Supporting information

Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycins in *Streptomyces roseosporus*

Wei-Ting Liu[1], Roland D. Kersten[2], Yu-Liang Yang[3], Bradley S. Moore[2,3], Pieter C. Dorrestein[1,2,3.*]

1) Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, USA.

2) Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA.

3) Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, California, USA.

* Corresponding author: Pieter C. Dorrestein (Email: pdorrest@ucsd.edu)

Material and Methods

Strains and materials

Streptomyces roseosporus NRRL 11379 and *Streptomyces roseosporus* NRRL 15998 was acquired from the Broad institute, MIT/Harvard, MA, USA. *Staphylococcus aureus* UAMS-1 and *Staphylococcus epidermidis* ATCC 35984 were acquired from Michael Fischbach from University of California at San Francisco. All chemicals used for ISP2 media were purchased from Sigma-Aldrich. Organic solvents were purchased from J. T. Backer.

Bacterial growth for the zone-of-inhibition and time course experiments

Lawns of *S. aureus* UAMS-1 and *S. epidermidis* ATCC 35984 were created by mixing 150 µl exponentially growing cells (OD_{600} = 0.4-0.6) with 10 ml 1.5% ISP2 agar at temperature below 60^{o}C and pouring the mixture onto 100 O.D. x 25 mm Petri dishes (Fisherbrand). 5 µl of *S. roseosporus* starter culture were spotted on top of the lawn and allowed to dry. The plates were then incubated at 28°C. After 36 hours incubation, an inhibition zone was observed in both *S. aureus* and *S. epidermidis* lawn. A rectangular section containing the colonies was transferred to a MALDI target plate and then subjected to MALDI-IMS analysis. For time courses, 5 µl of *S. roseosporus* starter culture were spotted on ISP2 agar plates without lawns. At defined time intervals 1cm to 1cm area centering the colony was excised and extracted using MeOH and then analyzed by MALDI-TOF MS.

MALDI-IMS

Sample preparation for MALDI-IMS experiments was performed as described previously[1]. After culturing, a rectangular section containing the colonies was transferred to a MALDI MSP 96 anchor plate. After taking a photograph, universal MALDI matrix (Sigma-Aldrich) was sprinkled on top of the culture using a 20 µm sieve method, and was dried in a 37 ^{o}C oven for three hours. The detailed instrument parameters for collecting image data were described in reference 1. Briefly, the sample was subjected to Microflex Bruker Daltonics MALDI-TOF MS for imaging MS acquisition with scan range from 400-5000 *m/z* and the data was analyzed using the FlexImaging 2.0 software. The sample was run in positive reflectron mode, with 800 µm laser intervals in XY.

Purification of arylomycins (compound 1-3)

20 ISP2 agar plates were inoculated with *S. roseosporus* starter culture by 4 parallel streaks. The plates were incubated for 10 d at 28 °C. The agar was sliced into small pieces and then put in a 50 ml centrifuge tube and covered with equal amount of Milli-Q water and n-butanol for 12 h at 28 °C, shaken at 225 rpm. The n-butanol layer were taken out using transfer pipette, dried with rotoevaporator, re-dissolved and then fractioned via size exclusion using a Sephadex LH-20 column (25 cm x 30 mm) using MeOH at a flow rate of 0.5 mL/min. Each fraction (0.5 ml) was analyzed by MALDI MS and the fractions contained compounds **1-3** were collected and further purified by HPLC (C-18, 25 cm x 10 mm) with a gradient running from 40% solvent B to 95% solvent B in 25 minutes with flow rate 2mL/min. Solvent A is H_2O containing 0.1% TFA; solvent B is MeCN containing 0.1% TFA (aq). Purified compound **1** and **2** were lyophilized and stored at -80ºC before using for bioassay and structural elucidation. The yield for compound **1** and **2** were ~10 µg each per plate.

MALDI-TOF MS analysis for the time course experiment and MS-guided purification.

MALDI-TOF MS was used to detect target peptides in crude extracts and fractions of gel filtration and HPLC. The sample was mixed 1:1 with a saturated solution of Universal MALDI matrix in 78 % acetonitrile containing 0.1 % TFA and spotted on a Bruker MSP 96 anchor plate. The sample was dried and inserted into the Microflex mass spectrometer (Bruker Daltonics). Mass spectra were obtained with the FlexControl method as used for MALDI-imaging and a single spot acquisition of 80 shots. Single spot MALDI-TOF MS data was analyzed by FlexAnalysis software.

