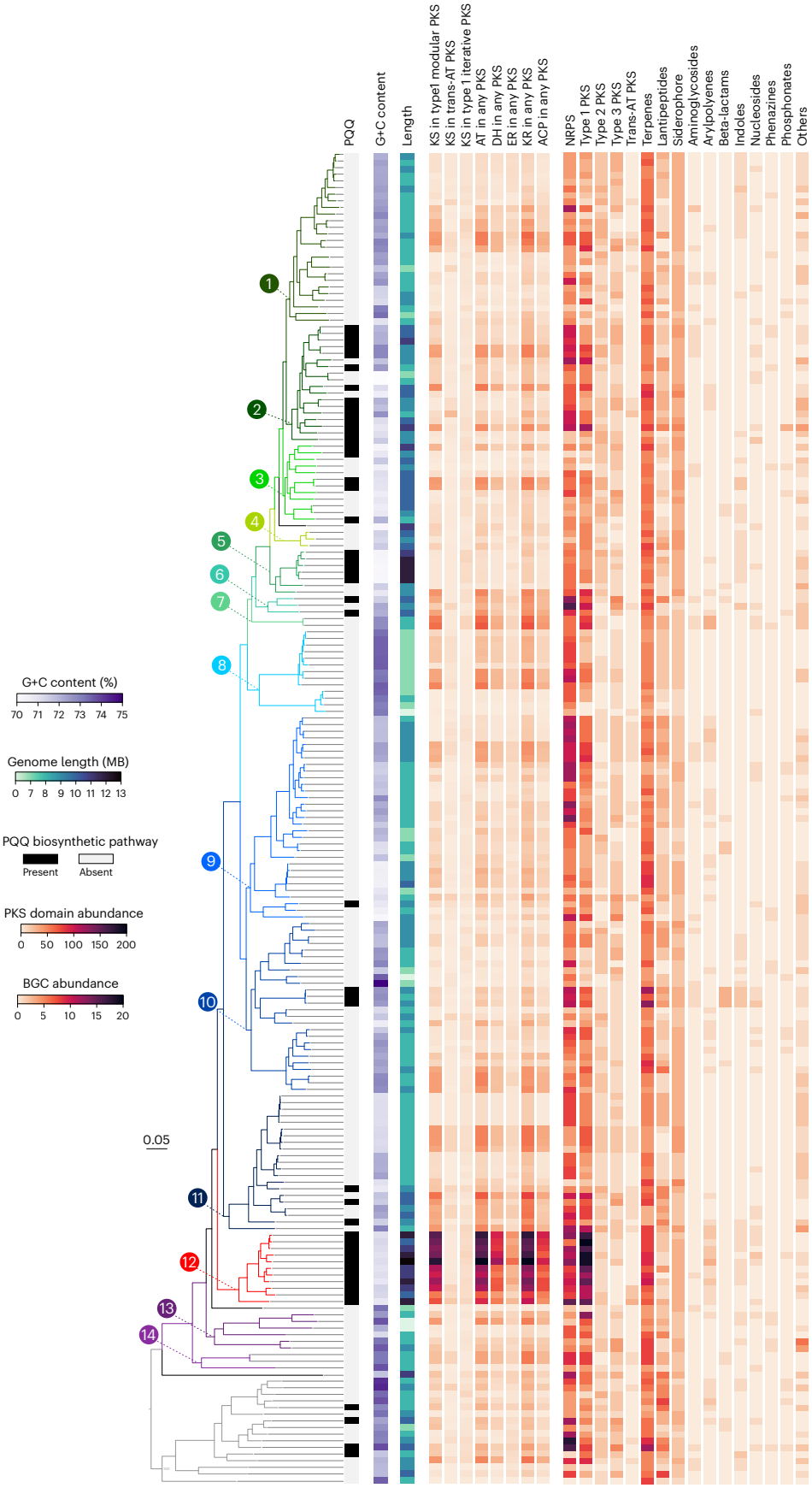

Fig. 1 | Phylogenomic analysis of the *Streptomyces* genus and the number of PKS gene clusters. The phylogenetic tree is divided into 14 clades highlighted in different colours. Genome length, G+C content, natural product BGCs, PKS-related domains and *pqq* BGCs are shown as coloured boxes to the right. The values are indicated by legends on the left. KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase; KR, ketoreductase; ACP, acyl-carrier protein.

https://doi.org/10.1038/s42255-024-01024-9

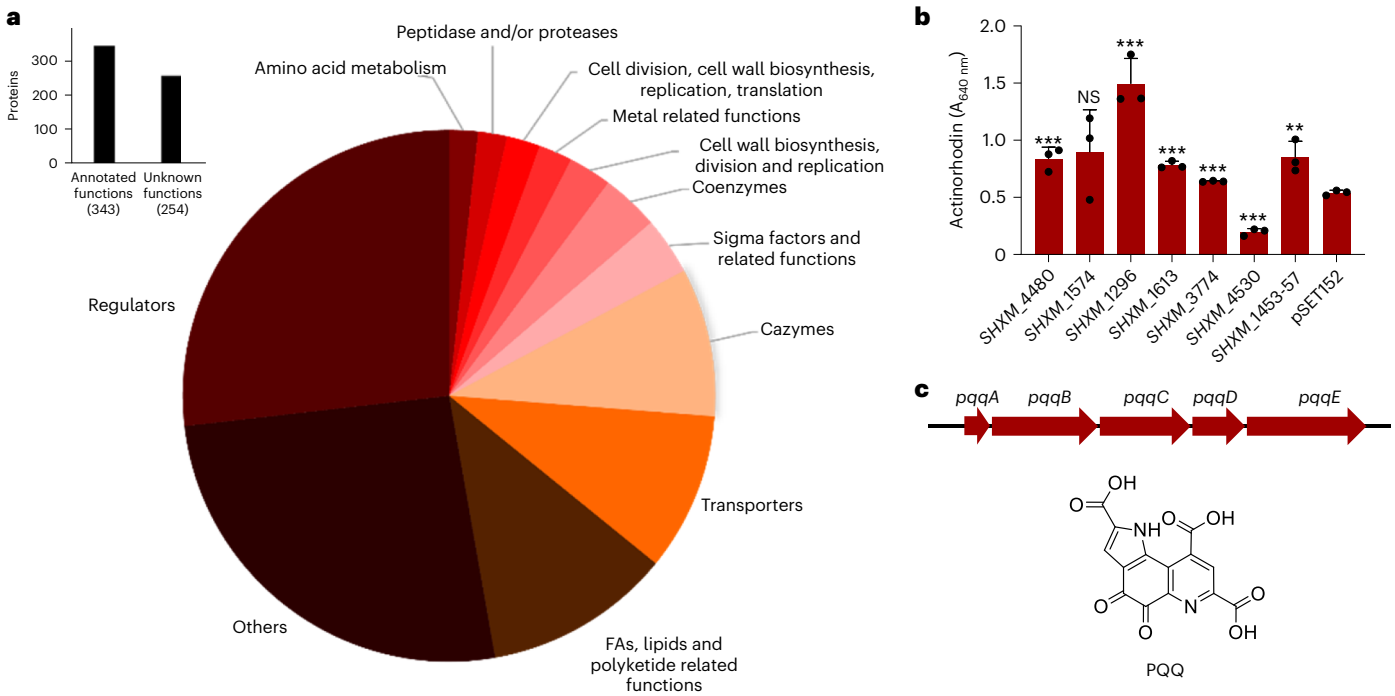

Fig. 2 | Exploration and evaluation of unique genes in PKS rich clade. **a**, Functional classification of the genes uniquely conserved in clade 12. Among the 597 co-evolved genes, 254 encoded for proteins with unknown functions. The rest of the 343 genes encoding for proteins with specific functions were mainly distributed in the family (from large to small) of other functions, regulators, FAs, lipids and polyketide related functions, transporters, cazymes, sigma factors, coenzymes, cell wall biosynthesis, metal related functions, cell division, peptidase and amino acid metabolism. FAs, fatty acids. **b**, Evaluation of genes in the 'coenzyme clade' by heterologous expression in *S. coelicolor*. The titre of actinorhodin was determined by the absorbance at 640 nm (A_{640}). *SHXM_4480*, *S. coelicolor* M145 expressing gene *SHXM_4480*; *SHXM_1574*, *S. coelicolor* M145 expressing gene *SHXM_1574*; *SHXM_1296*, *S. coelicolor* M145 expressing gene *SHXM_1296*; *SHXM_1613*, *S. coelicolor* M145 expressing gene *SHXM_1613*; *SHXM_3774*, *S. coelicolor* M145 expressing gene *SHXM_3774*; *SHXM_4530*, *S. coelicolor* M145 expressing gene *SHXM_4530*; *SHXM_1453-57*, *S. coelicolor* M145 expressing gene *SHXM_1453-57* and pSET152, control strain, *S. coelicolor* M145 introduced with the backbone vector pSET152. Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). Multiple comparison significance was tested to $^{**}P < 0.05$, $^{***}P < 0.01$ by an unpaired two-sided Student's *t*-test. NS, not significant. **c**, Organization of the *pqq* operon.

and other functions. The 'regulators' category is the most abundant; however, as regulators are already considered engineering targets, we did not further explore them (Fig. 2a). The genes within the 'fatty acids, lipids and polyketide related functions' category include enzymes such as phosphopantetheinyl transferases, methylmalonyl-coenzyme A (CoA) mutases, 3-oxoacyl-acyl-carrier-protein synthases, acyl-carrier proteins, long-chain-fatty-acid-CoA ligases and thioesterases (Supplementary Data 3), some of which have been already pinpointed as engineering targets for enhancing PKS productivity[3].

In addition, there are genes that have few reports associated with polyketide production, such as coenzymes, peptidase and/or proteases or proteins with metal related functions and so on (Fig. 2a). This discovery suggests connections between these genes and PKS production, which could be useful for developing tools to enhance polyketide production. As a proof of concept, we investigated the connection of genes in the 'coenzyme' family with PKS production. There are 11 genes inside the 'coenzyme' family, all of which were found to have a positive correlation with the PKS genes (Supplementary Data 4 and Supplementary Fig. 2). The 11 genes are responsible for biosynthesis of coenzyme B6, coenzyme Q, PQQ and folate (Supplementary Table 2).

