ARTICLE

Genomics and metabolomics-based assessment of the biosynthetic potential of the sponge-associated microorganism *Streptomyces cacaoi* strain R2A-843A from the Philippines

Zabrina Bernice L. Malto[a], Joeriggo M. Reyes[a], Bernard Isaiah D.C. Lo, Kevin Bossie S. Davis, Gisela P. Concepcion, Lilibeth A. Salvador-Reyes*

[a]Contributed equally

The Marine Science Institute, University of the Philippines, Diliman, Quezon City 1101, Philippines

ABSTRACT

The biosynthetic machinery of the sponge-associated *Streptomyces cacaoi* strain R2A-843A was assessed using a combined genomics and metabolomics approach. Whole genome sequencing and molecular networking showed the high biosynthetic potential of this actinomycete. A significant proportion of the genome is dedicated to secondary metabolite production, with biosynthetic gene clusters for nonribosomal peptides, polyketides, and terpenes being the most represented. Seven cyclic pentapeptides, including a putative new analogue, and a glycosylated lanthipeptide were identified using HRMS and untargeted MS/MS analysis. To validate our genome and metabolome analysis, we undertook a mass spectrometry-guided purification and confirmed the production of the known peptides BE-18257A (**1**) and BE-18257B (**2**). The production of **1** and **2** and the growth of the microorganism were monitored for eight days. Compound **2** was produced at a higher concentration, starting at 48 h post-incubation. Both compounds were noncytotoxic against colorectal and breast cancer cell lines.

*Corresponding author
Email Address: lsreyes@msi.upd.edu.ph
Date received: August 30, 2023
Date revised: September 18, 2023
Date accepted: September 19, 2023

INTRODUCTION

Actinomycetes in the genus *Streptomyces* are responsible for the production of half of the discovered bioactive secondary metabolites including antibiotics and antitumor agents (Bérdy 2005). *Streptomyces* from the marine environment are genetically diverse compared to their terrestrial counterparts, due in part to their adaptations to the physicochemical

KEYWORDS

Actinobacteria, BE-18257, genome, metabolome, pentaminomycins, *Streptomyces*, secondary metabolites

conditions in this environment (Tian et al. 2016). Marine-derived *Streptomyces* possess a smaller genome size and a slightly higher G+C content (Tian et al. 2016). Genome sequencing of *Streptomyces* revealed a diverse biosynthetic machinery for the production of secondary metabolites, with more than 20-50 secondary metabolic gene clusters per strain (Nett et al. 2009). More recently, 922 biosynthetic gene clusters (BGCs) were identified from 30 complete *Streptomyces* genome sequences (Lee et al. 2020). Most of the identified BGCs had unknown products.

The discovery of new natural products from microorganisms has been aided by metabolomic analysis using mass spectrometry (MS) (Palazzotto and Weber 2018). In particular, the Global Natural Product Social (GNPS) molecular networking platform has allowed for the rapid analysis of MS data through the dereplication of known compounds in the crude extract, thereby aiding in improved strain prioritization (Crüsemann et al. 2016; Jarmusch et al. 2021). New compounds from *Streptomyces*s spp. have been identified through molecular networking analysis using the GNPS platform (Mudalungu et al. 2019; Bai et al. 2021; L. Liu et al. 2020; Y. Liu et al. 2022; Ouchene et al. 2022). Advances in mass spectrometry, sequencing, and bioinformatics have enabled the development of methods in correlating metabolic data to genomic information for accelerated and rationalized natural products discovery (Soldatou et al. 2019; Crüsemann 2021). Methods range from manual targeted linking to automated approaches. Manual linking is performed by prioritizing strains based on genome mining information, targeting a specific predicted compound for purification (Soldatou et al. 2019). This method has been applied to isolate new compounds in *Streptomyces* spp. (Kaweewan et al. 2017, 2019; Son et al. 2018). Automated approaches can be performed through feature-based and correlation-based linking, both using tools to link predicted compounds from BGCs to the metabolomic data (Soldatou et al. 2019). The term *metabologenomics* was coined to describe the approach combining genome sequencing data and automated gene cluster prediction with MS-based metabolomics (Goering et al. 2016; Doroghazi et al. 2014). Combined omics methods have been proven as a powerful strategy to discover new compounds from bacteria, particularly in actinomycetes (Palazzotto and Weber 2018; Soldatou et al. 2019; Krause 2021; Crüsemann 2021).

As part of our ongoing efforts to discover new compounds from sponge-associated actinomycetes, we screened for the potential producers of peptides and modified peptides. Among those screened, the crude extract from *Streptomyces cacaoi* strain R2A-843A showed characteristic ^{1}H NMR chemical shifts corresponding to the α-protons of peptides and modified peptides. In this paper, a combined genomics and metabolomics approach was performed to comprehensively assess the biosynthetic potential of the sponge-associated *Streptomyces cacaoi* strain R2A-843A. MS-guided purification afforded two known cyclic peptides and a putative new compound was also identified. The planar structure was assigned via mass spectrometry and nuclear magnetic resonance spectroscopy. The production of the purified compounds by *S. cacaoi* strain R2A-843A was monitored for eight days using mass spectrometry to provide insights into the dynamics of the cyclic peptide expression in this actinomycete.

MATERIALS AND METHODS

General Experimental Procedure

Silica gel 60 (Merck, F_{254} Al) was used for the TLC. HPLC purification was done using Shimadzu Prominence System LC-20AT coupled with a diode array detector SPD-M20A. ^{1}H NMR spectra were acquired using a Varian-Inova 500 MHz NMR with DMSO-d_6 (CIL) as solvent and TMS as reference. Low resolution ESIMS was carried out using a Shimadzu LCMS-8040. High resolution mass analysis was carried out on Waters Xevo® G2-XS Quadrupole Time-of-Flight (QToF) mass spectrometer equipped with ESI ion source detector.

Biological material

R2A-843A was obtained from an unidentified sponge collected in Cebu, Philippines in October 2010. The strain was identified as *Streptomyces cacaoi* (99%), based on 16S rRNA and whole genome sequence similarity.

Microbial cultivation and extraction

S. cacaoi strain R2A-843A was fermented for seven days in 1.0 L of R2A broth (Himedia) supplemented with 20.0 g/L NaCl (J.T. Baker) at 30°C and 150 rpm. The fermentation broth was centrifuged at 4000 rpm, 20°C to separate the cell pellet and broth, and extracted using 5% (w/v) activated Diaion® HP-20 resin (Supelco). Separation was achieved by washing stepwise with 25% aq. MeOH (LCMS-grade, Merck), 50% MeOH, 75% MeOH, and 100% MeOH. A media blank containing 1.0 L R2A broth was similarly incubated and extracted.

Genome Sequencing and Prediction of Secondary Metabolites

DNA was extracted from the cell pellet using Invitrogen™ PureLink™ Genomic DNA Mini Kit following the manufacturer protocol and subjected to whole genome sequencing using MiSeq [2 x 300 V3 kit] at 100X coverage, with paired-end library size of 150 bp.

The quality of the raw reads was first assessed with FastQC v0.11.9 (Andrews 2010), and subsequent clipping, quality-based trimming and quality-based filtering were done with fastx-toolkit v0.0.14 (https://github.com/agordon/fastx_toolkit), with minimum quality score of 20 for all steps, minimum length of 36 bp and minimum average quality percentage of 90. Cleaned reads were paired using fastq-pair v1.0 (Edwards and Edwards 2019) and the resulting final reads were re-assessed with FastQC.

Assembly was done using SPADes v3.15.3 (Prjibelski et al. 2020) using only paired reads with the "--careful" flag enabled and coverage cutoff set to 10. Assessment of assembly quality was performed using Quast v5.0.1 (Mikheenko et al. 2018) with "--gene-finding", "--rna-finding" and "--conserved-genes-finding" flags enabled. Only contigs with length >=500 bp were used for subsequent analyses. Actual sequencing coverage for each contig/scaffold was calculated by mapping the reads to the assembly with bowtie2 v2.4.5 (Langmead and Salzberg 2012), samtools v1.14 (Danecek et al. 2021) and custom R scripts.