General LTQ/FT-ICR MS and MS/MS procedure

For the IT and FT MS data acquisition, each compound was dissolved in spray solvent 50:50 MeOH/H_2O containing 1% formic acid, and underwent nano-electrospray ionization on a biversa nanomate (Advion Biosystems, Ithaca, NY) using a back pressure of 0.3-0.5 p.s.i. and the spray voltage of 1.4 -1.5 kV. MS and MS/MS spectra were acquired on a 6.42 T Finnigan LTQ-FTICR MS or a Finnigan LTQ-MS (Thermo-Electron Corporation, San Jose, CA) running Tune Plus software version 1.0 and Xcalibur software version 1.4 SR1. The instrument was first auto-tuned on the *m/z* value of the ion to be fragmented. Then, the ions were isolated by the linear ion trap and fragmented by collision induced dissociation (CID) (isolation window: 3 *m/z*; collision energy: 30).

NMR measurement

200 µg compound **1** and **2** were dissolved in 50 µL of CD_3OD for NMR acquisition. ^{1}H-NMR spectra were recorded on Bruker Avance III 600 MHz NMR with 1.7 mm Micro-CryoProbe at 298 K, with standard pulse sequences provided by Bruker. The data was analyzed using the Topspin 2.1 software.

Spot assay

Lawns of *S. aureus* UAMS-1 and *S. epidermidis* ATCC 35984 were created by mixing 150 µl exponentially growing cells (OD_{600} = 0.4-0.6) with 10 ml 1.5% ISP2 agar and pouring the mixture onto 100 O.D. x 25 mm Petri dishes (Fisherbrand). 5 µl of 5% DMSO, or purified arylomycin A4 (compound **2**) and daptomycin A21978C2 dissolved in 5% DMSO were spotted on top of the lawn and allowed to dry. The plates were then incubated at 28°C for 1-2 days.

Phylogenetic analysis of C domains

Following the analysis done by Rausch *et al.*[2] and Imker *et al.*[3], a subset of condensation domain sequences were collected to represent the six condensation families (heterocyclization (*H*), epimerization (*E*), dual condensation/epimerization (*duet*), condensation of L amino acids to L amino acids (*L2L*), and condensation of D amino acids to L amino acids (*D2L*), and starter (*S*)). Sequences were annotated with the accession number and C domain type and then aligned using ClustalW2. The resulting unrooted guide-tree was downloaded and modified in PowerPoint for publication purposes.

Sequence definition of the arylomycin gene cluster in *S. roseosporus*

The sequence borders of the arylomycin gene cluster in *S. roseosporus* NRRL 15998 were defined from the NP.searcher-based location and manual assignment of the biosynthetic genes by Frameplot 4.0 and BLAST analysis as 6226494-6258861 bp in NCBI reference genome DS999644.1. As this supercontig was a draft genome, it was further compared with the homolog genome region (6272054-6304406 bp) in *S. roseosporus* NRRL 11379 (Figure S6) whose genome sequence was semi-finished in March 2010 (Broad Institute database, supercontig 4.1). The arylomycin gene cluster regions were 100% identical in sequenced parts. 1.6% of the NRRL 15998 gene cluster were gaps ((a)1C, 25998 bp, (b) 25441-25988 bp) and 1 insertion (1A, 9767 bp). Given the high similarity of the 2 arylomycin gene cluster regions, the NRRL 11379 sequence was used for gene annotation (Table S1) and proposal of a scheme of arylomycin biosynthesis (Scheme 1).

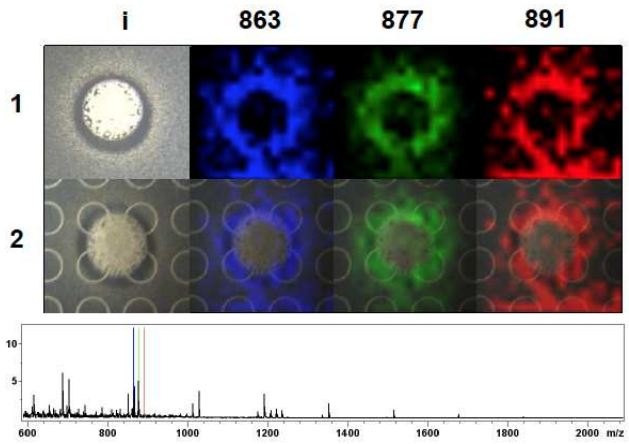


Figure S1: IMS of *S. roseosporus* spotted on top of a *S. aureus* lawn. (1) ion distribution of compound **1**-**3** (863, 877, 891) observed in IMS. 1i is a photograph showing *S. roseosporus* inhibit *Staphylococci* growth. (2) Superimposition of the photograph with IMS data on top of MALDI target plate. Average mass spectrum of each IMS experiment was shown below IMS images with signals correlated to compound **1**-**3** labeled with corresponding color as displayed in images.