To test the contribution of genes in 'coenzyme' family on natural product production, all 11 orthologous genes were cloned from the genome of *Streptomyces hygroscopicus* XM201 (a strain in clade 12), placed under control of the strong promoter *kasOp**, and introduced into the model strain *Streptomyces coelicolor* M145. These five PQQ biosynthetic genes (*pqqA–E*) were co-expressed as an operon. The blue-colour actinorhodin, a polyketide naturally produced by *S. coelicolor*, was followed as an indicator of titre for this class of compounds. As shown in Fig. 2b, introduction of *shxm_4480*, *shxm_1296*, *shxm_1613*, *shxm_3774* and *shxm1453-57* (the *pqq* biosynthetic cassette) had a positive effect on actinorhodin production, which resulted in 62, 189, 51, 20 and 85% increases in actinorhodin titre, respectively.

Given that all the functions needed to produce PQQ were found among the biosynthetic proficiency-correlated genes, and that their expression in *S. coelicolor* induces polyketide production, we reasoned that PQQ should play a conserved role in enhancing the metabolism that leads to polyketide production. The positive effect of PQQ on actinorhodin production was first confirmed by feeding PQQ molecule into the fermentation medium (Supplementary Fig. 3). For further confirmation, truncated *pqq* gene clusters were constructed. The biosynthesis of PQQ starts with the ribosomal production of a short peptide encoded in *pqqA*, which then undergoes a series of modifications by *pqqB–E* to form the characteristic pyridine and pyrrole rings in PQQ[20] (Fig. 2c). The genes *pqqA* and *pqqC* are responsible for the first and final biosynthetic steps, respectively. To confirm that the increase in actinorhodin production is caused by PQQ and not by the isolated function of one or more of the *pqq* genes, two truncated *pqq* gene clusters—one lacking *pqqA* and the other lacking *pqqC*—were constructed and introduced to *S. coelicolor*. As shown in Supplementary Fig. 4, their previously observed positive influence on actinorhodin was distinctly abolished, affirming that it was the mature PQQ molecule that enhanced actinorhodin biosynthesis.

Mechanism of *pqq* operon for enhanced actinorhodin yield

To understand why the presence of PQQ improves production of actinorhodin in *S. coelicolor*, and indeed other natural products in a wide range of species, as demonstrated later in this work (sections 'Effects

https://doi.org/10.1038/s42255-024-01024-9

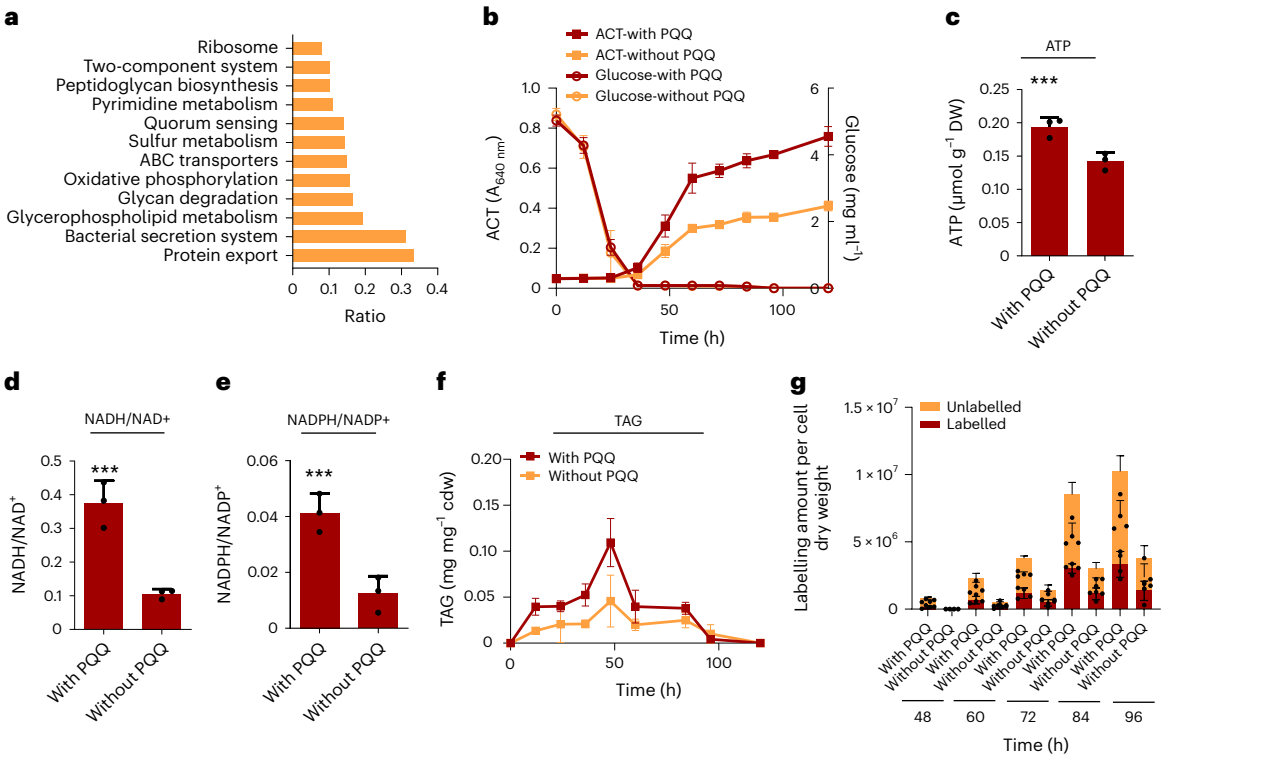

Fig. 3 | PQQ increases actinorhodin production by enhancing cofactor biosynthesis and modulating intracellular TAG level. a, Pathway enrichment analysis of the upregulated proteins. The ratio indicates the fraction of upregulated proteins enriched in a pathway versus all the proteins in the particular pathway. **b**, Actinorhodin production curve and glucose consumption curve for *S. coelicolor* with and without the *pqq* gene cluster. Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). ACT, actinorhodin. **c**–**e**, Introduction of the *pqq* BGC increased the levels of ATP (**c**), NADH/NAD$^+$ (**d**) and NADPH/NADP$^+$ (**e**). DW, dry weight. Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). Multiple comparison significance was tested to $^{**}P < 0.05$, $^{***}P < 0.01$ by an unpaired two-sided Student's *t*-test. **f**, Introduction of the *pqq* BGC increased TAG formation and degradation efficiency. Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). **g**, The labelled and unlabelled amount of actinorhodin at different time points during fermentation after supplementation of ^{13}C-labelled glucose. Mean concentrations with error bars showing s.d. are plotted ($n = 4$, three biologically independent samples). Multiple comparison significance was tested to $^{**}P < 0.05$, $^{***}P < 0.01$ by an unpaired two-sided Student's *t*-test.

of the *pqq* operon on several *Streptomyces*', '*pqq* operon boosts natural product biosynthesis' and '*pqq* operon improves natural products in industrial strains'), we then investigated the mechanism in the model strain *S. coelicolor* M145.

We first examined the changes in the host proteome in the presence of the *pqq* gene cluster by performing a fractionated proteomic analysis. The results revealed that 299 proteins were upregulated. We classified these proteins following functional categories defined by the Kyoto Encyclopedia of Genes and Genomes. The upregulated proteins (Supplementary Data 5) spanned 12 distinct pathways (ratio >0.08). Upregulation of bacteria secretion system (ko03070) and protein exports (ko03060) indicate more active transport of metabolic substrates and products (Fig. 3a). Notably, several pathways related to cofactor regeneration—including glycan degradation (ko00511), oxidative phosphorylation (ko00190) and pyrimidine metabolism (ko00240)—were upregulated, indicating the potential increase in intracellular cofactor levels. Moreover, the enhancement of glycerophospholipid metabolism (ko00564) indicated a more robust interconversion between lipids and glycerophospholipid or triacylglyerol (TAG), a potential carbon source for secondary metabolite production, especially for polyketides.

In *Streptomyces*, polyketides are commonly produced during the stationary phase, and it has been established that TAG and fatty acids are the direct carbon sources for actinorhodin[3]. Consistently, we observed production of actinorhodin in our strains with and without the *pqq* operon after the glucose was totally consumed (Fig. 3b). Given these results, we reasoned that the expression of the *pqq* operon should cause an increase in the production of cofactors such as ATP, NADH and NADPH via glycan degradation and oxidative phosphorylation, and that the mobilization of carbon stored as lipids should lead to an increase in the acyl-CoA pools. That extra ATP, NADH, NADPH and acyl-CoA should then account for the increase in polyketide production.

To test this idea, we examined the changes in the abundance of ten important cofactors involved in natural product biosynthesis on expression of the *pqq* operon. As shown in Fig. 3c–e, the intracellular level of ATP and the ratios of NADPH to NADP and NADH to NAD were found to increase by 32, 392 and 253%, respectively. No significant change was observed for other cofactors (Supplementary Fig. 5). These results indicate that NADPH/NADH and ATP assisted in improving the product levels by providing additional reducing power and cellular energy.

We also assessed the changes in TAG abundance during different growth phases on expression of the *pqq* operon. A significant increase in TAG accumulation was observed in the first 48 hours of growth with the highest at 48 hours, which was twofold higher than the strain without the *pqq* gene cluster. In addition, the consumption of TAG after 48 hours has also been accelerated, potentially providing a larger flux of acetyl-CoA for actinorhodin biosynthesis (Fig. 3f). Increased individual and overall fatty acids pools were higher in the *pqq*-expressing strain during actinorhodin production (Supplementary Figs. 6 and 7), which suggests that PQQ-enhanced accumulation and degradation of TAG leads to a generalized increase in acyl-CoA-dependent pathways.

To further study the link between this TAG reserve and production of actinorhodin, we performed a glucose feeding experiment, in which unlabelled glucose was fed during the first 24 hours after which the carbon source was switched to [^{13}C] glucose. The results indicate that in the strain harbouring the *pqq* gene cluster, more than half of the actinorhodin originates from unlabelled glucose, indicating the elevated product may be associated with TAG reservoir generated during the period where the cells were growing on unlabelled glucose (Fig. 3g).

