








Cite this: DOI: 10.1039/d6cb00137h

ArpA-like regulatory control and discovery of new biosynthetic genes driving scleric acid biosynthesis

Jingfan Zhang, ^a Lijiang Song, ^b Manuela Tosin, ^b Christophe Corre ^{*ab} and Fabrizio Alberti ^{*a}

Scleric acid is a natural product that was discovered by heterologous expression of the silent *scl* gene cluster from *Streptomyces sclerotialus* in *Streptomyces albidoflavus* and rational de-repression of the biosynthetic genes through deletion of the ArpA-like regulator *SclM4*. It is a weak inhibitor of the cancer-associated enzyme nicotinamide *N*-methyltransferase (NNMT). In this work we aimed to improve understanding of scleric acid biosynthesis and of gene regulation of the *scl* gene cluster. We performed CRISPR/Cas9-mediated deletion of the biosynthetic gene *sclG*, which confirmed the essential role of the corresponding ATP-grasp family enzyme in the biosynthesis of scleric acid. We also carried out whole-genome Nanopore long-read sequencing of *S. sclerotialus* to investigate the borders of the *scl* gene cluster, which led us to discover two additional genes, *sclC* and *sclK*, proposed to be involved in biosynthetic precursor supply. Transcriptomics analyses were performed to investigate the effect of the deletion of the ArpA-like gene *sclM4* on the expression of the *scl* biosynthetic genes, shedding light on the role of the corresponding transcriptional repressor. The present work improves understanding of scleric acid biosynthesis and provides new insights into the ArpA-like mediated regulation of biosynthetic gene expression in *Streptomyces* bacteria.

Received 20th April 2026,
Accepted 17th June 2026

DOI: 10.1039/d6cb00137h

rsc.li/rsc-chembio

Introduction

Secondary metabolites display enormous structural diversity, but the biosynthetic principles for many of them are highly conserved.¹ As a result, within each biosynthetic class (e.g. polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), ribosomally synthesised and post-translationally modified peptides (RiPPs)), many of the core and tailoring enzymes have highly conserved amino acid sequences.^{1,2} This significant sequence similarity facilitates their identification based solely on nucleotide sequences, enabling easier prediction and annotation of biosynthetic pathways.

The *scl* gene cluster was originally discovered in *Streptomyces sclerotialus* NRRL ISP-5269 through a multigene BLAST and ClusterTools search for homologous sequences of the *mmfR-mmfLHP* methylenomycin (Mm) regulatory cassette from *Streptomyces coelicolor* A3.³ MmyR and MmfR are both paralogues of ArpA with 32% amino acid identity, whereas MmfL is homologous to AfsA. However, the strain containing only MmfL lacks the ability to produce methylenomycin furan (MMF) signalling molecules, indicating that MmfH and MmfP are also significant

in the biosynthesis of MMFs.^{4,5} This was confirmed by further mutagenesis analysis.⁴ The *scl* cluster contains orthologues of all five genes as well as an additional LysR-family transcriptional regulator and two divergent sets of biosynthetic genes (Fig. 1a).³

The product of the *scl* gene cluster was characterised to be the novel molecule scleric acid, which was discovered using a combination of transformation-associated recombination (TAR) cloning and rational de-repression of the *scl* gene cluster by *sclM4* knockout. Scleric acid showed moderate antibacterial activity against *Mycobacterium tuberculosis* (H37Rv) and inhibitory activity against the cancer-associated enzyme nicotinamide *N*-methyltransferase (NNMT) with IC₅₀ of 178.0 μM (1-methylnicotinamide (MNAN) product detection) and 186.6 μM (*S*-adenosylhomocysteine (SAH) co-product detection).³ The two IC₅₀ values are within 5% of each other, indicating that scleric acid inhibits NNMT consistently, regardless of which reaction product is being monitored.

The involvement of *sclQ1-4*, *sclN* and *sclA* in scleric acid biosynthesis was confirmed through additional rounds of CRISPR/Cas9-mediated knockouts, alongside detection of the intermediate *L*-proline-oxyacetic acid in selected strains, which led to a proposed biosynthetic pathway (Fig. 1b).³ Three main building blocks were identified: a glycolic acid unit biosynthesised by *SclQ1-4*, an *L*-proline residue activated by *SclN*, which carries out a condensation reaction with glycolic acid, a benzoic acid unit biosynthesised by *SclADI*. The benzoic acid unit is

^a School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK.
E-mail: C.Corre@warwick.ac.uk, F.Alberti@warwick.ac.uk