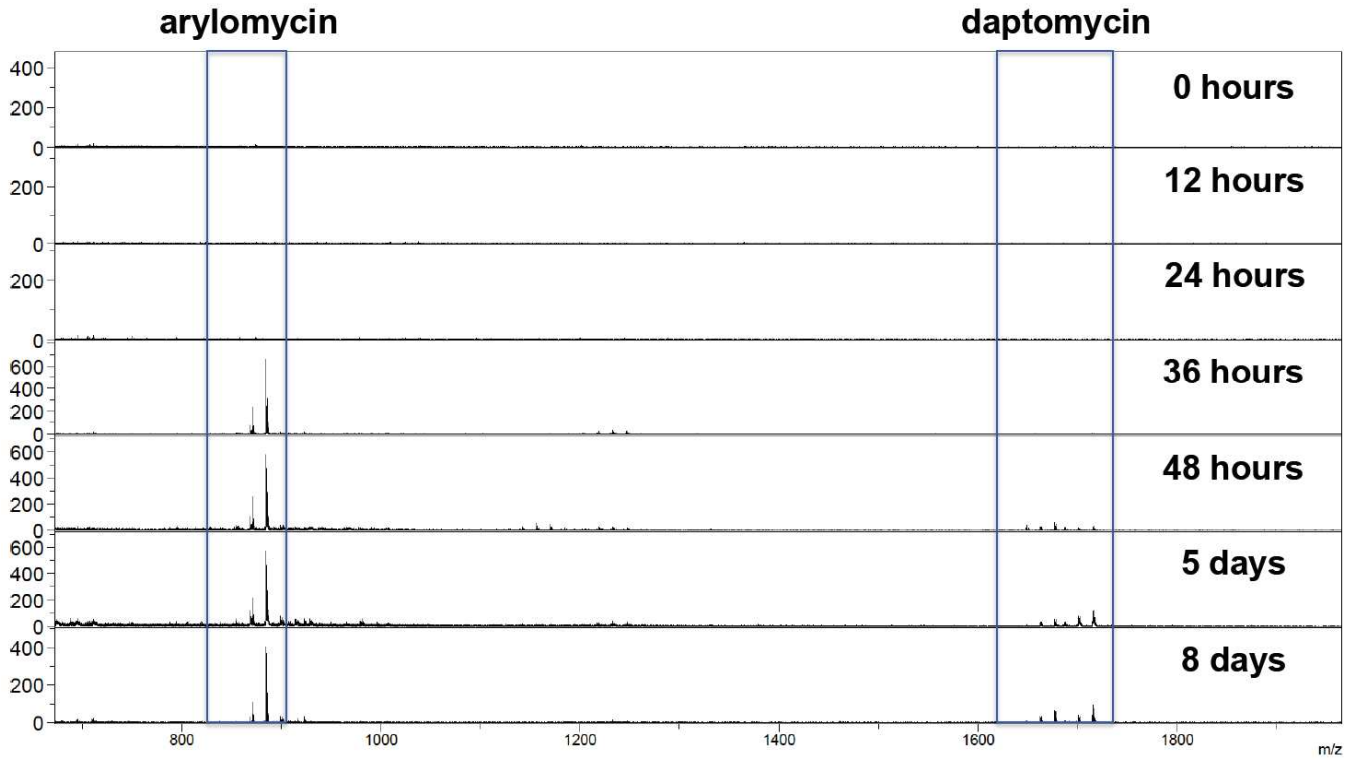


Figure S2: Time course for the production of arylomycin and daptomycin variants by *Streptomyces roseosporus*.

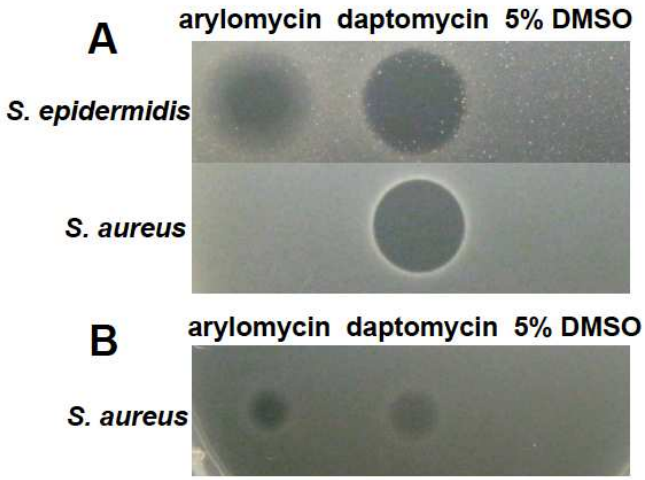


Figure S3: Spot assay showing arylomycin inhibit *S. epidermidis* and *S. aureus*. (A) 3µl of purified arylomycin (0.2 mg/ml), daptomycin (0.2 mg/ml), and DMSO (5%) or (B) 10µl of purified arylomycin (0.2 mg/ml), daptomycin (0.02 mg/ml), and DMSO (5%) were spotted on *staphylococcal* (indicated on the left) lawns.