We also determined the intracellular concentration of 121 compounds, including nucleotides, amino acids and various organic compounds that are building blocks of various natural products. The concentrations of compounds were quantified throughout the period of actinorhodin production aiming to provides a comprehensive and systematic overview of changes during the entire fermentation period. As shown in Supplementary Table 3 and Supplementary Data 6, the levels of 41 compounds changed significantly on PQQ treatment (fold change >1.0, $P < 0.05$). These compounds include amino acids, carbohydrates and tricarboxylic acid and Embden–Meyerhof–Parnas (EMP) pathway intermediates, purine and pyrimidine ribonucleotides (pyruvate, asparagine and UDP-glucose and so on), which are precursors to nucleoside antibiotics, alkaloids, peptides and sugar-derived natural products. We found the concentration of acetyl-CoA, the direct precursor to actinorhodin, and all polyketides (Supplementary Table 4) significantly increased from 103 to 126% between 36 and 60 h (Supplementary Fig. 8). This coincided with the degradation of lipids and the concurrent augmentation of actinorhodin.

In conclusion, our integrated proteomic and metabolomic investigation revealed that PQQ enhances the biosynthesis of natural products. This enhancement is achieved by augmenting ATP levels as well as elevating the ratios of NADPH/NADP and NADH/NAD, which are vital cofactors in natural product biosynthesis. Additionally, PQQ boosts the supply of substrates, notably acetyl-CoA—a fundamental precursor for numerous natural products—by promoting the accumulation of triacylglycerol (TAG) stores within the initial 48 hours, followed by accelerating their catabolism concurrent with the onset of actinorhodin production.

Effects of the *pqq* operon on several *Streptomyces*

Since PQQ was shown to be effective in increasing actinorhodin production in *S. coelicolor*, we speculated that PQQ production may enhance natural product biosynthesis in other *Streptomyces* species. To explore this idea, we introduced the *pqq* gene cluster into 11 *Streptomyces* strains. These strains were selected from eight different clades, encompassing the taxonomic diversity of the genus (Supplementary Fig. 9). The changes in growth and metabolites production were measured. There were no observable changes in growth of the 11 strains in which the *pqq* gene cluster was introduced, implying that the *pqq* gene had a neutral metabolic burden (Supplementary Fig. 10).

Global changes in metabolism were measured using high-performance liquid chromatography with quadrupole time of flight (HPLC–Q-TOF) (Fig. 4a). This analysis showed increases in the abundance of 16,385 metabolites, which is 54% of all the measured metabolites in 11 strains (Fig. 4b, Supplementary Fig. 9 and Supplementary Table 5). More than 50% of the metabolites analysed increased on introduction of the *pqq* gene cluster in 8 out of the 11 strains (Fig. 4c and Supplementary Table 6) and in 6 of the 11 strains, several metabolites increased more than 100,000-fold after integration of the *pqq* operon. To assess the gross metabolomic differences caused by PQQ production, we performed a principal component analysis on strains that showed changes larger than 1,000-fold (Supplementary Fig. 11). This analysis showed effective clustering of replicates for each strain and clear segregation of strains with and without the *pqq* operon. Overall, these results showed that introduction of the *pqq* operon or the addition of PQQ molecules could induce significant metabolic changes in a wide range of *Streptomyces* strains.

pqq operon boosts natural product biosynthesis

The general effect of PQQ on known natural products was also investigated. For this purpose, the metabolites whose production increased on integration of the *pqq* gene cluster were annotated with the StreptomeDB chemical dictionary, as in ref. 24. The annotated metabolites were further confirmed by manual inspection of the corresponding spectra (Fig. 4a).

For the 11 strains, introduction of the *pqq* operon successfully improved production of 34 known natural products including antibacterials, antifungals, anticancer, antiparasitic and immunosuppressive molecules (Fig. 5a and Supplementary Table 4). We also noted that 70% of the natural products whose production increased were PKS or PKS-hybrid products, and that 85% of them use acyl-CoAs as precursors. This is consistent with our observation that the acyl-CoA pools are increased on incorporation of *pqq* operon in the model strain *S. coelicolor*.

In the following section, we elaborate on two examples that illustrate the effects of PQQ in natural product biosynthesis. Detailed information on the upregulated natural products found in this study is available in Supplementary Table 4 and Supplementary Data 7.

(1) *Streptomyces rapamycinicus* is a member of clade 12. This clade includes the largest genomes and biosynthetic repertoires among *Streptomyces*, and all its members have endogenous *pqq* operons. In this strain, all known polyketides detected were increased after integration of an extra copy of the *pqq* operon (Fig. 5a and Supplementary Table 4). For instance, rapamycin, a macrolide drug used as immunosuppressive agent after organ transplantation, increased by 12.5-fold (Fig. 5a); this is the greatest single increase in production of rapamycin reported in the literature (Supplementary Table 7). Nigericin, a potassium ionophore that triggers the NLRP3 inflammasome and is used in inflammation research, increased by 31.5-fold.

(2) In *Streptomyces albus*, introduction of the *pqq* operon significantly enhanced production of seven antifungal compounds derived from three independent endogenous BGCs, namely alteramide A (11.5-fold) and B (2.5-fold), antimycin A1 (16.9-fold), A2 (11.4-fold) and A3 (2.1-fold), and surugamide A (38.4-fold) and E (36.7-fold) (Fig. 5a and Supplementary Table 4). *S. albus* is a member of clade 8, which includes the smallest genomes within the genus and cannot produce PQQ naturally. *S. albus* is a popular host for heterologous expression of BGCs and only basal expression of its endogenous products has been reported[25,26].

In summary, our results showed that the introduction of the *pqq* operon can significantly enhance endogenous natural product production across the *Streptomyces* genus. In addition, supplementation of PQQ increased biosynthesis of natural products in ten out of 11 strains (Supplementary Figs. 3 and 12), indicating exogenous supplementation of PQQ can also be an effective way to natural product titre improvement.

pqq operon improves natural products in industrial strains

Industrial strains are usually heavily modified for optimal performance[27], thus additional modifications usually lead to only small increases in productivity[28]. Given the successful improvement of natural product production in various *Streptomyces* strains through the introduction of the *pqq* operon, we sought to explore whether PQQ could similarly enhance production in industrial antibiotic producer strains. Our investigation extended to two industrial *Actinobacteria*: *Actinosynnema* sp. XYF21, an ansamitocin-P3 overproducer, and *Micromonospora echinospora* Q11, an overproducer of gentamicin aminoglycosides. As show in Fig. 5b, expression of the *pqq* gene cluster significantly increased the efficiency of gentamicin production (measured as the mix of the four main gentamicins), resulting in a 58% increase. While introduction of the *pqq* operon in *Actinosynnema pretiosum* XYF21 resulted in a 40% increase in ansamitocin-P3 production

https://doi.org/10.1038/s42255-024-01024-9

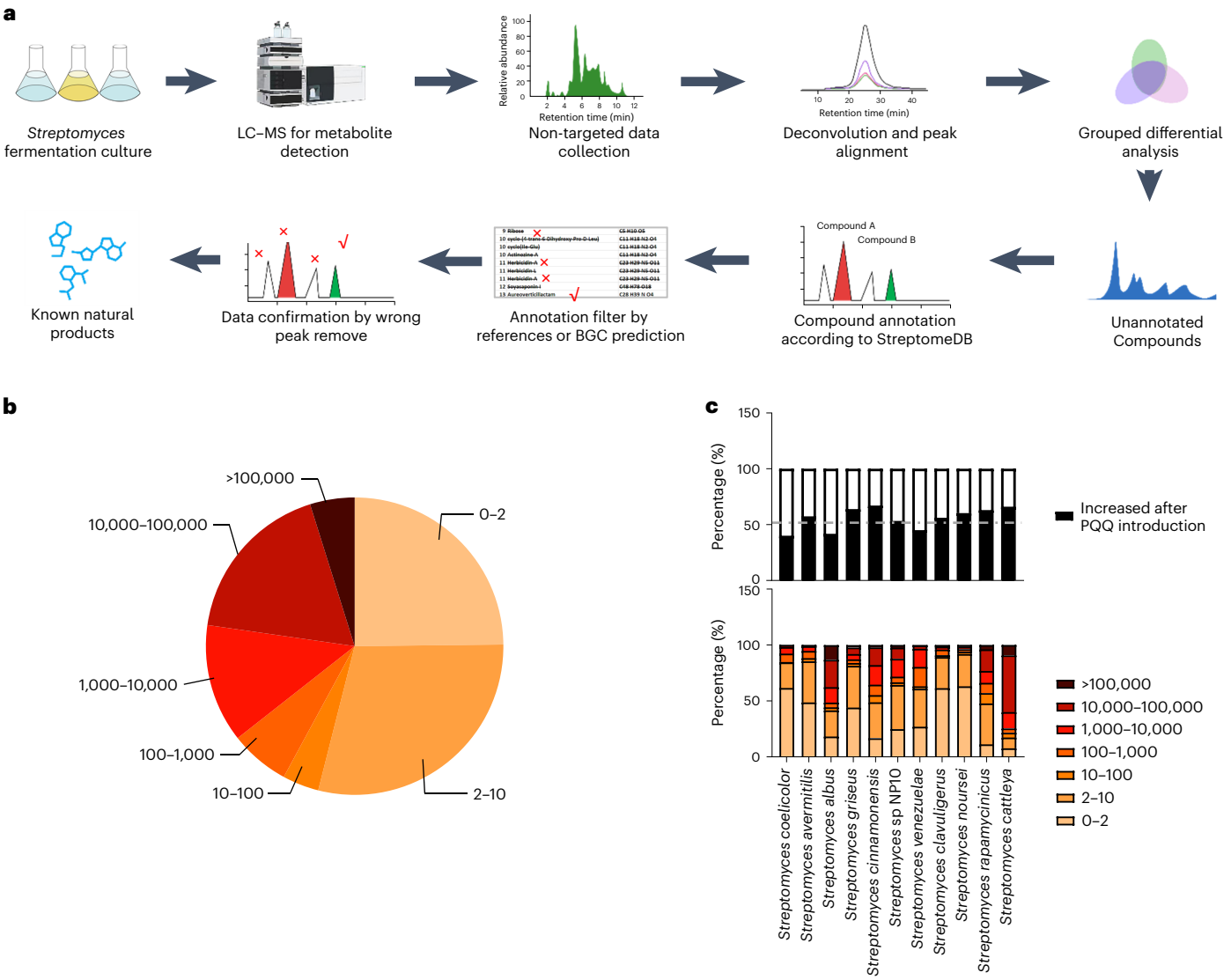

Fig. 4 | Introduction of the *pqq* gene cluster results in a positive impact on the metabolome of 11 *Streptomyces* strains. a, Overall experimental flow chart of metabolomics analysis. The upper panel is the process from cultivation to determining differentially changed metabolites. The lower panel is the process of compound annotation and confirmation. **b**, The fraction of compounds that showed various fold increases in concentrations. 1–2 indicates a fold increase of 1–2; 2–10, a fold increase of 2–10 and so on. **c**, The impact of incorporating the *pqq* gene cluster into 11 *Streptomyces*. The upper panel is the percentage of compounds that increased in concentration, whereas the lower panel is the percentage of compounds in each strain by the fold increase in concentration.