^b Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK



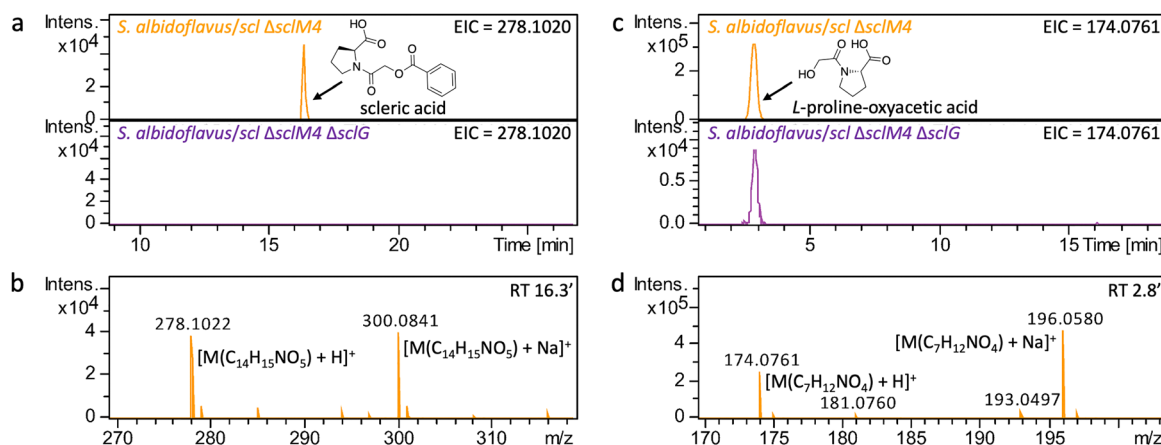


Fig. 2 LC-HRMS detection of metabolites produced in *S. albidoflavus/scl ΔsclM4* (orange) and *S. albidoflavus/scl ΔsclM4ΔsclG* (purple). (a) Extracted ion chromatograms (EICs) in positive mode for m/z 278.1020 are shown highlighting scleric acid accumulation in *S. albidoflavus/scl ΔsclM4* but not in *S. albidoflavus/scl ΔsclM4ΔsclG*. (b) Mass spectrum in positive mode of scleric acid. (c) EICs in positive mode for m/z 174.0761 are shown highlighting L-proline-oxyacetic acid accumulation in both *S. albidoflavus/scl ΔsclM4* and *S. albidoflavus/scl ΔsclM4ΔsclG*. (d) Mass spectrum in positive mode of L-proline-oxyacetic acid.

A high-quality assembled and polished genome was produced, resulting in two contigs, and with a BUSCO score of 99.5% (Fig. S2 and S3). The genome was deposited at DDBJ/ENA/GenBank under the accession JBF0HP000000000. Annotations of the assembled and polished *S. sclerotialis* NRRL ISP-5269 genome against the Nucleotide collection (nr/nt) database were performed for the general features of the chromosome sequence (Table S2).

Using antiSMASH bacterial version 8.0.4,^{7,8} we were able to identify 36 biosynthetic gene clusters (BGCs) for secondary metabolites in the newly assembled *S. sclerotialis* NRRL ISP-5269 genome, including the *scl* BGC (Fig. S4 and Table S3). Nearly 1/3 of the *S. sclerotialis* NRRL ISP-5269 clusters are predicted to produce antibiotic-like molecules. Most of the BGCs are predicted to encode for NRPS, PKS or NRPS/PKS hybrids, followed by terpene synthases, RiPPs and siderophores. It is interesting that four BGCs are proposed to encode for siderophore biosynthesis, which implies that *S. sclerotialis* NRRL ISP-5269 is likely to undergo strong selective pressure in low iron environments.⁹ There are also a few regions containing ectoine, alkaloid and saccharide type BGCs. In conclusion, *S. sclerotialis* NRRL ISP-5269 has the potential to produce a plethora of natural products with different biological functions that are still to be explored.

Border definition of the *scl* gene cluster and proposed role of *sclC* and *sclK*

The *scl* BGC identified from the antiSMASH analysis of the assembled and polished genome sequence revealed the presence of additional genes outside of the original *scl* gene cluster characterised in ref. 3 (Fig. 1a). Genes adjacent to the *scl* BGC were further investigated, and their amino acid sequences were analysed through BlastP for homology and putative function investigation (Table S4).¹⁰ Two genes downstream of *sclE* – *sclC* and *sclK* – were predicted to be part of the *scl* gene cluster and contribute to the biosynthesis of scleric acid precursors based on their putative functions (Tables S5 and S6).

SclC is proposed to be a chorismate synthase involved in the biosynthesis of benzoic acid, along with *SclADI*. *SclA* is a proposed anthranilate synthase, *SclD* is a proposed DAHP synthase and *SclI* is a proposed isochorismatase.³ These four enzymes are proposed to work closely together to provide benzoic acid (Fig. 1b).^{11–13}

One of the main microbial pathways to produce L-proline is through sequential catalysis by ProB (glutamate 5-kinase), ProA (glutamate 5-phosphate reductase) and ProC (pyrroline-5-carboxylate reductase).^{14,15} Within the *S. sclerotialis* NRRL ISP-5269 genome, two genes were found to be orthologues of *proA*, and two genes were found to be orthologues of *proB*, including *sclK* within the *scl* gene cluster, and one gene was found to be an orthologue of *proC*. This further suggests that L-proline is produced in *S. sclerotialis* NRRL ISP-5269 via this pathway and that *SclK* is likely to be part of the *scl* gene cluster and to be involved in the biosynthesis of L-proline that is then used as a building block for scleric acid (Fig. 1b).