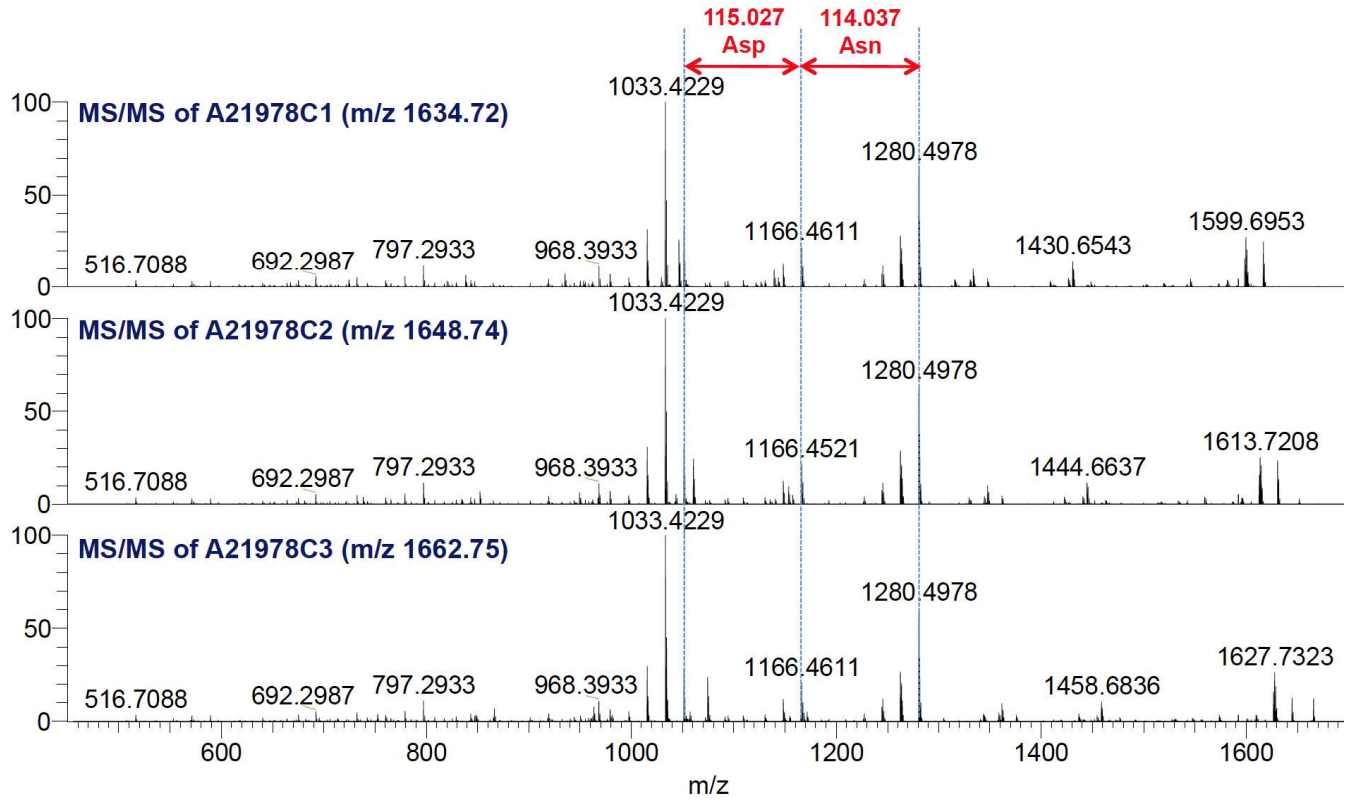
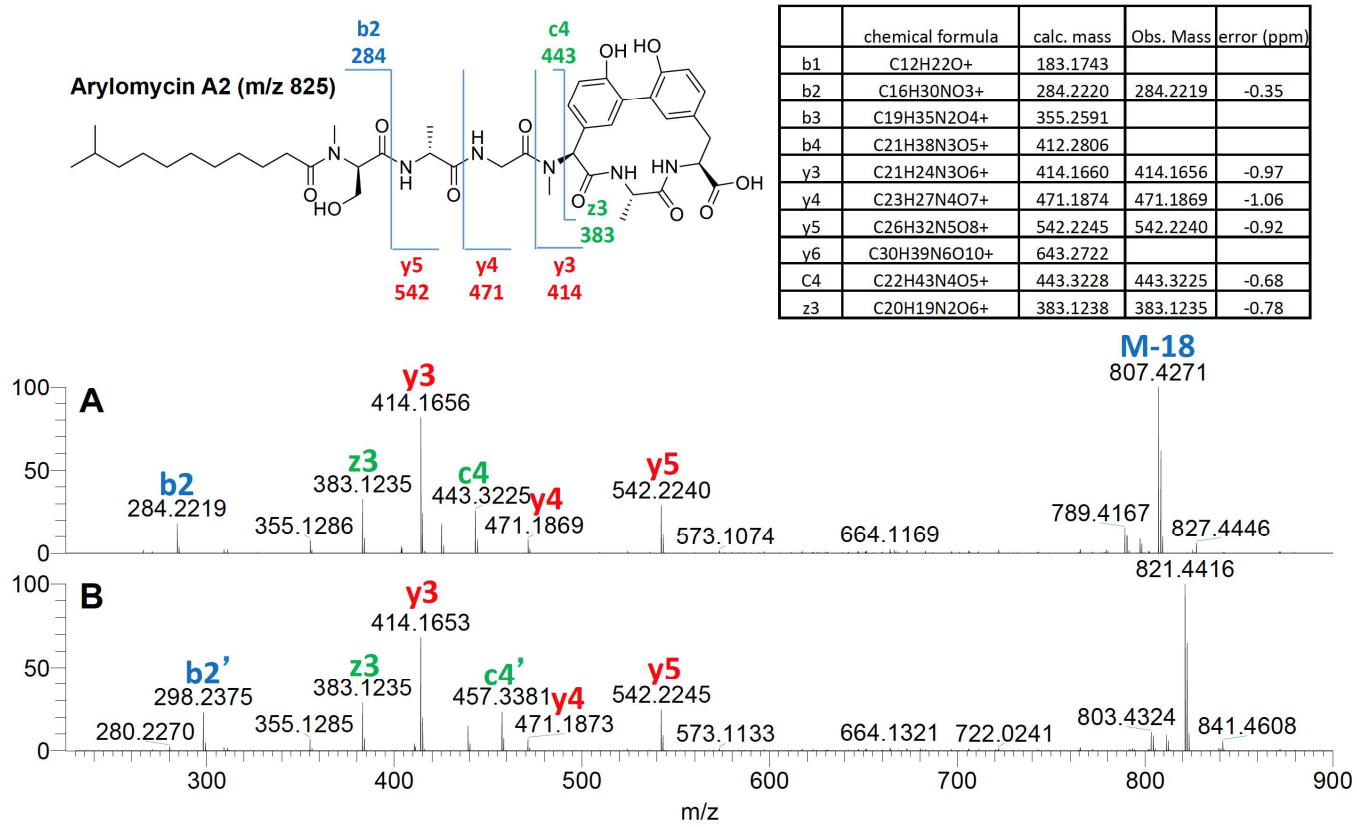