(Fig. 5c). These results further expand the potential application of PQQ as an enhancer of natural product biosynthesis, beyond polyketides, even in strains from other actinobacterial genera that are already optimized.

pqq operon induces the expression of silent BGCs

Since the introduction of the *pqq* operon enhanced production of known natural products, we wondered whether it can also activate silent BGCs. To explore this idea, we analysed the metabolomes of *S. rapamycinicus* and *S. griseus* in the search for unknown molecules. In *S. rapamycinicus*, integration of the *pqq* operon resulted in the appearance of a peak at a retention time of 6.65 min in the total ion chromatogram (Fig. 6a). The induced metabolite has a mass to charge ratio of 1,175.647 *m/z*, which corresponds to the formula $C_{62}H_{97}NO_{18}S$. This formula matches that of the polyketide mediomycin-A (mass error 3.2 ppm), a linear polyene polyketide with potent antifungal activity. Consistently, we then identified a BGC in the *S. rapamycinicus* genome matching the previously reported BGC for mediomycins in *S. mediocidicus* ATCC (American Type Culture Collection) 23936 (ref. 29) (Supplementary Fig. 13). We also found that this BGC is conserved among other strains in clade 12 (Supplementary Fig. 14).

In *S. griseus*, introduction of the *pqq* gene cluster resulted in the appearance of a peak with retention time of 7.33 min and an *m/z* of 303.1794 (Fig. 6b and Supplementary Fig. 15), which is consistent with the *m/z* value of the γ-butyrolactone SCB3. We confirmed this match using its MS2 fragmentation patterns (Supplementary Fig. 16). SCB3 is a variant of the γ-A-factor, a butyrolactone previously described in *S. griseus* featuring a ketone instead of a hydroxyl moiety at C-6. Until now, no reduced form of A-factor analogue (that is, SCB-like γ-butyrolactones) has been detected in *S. griseus*. The enzyme responsible for the reduction of the ketone group (*scbB*) in *S. coelicolor*, a producer of SCB3, is missing in the gene cluster of *S. griseus* (Supplementary Fig. 17). However, we found homologues of ScbB that may be involved in the generation of the SCB series molecules in *S. griseus* (Supplementary Table 8).

To obtain a wider perspective on the changes that may account for unidentified products in these two strains, we ranked their metabolites according to fold changes. We found that 71 and 226 metabolites in

https://doi.org/10.1038/s42255-024-01024-9

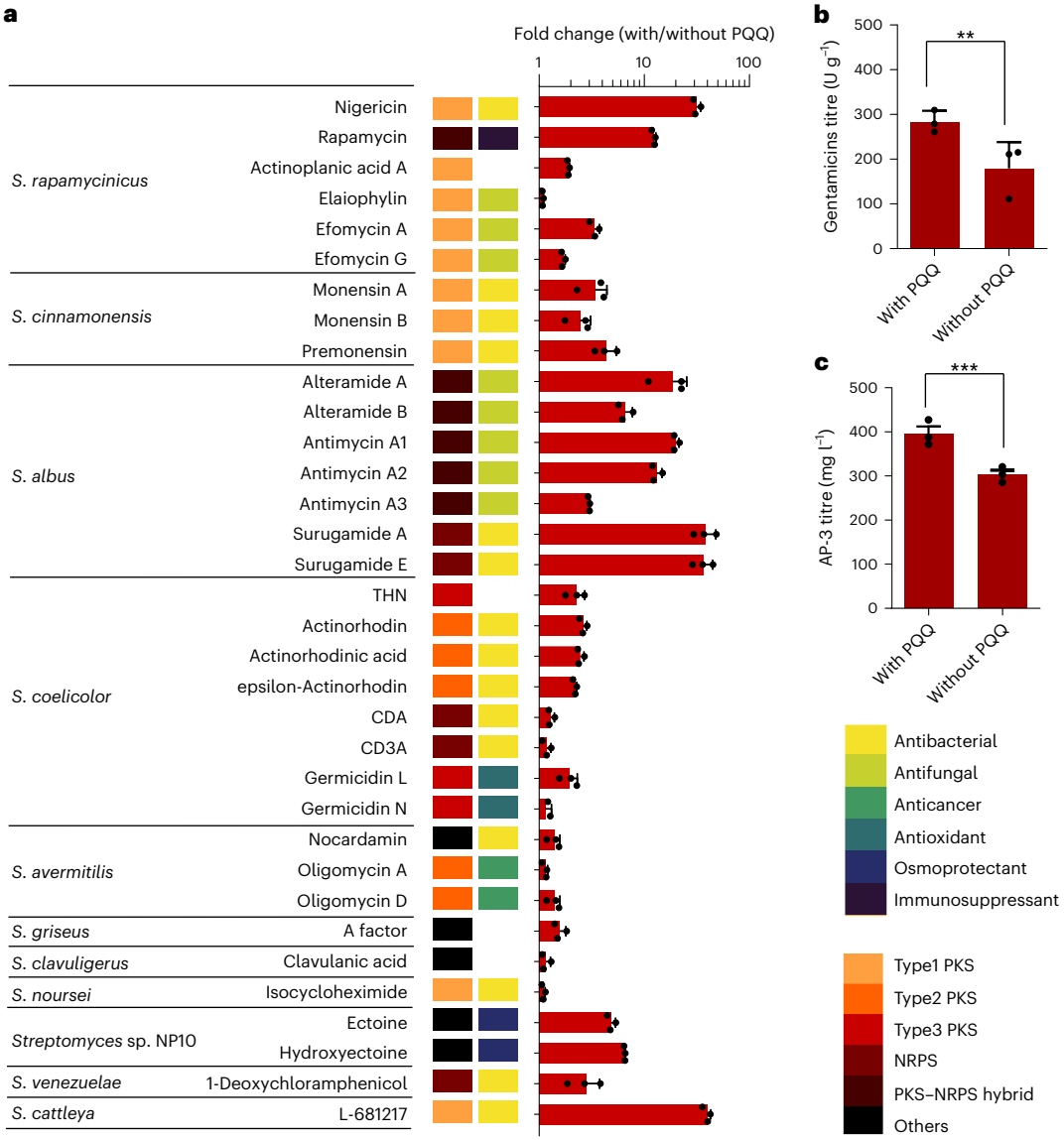

Fig. 5 | Introduction of *pqq* gene cluster resulted in increased production of known natural products. a, Introduction of *pqq* gene cluster increased production of 34 known natural products in 11 *Streptomyces* strains. The bioactivities and types of natural product are shown in the right panel next to the compound name ($n = 3$, three independent cultures analysed). THN, 1,3,6,8-tetrahydroxynaphthalene. **b**, Introduction of *pqq* gene cluster increased gentamicin production in gentamicin industrial strain; gentamicin production (U g^{-1}) was calculated as: peak area/dry weight (in g). Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). Multiple comparison significance was tested to $^{**}P < 0.05$, $^{***}P < 0.01$ by an unpaired two-sided Student's *t*-test. **c**, Introduction of *pqq* gene cluster increased ansamitocin-P3 (AP-3) production in an ansamitocin industrial strain. The ansamitocin titre (mg l^{-1}) was determined by comparing peak area with the standard curve. Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). Multiple comparison significance was tested to $^{**}P < 0.05$, $^{***}P < 0.01$ by an unpaired two-sided Student's *t*-test.

S. griseus and *S. rapamycinicus* went from not having any signal to intensities higher than 10^5 (Supplementary Tables 9 and 10), respectively. These results showed that introduction of the *pqq* gene cluster leads to the production of many compounds that could not be detected otherwise. There are 46 (65% *S. griseus*) and 132 (58% *S. rapamycinicus*) of these induced compounds that cannot be matched to those in current databases (Supplementary Tables 9 and 10), indicating the possibility of unidentified antibiotics generated by introducing the *pqq* gene cluster.

Besides the known compounds in *S. rapamycinicus*, we found increased transcription of key structural genes in unknown BGCs 4, 6, 23, 26, 27, 29, 35, 36, 38, 41, 46 and 47 by two- to 32-fold after *pqq* gene cluster introduction (Fig. 6c). These BGCs encode for PKSs, nonribosomal peptide synthetases (NRPSs) or terpenes that have not been previously characterized (Supplementary Table 11). We also found increased transcription of key structural genes of BGCs 5, 14, 16 and 24 in *S. griseus* (Fig. 6d), and these BGCs encode for melanin, siderophore, lanthipeptide and thiopeptide products (Supplementary Table 12). Upregulation of these BGCs is very likely to result in generation of previously uncharacterized natural products. Additionally, we conducted a bioactivity assay of the 11 strains with and without *pqq* cluster to assess the potential generation of unidentified bioactive compounds. As shown in Fig. 6e and Supplementary Table 13, introduction of the *pqq* gene cluster into *S. venezuelae*, *S. albus*, *S. rapamycinicus*, *Streptomyces* sp. NP10 and *S. avermitilis* all resulted in strains exhibiting distinguished bioactivities. Importantly, introducing *pqq* gene cluster to *S. venezuelae* results in bioactivities against *A. baumannii* ATCC 19606 and *K. pneumoniae* ATCC 13883, indicating the possibility of developing antibiotics and therapies against severe clinic infection caused by these bacteria.