S. albidoflavus/scl ΔsclM4 appeared to be able to produce scleric acid without the need for *sclC* and *sclK*, since these two genes were not included in the original heterologously-expressed *scl* cluster. This is likely due to the expression of chorismate synthase and glutamate 5-kinase in the heterologous host *S. albidoflavus*, which complement the lack of these two genes in the *scl* heterologously expressed gene cluster.³ Indeed, BlastP searches confirmed the presence of *sclC* and *sclK* homologues in the *Streptomyces albidoflavus* genome (Table S6).

Knockout of *sclM4* leads to transcriptional changes in the *scl* gene cluster

Transcriptional changes in the *scl* gene cluster following inactivation of *sclM4* via a 20-bp deletion were investigated to assess how the biosynthetic genes and the regulatory genes are controlled by *SclM4*. The newly discovered genes *sclC* and *sclK* were not included in the transcriptional study due to their absence from the plasmid originally integrated in *S. albidoflavus*. Transcript



coverage analysis was also performed to find the promoter region(s) and thus the operon structure of the *scl* gene cluster.

RNA sequencing of *S. albidoflavus/scl* Δ *sclM4* and *S. albidoflavus/scl* was performed and the normalised counts were collected, compared (see Tables S7 and S8 for the comparison of the genes belonging to the *scl* gene cluster) and visualised (Fig. S5 and Fig. 3). Upon knocking out *sclM4*, all the genes taking part in the biosynthetic pathway of scleric acid appeared to be upregulated, which is in agreement with the observed production of scleric acid in *S. albidoflavus/scl* Δ *sclM4*.³

In contrast, *sclM5* and *sclE* did not appear to be affected by the deletion of *sclM4*. The Integrative Genomics Viewer (IGV) visualization of *sclM5* suggests that the RNAseq reads available for *SclM5* do not cover the full gene sequence in both the *S. albidoflavus/scl* Δ *sclM4* strain and the *S. albidoflavus/scl* strain (Fig. S6). This could be a pseudo-truncation arising from sequencing artifacts associated with high-GC RNA. The lack of upregulation of *sclM5* in the *S. albidoflavus/scl* Δ *sclM4* strain may be due to the 20-bp deletion of *sclM4* that is present in this strain, which can be observed from the visualised transcriptome map of *sclM4* (Fig. S7). The 20-bp deletion may be responsible for decreasing *sclM5* gene expression driven by the promoter located upstream of *sclM2*, which is known to direct the expression of the whole *mmfLHP* regulatory cassette in *S. coelicolor*.⁴

The transporter *sclE* belongs to the major facilitator superfamily (MFS), which uses the proton-motive force in the transport

process.¹⁶ The expression levels of *sclE* remained consistently high both before and after the deletion of *sclM4* (Table S7). Thus, it is possible that it is not affected by *SclM4*, yet it can still contribute to scleric acid secretion.

SclM1 is homologous to the *S. coelicolor* *MmfR* (56% identity/73% similarity) and *SclM4* is homologous to *MmyR* (41% identity/56% similarity). *MmfR* and *MmyR* are well studied ArpA-family repressors that were confirmed to each act as a repressor. Each of *MmfR* and *MmyR* homodimers are proposed to bind independently to the same DNA motif, the MARE (MMF-autoregulator responsive elements) sequence.^{4,17,18} However, only *MmfR* is sensitive to the signalling molecule MMF.⁴ Due to the homology of *SclM1/MmfR* and *SclM4/MmyR*, *SclM1* and *SclM4* are hypothesised to bind to the same ARE motif. Operon determination in *Streptomyces* is complicated by the prevalence of leaderless transcripts. Nonetheless, by focusing on intergenic regions which are present regardless of transcript architecture, multiple expectation maximizations for motif elicitation (MEME) enrichment analysis was performed on the *scl* gene cluster to identify the proposed ArpA-family repressor binding site for *SclM1* and *SclM4* (Fig. 3b and Table S9).¹⁸ The location of the four ARE sites and the six associated start codons are listed (Table S9). Promoter regions near these ARE motifs were all confirmed by visualization of the RNA sequencing reads on IGV (Fig. S8), further confirming the proposed *SclM4* binding sites.