Whole-genome annotation was done using the RAST Online Server (Aziz et al. 2008; Overbeek et al. 2014; Brettin et al. 2015) with default parameters. Taxonomic identification of this strain was done with 16S rRNA sequence extracted from the annotations searched against the NCBI 16S RNA database with BLASTn (accessed March 13, 2023) (https://ncbi.nlm.nih.gov/blast), and with the whole genome assembly using Tetranucleotide Correlation Score search in JSpeciesWS GenomesDB v3.9.8 (Richter et al. 2015).

Prediction of secondary metabolite gene clusters was done with AntiSMASH v7 (beta) (Blin et al. 2023) and PRISM v4 (Skinnider et al. 2020) with all types of biosynthetic clusters enabled. Comparison with the biosynthetic gene cluster for pentaminomycin C and BE-18257A previously described by Roman-Hurtado et al. (2021B) was done with BLAST+ v2.5.0 and Mauve v2015-02-13 (Darling et al. 2009) and visualized with EasyFig v2.2.2 (Sullivan et al. 2011).

Untargeted Metabolite Analysis

The 100% MeOH Diaion® extracts from *S. cacaoi* strain R2A-843A and R2A media blank were dried *in vacuo* and reconstituted in 50% MeOH in H_2O to 1.0 mg/mL final concentration. Chromatographic analysis was performed by injecting a 1.0 µL aliquot of the extracts in a Phenomenex Kinetex 2.6 µm C18 100Å column (50 × 2.1 mm) and eluted at 0.35 mL/min, using a gradient program of CH_3CN/H_2O (both with 0.1% formic acid modifier): 20-100% CH_3CN for 4.5 min, and 100% CH_3CN for 3.0 min. Untargeted metabolite analysis was carried out on a Waters Xevo® G2-XS Quadrupole Time-of-Flight (QToF) mass analyzer equipped with an ESI ion source detector. The instrument was set using the following parameters: capillary voltage at 3.0 kV, cone voltage at 42 kV, and source offset of 80 kV. The source temperature was maintained at 150°C, and the desolvation gas temperature was at 500°C. The mass spectrometer was set to observe at m/z 100-2000 for the MS^1 scan in the positive mode. Ions of interest with a minimum signal intensity of 50,000 cps were data dependently subjected to MS/MS with ramp collision energy 20-40 eV. Three DDA scans were acquired for each sample including the solvent and media blanks.

Chromatograms were converted to mzxml format using freely available MSConvert software (www.proteowizard.sourceforge.net) (Holman et al. 2014). A molecular network was created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu) (M. Wang et al. 2016). Possible impurities detected from the solvent blank and media components were filtered out by subtracting the DDA data of the solvent and R2A media blanks. The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.5 Da and a MS/MS fragment ion tolerance of 0.1 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than six matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least six matched peaks. GNPS analysis was done according to the standard protocol of M. Wang et al. (2016). The generated molecular network was viewed in Cytoscape software (Shannon et al. 2003). MS and tandem MS spectra were viewed using MZmine 2.53 (Pluskal et al. 2010).

Purification of BE-18257A (1) and BE-18257B (2)

The crude 100% Diaion extract (240 mg) was chromatographed on a Sephadex LH-20 column with an isocratic elution of MeOH. Subfractions 4-7 were pooled based on TLC profile. The pooled fraction (90 mg) was further chromatographed on a reversed-phase HPLC using a semipreparative C18 column (Phenomenex® Synergi™ Hydro-RP, 250 x 10 mm, 4 µm, 2.0 mL/min) with a linear gradient of $MeOH/H_2O$ (40-100% MeOH for 10 min, 100% MeOH for 10 min) to afford BE-18257A (**1**, 0.8 mg, t_R 17.8 min) and BE-18257B (**2**, 1.5 mg, t_R 19.2 min).

BE-18257A (**1**). Pale yellow powder; UV (MeOH) λ_{max} 222, 273, 290 nm; LRESIMS m/z 599.0 $[M+H]^+$, m/z 597.0 $[M-H]^-$. ^1H NMR (500 MHz, DMSO-d_6) δ 10.8 (s, 1H), 8.78 (d, J = 7.8 Hz, 1H), 8.49 (s, 1H), 7.52 (d, J = 8.0 Hz, 1H), 7.30 (d, J = 8.0 Hz, 1H), 7.03 (t, J = 7.4 Hz, 1H), 6.96 (t, J = 7.4 Hz, 1H), 4.39 (m, 1H), 4.22 (m, 3H), 4.15 (m, 1H), 2.92 (m, 1H), 2.16 (m, 1H), 1.82 (m, 1H), 1.80 (m, 1H), 1.20 (m, 2H), 1.12 (d, J = 6.9, 3H), 1.08 (m, 2H), 0.83 (d, J = 6.3 Hz, 3H), 0.82 (d, J = 6.3 Hz, 3H), 0.73 (d, J = 6.4 Hz, 3H), 0.63 (d, J = 6.4 Hz, 3H). UV, MS, and ^1H NMR data are in agreement with Kojiri et al. (1991) and Nakajima et al. (1991).

BE-18257B (**2**). Pale yellow powder; UV (MeOH) λ_{max} 222, 273, 290 nm; LRESIMS m/z 613.0 $[M+H]^+$, m/z 611.0 $[M-H]^-$. ^1H NMR (500 MHz, DMSO-d_6) δ 10.8 (s, 1H), 8.78 (d, J = 7.8 Hz, 1H), 8.49 (d, J = 6.6 Hz, 1H), 7.53 (d, J = 7.6 Hz, 1H), 7.30 (d, J = 8.0 Hz, 1H), 7.15 (d, J = 7.9 Hz, 1H), 7.03 (t, J = 7.4 Hz, 1H), 6.96 (t, J = 7.3 Hz, 1H), 4.38 (dq, J = 6.3 Hz, 1H), 4.22 (m, 3H), δ 4.14 (m, 1H), 2.16 (m, 2H), 1.89 (m, 2H), 1.21 (m, 2H), 1.12 (d, J = 6.6 Hz, 3H), 1.07 (m, 2H), 0.85 (t, J = 6.3 Hz, 3H), 0.78 (d, J = 6.7 Hz, 3H), 0.73 (d, J = 6.4 Hz, 3H), 0.63 (d, J = 6.4 Hz, 3H). UV, MS, and ^1H NMR data are in agreement with Kojiri et al. (1991) and Nakajima et al. (1991).

Antiproliferative assay

Cytotoxicity was tested against human colorectal cancer cells (HCT116) and human breast adenocarcinoma cells (MCF-7) using a tetrazolium-based colorimetric assay as described by Susana and Salvador-Reyes (2022). Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% v/v fetal bovine serum (FBS) and 1% antibiotic-antimycotic solution in a humidified atmosphere at 37°C containing 5% CO_2. Taxol was used as the positive control. The assay was terminated at 48 h post-treatment and the growth of the cells was assessed using the tetrazolium reagent. Cell viability was calculated relative to the solvent control.

Cell density monitoring of *S. cacaoi* strain R2A-843A

The cell density of *S. cacaoi* strain R2A-843A was monitored daily for eight days. A 1% (v/v) of the seed culture was inoculated into three 2.8 L Fernbach flasks containing 1 L of the production medium and incubated for eight days using the same incubation condition. A 10-mL aliquot of the culture broth was collected every 24 h. The biomass was collected by centrifugation at 4000 rpm, 20°C. The cell pellet biomass was freeze-dried and weighed to assess the microbial growth.

Time-course monitoring of BE-18257A (1) and BE-18257B (2) production by *S. cacaoi* strain R2A-843A

Supernatants from cultures of *S. cacaoi* strain R2A-843A were retrieved every 24 h and extracted with an equal volume of EtOAc. The organic extract was dried *in vacuo*, reconstituted in 50% aq MeOH and subjected to targeted quantitative analysis of BE-18257A (**1**) and BE-18257B (**2**) using multiple reaction monitoring. Data analysis and statistical evaluation were performed using GraphPad Prism 7.