https://doi.org/10.1038/s42255-024-01024-9

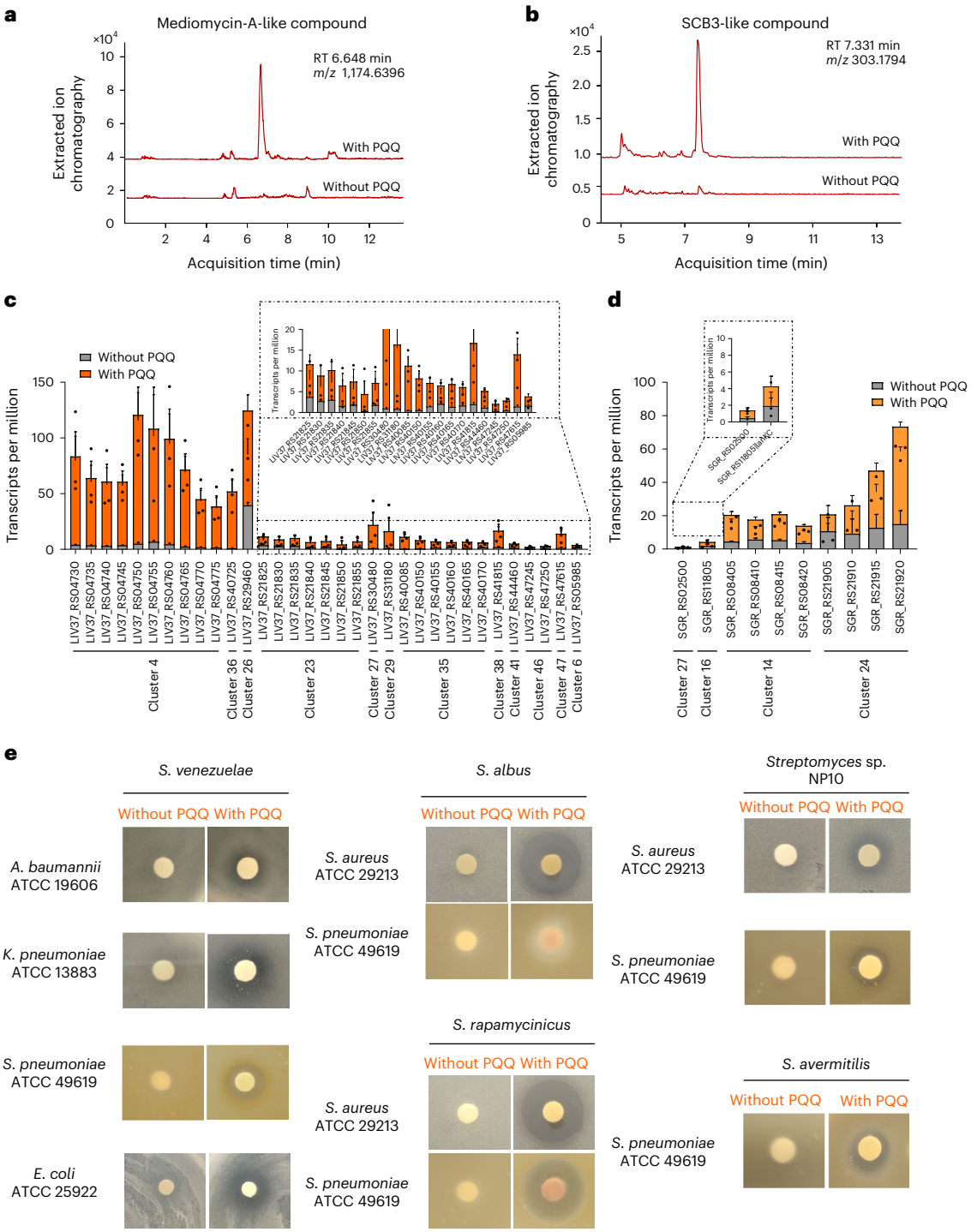

Fig. 6 | Natural products activated by introduction of the *pqq* BGC. **a**, Extracted ion chromatography showed a PQQ activated mediomycin-A-like compound in crude extracts of *S. rapamycinicus* before and after *pqq* cluster introduction. **b**, Extracted ion chromatography showed a PQQ activated SCB3-like compound from crude extracts of cultures of *S. griseus* before and after *pqq* cluster introduction. **c**,**d**, Changes in transcription of unknown BGCs increased with *pqq* gene cluster in *S. rapamycinicus* (**c**) and *S. griseus* (**d**). The cluster number was called according to antiSMASH cluster prediction with corresponding genome sequences (Supplementary Tables 9 and 10). Mean concentrations with error bars showing s.d. are plotted ($n = 3$, three biologically independent samples). **e**, Comparative bioactivity assays are shown for *Streptomyces* with and without *pqq* gene cluster. Without PQQ refers to a bioactivity assay using fermentation crude extracts by *Streptomyces* introducing the empty vector. With PQQ refers to a bioactivity assay using fermentation crude extracts by *Streptomyces* introducing the vector containing *pqq* gene cluster. RT, retention time.

Discussion

Streptomyces encodes the largest repertoire of natural product BGCs in bacteria[5]. Efforts have been made to engineer *Streptomyces*, but because of the complexity of natural product biosynthesis, only a handful of strategies have been developed[30].

Currently, only a few engineering strategies are commonly used in *Streptomyces*, and they are typically identified in one strain and then applied to others. For example, Chen et al. established a strategy using transfer RNA-Asp-AUC to circumvent inefficient wobble base-pairing to enhance actinorhodin and four other antibiotics[31]. Gessner et al.

expressed *bldA* in 15 *Streptomyces* strains, and seven showed changed HPLC profiles[32]. Introduction of *bldA* activated potential products and resulted in overproduction of eight known natural products. In this study, we used phylogenomic and pan-genomic approaches to identify functions that co-evolved with polyketide BGCs. We also developed a universal engineering strategy by using the *pqq* gene cluster to increase natural product biosynthesis. This strategy successfully improved the production of 34 natural products in 11 *Streptomyces* families, making it the most widely applicable engineering strategy developed in *Streptomyces* so far. It is to be noted that current approach diverges from previous strategies that focused on singular genetic modifications. Instead, this method simultaneously influences multiple pathways, which is a distinctive feature enabling the universal enhancement of natural product biosynthesis

Our study presents an approach for developing engineering strategies for *Streptomyces* and other species, as well as for other natural products such as non-ribosomal peptides or terpenes. Specifically, we identified *Streptomyces* strains such as *Streptomyces silaceus* ACCC 40021, *Streptomyces* sp. CFMR 7, *Kitasatospora albolonga* YIM 101047, *Streptomyces bingchenggensis* BCW1, *Streptomyces* sp. MUM8645 and *Streptomyces* sp. TLi053 as having more NRPS gene clusters than other strains, which could lead to strategies for NRPS engineering by identifying co-evolved genes. Additionally, *Streptomyces clavuligerus* ATCC 27064, *Streptomyces* sp. TLi053 and *S. hygroscopicus* XM201 encode more terpene gene clusters, which could also benefit from this approach. Our strategy can be extended to other natural product-producing bacteria and fungi with the help of massive sequencing data.

Furthermore, our study identified 597 genes from different metabolic pathways that co-evolved with polyketide BGCs, and we tested the *pqq* BGC as a proof of concept. Besides PQQ, other functions encoded within these genes may also be beneficial for polyketide or natural product production. For example, co-evolution of peptidase and/or protease may be involved in the degradation of misfolded proteins caused by ribosomal stress due to production of natural products; co-evolution of enzymes with metal related functions may be driven by the need for metal ligands needed by natural product biosynthetic enzymes. These observations open the door for future research.

In our study, phylogenomic analysis indicated that the *pqq* gene cluster evolved in parallel with PKS gene clusters, and the introduction of the *pqq* gene cluster increased biosynthesis of several polyketides in the 11 strains. We found that biosynthesis of other types of compound were also elevated. We reasoned that this effect is due to the interconnection between the central metabolic pathways that feed cofactors and precursors to the pathways for various natural products[33]. First, many natural products share common cofactors and precursors such as ATP, NADH, NADPH and acetyl-CoA. Second, the biosynthetic precursors from central metabolism are interconnected. For example, alanine, a precursor to peptides, and acetyl-CoA, a precursor to polyketides, are derived from pyruvate. Their pools may be affected simultaneously through the change in pyruvate, which will, in turn, alter the biosynthesis of downstream natural products. Finally, the biosynthetic pathways for various natural products may be controlled by global regulators, such that a change in one pathway may influence the expression of global regulators that alter production of other natural products[4]. Therefore, although the prevalence of the *pqq* gene cluster correlates directly with abundance of PKS gene clusters, the contribution of the *pqq g*ene cluster to polyketide production can be extended to other natural products through the shared and interconnected precursors and cofactors, as well as the influence of global regulators.

The PQQ-induced changes that we found in this study are consistent with effects on mitochondrial respiration caused by PQQ in mammalian cells that have been previously reported[34]. However, the mechanistic details behind these effects on natural product biosynthesis require further investigation. We found that the *pqq* gene cluster in *Streptomyces* is often associated with carbohydrate-active enzymes such as chitinases and with some dehydrogenases and catalases; we speculate that PQQ may also play a role in these catabolic pathways.

Before this study, the only case in which PQQ has been linked to natural products is in lankacidin biosynthesis. In this case the *pqq* gene cluster is located on the chromosome near the lankacidin BGC in in *S. rochei* 7434AN4. PQQ works as a cofactor for the quinoprotein dehydrogenase (*orf23*), which catalysers the dehydrogenation of the C23-C25 lactate moiety to pyruvate during lankacidin biosynthesis[22]. However, there are no quinoprotein dehydrogenase homologues that could use PQQ as a cofactor in *S. coelicolor* and other *Streptomyces*. The mechanism by which PQQ affects the intracellular redox state, whether through specific binding of unknown quinoproteins or by the means of its electron transfer nature, will be investigated in future.

In conclusion, we combined genomics, proteomics and metabolomics to evaluate the impact of the PKS co-evolved *pqq* gene cluster on natural product biosynthesis in *Streptomyces* genus. Our results indicate presence of the *pqq* operon correlates with biosynthetic proficiency in the *Streptomyces* genus and we showed that PQQ enhances the biosynthesis of natural products and activates BGCs for unidentified products across the genus *Streptomyces* and beyond. We believe that this effect can be exploited to discover bioactive natural products in the future.

Methods

Genomes database integration

We collected the genomes from a previously compiled database with 7,762 genomes retrieved from the GenBank. These sequences were processed with ad-hoc annotation and formatting tools[23]. To create a high-quality *Streptomyces* genome database, we first selected sequences assigned to the *Actinomycetota* phylum (formerly *Actinobacteria*). For this step, we parsed the taxonomic labels in our database against the genera listed in the entry for the *Actinobacteria* phylum at the National Center for Biotechnology Information (NCBI) taxonomy database (ID 201174). This selection step resulted in more than 5,000 genomes. Then, we selected for high-quality genomes by filtering-out assemblies with more than ten contigs. We reasoned that this cut off would leave high-quality genomes that may have one or more plasmids or chromosomes. Finally, we assessed the completeness of the selected genomes with BUSCO using the actinobacteria_class_odb10 reference lineage dataset[35]. We kept genomes with a minimum completeness of 85%. The final dataset included 720 genomes (Supplementary Data 1).