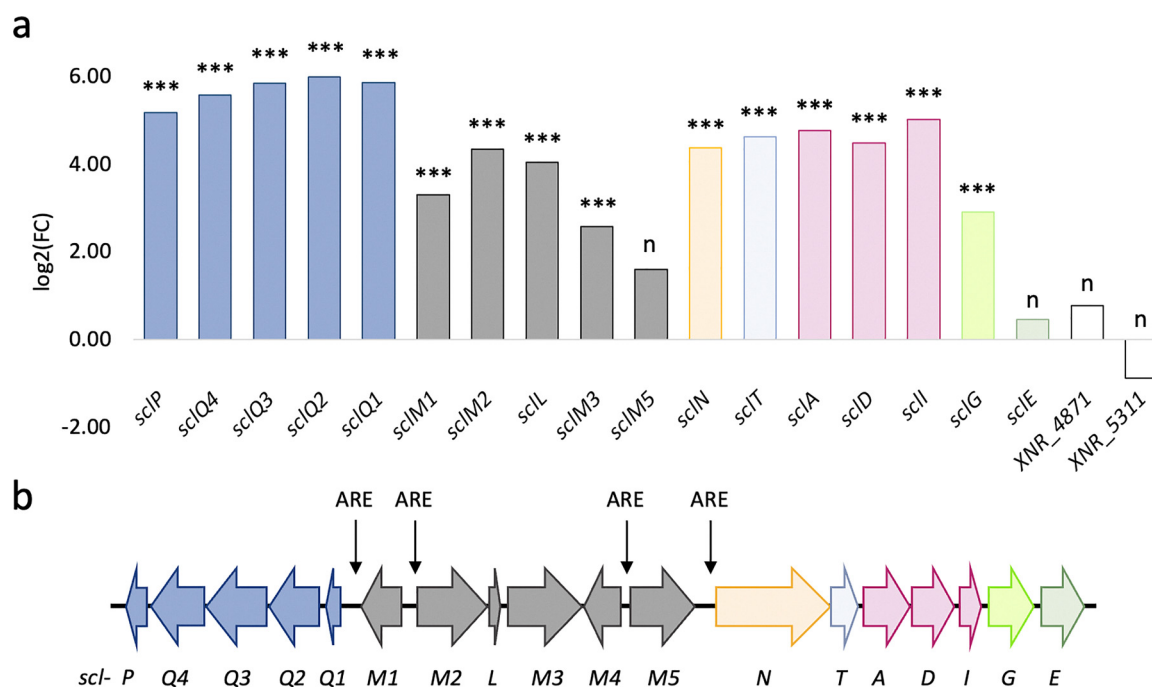


Fig. 3 (a) Log₂(FoldChange(FC)) of normalised gene counts from *S. albidoflavus/scl* Δ *sclM4* compared to normalised gene counts from *S. albidoflavus/scl*. The biosynthetic genes (*sclP*-*Q1*, *sclN*-*G*) were upregulated upon deletion of *sclM4*. *XNR_4871* and *XNR_5311* are two reference genes used for comparison. The log₂(FC) of the two reference genes between *S. albidoflavus/scl* Δ *sclM4* and *S. albidoflavus/scl* have small *P*-adj but their |log₂(FC)| falls within the cutoff of 2, therefore the transcription does not change significantly. Precise normalised counts, log₂(FC) value, *P*-value and *P*-adj can be found in Tables S7 and S8. *P*-adj higher than 0.05 is marked with (n). *P*-adj lower than 0.001 is flagged with (***). *n* = 3. (b) The *scl* biosynthetic gene cluster heterologously expressed in *S. albidoflavus/scl*. The arrows point to the proposed autoregulatory elements (AREs) in the gene cluster, which are locations proposed to be ArpA-family regulator binding sequences.



Discussion

Streptomyces are one of the most important sources of medicines and agrochemicals due to their ability to produce a wealth of bioactive natural products, such as antibiotics, herbicides, anti-cancer agents and immunosuppressants, through complex secondary metabolic biosynthesis pathways.⁹ In this study, we set out to investigate the biosynthesis of scleric acid and the regulation of its biosynthetic gene expression. The involvement of SclG in scleric acid biosynthesis was investigated. SclG is proposed to be an ATP-grasp domain-containing enzyme that activates benzoic acid and catalyses a condensation reaction between the activated benzoic acid and L-proline-oxyacetic acid to form scleric acid.³ In this study, *sclG* was knocked out from the scleric acid producing strain *S. albidoflavus/scl ΔsclM4*, leading to production of scleric acid to cease. L-Proline-oxyacetic acid was still detected from the metabolite extracts of the *S. albidoflavus/scl ΔsclM4 ΔsclG* strain, indicating that SclG does not take part in the biosynthesis of L-proline-oxyacetic acid but is needed in the last step of scleric acid biosynthesis.

Prior to the present work, the genome of *S. sclerotialis* NRRL ISP-5269 had been sequenced using Illumina technology, resulting in 92 contigs. In this work, the genome of *S. sclerotialis* was sequenced *via* Nanopore long-read technology, resulting in 2 contigs and a complete chromosome. Two genes, *sclC* (encoding a chorismate synthase) and *sclK* (encoding a glutamate 5-kinase), adjacent to *sclE*, were found to likely be part of the scleric acid BGC. They are proposed to be involved in the biosynthesis of building blocks (benzoic acid and L-proline, respectively) of scleric acid, and their absence in the scleric acid-producing heterologous strain³ was proposed to be complemented by genes with homologous functions present in the *S. albidoflavus* genome. Chorismate synthase and glutamate 5-kinase are typically associated with primary metabolism. Chorismate is often the obligatory precursor for benzoic acid, and housing a chorismate synthase (*i.e.* SclC) within the BGC would ensure a localised supply of chorismate for benzoic acid production. Glutamate 5-kinase (SclK) is proposed to catalyse the first committed step in L-proline biosynthesis. The duplication of primary metabolic enzymes into BGCs ensures the precursor supply and enables pathway self-sufficiency for secondary metabolism, and also helps maintain cell functionality during stress such as secondary metabolites production or nutrient limitation.¹⁹