Targeted MRM analysis

Samples and standard solutions were injected into a LCMS-8040 triple quadrupole mass spectrometer (Shimadzu, Japan) with positive-mode electrospray ionization (ESI) source. Cyclic peptides in the crude organic extract were detected using multiple reaction monitoring (MRM). Separation was achieved on a Kinetex C18 column (Phenomenex, 50 × 2.1 mm, 2.6 µm) with a flow rate of 0.2 mL/min. The mobile phase consisted of H_2O and CH_3CN (both with 0.1% formic acid). A linear gradient elution program was applied as follows: initial 5% CH_3CN for 1 min, 5-100% CH_3CN for 14 min, 100% CH_3CN for 3 min. ESI-MS conditions were operated at a spray voltage of 4.5 kV, nebulizing gas flow of 3 L/min, DL temp of 250°C, heat block temp 400°C and drying gas flow of 15 L/min.

Standard peptides and test samples were spiked with 50 ppb pyoluteorin as an internal standard prior to analysis. MRM experiment was conducted with retention times and precursor to fragment ion transitions as follows: BE-18257A (7.3 min; 599.0 > 387.0), (7.3 min; 599.0 > 300.0), (7.3 min; 599 > 86.2), and BE-18257B (7.7 min; 613.0 > 314.1), (7.7 min; 613.0 > 159.1), (7.7 min; 613 > 86.2). Analysis was done in triplicate injections.

Data deposition. The mass spectrometry data is deposited in the MassIVE public repository (MSV000092786). The molecular networking job can be publicly accessed with the task ID: task=1c67465868fe4b6186771198adb6ad0f. The Whole Genome Shotgun project is deposited at DDBJ/ENA/GenBank under the accession number JAUZGE000000000. The version described in this paper is version JAUZGE010000000.

RESULTS

Genome mining for biosynthetic gene clusters

We sequenced and assembled an 8,551,568-bp genome consisting of 861 contigs from a three-day culture of strain R2A-843A using MiSeq with a library size of 150-bp and annotated using the RAST platform (Supplementary Figure S1). Based on sequence similarity to the NCBI 16S rRNA database of the predicted 16S rRNA and tetranucleotide correlation score (TCS) search against the JSpeciesWS GenomDB using our whole-genome assembly, we identified strain R2A-843A as *Streptomyces cacaoi* with e-value of 0.0 and Z-score of >0.999% in the BLAST and TCS search, respectively (Tables S1, S2). The average nucleotide identity with *S. cacaoi* NBRC-12748, its best match in both 16S BLAST and TCS search, showed >98.6% identity and >82.5% alignment coverage across both genomes. Complete assembly and annotation statistics are shown in Supplementary Table S3 and Supplementary Figure S1.

Search for biosynthetic gene clusters using AntiSMASH and PRISM showed that ~12% of its genome (by length) is associated with secondary metabolite production, with the majority of its biosynthetic potential dedicated to nonribosomal peptide (NRPS) synthesis (Figure 1). AntiSMASH identified a total of 70 potential clusters, 34 (44%) of which are NRPS or NRPS-like, although some of these clusters may further be combined into contiguous units due to contig fragmentation in the genome assembly. *S. cacaoi* NBRC-12748, in comparison, has a genome size of 8.55 Mbp and a similar BGC genome fraction of 16%, with 37% annotated as NRPS or NRPS-like clusters and is the most abundant BGC as well (Supplementary Figure S2).

Previous genome mining and metabolite profiling of the top BLAST hit *S. cacaoi* NBRC-12748T identified a biosynthetic cluster with two tandem NRPS genes for the synthesis of BE-18257 analogues and pentaminomycins for the first and second NRPS, respectively, each of which has five adenylation modules (Kaweewan et al. 2020; Carretero-Molina et al. 2021; Roman-Hurtado et al. 2021B). We mapped six contigs over the entire length of the *cpp* BGC from *S. cacaoi* CA-170360 using blastn (Supplementary Table S4) at >96.5% nucleotide identity with only small gaps corresponding to contig breaks. None of the contig breaks were juxtaposed with predicted functional domains in CA-170360. Five of these contigs were annotated in AntiSMASH and PRISM to contain predicted NRPS clusters (Figure 2). A closer inspection of amino acid sequence alignments showed >98.5% amino acid identity for all adenylation domains in both NRPS genes, except for adenylation domain 2 in the BE-18257 NRPS (97.3% identity), suggesting highly-conserved specificity for adenylation substrates (Supplementary Table S5). However, M2 in the BE-18257 NRPS (*cppB*) of our genome is missing an epimerization domain, likely attributable to an assembly artifact since the module M2 is at the terminal end of a contig and adjacent to a short contig with no predicted functional domains.

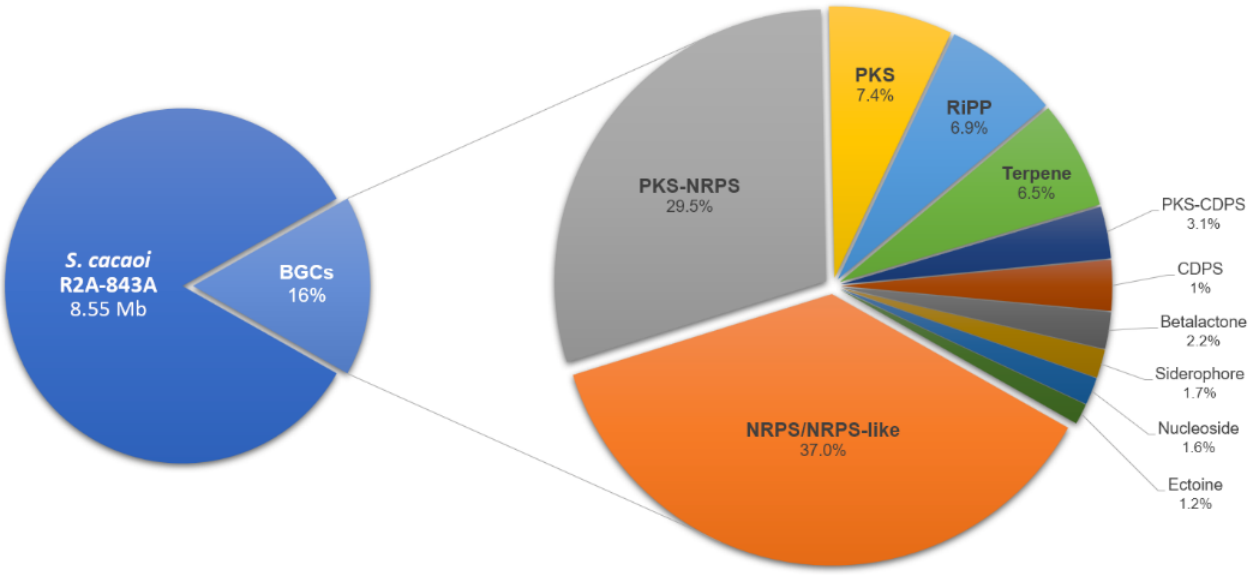


Figure 1: Genome fraction of *S. cacaoi* strain R2A-843A dedicated to secondary metabolite production based on AntiSMASH prediction (by BGC length).