Taxonomic classification of *Streptomyces* strains

We calculated the core genome for the 720 selected strains. For this, we used the predicted proteomes of the 720 genomes to identify sets of conserved orthologues at different sequence identity cut offs using the BPGA pipeline[36]. We identified a set of 18 *Actinomycetota*-wide conserved protein sequences at a minimum sequence identity cut off of 20% (Supplementary Table 1). The set of 18 orthologues was then collected from each genome, and the amino acid sequences were aligned and concatenated to generate a super matrix with 6,517 characters for phylogenetic analysis using IQtree 2 (ref. 37) with the following parameters -m TEST, -bb 10000, alrt 10000. A substitution model was calculated for each partition in the super matrix and then a phylogenetic tree was calculated (Supplementary Fig. 1). Branch support was obtained from 10,000 ultrafast bootstrap replicates and SH-like approximate likelihood ratio test (SH-aLRT), which was consistent with bootstrap-based support, thus, only bootstrap support is reported herein. The tree showed excellent support for most nodes, and it is consistent with the known general taxonomy of the phylum. The tree features a clade with 201 taxa that we could safely assign to the Streptomycetaceae family (NCBI taxID 2062). This clade includes strains assigned to the genera *Streptomyces* (190 strains), *Kitasatospora* (eight strains) and *Streptacidiphilus* (three strains).

Within this group only a few genomes were misclassified: one *Streptomyces* strain classified as *Kitasatospora* and six *Kitasatospora* strains that were classified as *Streptomyces*; finally, one strain at the base of this clade (classified as *Streptomyces scabrisporus* NF3) did not group with the other genera and may belong to a different genus within the Streptomycetaceae family. We used the 201 genomes to construct a species tree for the *Streptomyces* genus using the same approach described above for the *Actinomycetota* species tree. For the *Streptomyces* tree, we identified a set of 619 conserved orthologues across the 201 genomes at a sequence identity cut off of 0.4. This set of orthologues was processed leading to a super matrix with 604 partitions (Supplementary Data 2) that was used to calculate a species tree that was rooted using the 17 non-*Streptomyces* strains as outgroup. This improved version is available at https://github.com/WeMakeMolecules/myCORASON. The *Actinomycetota* and *Streptomyces* species tree were constructed using code available at https://github.com/WeMakeMolecules/Core-to-Tree.

Functional analysis of conserved traits in *Streptomyces*

The number and type of BGCs for each strain was obtained using AntiSMASH[38]. To calculate the conserved set of orthologues of each of the 14 clades we used BPGA[36]. To identify the clade-specific conserved genes, we compared the clade-level cores with BPGA and extracted the genes that are uniquely conserved in clade 12 and their annotation. For the identification of *pqq* BGCs, we used PqqB (UniProt ID A0A0C6FBE1) as query for a search using Corason3, which has been modified from https://doi.org/10.1038/s41589-019-0400-9.

Point biserial correlations between KS abundance (continuous data) and genes uniquely conserved in clade 12 (dichotomous data) were calculated using a matrix of presence (recorded as 1) or absence (recorded as 0) of at least one orthologue of the 597 genes. For this, the amino acid sequences encoded in the 597 genes were used as query for a blastP search across the 203 genomes dataset. The search was performed with a cut off of 1×10^{-6}, the score cut off was adjusted depending on the length of the query automatically using script available at https://github.com/WeMakeMolecules/Core-to-Tree/raw/main/dichotomizer.pl. Point biserial correlations (r) were calculated using Microsoft Excel using the function =CORREL (dichotomous data, continuous data); t-test statistics were calculated using the formula $=r*\text{SQRT}(\text{sample size}-2/\text{SQRT}(1-r^2)$ and P values were obtained with the function =T.DIST.2 T(T, sample size −2), in all cases the simple size was 202 genomes.

Strains and growth conditions

Strains and plasmids used in this study are listed in Supplementary Table 14. *Escherichia coli* DH5α was used for plasmids construction, and ET12567 (pUZ8002) was used for conjugation. All *E. coli* strains was cultivated in Luria-Bertani medium with appropriate antibiotics at 37 °C.

For cultivation and harvesting spores, *S. coelicolor* M145, *S. albus* J1074, *S. cinnamonensis* ATCC 15413 (recently reclassified as *S. virginiae* ATCC 15413 (ref. 39)), *S.* sp. NP10 and *S. venezuelae* ISP5230 (ATCC 10712) were cultivated on SFM medium (20 g l^{-1} soybean powder, 20 g l^{-1} mannitol, 2% agar) plates. *S. avermitilis* NRRL 8165 (MA-4680), *S. clavuligerus* ATCC 27064, *S. noursei* ATCC 11455, *S. cattleya* NRRL 8057 were cultivated on YMG medium (4 g l^{-1} yeast extract, 10 g l^{-1} malt extract, 4 gl^{-1} glucose, 2% agar); *S. rapamycinicus* NRRL 5491 was cultivated on oatmeal medium (20 g l^{-1} oatmeal, 2% agar) and *S. griseus* DSM 40236 was cultivated on ISP-4 medium (37 g l^{-1} ISP-4 medium powder (BD DIFCO), 2% agar).

For conjugation, *S. coelicolor*, *S. albus*, *S. cinnamonensis*, *Streptomyces* sp. NP10, *S. venezuelae*, *S. avermitilis* and *S. rapamycinicus* were performed on SFM medium supplemented with 10 mM $MgCl_2$. *S. clavuligerus*, *S. noursei* and *S. cattleya* were performed on YMG medium supplement of 10 mM $MgCl_2$. *S. griseus* was performed on ISP-4 medium supplement of 10 mM $MgCl_2$.

Unless stated, cultivation of *Streptomyces* strains were performed according to the publications, with only a few modifications to accommodate to the present media component stock in our laboratory. For fermentation, strains were first inoculated on corresponding plates at 30 °C for 2 days, then 1 cm^2 medium was inoculated into seed medium for growth of mycelium, and then 10% seed medium was transferred into fermentation medium for another round of cultivation, accumulation for secondary metabolites. Both seed cultivation and production cultures were performed in 250 ml flask filled with spring steels circle. All medium was adjusted to pH 7.0–7.5 before sterilization.

Construction of plasmids and strains

The primers used in this study are listed in Supplementary Table 14. For genes in the 'cofactor' category and the truncated *pqq* cluster without *pqqA*, the open reading frames of corresponding genes were amplified from genome sequence of *S. hygroscopicus* XM201 and were introduced between the *Nde*I/*Eco*RI restriction sites of pLQ646 by Gibson Assembly. For the truncated *pqq* gene cluster without *pqqC*, the whole cassette was directly synthesized and inserted between the *Nde*I/*Eco*RI restriction sites at GenScript Co., Ltd. The plasmid pLQ646 was constructed previously by introduction of *kasOp** promoter in the multiple cloning site of pSET152 (ref. 40). The resulting plasmid from pXR-1 to pXR-7 was then introduced into *E. coli* ET12567 (pUZ8002) for conjugation with different *Streptomyces*. After conjugation, the exo-conjugants were confirmed both by apramycin resistance and PCR verification.

Metabolomic analysis

After fermentation, the culture from three biological replicates was diluted by mixing with two volumes of methanol and sonicated for 30 min. The samples were centrifuged at 13,800*g* for 5 min to remove insoluble particles and then filtered using a 0.22 μm filter membrane. The filtrate was injected to Agilent 1290 series HPLC (Agilent Technologies) equipped with an Agilent 6546 Accurate-Mass Q-TOF liquid chromatography with mass spectrometry (LC–MS) (Agilent Technologies) via an electrospray ionization source. LC conditions were as follows: A, water + 0.1% formic acid; B, acetonitrile + 0.1% formic acid; 0.4 ml min^{-1}. The data were collected using Agilent Data Acquisition software (v.10.1) (Agilent Technologies). Then the collected raw data were further processed by Agilent Profinder (v.10) (Agilent Technologies) for peak alignment and Agilent Mass Profiler Professional (v.15.0) (Agilent Technologies) for grouped differential analysis and compound annotation.

For compound annotation, the compounds were annotated using the dictionary of the StreptomeDB database covering most natural products identified in *Streptomyces* (so far) in Agilent Mass Profiler Professional (v.15.0) according to the molecular weight and predicted addiction. The annotated compound was embedded into Agilent Qualitative Analysis (v.0.0) (Agilent Technologies) for peak confirmation to make sure the difference from the target molecule is <5 ppm, and the shapeless peaks were also removed. The detailed information for known natural products is shown in Supplementary Fig. 18.

Proteomic analysis

Protein extracts were prepared as previously described with minor modifications[41]. To prepare protein extracts, strains were grown in YMG medium for 24 h and cells were collected. Cells were then resuspended in a solution containing 50 mM NH_4HCO_3 and 8 M urea, along with 0.5 mm glass beads (Biospec), and mechanically disrupted using a tissue grinding machine (ten times for 30 s each). The resulting cell debris was pelleted by centrifugation at 12,000*g* for 10 min at 4 °C. Protein concentration was determined using a BCA protein assay kit (Solarbio) and 100 μg of protein was used for digestion. The samples were first treated with 5 mM reducing dithiothreitol for 30 min at 56 °C, followed by 25 mM iodoacetamide for alkylation in the dark at

room temperature for 1 h. The samples were then diluted with 50 mM NH_4HCO_3 to a final urea concentration of 2 M and digested overnight with trypsin (Promega) at an enzyme-to-protein ratio of 1:50 (w/w) at 37 °C. The resulting tryptic peptides were desalted using C18 Sep-Pak cartridges (Welch), vacuum dried and stored at −80 °C before LC with tandem MS (LC–MS/MS) analysis.