We next turned our attention to the regulatory cassette SclM1-5 from the scleric acid gene cluster, which is homologous to the well-studied regulatory cassette (MmfR/MmyR/MmLHP) of the methylenomycin biosynthetic gene cluster in *S. coelicolor*.³ The knockout of *mmyR* is known to trigger the production of Mm.⁴ However, no studies were performed on the effects of the deletion of MmyR, or of its homologue SclM4, on the transcription of the biosynthetic genes from their respective BGCs. In this study, comparative RNA sequencing analysis was carried out between *S. albidoflavus/scl ΔsclM4* and *S. albidoflavus/scl*. Generally, it is assumed that the disruption of the AfsA homologues (*e.g.* MmLHP and SclM2/3/5) leads to reduced or ceased production of the target natural products, whereas mutation of the ArpA

homologues (*e.g.* MmyR/SclM4) leads to the opposite effect.²⁰ For example, the inactivation of *afsA* homologues *shbA1*, *shbA2* or *shbA3* in *S. hygrosopicus* 5008 decreased the biosynthesis of validamycin whereas the deletion of *arpA* homologues *shbR1* or *shbR3* increased the production of validamycin.²¹ This hypothesis was confirmed in our study, where the knockout of *sclM4* led to the upregulation of *sclM2* and *sclM3*, and further resulted in the initiation of scleric acid production. However, exceptions have also been observed, such as the AfsA/ArpA-like system SbbA/SbbR in *S. bingchenggensis*, in which the inactivation of the signal molecule synthase gene *sbbA* resulted in an increased production of the final secondary metabolite, whereas mutation of *sbbR* led to reduced production of the final product.²⁰

It emerged that upon *sclM4* knockout, all the biosynthetic genes from the *scl* gene cluster were upregulated. The gene *sclM5*, involved in the biosynthesis of signalling molecules, was not significantly upregulated, which may be due to a biological sequencing error or to the upstream 20-bp deletion made within the *sclM4* sequence in *S. albidoflavus/scl ΔsclM4*³ affecting the expression of *sclM5*. The transporter *sclE* was also not upregulated, though still expressed at high levels in both strains.

Conclusions

In this study, we improved understanding of scleric acid biosynthesis and of gene regulation in the *scl* gene cluster. We identified genes putatively involved in precursor supply at the edge of the *scl* gene cluster. We also showed that the gene cluster-specific ArpA-like transcriptional regulator can influence the expression of both scleric acid biosynthetic genes and signalling molecule biosynthesis genes.

Materials and methods

Culturing conditions

Streptomyces strains were grown in Soy Flour Mannitol (SFM) (20 g soya flour, 20 g mannitol, 20 g agar, tap water to a final volume of 1000 mL) at 30 °C.

sgRNA design

The nucleotide sequence of *S. sclerotialis* NRRL ISP-5269 was uploaded onto the CRISPY-web tool.²² The CRISPY-web tool then generated sgRNAs based on the knockout's target gene, which included a 20 nt protospacer and a 3 nt PAM sequence. Because the *scl* gene cluster is heterologously expressed in *S. albidoflavus* J1074, the result from the CRISPY-web tool was manually aligned against the *S. albidoflavus* J1074 genome using BLAST, to avoid off-target effects on the host genome.

The sgRNA close to the centre of the target gene was chosen. Its DNA oligonucleotides were then ordered from Integrated DNA Technologies, with BbsI-compatible overhangs for the Golden Gate assembly (5'-ACGC for forward and 5'-AAAC for reverse).



pCm2 plasmid construction

Construction of pCm2-sgRNA-HRs plasmid able to perform HDR knockouts *in vivo* involved two steps. Firstly, Golden Gate assembly was performed to insert the protospacer into the sgRNA cassette of the plasmid. Secondly, Gibson assembly was performed to insert the HR arms downstream of the sgRNA cassette.²³ The linearisation of pCm2-sgRNA vectors for Gibson assembly was performed with XbaI (NEB).

Intergeneric conjugation of plasmids from *E. coli* to *Streptomyces*

Plasmid was transformed into *E. coli* ET12567/pUZ8002 electro-competent cells *via* electroporation. The *E. coli* ET12567/pUZ8002 plasmid mixtures were grown to an OD₆₀₀ of 0.4. Whilst *E. coli* was growing, spores from the target *S. albidoflavus* strain were harvested and resuspended in 1 mL of 2 × YT broth (16 g tryptone, 10 g yeast extract, 5 g NaCl, dH₂O for a final volume of 1000 mL). They were preserved at 4 °C before the OD₆₀₀ of *E. coli* reached 0.4. Once the OD₆₀₀ of *E. coli* ET12567/pUZ8002 reached 0.4, they were harvested by centrifugation and washed in LB twice to remove residual antibiotics. The *E. coli* cells were then resuspended in 1 mL LB. When *E. coli* was ready, *Streptomyces* spores were heat shocked at 50 °C for 10 minutes for germination and left at room temperature to cool.