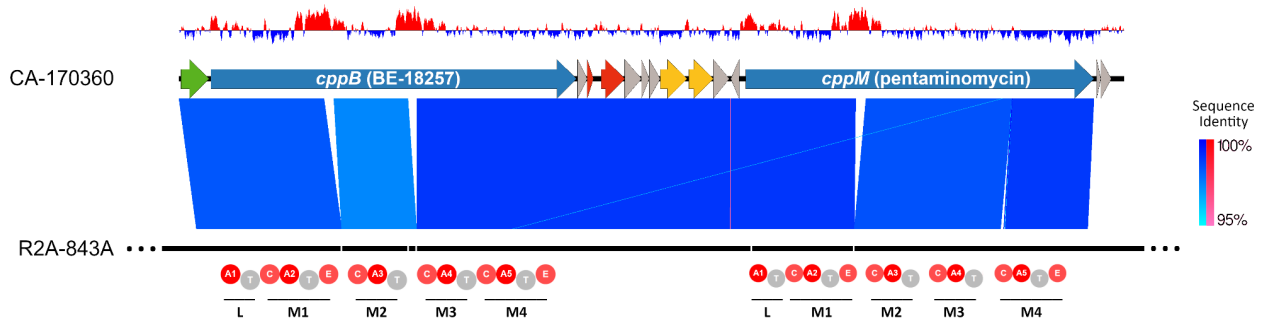


Figure 2: Alignment of *cpp* BGC in *S. cacaoi* CA-170360 with the draft genome assembly of *S. cacaoi* strain R2A-843A. Top graph shows the GC skew across the BGC. Solid blocks show alignment between regions of the *cpp* BGC for CA-170360 and R2A-843A contigs, with gene annotations for CA-170360 shown in various colors (*teal* - NRPS; *green* - penicillin-binding protein-type thioesterase; *red* - regulatory genes; *yellow* - cytochrome P450; *gray* - hypothetical and other proteins). Color coding of the open-reading frames (ORFs) follows that of Roman-Hurtado et al. (2021B). Contig breaks in *S. cacaoi* strain R2A-843A are shown as white vertical lines, coinciding with inversions in the GC skew plot. Domain and module arrangement for the two NRPS genes is shown at the bottom (*domains*: *A1-5* - adenylation; *C* - condensation; *T* - thiolation; *E* - epimerization; *modules*: *L* - loading module; *M1-4* - modules 1-4).

Additionally, we were also able to map three contigs to the cacaoidin biosynthetic cluster of *S. cacaoi* CA-170360 (Roman-Hurtado et al., 2021) at 98.97% nucleotide identity over a single alignment block (Supplementary Table S6), although this was truncated to cao25 (Figure 3A). Sequence alignment of the CaoA precursor peptide using tblastn between the two strains showed 100% identity at the amino acid level (Figure 3B). No other significant alignments were made to our draft genome covering the last ~3.5kbp of the cacaoidin BGC containing a AfsR/SARP family transcriptional regulator, and a closer inspection of the annotations for *S. cacaoi* strain R2A-843A indicates that this region was replaced by an amino acid permease, a hypothetical protein, and a LysR transcriptional regulator (Supplementary Figure S3). The conserved BGC, including the precursor peptide and glycosyltransferases (colored as yellow ORFs in Figure 3A), indicates the capacity for the production of glycosylated cacaoidin, albeit under different regulatory control.

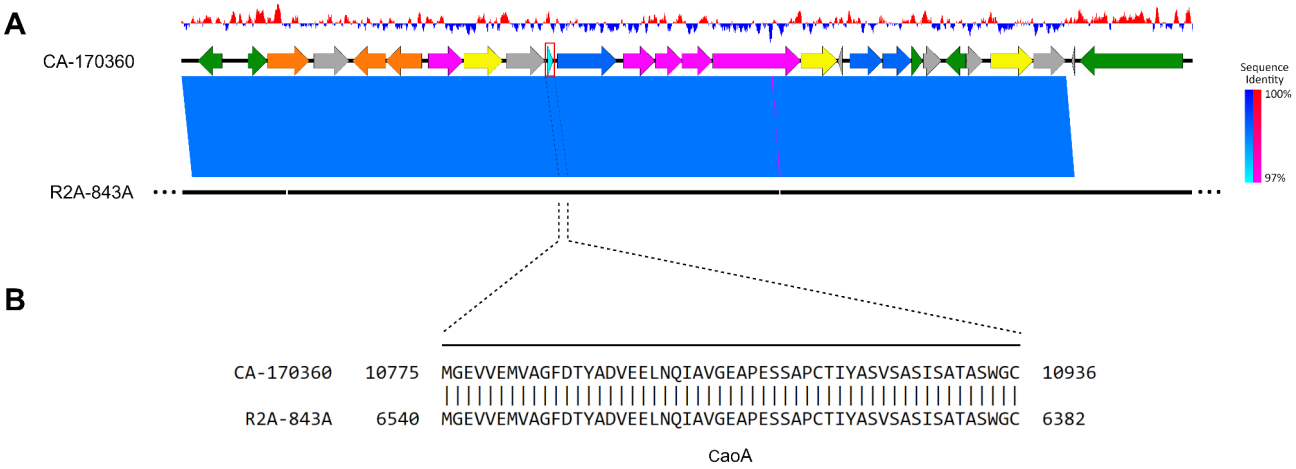


Figure 3: (A) Alignment of cacaoidin BGC in *S. cacaoi* CA-170360 with the draft genome of *S. cacaoi* strain R2A-843A. Top graph shows the GC skew across the BGC. Solid blocks show alignment between regions of the cacaoidin BGC for CA-170360 and R2A-843A contigs, with open-reading frames for CA-170360 shown as block arrows following the color scheme in Roman-Hurtado, et al. (2021A). *Cyan* - cacaoidin precursor peptide (*caoA*), emphasized in red box; *dark green* - regulatory genes; *orange* - L-rhamnose synthesis; *magenta* - lanthipeptide synthesis; *yellow* - glycosyl-transferase; *blue* - ABC transporter; *gray* - hypothetical protein. Contig breaks are shown as white vertical lines along R2A-843A. (B) Amino acid alignment of CaoA with contig191 using tblastn (frame -2) shows 100% identity between CA-170360 and R2A-843A.

Untargeted metabolite analysis of *Streptomyces cacaoi* strain R2A-843A

To validate the results of the genome sequencing, we undertook a preliminary secondary metabolite characterization of *S. cacaoi* strain R2A-843A using LC-MS-based metabolomics and molecular networking. Untargeted LC-MS/MS analysis was coupled with the Global Natural Products Social Molecular Networking Platform (GNPS), which is based on spectral alignment to assess the similarities and relationships among molecules (M. Wang et al. 2016) and visualized using a molecular network. GNPS has been used for the discovery of new natural products from cyanobacteria, marine sponges, fungi, and bacteria (Bai et al. 2021; Ding et al. 2018; Hautbergue et al. 2019; Khushi et al. 2020; Naman et al. 2017; Teta et al. 2015; Via et al. 2018; X. Wang et al. 2021). The molecular network generated from the crude extract of *S. cacaoi* strain R2A-843A is presented in Figure 4 (full molecular network is shown in Supplementary Figure S4). From the molecular network, GNPS was only able to putatively identify three compounds, BE-18257A (cosine score 0.78), indole-3-carbinol (cosine score 0.71), and 5-methylbenzotriazole (cosine score 0.78). This corroborated the results from the BGC analysis of *S. cacaoi* strain R2A-843A.