To prepare for LC–MS/MS analysis, the vacuum dried samples were reconstituted in an aqueous solution containing 0.1% formic acid. Peptides were then loaded onto a 3 µm × 0.2 cm precolumn (P/N 164535, Thermo Scientific) and separated on a 2 µm × 15 cm capillary column (P/N 164534, Thermo Scientific) using the following HPLC gradient: 3–8% B in 5 min, 8–20% B in 65 min, 20–30% B in 20 min, 30–40% B in 10 min, 40–99% B in 5 min, held at 99% B for 5 min and then held at 3% B for 10 min (A, 0.1% formic acid in water; B, 0.1% formic acid in 80% acetonitrile) at a flow rate of 300 nl min^{-1}. The peptide samples were analysed using an Ultimate 3000 RSLC nano system (Thermo Scientific) coupled to a Q-Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in data-dependent mode with a full MS scan (350–2,000 *m/z*) at a resolution of 60,000 followed by higher-energy collisional dissociation fragmentation at a normalized collision energy of 28%. The MS2 automatic gain control target was set to 45 s.

The raw MS files were processed by MaxQuant (http://maxquant.org/, v.1.6.10.43). MS/MS spectra were searched against the *S. coelicolor* (strain ATCC BAA-471/A3(2)/M145) protein database (downloaded from UniProt, v.2002) using the Andromeda search engine embedded in MaxQuant. Both peptide and protein assignments were filtered to achieve a false discovery rate <1%. After searching, the reverse hits, contaminants and proteins only identified by one site were removed.

Transcriptome analysis

To perform transcriptome analysis, strains of *S. rapamycinicus* and *S. griseus* introduced *pqq* gene cluster as well as the empty vectors were cultured under normal fermentation condition for 12 h. The total RNA of each sample was extracted using Total RNA Extractor kit (catalogue no. B511311, Sangon) according to the manufacturer's protocol. The extracted RNA was then sent to the Sangon Biotech Co., Ltd for library construction and sequencing. Gene expression level was calculated by the DESeq2 R package (v.1.20.0), showing average values of three clones from each group.

Growth determination

Unless noted, the strains were cultivated in 24-well plates filled with 2.5 ml of corresponding fermentation medium and the cultivation was carried out in 30 °C at 800 rpm. For sampling, 50 µl of fermentation culture was collected every 12 h and used for growth determination.

For *S. albus* J1074, *S. virginiae* ATCC 15413, *S.* sp. NP10, *S. avermitilis* NRRL 8165 (MA-4680), *S. clavuligerus* ATCC 27064, *S. noursei* ATCC 11455, *S. cattleya* NRRL 8057 and *S. rapamycinicus* NRRL 5491, growth was determined by the diphenylamine colorimetric method following the protocol in ref. 42.

For *S. venezuelae* ISP5230 (ATCC 10712), growth was determined by absorption at $A_{600\,nm}$ using a microplate reader.

For *S. coelicolor* M145, growth was determined by weighing the cell dry weight. The strains were cultivated in 250 ml flasks, and 1 ml of fermentation culture was collected every 12 h and used for growth determination.

PQQ addition experiment

Streptomyces strains were cultivated in 24-well plates and each well contained 2.5 ml of the corresponding fermentation medium. The cultivation temperature was maintained at 30 °C at 800 rpm (ZQZY-85CH, Zhichu). PQQ (Macklin, P863945) was added to the fermentation cultures at different final concentrations. The concentrations were as follows: 0 nM, 0.4 nM, 400 nM, 40 µM, 200 µM and 400 µM. The supplementation time for *S. coelicolor*, *S. albus, S. avermitilis*, *Streptomyces* sp. NP10, *S. cattleya*, *S. clavuligerus* and *S. griseus* was 0 h of fermentation, for *S. noursei* was 24 h, for *S. cinnamonensis* was 48 h and for *S. venezuela* was 72 h. Each concentration was added to four replicates. For sampling, 300 µl of fermentation culture was collected corresponding to fermentation time and analysed by LC–MS (Q-TOF, Agilent Technologies).

Key metabolites determination

For determination of precursors and fatty acids, the mycelia were collected at different time points by centrifuge at 4 °C, 13,800*g* for 2 min followed by removal of the supernatant medium. To extract metabolites from samples, 400 µl of cold methanol and acetonitrile (1:1, v:v) extraction solvent was added to remove the protein and extract the metabolites, then adequately vortexed. For absolute quantification of the metabolites, stock solutions of stable-isotope internal standards were added to the extraction solvent simultaneously. The mixture was collected into a new centrifuge tube, and centrifuged at 13,800*g* for 20 min at 4 °C to collect the supernatant. The supernatant was dried in a vacuum centrifuge. For LC–MS analysis, the samples were redissolved in 100 µl of acetonitrile:water (1:1, v:v) solvent and centrifuged at 13,800*g* at 4 °C for 20 min, then the supernatant was injected. Analyses were performed using an UHPLC (1290 Infinity LC, Agilent Technologies) coupled to a QTRAP MS (AB 6500+, AB Sciex) in Shanghai Applied Protein Technology Co., Ltd. The metabolites were separated using hydrophilic interaction chromatography (Waters UPLC BEH Amide column, 2.1 × 100 mm, 1.7 µm) and C18 columns (Waters UPLC BEH C18x2.1 ×100 mm, 1.7 µm), and were quantified on a 6500+ QTRAP (AB SCIEX) in MRM mode, and in positive and negative switch mode.

For determination of TAG, mycelium of *S. coelicolor* M145 was collected at different time points from 12 to 96 h for the intracellular TAG analysis. Intracellular TAG levels were analysed using a commercialized kit (Solarbio, catalogue no. BC0625) following the manufacturer's instructions.

For determination of intracellular cofactors, mycelium of *S. coelicolor* M145 was collected at 12 h as this time point had shown the most enrichment concentration of these molecules. The mycelia were collected by centrifuge at 4 °C, 13,800*g* for 2 min followed by removal of the supernatant medium. The levels of ATP (ab113849), NADH/NAD^+ (ab65348) and NADPH/$NADP^+$ (ab65349) were analysed by commercialized kits (Abcam) following the manufacturer's instructions. For other cofactors, analyses were performed using an ultra-HPLC (UHPLC) (1290 Infinity LC, Agilent Technologies) coupled to a QTRAP MS (AB 6500+, AB Sciex) in Shanghai Applied Protein Technology Co., Ltd. The cofactors were separated using hydrophilic interaction chromatography (Waters UPLC BEH Amide column, 2.1 × 100 mm, 1.7 µm) and C18 columns (Waters UPLC BEH C18-2.1 ×100 mm, 1.7 µm), and were quantified on a 6500+ QTRAP (AB SCIEX) in MRM mode, and in positive and negative switch mode.

C^{13}-labelling experiments

During cultivation, the stable uniformly ^{13}C-labelled glucose (*U*-$^{13}C6$, MCE) was fed into the fermentation culture of *S. coelicolor* M145 at 24 h. Samples were collected from 48 to 96 h to detect the labelling fraction of actinorhodin. To prepare samples, the culture from three biological replicates was diluted by mixing with two volumes of methanol and sonicated for 30 min. The samples were centrifuged at 13,800*g* for 5 min to remove insoluble particles and then filtered using a 0.22 µm filter membrane. The filtrate was injected to Agilent 1290 series HPLC (Agilent Technologies) equipped with an Agilent 6546 Accurate-Mass Q-TOF LC–MS (Agilent Technologies) via an electrospray ionization source. LC conditions were as follows: A, water + 0.1% formic acid; B, acetonitrile + 0.1% formic acid; 0.4 ml min^{-1}. The data were collected using Agilent Data Acquisition software (v.10.1) (Agilent Technologies). Then the collected raw data were further processed by Agilent Qualitative Analysis (v.10) (Agilent Technologies). Actinorhodinic acid was

chosen as a representative for C^{13}-labelling analysis among the different forms of actinorhodin, as the producing curve was most close to the total actinorhodin changes determined by colorimetric analysis. The peaks different isotope forms were extracted by corresponding exact molecular weight in Agilent Qualitative Analysis (v.10) (Agilent Technologies).

The data were further processed and corrected for natural abundance according to the reported method in ref. 43. The unlabelled amount was calculated as the amount of the [M+0] fraction, and the labelled amount was calculated as the sum of all the other isotopic fractions.

Statistics and reproducibility

Statistical analysis was performed using Microsoft Excel software using a two-tailed *t*-test analysis of variance hypothesis. Significant differences are marked as $^{**}P < 0.05$ and $^{***}P < 0.01$. All data are presented as mean ± s.d. The number of biologically independent samples for each panel was three unless otherwise stated in the figure legends.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The transcriptomic data of *S. griseus* and *S. rapamycinicus* are available at NCBI GEO database (GSE256209). The mass spectrometry proteomics data are available at the ProteomeXchange Consortium via the iProX repository (PXD049454). Other data supporting the findings of this study are included in the published article and Supplementary Information. Requests for any additional information can be made to the corresponding authors. Source data are provided with this paper.

Code availability

The code of pan-genomic analysis was described in the Methods section of the paper. All the code is openly available in GitHub.