500 µL of *E. coli* ET12567/pUZ8002 cells containing the target plasmid and 500 µL of *Streptomyces* spores were mixed and centrifuged at 4000 rpm for 2 minutes. 900 µL of the supernatant was removed and the *Streptomyces/E. coli* mixture was resuspended in the remaining 100 µL by gently pipetting up and down. Both the ten-fold dilution and the neat mixture were then spread on SFM agar plates containing 10 mM MgCl₂. After an overnight (16–20 hours) incubation at 30 °C, the conjugation plates were overlaid with 1 mL sterile solution of 0.5 mg nalidixic acid and 1 mg kanamycin. The plates were then incubated at 30 °C for 2–5 days until exconjugants appeared. Exconjugants were further cultured on SFM plates containing 50 µg mL⁻¹ kanamycin and 25 µg mL⁻¹ nalidixic acid. The plates were incubated for an additional 4–6 days until single colonies appeared. Single colonies were sub-cultured onto SFM plates containing 50 µg mL⁻¹ kanamycin.

Screening for clones containing the expected plasmid using colony PCR

Amplification of the target knock-out regions was performed using colony PCR. The primers used were pCm2-*sclG* forward and pCm2-*sclG* reverse (Table S10).

The PCR results were then analysed by gel electrophoresis in a 1% (w/v) agarose gel in 1X TAE buffer containing 1X GelRed. The gel was run at 100 V for 60 minutes. It was then visualized and analysed. Six of the potential knockout strains were then picked and their genomic DNA was extracted using MP Biomedicals™ FastDNA™ SPIN Kit for Soil. PCR amplification was carried out using primers pCm2-*sclG* *scl* FF and pCm2-*sclG* *scl* RR (Table S10). The PCR was then run using gel electrophoresis in a 1% (w/v) agarose gel in 1X TAE buffer containing 1X

GelRed. The gel was run at 100 V for 60 minutes. Amplified gDNA fragments were extracted from the gel using a GeneJET Gel Extraction Kit. Deletion was then confirmed through Sanger sequencing.

Metabolite extraction and analysis

Streptomyces strains were cultured on 33 mL of supplemented minimal media solid (SMMS), prepared according to Strauch *et al.*²⁴ After ten days of growth, metabolites were extracted with 50 mL ethyl acetate. Five drops of HCl were added to 50 mL of ethyl acetate and the agar culture mixture to acidify the extract. The mixture was stirred at room temperature for ten minutes and then anhydrous magnesium sulfate was added to absorb the remaining water. The mixture was filtered and then dried *via* a SpeedVac (using a low boiling point setting). The dried metabolite extract was then dissolved in a 50:50 acetonitrile:water mixture. LC-HRMS analyses were carried out with 5 µL of prepared extracts injected through a reverse phase column (Zorbax Eclipse Plus C18, size 2.1 × 100 mm, particle size 1.8 µm) connected to a Dionex 3000RS UHPLC coupled to a Bruker Ultra High Resolution (UHR) Q-TOF MS MaXis II mass spectrometer with an electrospray source. Sodium formate (10 mM) was used for internal calibration and a *m/z* scan range of 50–2500 was used. The mobile phase was a gradient mixture of acetonitrile and formic acid in water following the elution programme reported in Table S11.

DNA purification

The MagAttract HMW DNA Kit (Qiagen) was used for extraction and purification of high-molecular weight genomic DNA of *S. sclerotialis* NRRL ISP-5269, following the manufacturer's instructions.

Nanopore sequencing

A MinION Flow Cell (FLO-MIN112), native barcoding expansion (EXP-NBD104), native barcoding kit (EXP-NBD114), and protocol SQK-LSK109 from ONT were used to perform Nanopore sequencing of the genome of *S. sclerotialis* NRRL ISP-5269, following the manufacturer's instructions.

Genome sequencing and assembly

The basecalling process was carried out on the raw data using Guppy (Guppy version: 5.0.7 + 2332e8d65).²⁵ Assembly of the genome was performed *via* flye (flye version: 2.8.2-b1689).²⁶ The assembled data was polished using minimap2 (version: 2.11-r797),²⁷ racon (v1.4.20)²⁸ and Medaka (version: 1.2.1).²⁹ A quality check of the polished data was then carried out using BUSCO (v5.0.0).³⁰ The structural and functional annotation was performed with PROKKA (1.14.6).³¹

Total RNA extraction from *S. albidoflavus*

The Monarch Total RNA Miniprep Kit (New England Biolabs) was used to purify RNA from *S. albidoflavus/scl* and *S. albidoflavus/scl ΔsclM4*, following the manufacturer's instructions.



Whole transcriptome sequencing and analysis

Extracted total RNA was sequenced by Genewiz Azenta using Illumina NovaSeq 2 × 150 bp sequencing, 10M read pairs, value package.