Figure S4: Alignment of FT MS/MS of daptomycin A21978C1-C3, the proof-of-principle experiment for SST, revealing the sequence tag Asp-Asn.



	chemical formula	calc. mass	Obs. Mass	error (ppm)
b1	C12H22O+	183.1743		
b2	C16H30NO3+	284.2220	284.2219	-0.35
b3	C19H35N2O4+	355.2591		
b4	C21H38N3O5+	412.2806		
y3	C21H24N3O6+	414.1660	414.1656	-0.97
y4	C23H27N4O7+	471.1874	471.1869	-1.06
y5	C26H32N5O8+	542.2245	542.2240	-0.92
y6	C30H39N6O10+	643.2722		
C4	C22H43N4O5+	443.3228	443.3225	-0.68
z3	C20H19N2O6+	383.1238	383.1235	-0.78

Figure S5: Annotated FT MS/MS spectra of compound 1 and 2. (A, B) FT MS/MS of compound **1** (arylomycin A2) and compound **2** (arylomycin A4), respectively. Ion b2 and c4 showed 14 Da shift due to the different length in fatty acid portion.

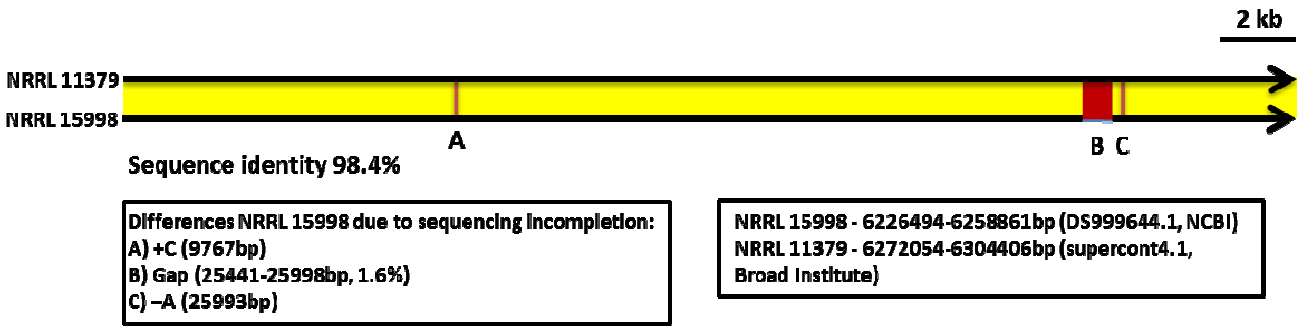


Figure S6: Sequence comparison of arylomycin gene cluster region in *S. roseosporus* NRRL 11379 and NRRL 15998. Sequences were compared by nblast-alignment. The total sequence identity was 98.4% due to a frameshift and missing sequencing information in NRRL 15998. The sequence identity in non-gap regions was 100%. Thus, the gene cluster annotation (Table S1) was done on the complete sequence of the NRRL 11379 arylomycin gene cluster region.