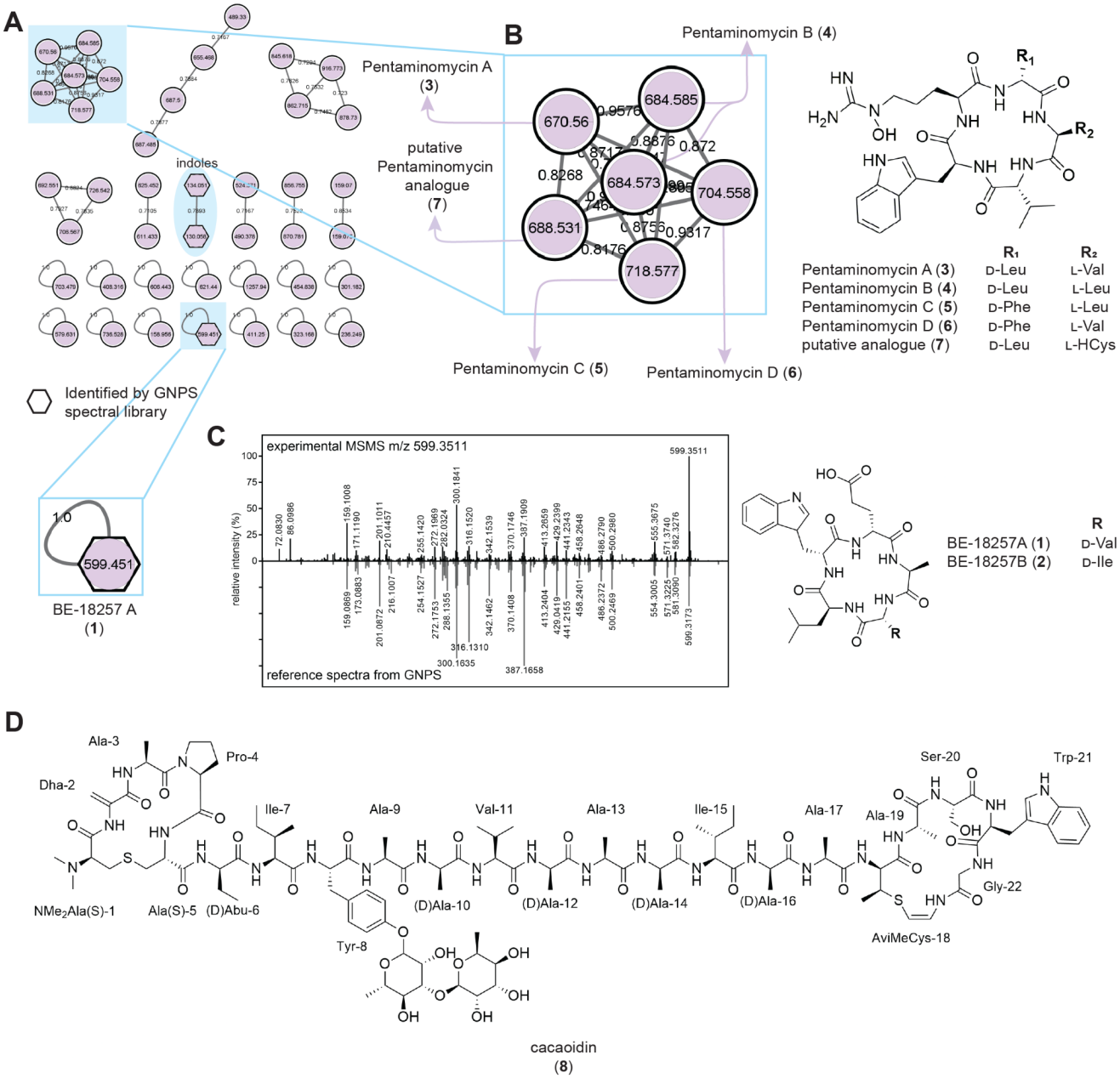


Figure 4: Molecular network generated from the HRMS/MS spectra of the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A using the Global Natural Products Social Molecular Networking (GNPS) platform. Part of the molecular network is shown in A wherein nodes annotated using the GNPS spectral library through MS/MS matching are shown as hexagons. Manual annotation of the main compound cluster showed that the masses correspond to pentaminomycins A-D (3-6) and a putative new analogue (7) (B). The cyclic pentapeptide BE-18257A (1) was putatively identified and demonstrated good tail-to-tail matching with the reference spectrum (C). BE-18257B (2) and cacaoidin (8) (D) were putatively identified from the crude extract but were not observed in the molecular network.

Tail-to-tail matching of the MS/MS spectrum of **1** in the crude extract of *S. cacaoi* strain R2A-843A and from the GNPS reference spectra demonstrated good matching (Figure 4C). The observed m/z 599.3199 (Supplementary Figure S5A) also corresponds to the $[M+H]^+$ of BE-18257A ($C_{30}H_{43}N_6O_7^+$, calcd. 599.3193, Δ + 1.00 ppm). The structural analogue **2** was not observed in the molecular network, but closer inspection of the HRMS chromatogram of the crude extract (Supplementary Figures S6C, S6D) showed an m/z of 613.3378 (Supplementary Figure S5B) ($[M+H]^+$ $C_{30}H_{43}N_6O_7^+$, calcd. 613.3349, Δ + 4.56 ppm). MS/MS analysis of the putative BE-18257B (**2**) showed a similar fragmentation pattern with putative BE-18257A (**1**) (Figures S7A, S7B) suggesting structural similarity.

Recent literature on *S. cacaoi* demonstrated that the cyclic pentapeptides BE-18257A–C and pentaminomycins A–H are synthesized by two nonribosomal peptide synthetases encoded in tandem within the same biosynthetic gene cluster (Román-Hurtado et al. 2021B) and shown in our genome mining analysis. Pentaminomycins are hydroxyarginine-containing cyclic pentapeptides produced by several *Streptomyces* species (Carretero-Molina et al. 2021; Hwang et al. 2020; Jang et al. 2018; Kaweewan et al. 2020). Pentaminomycin A has antimelanogenic activity against alpha-melanocyte stimulating hormone (α-MSH)-stimulated B16F10 melanoma cells (Jang et al. 2018), pentaminomycin C was reported to show bioactivity against Gram-positive bacteria including *Bacillus subtilis*, *Micrococcus luteus*, and *Staphylococcus aureus* (Kaweewan et al. 2020), pentaminomycins C and D also showed autophagy-inducing activity in HEK293 cells (Hwang et al. 2020), while pentaminomycins F and G showed modest activity against *Acinetobacter baumannii* (Carretero-Molina et al. 2021).

A manual inspection of the other nodes in the molecular network was performed in an effort to characterize the other secondary metabolites produced by *S. cacaoi* strain R2A-843A. A cluster

of six nodes corresponded to the masses of pentaminomycins A, B, C/H, D, and an unassigned node (Figure 4B). Pentaminomycin A (**3**) was observed as m/z 670.4044 ($[M+H]^+$ $C_{33}H_{52}N_9O_6^+$, calcd. 670.4041, Δ + 0.45 ppm), pentaminomycin B (**4**) was observed as m/z 684.4224 ($[M+H]^+$ $C_{34}H_{54}N_9O_6^+$, calcd. 684.4197, Δ + 3.94 ppm), and pentaminomycin D (**6**) was observed as m/z 704.3900 ($[M+H]^+$ $C_{36}H_{50}N_9O_6^+$, calcd. 704.3884, Δ + 2.27 ppm). The detected m/z corresponding to either pentaminomycin C or H (**5**) was 718.4061 ($[M+H]^+$ $C_{37}H_{52}N_9O_6^+$, calcd. 718.4040, Δ + 2.92 ppm).

Fragmentation analysis of the putative pentaminomycins was performed to confirm their identity. Supplementary Figure S8 shows the fragmentation of putative pentaminomycin A–D with the MS/MS data. The m/z of the fragment corresponding to the removal of the hydroxyguanidine group of the N5-hydroxyarginine amino acid common to the pentaminomycins was consistently observed as the base peak in all the acquired MS/MS spectra. Subsequent fragmentation of the amide bonds is shown, and fragment ions accounted for in the acquired MS/MS data are shown. For the node corresponding to structural isomers pentaminomycin C or H, a diagnostic fragment ion at 427 m/z for pentaminomycin C was observed due to a break between the alkyl carbonyl bond of Phe and Leu residues. If these residues are switched as in the case of pentaminomycin H, a 461 m/z peak that corresponds to this fragment should be observed. Thus, pentaminomycin C was assigned as compound **5**.

For the putative pentaminomycin analogue with a protonated m/z of 688.3629 (Supplementary Figure S9A), the acquired MS/MS spectrum (Figure S9B) is most similar to **3** and **4** (Supplementary Figures S8A-B) which suggests that the residue at the R_2 position was substituted to another amino acid (Figure 4B). Based on the observed mass, there are only two possible amino acids: 4-hydroxythreonine (4-OH-Thr) or homocysteine (HCys). Comparison of the theoretical protonated m/z of the 4-OH-Thr ($C_{32}H_{50}N_9O_8^+$, calcd. 688.3782, Δ - 22.22 ppm), and the HCys ($C_{32}H_{50}N_9O_6S^+$, calcd. 688.3605, Δ + 3.49 ppm) derivatives with the observed m/z suggests that compound **7** most probably contains the HCys amino acid. The proposed structure and fragmentation analysis of the putative pentaminomycin analogue is shown in Supplementary Figure S9B.