Galaxy (<https://usegalaxy.org>, Galaxy Version 1.14.6 + galaxy1) was used for whole transcriptomics analysis.³² Trim Galore! was performed for trimming adapters from sequence data.³³ FastQC was then used for quality control.³⁴ The sequence of the heterologously expressed plasmid pCAP03-*scl* was added at the end of the *S. albidoflavus* genome (GenBankCP04370) sequence to reflect insertion of the *scl* gene cluster in the heterologous host, and the combined sequence was used as the reference genome. Bowtie2 was used to align the paired-end, trimmed RNA sequencing reads with the reference genome, with very sensitive local (–very-sensitive-local) parameter presets.^{35,36} FeatureCounts was performed to align the alignments from Bowtie2 with annotated reference genome, with both multi-mapping and multi-overlapping features included (–M –O).³⁷ DESeq2 was finally used to normalise the counts of each gene from each sample for the final data analysis.³⁸

Author contributions

J. Z., F. A. and C. C. were responsible for project conceptualisation, experimental design and analysis, with contributions from M. T. J. Z. performed experimental work, analysed and interpreted the data. L. S. performed mass spectrometry analysis. J. Z. wrote the initial manuscript. F.A. reviewed and edited the manuscript with contributions from C. C. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no competing interests to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: Fig. S1–S8 and Tables S1–S11, oligonucleotides sequences, DNA/RNA sequencing data, gel electrophoresis images, homology search results, bioinformatics data, LC-HRMS conditions. The genome sequence was deposited at DDBJ/ENA/GenBank under the accession JBF0HP000000000. Raw RNA sequencing data was deposited in the NCBI sequence read archive (SRA) under BioProject accession PRJNA1474967 and SRA accessions SRX33758471-6. See DOI: <https://doi.org/10.1039/d6cb00137h>.

Acknowledgements

J. Z. was supported by a scholarship from the China Scholarship Council [202008060166] and the University of Warwick. F. A. was supported by a UKRI Future Leaders Fellowship [MR/V022334/1]. L. S. acknowledges funding from EPSRC [EP/P030572/1] for LC-HRMS equipment. The authors further

acknowledge a UK Synthetic Biology Research Centre grant from EPSRC and BBSRC [BB/M017982/1] that funded vacuum concentrator equipment. We thank Huimin Zhao for providing the pCRISPomyces-2 plasmid (Addgene plasmid # 61737). We thank Vincent Poon for the identification of the proposed ArpA-family repressor binding site on the *scl* gene cluster for ScLM1 and ScLM4 using MEME enrichment analysis. We thank Jack Weaver for assistance running Guppy on the assembled whole genome sequence.

References

- 1 N. Ziemert, M. Alanjary and T. Weber, The evolution of genome mining in microbes – a review, *Nat. Prod. Rep.*, 2016, **33**(8), 988–1005.
- 2 N. Lee, S. Hwang, J. Kim, S. Cho, B. Palsson and B.-K. Cho, Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 1548–1556.
- 3 F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin and C. Corre, Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery, *Chem. Sci.*, 2019, **10**(2), 453–463.
- 4 S. O'Rourke, A. Wietzorrek, K. Fowler, C. Corre, G. L. Challis and K. F. Chater, Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*, *Mol. Microbiol.*, 2009, **71**(3), 763–778.
- 5 S. Zhou, N. R. Malet, L. Song, C. Corre and G. L. Challis, MmFL catalyses formation of a phosphorylated butenolide intermediate in methylenomycin furan biosynthesis, *Chem. Commun.*, 2020, **56**, 14443–14446.
- 6 R. E. Cobb, Y. Wang and H. Zhao, High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system, *ACS Synth. Biol.*, 2015, **4**(6), 723–728.
- 7 K. Blin, S. Shaw, L. Vader, J. Szenei, L. Reitz Zachary and E. Augustijn Hannah, *et al.*, antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation, *Nucleic Acids Res.*, 2025, **53**(W1), W32–W38.
- 8 M. Zdouc Mitja, K. Blin, L. L. Louwen Nico, J. Navarro, C. Loureiro and D. Bader Chantal, *et al.*, MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration, *Nucleic Acids Res.*, 2024, **53**(D1), D678–D690.
- 9 S. D. Bentley, K. F. Chater, A. M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson and K. D. James, *et al.*, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, 2002, **417**(6885), 141–147.
- 10 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, Basic local alignment search tool, *J. Mol. Biol.*, 1990, **215**(3), 403–410.
- 11 A. R. Alcántara, M. J. Hernaiz and J. V. Sinisterra, 3.28 - Biocatalyzed Production of Fine Chemicals, in *Comprehensive*