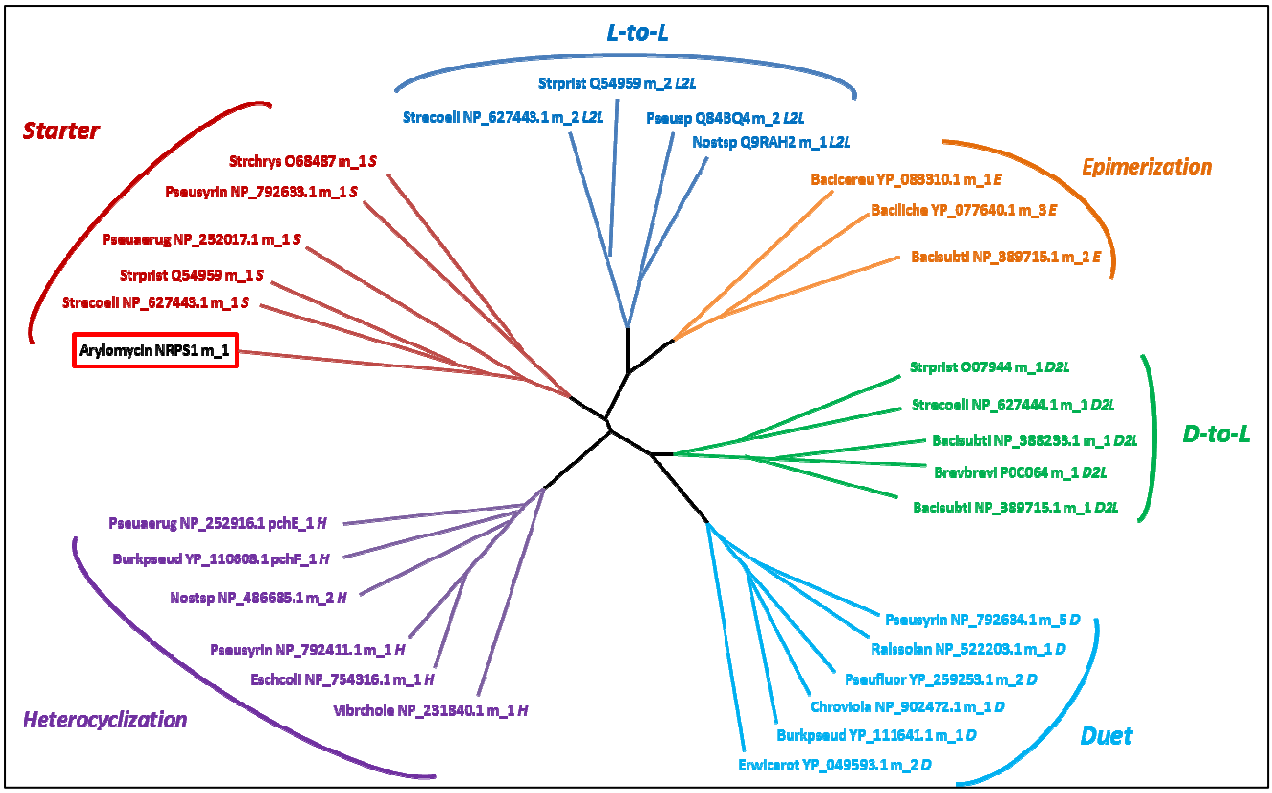


Figure S7: Phylogenetic analysis of arylomycin NRPS1m_1 C domain by ClustalW2 analysis.