The nonproteinogenic amino acid homocysteine has been shown to be incorporated in compounds from *Streptomyces* spp. as thiolactones (Jizba et al. 1985; Jang et al. 2021), and in the thioxylofuranose moiety of albomycin (Zeng et al. 2012; Kulkarni et al. 2016), however its presence as an unmodified amino acid in peptidic natural products has not been reported in literature. Homocysteine is biosynthesized by bacteria through the activated methyl cycle which recycles methionine and adenosine from the S-adenosylmethionine (SAM)-mediated methylation reactions (Parveen and Cornell 2011). In *S. griseus*, methionine is converted to SAM, then to S-adenosylhomocysteine (SAH) through a series of enzymes, and SAH is broken down into adenosine and homocysteine (Kulkarni et al. 2016). Homocysteine and a hydroxy-activated serine are combined to form the unusual thioxylofuranose moiety of albomycin (Kulkarni et al. 2016). SAM is shared by many biological pathways, thus either an exogenous supply of SAM, or an overexpression of SAM synthetase is necessary for the production of these homocysteine-based compounds (Kulkarni et al. 2016).

Another secondary metabolite that is encoded by *S. cacaoi* is the glycosylated lanthipeptide cacaoidin (**8**, Figure 4D) which showed potent activity against several Gram-positive pathogens, including a clinical strain of methicillin-resistant *Staphylococcus aureus* (MRSA) (Ortiz-Lopez et al. 2020). Manual inspection of the MS^1 data of the crude extract of *S. cacaoi* strain R2A-843A showed masses corresponding to **8** at t_R = 3.32 min (Supplementary Figure S10). We were able to annotate a triply charged adduct $[M+3H]^{3+}$ ($C_{107}H_{165}N_{24}O_{32}S_2^{3+}$, calcd. Δ - 3.63 ppm) (Supplementary Figure S10B), and three doubly charged adducts $[M+2H]^{2+}$($C_{107}H_{164}N_{24}O_{32}S_2^{2+}$, calcd. Δ - 4.90 ppm), $[M+H+Na]^{2+}$ ($C_{107}H_{163}N_{24}O_{32}S_2Na^{2+}$, calcd. Δ - 5.51 ppm), and $[M+2Na]^{2+}$ ($C_{107}H_{162}N_{24}O_{32}S_2Na_2^{2+}$, calcd. Δ + 1.88 ppm) (Supplementary Figure S10C). All four detected adducts had ion intensities below the threshold for MS/MS analysis, thus no fragmentation evidence for **8** can be established. The MS^1 corroborates the 100% sequence identity of the CaoA precursor peptide with the published sequence, as shown in Figure 3B.

Purification and biological activity evaluation of BE-18257A (1) and BE-18257B (2)

Validation of the genome and metabolome analysis was done using an MS-guided purification. The known cyclic peptides BE-18257A (**1**) and BE-18257B (**2**) were the major isolable products from the culture broth of *S. cacaoi* strain R2A-843A. The structures of the purified compounds were confirmed by MS and ^{1}H NMR spectroscopy (Supplementary Figure S11).

BE-18257A (**1**) was purified as a pale yellow powder. This compound gave an $[M+H]^+$ at m/z 599.0 and $[M-H]^-$ at m/z = 597.0, indicating a molecular weight of 598 Da. The ^{1}H NMR spectrum of **1** showed the presence of amide NHs (δ_H 10.8, 8.78, 8.49), five aromatic protons (δ_H 7.53, 7.30, 7.15, 7.03, 6.96), α-protons of amino acids (δ_H 4.39, 4.22, 4.15) and five methyl groups (δ_H 1.12, 0.85, 0.78, 0.73, 0.63) (Supplementary Figure S11). The ^{1}H NMR data is in agreement with literature values (Kojiri et al. 1991; Nakajima et al. 1991) for BE-18257A (**1**), thus, confirming the presence of the amino acids D-Trp, L-Ala, L-Leu, D-Glu, and D-Val. BE-18257B (**2**) gave a protonated peak at $[M+H]^+$ 613.0 and $[M-H]^-$ 611.0 which suggested a molecular weight of 612 Da. Similarly, ^{1}H NMR spectrum of **2** showed similar proton resonances, with the main difference in the ^{1}H NMR signals at δ_H 0.85 (3H), 1.06 (1H), and 1.31 (1H) corresponding to Ile-γCH_3 and Ile-γCH_2, confirming the D-Val to allo-D-Ile substitution in **2** (Supplementary Figure S11). The ^{1}H NMR data for 2 is also in agreement with the reports of Kojiri et al. (1991) and Nakajima et al. (1991).

The cyclic pentapeptides BE-18257A (**1**) and BE-18257B (**2**) were first structurally characterized by Nakajima et al. (1991). Bioactivity assessment showed these to be endothelin-binding inhibitors. BE-18257B is a competitive endothelin (ET) antagonist, binding to ET_A receptors (ET-1 selective) in cardiovascular tissues (Nakajima et al. 1991). The complex ET system plays a role in hypertension and proteinuric kidney disease, including the micro- and macro-vascular complications of diabetes (Meyers and Sethna, 2013). A study by Kaweewan et al. (2020) showed that BE-18257A did not exhibit any antibacterial activity against *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Bacillus subtilis*, and *Micrococcus luteus*. No other bioactivity has since been reported for these compounds.

We tested the cytotoxicity of **1** and **2** against HCT116 human colorectal cancer cells and MCF-7 human breast cancer cells. Compounds **1** and **2** did not affect the growth of HCT116 cancer cells at all concentrations tested (Supplementary Table S7). A modest inhibition of MCF-7 adenocarcinoma cell growth was observed with 10 µg/mL treatment of **2** (Supplementary Table S7).

Time-course monitoring of BE-18257A (1) and BE-18257B (2) production by *S. cacaoi* strain R2A-843A

Time course monitoring of the concentration of **1** and **2** in the fermentation broth of *S. cacaoi* strain R2A-843A was done every 24 h over eight days. Based on our initial observation, BE-18257A (**1**) and BE-18257B (**2**) are mainly present in the exometabolome of *S. cacaoi* strain R2A-843A. Hence, to quantify the concentration of **1** and **2** produced by the microorganism, 10-mL aliquots of the fermentation broth were obtained at each time point. This was centrifuged to separate the cell pellet and the fermentation broth. The cell pellet was weighed to quantify the cell density, while the broth was further extracted with an equal volume of EtOAc. BE-18257A (**1**) and BE-18257B (**2**) in the dried organic extract were quantified using HPLC-MS via multiple reaction monitoring.

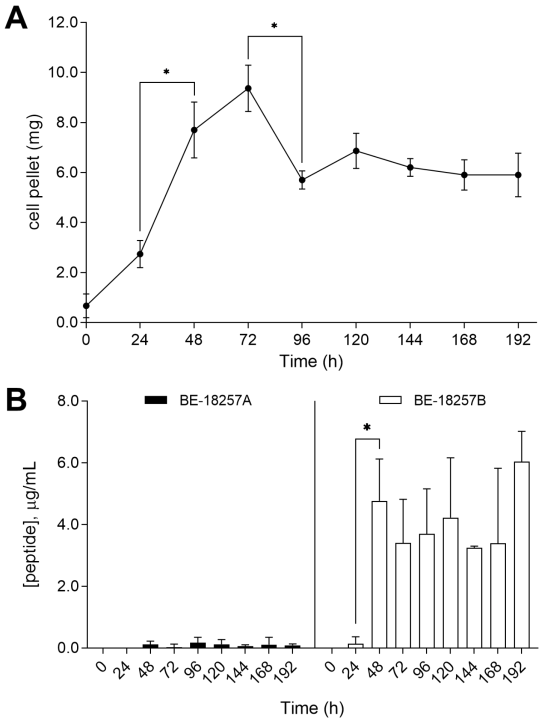


Figure 5: Time course of biomass (A) and cyclic peptide (B) production by *S. cacaoi* strain R2A-843A. The amount of biomass during specific time points is shown in A, values expressed as mean ± SD of three technical replicates. Metabolite concentrations of BE-18257A (1) (black) and BE-18257B (2) (white) in µg/mL determined using MRM at the corresponding time point are shown in B, values expressed as mean + SD of three technical replicates. Data representative from two biological replicates. Statistical analysis of two consecutive time points was performed using t-test (* denotes $p < 0.05$).