References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
2. Li, S., Li, Z., Pang, S., Xiang, W. & Wang, W. Coordinating precursor supply for pharmaceutical polyketide production in *Streptomyces*. *Curr. Opin. Biotechnol.* **69**, 26–34 (2021).
3. Wang, W. et al. Harnessing the intracellular triacylglycerols for titer improvement of polyketides in *Streptomyces*. *Nat. Biotechnol.* **38**, 76–83 (2020).
4. Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77**, 112–143 (2013).
5. Gavriilidou, A. et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **7**, 726–735 (2022).
6. Gao, L. et al. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **51**, 1044–1051 (2019).
7. de Vries, R. P. et al. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* **18**, 28 (2017).
8. Cruz-Morales, P. et al. Revisiting the evolution and taxonomy of Clostridia, a phylogenomic update. *Genome Biol. Evol.* **11**, 2035–2044 (2019).
9. Chandra, G. & Chater, K. F. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* **38**, 345–379 (2014).
10. Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* https://doi.org/10.1093/database/baaa062 (2020).
11. Bentley, S. D. et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147 (2002).
12. Kim, J. N. et al. Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. *J. Microbiol. Biotechnol.* **25**, 1599–1605 (2015).
13. Bu, Q. T. et al. Comprehensive dissection of dispensable genomic regions in *Streptomyces* based on comparative analysis approach. *Micro. Cell Fact.* **19**, 99 (2020).
14. Zhou, Z., Gu, J., Li, Y. Q. & Wang, Y. Genome plasticity and systems evolution in *Streptomyces*. *BMC Bioinf.* **13**, S8 (2012).
15. Belknap, K. C., Park, C. J., Barth, B. M. & Andam, C. P. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* **10**, 2003 (2020).
16. Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V. & Luzhetskyy, A. Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genom.* **15**, 97 (2014).
17. Chung, Y. H. et al. Comparative genomics reveals a remarkable biosynthetic potential of the *Streptomyces* phylogenetic lineage associated with rugose-ornamented spores. *mSystems* **6**, e0048921 (2021).
18. Jonscher, K. R., Chowanadisai, W. & Rucker, R. B. Pyrroloquinoline-quinone is more than an antioxidant: a vitamin-like accessory factor important in health and disease prevention. *Biomolecules* **11**, 1441 (2021).
19. Wagh, J., Shah, S., Bhandari, P., Archana, G. & Kumar, G. N. Heterologous expression of pyrroloquinoline quinone (PQQ) gene cluster confers mineral phosphate solubilization ability to *Herbaspirillum seropedicae* Z67. *Appl. Microbiol. Biotechnol.* **98**, 5117–5129 (2014).
20. Zhu, W. & Klinman, J. P. Biogenesis of the peptide-derived redox cofactor pyrroloquinoline quinone. *Curr. Opin. Chem. Biol.* **59**, 93–103 (2020).
21. Shen, Y. Q. et al. Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone. *Biochemistry* **51**, 2265–2275 (2012).
22. Yamauchi, Y. et al. Quinoprotein dehydrogenase functions at the final oxidation step of lankacidin biosynthesis in *Streptomyces rochei* 7434AN4. *J. Biosci. Bioeng.* **126**, 145–152 (2018).
23. Cruz-Morales, P. et al. Biosynthesis of polycyclopropanated high energy biofuels. *Joule* **6**, 1590–1605 (2022).
24. Moumbock, A. F. A. et al. StreptomeDB 3.0: an updated compendium of streptomycetes natural products. *Nucleic Acids Res.* **49**, D600–D604 (2021).
25. Olano, C. et al. Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074. *Microb. Biotechnol.* **7**, 242–256 (2014).
26. Xu, F., Nazari, B., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *J. Am. Chem. Soc.* **139**, 9203–9212 (2017).
27. Beganovic, S. et al. Systems biology of industrial oxytetracycline production in *Streptomyces rimosus*: the secrets of a mutagenized hyperproducer. *Microb. Cell Fact.* **22**, 222 (2023).
28. Abbate, E. et al. Optimizing the strain engineering process for industrial-scale production of bio-based molecules. *J. Ind. Microbiol. Biotechnol.* **50**, kuad025 (2023).
29. Sun, F., Xu, S., Jiang, F. & Liu, W. Genomic-driven discovery of an amidinohydrolase involved in the biosynthesis of mediomycin A. *Appl. Microbiol. Biotechnol.* **102**, 2225–2234 (2018).
30. Whitford, C. M., Cruz-Morales, P., Keasling, J. D. & Weber, T. The design-build-test-learn cycle for metabolic engineering of *Streptomycetes*. *Essays Biochem.* **65**, 261–275 (2021).

31. Chen, X. et al. A new bacterial tRNA enhances antibiotic production in *Streptomyces* by circumventing inefficient wobble base-pairing. *Nucleic Acids Res.* **50**, 7084–7096 (2022).
32. Gessner, A. et al. Changing biosynthetic profiles by expressing bldA in *Streptomyces* strains. *Chem. Bio. Chem.* **16**, 2244–2252 (2015).
33. Tsunematsu, Y. et al. Distinct mechanisms for spiro-carbon formation reveal biosynthetic pathway crosstalk. *Nat. Chem. Biol.* **9**, 818–825 (2013).
34. Saihara, K., Kamikubo, R., Ikemoto, K., Uchida, K. & Akagawa, M. Pyrroloquinoline quinone, a redox-active o-quinone, stimulates mitochondrial biogenesis by activating the SIRT1/PGC-1alpha signaling pathway. *Biochemistry* **56**, 6615–6625 (2017).
35. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
36. Chaudhari, N. M., Gupta, V. K. & Dutta, C. BPGA—an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **6**, 24373 (2016).
37. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
38. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
39. Komaki, H. & Tamura, T. Reclassification of *Streptomyces cinnamonensis* as a later heterotypic synonym of *Streptomyces virginiae*. *Int. J. Syst. Evol. Microbiol.* **71**, 004813 (2021).
40. Wang, X. R., Wang, R. F., Kang, Q. J. & Bai, L. Q. The antitumor agent ansamitocin P-3 binds to cell division protein FtsZ in *Actinosynnema pretiosum*. *Biomolecules* **10**, 699 (2020).
41. Wu, M. et al. Phosphoproteomics reveals novel targets and phosphoprotein networks in cell cycle mediated by Dsk1 kinase. *J. Proteome Res.* **19**, 1776–1787 (2020).
42. Zhao, Y., Xiang, S., Dai, X. & Yang, K. A simplified diphenylamine colorimetric method for growth quantification. *Appl. Microbiol. Biotechnol.* **97**, 5069–5077 (2013).
43. Nanchen, A., Fuhrer, T. & Sauer, U. *Determination of Metabolic Flux Ratios From 13C-Experiments and Gas Chromatography-Mass Spectrometry Data* (Humana, 2007).

Acknowledgements

We acknowledge the financial support provided by National Key R&D Program of China (to X.L., grant no. 2018YFA0903200), National Natural Science Foundation of China (to X.L., grant no. 32071421 and to H.D., grant no. 32201203), Shenzhen Science and Technology Program (to X.L., grant nos. ZDSYS20210623091810032 and RCYX20200714114736026), China Postdoctoral Science Foundation (to X.W., grant no. 31800023), Shenzhen Medical Research Fund (to X.W., grant no. D2301005), Shenzhen Science and Technology Program (to X.W., grant no. JCYJ20220531100207017), Shenzhen Bay Scholar Fellowship (to X.L. and X.T., grant no. 229100002), Novo Nordisk Foundation (to P.C.-M., grant no. CFB 2.0, NNF20CC0035580), which made this research possible. We thank J. Nikodinovic-Runic for kindly providing the strain *Streptomyces* sp. NP10, G. Liu, F. Ni and L. Zhou for help in industrial strains experiments, M. Wu for technique support on proteomics analysis, H. He and G. Zhang for project discussion and Z. Wei for meeting organization for this project.

Author contributions

X.W., P.C.-M., X.T., J.D.K. and X.L. conceived and supervised the project. X.W. and N.C. designed and performed the main experiments. P.C.-M. and S.A. performed the bioinformatics analysis. Y.S., L.B., J.W., Y.X. and X.F. designed and participated in the industry strain experiments. X.T. and Y.L. designed and participated in the bioassay test. B.Z. participated in the metabolomics analysis. Y.Z. participated in RNA sequencing data analysis. Z.L. and H.D. participated in fermentation data analysis. X.W., P.C.-M., J.D.K. and X.L. wrote and revised the paper.

Competing interests

J.D.K. has financial interests in Amyris, Ansa Biotechnologies, Apertor Pharma, Berkeley Yeast, Cyklos Materials, Demetrix, Lygos, Napigen, ResVita Bio and Zero Acre Farms. X.L. has financial interests in Demetrix and Synceres. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42255-024-01024-9.

Correspondence and requests for materials should be addressed to Xiaoyu Tang, Jay D. Keasling or Xiaozhou Luo.

Peer review information *Nature Metabolism* thanks Kenji Arakawa, Hyun Uk Kim and Lixin Zhang for their contribution to the peer review of this work. Primary Handling Editor: Alfredo Giménez-Cassina, in collaboration with the *Nature Metabolism* team.

nature portfolio

Corresponding author(s): Jay Keasling

Last updated by author(s): Feb 17, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Metabolomic analysis was performed on Agilent 1290 series HPLC (Agilent Technologies, Santa Clara, CA, USA) equipped with an Agilent 6546 Accurate-Mass Q-TOF LC-MS (Agilent Technologies, Santa Clara, CA, USA) via an electrospray ionization source (ESI). Data were collected using Agilent Data acquisition software (version 10.1) (Agilent Technologies, Santa Clara, CA, USA). Proteome analysis was performed on Ultimate™ 3000 RSLC nano system (Thermo Scientific, US) coupled to a Q-Exactive™ HF Orbitrap mass spectrometer (Thermo Fisher Scientific, US). Precursor determination were performed using an UHPLC (1290 Infinity LC, Agilent Technologies) coupled to a QTRAP MS (AB 6500+, AB Sciex) in Shanghai Applied Protein Technology Co., Ltd. The cofactors were quantified on 6500+ QTRAP (AB SCIEX).
Data analysis	Phylogenetic analysis was constructed using IQtree 2. Biosynthetic gene clusters for each strain was obtained using AntiSMASH. Point biserial correlations (r) were calculated using Microsoft Excel (version 2013). Metabolomic raw data wereprocessed by Agilent Profinder (version 10) (Agilent Technologies, Santa Clara, CA, USA) for peak alignment, and Agilent Mass Profiler Professional (version 15.0) (Agilent Technologies, Santa Clara, CA, USA) for grouped differential analysis and compound annotation. Proteome analysis was processed by MaxQuant (http://maxquant.org/, version 1.6.10.43). RNA-seq was calculated by DESeq2 R package (v.1.20.0). The peaks different isotope forms were extracted by corresponding exact molecular weight in Agilent Qualitative Analysis (version 10) (Agilent Technologies, Santa Clara, CA, USA).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The transcriptomic data of S.griseus and S.rapamycinicus are available at NCBI GEO database (GSE256209). The mass spectrometry proteomics data are available at the ProteomeXchange Consortium via the iProX repository (PXD049454). Other data supporting the findings of this study are included in the published article and supplementary information. Requests for any additional information can be made to the corresponding authors.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable.
Population characteristics	Not applicable.
Recruitment	Not applicable.
Ethics oversight	Not applicable.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Generally, three independent samples were adopted for data analysis due to relatively small deviation of microbial cells, which is also standard in our field that is deeded sufficient. For some special experiments, more sample sizes are adopted, as stated in the figure legends.
Data exclusions	No data were excluded.
Replication	All the biochemical and biological experiments were performed in three replications or more.
Randomization	Generally, it is not relevant to our study because the cultivation of microbial cells dose not require allocated experimental groups. In our study, colonies from transformations were randomly selected for further analysis.
Blinding	Investigators performing the experiment were blinded to data collection. The researchers performing analytics were blinded for strain information.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable.
Novel plant genotypes	Not applicable.
Authentication	Not applicable.