Arylomycin NRPS1 m_1C-domain clusters closest to starter C domains.

```
OxyC            MGHDIDQVAPLLREPANFQLRTNCDPHEDNFGLRAHGPLVRIVGESSTQLGRDFVWQAHG 60
AryC            ----------------------------MYLGGRRG---TEAVGESREPG----VWEVFR 25
                                             :* *     .. ****        **:..

OxyC            YEVVRRILGDHEHFTTRPQFTQSKSGAHVEAQFVGQISTYDPPEHTRLRKMLTPEFTVRR 120
AryC            YDEAVQVLGDHRTFSSDMNHFIPEEQRQLARAARGNFVGIDPPDHTQLRGLVSQAFSPRV 85
                *: . ::****. *::  :.  .:.  ::     *::   ***:**:** :::  *: *

OxyC            IRRMEPAIQSLIDDRLDLLEAEG---PSADLQGLFADPVGAHALCELLGIPRDDQREFVR 177
AryC            TAALEPRIGRLAEQLLDDIVAERGDKASCDLVGEFAGPLSAIVIAELFGIPESDHTMIAE 145
                   :** *  * :: ** : **    .*.** * **.*:.* .:.**:***..*:  :..

OxyC            RIRRNADLSRGLKARAADSAAFN----------RYLDNLLARQRADPDDGLLGMIVRDHG 227
AryC            WAKALLGSRPAGELSIADEAAMQNTADLVRRAGEYLVHHITERRARPQDDLTSRLATTEV 205
                  :   .   . :   **.**::          .** : ::.:** *:*.* . :.  .

OxyC            DN--VTDEELKGLCTALILGGVETVAGMIGFGVLALLDNPGQIELLFESPEKAERVVNEL 285
AryC            DGKRLDDEEIVGVIGMFLIAGYLPASVLTANTVMALDEHPAALAEVRSDPALLPGAIEEV 265
                *.  : ***: *:   :::.*  ..: : .  *:** ::*. :  : ..*     .::*:

OxyC            VRYLSPVQAPNPRLAIKDVVIDGQLIKAGDYVLCSILMANRDEALTPDPDVLDANRAAVS 345
AryC            LRWRPPLVR-DQRLTTRDADLGGRTVPAGSMVCVWLASAHRDPFRFENPDLFDIHRNAGR 324
                :*: .*:   : **: :*. :.*: : **. *   :  *:**     :**::* :* *

OxyC            DVGFGHGIHYCVGAALARSMLRMAYQTLWRRFPGLRLAVPIEEVKYRSAFVDCPDQVPVT 405
AryC            HLAFGKGIHYCLGAPLARLEARIAVETLLRRFERIEIPRDESVEFHESIGVLGPVRLPTT 384
                .:.**:*****:**.***   *:* :** ***  :.:.   .   :.*  *  * ::*.*

OxyC            W---- 406
AryC            LFARR 389
```

Figure S8: CLUSTALW2 alignment of Cytochrome P450 AryC and biaryl C-C coupling OxyC from the vancomycin biosynthetic pathway in *Amycolatopsis orientalis*. AryC and OxyC share the signature sequence FGHG*X*H*X*CLG (underlined, red – active site cysteine) of cytochrome P450 enzymes and are 49% similar and 30% identical based on blastp alignment.