From the growth curve, the exponential growth phase or trophophase occurred until 72 h and signaled the occurrence of primary metabolism (Figure 5A). A significant decline in microbial biomass was observed at 72-96 h post-incubation. The stationary phase or idiophase, where microbial biomass production has peaked and is constant ($p > 0.05$), was observed up to 192 h of incubation. This microorganism showed a similar growth pattern as *Streptomyces rochei*, where the mycelium growth increased up to 72 h and afterward entered the stationary phase (Reddy 2011). For other actinomycetes such as *Pseudonocardia* sp., the stationary phase extends from 72-120 h of incubation which corresponds to the phase of highest antimicrobial activity (Kiranmayi 2011).

While **1** and **2** are structurally related compounds, the dynamics of production were observed to vary. BE-18257B (**2**) was consistently produced at a higher concentration compared to BE-18257A (**1**) at all time points (Figure 5B). Maximum BE-18257B (**2**) production was observed at 48 h post-incubation with a concentration range of 3.2-6.7 µg/mL and showed no significant change in the concentration until the end of the incubation period (Figure 5B). The highest concentration of BE-18257A (**1**) was observed at only 0.01-0.32 µg/mL starting at 24 h post-incubation and showed no significant change until the end of the incubation period.

DISCUSSION

The whole genome sequence of *S. cacaoi* strain R2A-843A showed a significant proportion of the genome dedicated to secondary metabolite production, with nonribosomal peptide synthesis being the major biosynthetic pathway. The biosynthetic machinery of *S. cacaoi* strain R2A-843A parallels that of *S. cacaoi* subsp. *cacaoi*. While the BGC for the nonribosomal peptides BE-18257 and pentaminomycins are conserved among *S. cacaoi*, the cacaoidin pathway is differentiating among the strains. Based on the analysis of cacaoidin BGC in public databases, this has been confined to strains of the *S. cacaoi* subsp. *cacaoi* (Roman-Hurtado et al. 2021A). In our study, we observed the $[M+3H]^{3+}$, $[M+2H]^{2+}$, $[M+H+Na]^{2+}$, and $[M+2Na]^{2+}$ molecular ions corresponding to cacaoidin and mapped the precursor peptide in the genome. However, the concentration of cacaoidin from the culture fermentation was deemed too low for further analysis.

The metabolome analysis of *S. cacaoi* strain R2A-843A corroborated with the genome analysis, which showed predominantly peptide-based metabolites. The pentaminomycins formed the largest cluster in the molecular network, indicating the production of the most number of analogues from this compound class. Purification of the compounds from the culture broth of *S. cacaoi* strain R2A-843A showed that BE-18257A (**1**) and BE-18257B (**2**) were the major isolable peptides. The ^{1}H NMR data for the two peptides were in agreement with the values indicating that the planar structure and absolute configurations of the stereocenters are identical for the purified compounds in this study and the previous reports by Kojiri et al. (1991) and Nakajima et al. (1991). This validated that the lack of epimerase in the genome prediction was an artifact of the assembly.

Despite numerous attempts to purify the pentaminomycins, the yield was very low and insufficient for structure characterization as single compounds. This is similar to the results of Roman-Hurtado et al. (2021B), where the pentaminomycins were shown to be produced at low concentrations, with cell culture media optimization having a limited effect on pentaminomycin production. Using the one-strain-many compounds (OSMAC) approach, the production of BE-18257A and pentaminomycins were monitored for 7, 14, and 21 days post-inoculation (Roman-Hurtado et al. 2021B). While pentaminomycins were produced using YEME and R2YE after 14 and 21 days, the yield remained low (Roman-Hurtado et al. 2021B). Despite BE-18257 and pentaminomycins being encoded by the same BGC, the major product is dependent on the microbial producer and not solely on the BGC (H. Wang et al. 2023). H. Wang et al. (2023) increased the production of pentaminomycins by preparing a double deletion mutant, with inactive BE-18257 and candicin production.

Previous work on *S. cacaoi* also indicated the significant influence of the culture medium on compound production. Cacaoidin for example was grown in MPG medium for 13 days (Ortiz-Lopez et al. 2020). A series of ionophores from *S. cacaoi* Waksman and Henrici (Streptomycetaceae) strain JX912350.1 were obtained using an optimized culture medium (Gezer et al.

2022). From the response surface methodology, 2.25% glycerol, 1% peptone water, 0.2% $CaCO_3$ and 0.1% $MgCl_2$ in distilled water at 30℃, 150 rpm for 10 days favored the production of four new ionophores belonging to the K41-A class of compounds. The biosynthetically talented *S. cacaoi* strain R2A-843A from the Philippines can hence be a good candidate for further elicitation of other potential cryptic compounds encoded in the genome.

CONCLUSIONS

We demonstrate the biosynthetic potential of the sponge-associated *Streptomyces cacaoi* strain R2A-843A. Whole genome sequencing showed the capability of this microorganism to produce a diverse set of compounds including nonribosomal peptides, polyketides, and terpenes. Metabolomics and MS-guided purification corroborated the BGC analysis, with peptide-based compounds being the major products when grown for seven days in R2A medium.

ACKNOWLEDGMENT

The authors acknowledge funding support from the Department of Science and Technology - Philippine Council for Health Research and Development through the Discovery and Development of Health Products - Marine Component Program. Genome sequencing was made possible through the CHED-PCARI IHITM63 Project. We thank Ms. Shalice R. Susana-Guevarra for conducting the bioactivity assay. This work was done under the supervision of the Bureau of Fisheries and Aquatic Resources under Gratuitous Permit No. FBP-0035-10. This is MSI Contribution No. 501.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

ZBLM- Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization; JMR- Conceptualization, Methodology, Formal Analysis, Investigation, Data Curation, Writing – original draft, Writing – review & editing, Visualization; BDL and KBD- Formal analysis, Writing – review & editing; GPC- Resources, Writing – review & editing; LAS-R- Conceptualization, Methodology, Validation, Formal Analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project Administration, Funding Acquisition.

REFERENCES

Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008; 9:1-15. doi: 10.1186/1471-2164-9-75.

Bai, Y, Yi, P, Zhang, Y, Hu, J, Wang, Y, Ju, J, Pan, H. Structure-based molecular networking for the target discovery of novel germicidin derivatives from the sponge-associated *Streptomyces* sp. 18A01. J. Antibiot. 2021; 74:799–806. doi:10.1038/s41429-021-00447-w

Bérdy, J. Bioactive microbial metabolites. J. Antibiot. 2005; 58:1–26. doi: 10.1038/ja.2005.1.

Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023; 51:46-50. doi: 10.1093/nar/gkad344

Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA 3rd, Stevens R, Vonstein V, Wattam AR, Xia F. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015; 5:1-6. doi: 10.1038/srep08365.

Carretero-Molina, D, Ortiz-Lopez, FJ, Martin, J, Gonzales, I, Sanchez-Hidalgo, M, Roman-Hurtado, F, Diaz, C., de la Cruz, O, Genilloud, O, Reyes F. Pentaminomycins F and G, nonribosomal peptides containing 2-pyridylalanine. J. Nat. Prod. 2021; 84:1127-1134. doi: 10.1021/acs.jnatprod.0c01199.

Crüsemann, M, O'Neill, EC, Larson, CB, Melnik, AV, Floros, DJ, Da Silva, RR, Jensen, PR, Dorrestein, PC, Moore, BS. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. J. Nat. Prod. 2017; 80:588-597. doi: 10.1021/acs.jnatprod.6b00722.