- Biotechnology*, ed. M. Moo-Young, Academic Press, Burlington, 2nd edn, 2011, pp. 309–31.
- 12 S. Wang, D. Liu, M. Bilal, W. Wang and X. Zhang, Uncovering the Role of PhzC as DAHP Synthase in Shikimate Pathway of *Pseudomonas chlororaphis* HT66, *Biology*, 2022, **11**(1), 86.
 - 13 R. Widhalm Joshua and N. Dudareva, A Familiar Ring to It: Biosynthesis of Plant Benzoic Acids, *Mol. Plant*, 2015, **8**(1), 83–97.
 - 14 Z. Zhang, P. Liu, W. Su, H. Zhang, W. Xu and X. Chu, Metabolic engineering strategy for synthesizing trans-4-hydroxy-L-proline in microorganisms, *Microb. Cell Fact.*, 2021, **20**(1), 87.
 - 15 L. T. Smith, Characterization of a gamma-glutamyl kinase from *Escherichia coli* that confers proline overproduction and osmotic tolerance, *J. Bacteriol.*, 1985, **164**(3), 1088–1093.
 - 16 P. K. Chang, J. Yu and J. H. Yu, aflT, a MFS transporter-encoding gene located in the aflatoxin gene cluster, does not have a significant role in aflatoxin secretion, *Fungal Genet. Biol.*, 2004, **41**(10), 911–920.
 - 17 S. Zhou, *Antibiotic Biosynthesis and its Transcriptional Regulation in Streptomyces Bacteria*, University of Warwick, 2016.
 - 18 S. Zhou, H. Bhukya, N. Malet, P. J. Harrison, D. Rea and M. J. Belousoff, *et al.*, Molecular basis for control of antibiotic production by a bacterial hormone, *Nature*, 2021, **590**(7846), 463–467.
 - 19 J. K. Schniete, P. Cruz-Morales, N. Selem-Mojica, L. T. Fernández-Martínez, I. S. Hunter and F. Barona-Gómez, *et al.*, Expanding Primary Metabolism Helps Generate the Metabolic Robustness To Facilitate Antibiotic Biosynthesis in *Streptomyces*, *mBio*, 2018, **9**(1), e02283–17.
 - 20 H. He, L. Ye, C. Li, H. Wang, X. Guo and X. Wang, *et al.*, SbbR/SbbA, an Important ArpA/AfsA-Like System, Regulates Milbemycin Production in *Streptomyces bingchenggensis*, *Front. Microbiol.*, 2018, **9**, 1064.
 - 21 G.-Y. Tan, Y. Peng, C. Lu, L. Bai and J.-J. Zhong, Engineering validamycin production by tandem deletion of γ -butyrolactone receptor genes in *Streptomyces hygroscopicus* 5008, *Metab. Eng.*, 2015, **28**, 74–81.
 - 22 K. Blin, L. E. Pedersen, T. Weber and S. Y. Lee, CRISPy-web: An online resource to design sgRNAs for CRISPR applications, *Synth. Syst. Biotechnol.*, 2016, **1**(2), 118–121.
 - 23 A. Chhun and F. Alberti, CRISPR/Cas9-Based Methods for Inactivating Actinobacterial Biosynthetic Genes and Elucidating Function, in *Engineering Natural Product Biosynthesis: Methods and Protocols*, ed. E. Skellam, Springer US, New York, NY, 2022, pp. 201–222.
 - 24 E. Strauch, E. Takano, H. A. Baylts and M. J. Bibb, The stringent response in *Streptomyces coelicolor* A3(2), *Mol. Microbiol.*, 1991, **5**(2), 289–298.
 - 25 Technologies ON, Guppy protocol 2018. Available from: <https://nanoporetech.com/document/Guppy-protocol>.
 - 26 M. Kolmogorov, J. Yuan, Y. Lin and P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs, *Nat. Biotechnol.*, 2019, **37**(5), 540–546.
 - 27 H. Li, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, 2018, **34**(18), 3094–3100.
 - 28 R. Vaser, I. Sović, N. Nagarajan and M. Šikić, Fast and accurate de novo genome assembly from long uncorrected reads, *Genome Res.*, 2017, **27**(5), 737–746.
 - 29 Ltd. ONT. nanoporetech/medaka 2024. Available from: <https://github.com/nanoporetech/medaka>.
 - 30 M. Manni, M. R. Berkeley, M. Seppey, F. A. Simão and E. M. Zdobnov, BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes, *Mol. Biol. Evol.*, 2021, **38**(10), 4647–4654.
 - 31 T. Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics*, 2014, **30**(14), 2068–2069.
 - 32 T. G. Community, The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update, *Nucleic Acids Res.*, 2024, **52**(W1), W83–W94.
 - 33 F. Krueger, F. James, P. Ewels, *et al.*, FelixKrueger/TrimGalore: v0.6.10 - add default decompression path 2023 [updated 02/02/2023. Available from: <https://zenodo.org/records/7598955>.
 - 34 Andrews Snd. FastQC A Quality Control tool for High Throughput Sequence Data. 2023 [updated 01-03-23. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 - 35 B. Langmead, C. Trapnell, M. Pop and S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol.*, 2009, **10**(3), R25.
 - 36 B. Langmead and S. L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods*, 2012, **9**(4), 357–359.
 - 37 Y. Liao, G. K. Smyth and W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, 2013, **30**(7), 923–930.
 - 38 M. I. Love, W. Huber and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.*, 2014, **15**(12), 550.