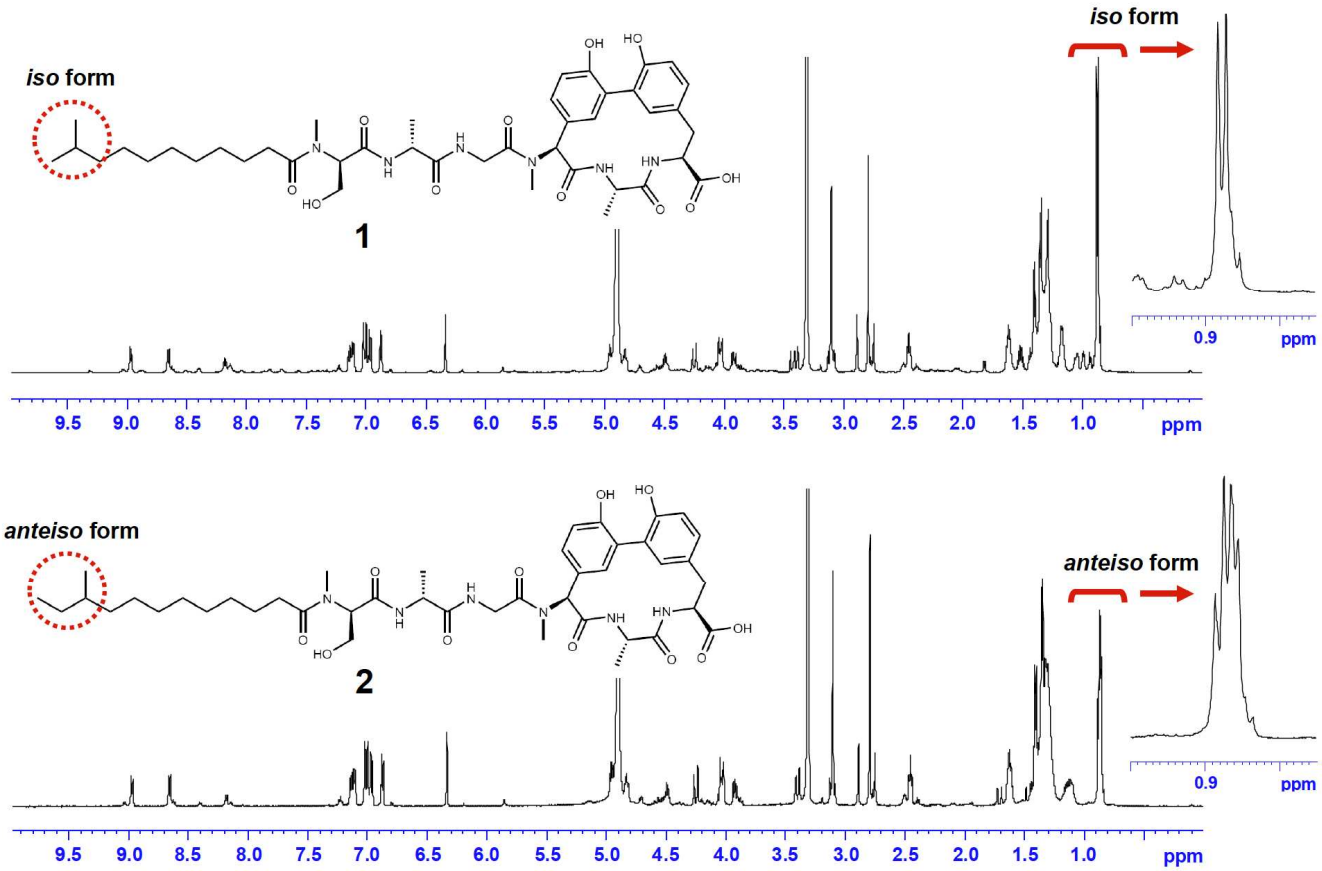


Figure S9: ^{1}H NMR spectra (600 MHz, CD_3OD) to confirm the discovery of arylomycins from *S. roseosporus*. NMR spectra are identical to arylomycin A2 (compound **1**) and A4 (compound **2**)[4,5].

Table S1. Genes in the arylomycin biosynthetic cluster from *S. roseosporus* NRRL 11379 and predicted functions based on sequence homology and protein domain analysis

Gene	Size [aa]	Predicted function	Protein homolog*	Accession number	Protein similarity/ identity [%/%]
aryA	584	NRPS2 (C)	amino acid adenylation domain protein [*Streptomyces violaceusniger* Tu 4113]	ZP_07611402.1	64/51
aryB	3485	NRPS1 (C^S-A-T-MT-E-C-A-T-E)	amino acid adenylation [*Streptomyces ghanaensis* ATCC 14672]	ZP_06575791.1	61/48
aryC	389	Cytochrome P450	cytochrome P450 [*Streptomyces steffisburgensis*]	CAJ42333.1	61/46
aryD	4297	NRPS3 (A-T-C-A-MT-T-C-A-T-C-A-T-TE)	putative pristinamycin I peptide synthase 3 and 4 [*Streptomyces pristinaespiralis*]	CBH31051.1	60/459
aryE	73	MbtH domain-containing protein	MbtH domain-containing protein [*Frankia* sp. EAN1pec]	YP_001510193.1	79/66
aryF	354	4-hydroxyphenylpyruvate dioxygenase	4-hydroxyphenylpyruvate dioxygenase [*Streptosporangium roseum* DSM 43021]	YP_003342414.1	69/57
aryG	371	FMN-dependent α-hydroxy acid dehydrogenase	FMN-dependent alpha-hydroxy acid dehydrogenase [*Streptomyces bingchenggensis* BCW-1]	ADI04246.1	72/61
aryH	4 9	Aminotransferase	aminotransferase class I and II [*Micromonospora* sp. ATCC 39149]	ZP_04604097.1	76/66
aryI	290	ABC transporter	ATP/GTP binding protein NosF [*Streptomyces albus* J1074]	ZP_04704494.1	70/62
aryJ	207	ABC transporter	hypothetical protein SalbJ_21220 [*Streptomyces albus* J1074]	ZP_04704495.1	71/56

* pBLAST analysis of manually annotated ORFs, *S. roseosporus* NRRL 15998 excluded (best homologs for all proteins)

Reference

(1) Yang, Y. L.; Xu, Y.; Straight, P.; Dorrestein, P. C. *Nat Chem Biol* **2009**, *5*, 885.

(2) Rausch, C.; Hoff, I.; Weber, T.; Wohlleben, W.; Huson, D.H. *BMC Evol Biol* **2007**, *78*, 7.

(3) Imker, H.J.; Krahn, D.; Clerc, J.; Kaiser, M.; Walsh, C.T. *Chem Biol* **2010**, *17*, 1077-1083.

(4) Holtzel, A.; Schmid, D. G.; Nicholson, G. J.; Stevanovic, S.; Schimana, J.; Gebhardt, K.; Fiedler, H. P.; Jung, G. *J Antibiot (Tokyo)* **2002**, *55*, 571.

(5) Roberts, T. C.; Smith, P. A.; Cirz, R. T.; Romesberg, F. E. *J Am Chem Soc* **2007**, *129*, 15830.