Crüsemann, M. Coupling mass spectral and genomic information to improve bacterial natural product discovery workflows. Mar. Drugs, 2021; 19: 1-10. doi: 10.3390/MD19030142.

Danecek, P, Bonfield, JK, Liddle, J, Marshall, J, Ohan, V, Pollard, MO, Whitwham, A, Keane, T, McCarthy, SA, Davies, RM, Li, H. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10:1-4. doi: 10.1093/gigascience/giab008.

Darling, AE, Mau, B, Perna, NT. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010; 5:1-17. doi: 10.1371/journal.pone.0011147.

Ding, CYG, Pang, LM, Liang, ZX, Goh, KKK, Glukhov, E, Gerwick, WH, Tan, LT. MS/MS-based molecular networking approach for the detection of aplysiatoxin-related compounds in environmental marine cyanobacteria. Mar. Drugs, 2018; 16: 1–15. doi: 10.3390/md16120505.

Doroghazi, JR, Albright, JC, Goering, AW, Ju, KS, Haines, RR, Tchalukov, KA, Labeda, DP, Kelleher, NL, Metcalf, WW. A roadmap for natural product discovery based on large-scale

genomics and metabolomics. Nat. Chem. Biol. 2014; 10:963–968. doi: 10.1038/nCHeMBIO.1659.

Edwards, RA, Edwards, JA. fastq-pair: efficient synchronization of paired-end fastq files. bioRxiv. 2019; 1-5. doi: 10.1101/552885.

Gezer, E, Üner, G, Küçüksolak, M, Kurt, MÜ, Doğan, G, Kırmızıbayrak, PB, Bedir, E. Undescribed polyether ionophores from *Streptomyces cacaoi* and their antibacterial and antiproliferative activities. Phytochem. 2022; 195: 1-9. doi: 10.1016/j.phytochem.2021.113038.

Goering, AW, McClure, RA, Doroghazi, JR, Albright, JC, Haverland, NA, Zhang, Y, Ju, KS, Thomson, RJ, Metcalf, WW, Kelleher, NL. Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. ACS Cent. Sci. 2016; 2: 99–108. doi: 10.1021/acscentsci.5b00331.

Hautbergue, T, Jamin, EL, Costantino, R, Tadrist, S, Meneghetti, L, Tabet, JC, Debrauwer, L, Oswald, IP, Puel, O. Combination of isotope labeling and molecular networking of tandem mass spectrometry data to reveal 69 unknown metabolites produced by *Penicillium nordicum.* Anal. Chem. 2019; 91: 12191–12202. doi: 10.1021/acs.analchem.9b01634.

Hwang, S, Luu Le, LTH, Jo, SI, Shin, J, Lee, MJ, Oh, DC. Pentaminomycins C–E: Cyclic pentapeptides as autophagy inducers from a mealworm beetle gut bacterium. Microorganisms. 2020; 8: 1–16. doi: 10.3390/microorganisms8091390.

Jang, JP, Hwang, GJ, Kwon, MC, Ryoo, IJ, Jang, M, Takahashi, S, Ko, SK, Osada, H, Jang, JH, Ahn, JS. Pentaminomycins A and B, hydroxyarginine-containing cyclic pentapeptides from *Streptomyces* sp. RK88-1441. J. Nat. Prod. 2018; 81: 806–810. doi: 10.1021/acs.jnatprod.7b00882.

Jang, JP, Kwon, MC, Nogawa, T, Takahashi, S, Osada, H, Ahn, JS, Ko, SK, Jang, JH. Thiolactomide: A new homocysteine thiolactone derivative from *Streptomyces* sp. with neuroprotective activity. J. Microbiol. Biotechnol., 2021; 31: 1667–1671. doi: 10.4014/jmb.2108.08015.

Jarmusch, SA, Van Der Hooft, JJJ, Dorrestein, PC, & Jarmusch, AK. Advancements in capturing and mining mass spectrometry data are transforming natural products research. Nat. Prod. Rep. 2021; 38:2066–2082. doi: 10.1039/d1np00040c.

Jizba, JV, Sedmera, P, Vanek, Z, Drautz, H, Zahner, H. Two thiolactones from *Streptomyces* TU 2476. J. Antibiot. 1985; 38: 111–112. doi: 10.7164/antibiotics.38.111.

Kaweewan, I, Komaki, H, Hemmi, H, & Kodani, S. Isolation and structure determination of new antibacterial peptide curacomycin based on genome mining. Asian J. Org. Chem. 2017; 6: 1838–1844. doi: 10.1002/ajoc.201700433.

Kaweewan, I, Komaki, H, Hemmi, H, Hoshino, K, Hosaka, T, Isokawa, G, Oyoshi, T, Kodani, S. Isolation and structure determination of a new cytotoxic peptide, curacozole, from *Streptomyces curacoi* based on genome mining. J. Antibiot. 2019; 72: 1–7. doi: 10.1038/s41429-018-0105-4.

Kaweewan, I, Hemmi, H, Komaki, H, Kodani S. Isolation and structure determination of a new antibacterial peptide pentaminomycin C from *Streptomyces cacaoi* subsp. *cacaoi*. J. Antibiot. 2020; 73:224–229. doi: 10.1038/s41429-019-0272-y.

Khushi, S, Nahar, L, Salim, A, Capon, RJ. Trachycladindoles H–M: Molecular networking guided exploration of a library of Southern Australian marine sponges. Aust. J. Chem. 2020. 73(4), 338–343.

Kiranmayi, MU, Sudhakar, P, Sreenivasulu, K, Vijayalakshmi, M. Optimization of culturing conditions for improved production of bioactive metabolites by *Pseudonocardia* sp. VUK-10. Mycobiology. 2011; 39: 174–181. doi: 10.5941/MYCO.2011.39.3.174.

Kojiri, K, Ihara, M, Nakajima, S, Kawamura, K, Funaishi, K, Yano, M, Suda, H. Endothelin-binding inhibitors, BE-18257A and BE-18257B (I. Taxonomy, Fermentation, Isolation, and Characterization). J. Antibiot. 1991; 44:1342-1347.

Krause, J. Applications and restrictions of integrated genomic and metabolomic screening: An accelerator for drug discovery from actinomycetes? Molecules. 2021; 26:1-11. doi: 10.3390/molecules26185450

Kulkarni, A, Zeng, Y, Zhou, W, Van Lanen, S, Zhang, W, Chen, S. A branch point of *Streptomyces* sulfur amino acid metabolism controls the production of albomycin. Appl. Environ. Microbiol. 2016; 82;467–477. doi: 10.1128/AEM.02517-15.

Langmead, B, Salzberg, S. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 2012; 9:357–359. doi: 10.1038/nmeth.1923.

Lee, N, Hwang, S, Kim, J, Cho, S, Palsson, B, Cho, B-K. Mini review: Genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. Comp. Struct. Biotech. 2020; 18: 1548–1556. doi: 10.1016/j.csbj.2020.06.024.

Liu, LL, Chen, ZF, Liu, Y, Tang, D, Gao, HH, Zhang, Q, Gao, JM. Molecular networking-based for the target discovery of potent antiproliferative polycyclic macrolactam ansamycins from *Streptomyces cacaoi* subsp. *asoensis*. Org. Chem. Front. 2020; 7: 4008–4018. https://doi.org/10.1039/d0qo00557f

Liu, Y, Ding, L, Deng, Y, Wang, X, Cui, W, He, S. Feature-based molecular networking-guided discovery of siderophores from a marine mesophotic zone *Axinellida* sponge-associated actinomycete *Streptomyces diastaticus* NBU2966. Phytochem. 2022; 196: 1-11. doi: 10.1016/j.phytochem.2021.113078

Meyers, KE, Sethna C. Endothelin antagonists in hypertension and kidney disease. Pediatr Nephrol. 2013; 28: 711-720. doi: 10.1007/s00467-012-2316-4.

Mikheenko, A, Prjibelski, A, Saveliev, V, Antipov, D Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018; 34: 1-9. doi: 10.1093/bioinformatics/bty266.

Mudalungu, CM, Von Törne, WJ, Voigt, K, Rückert, C, Schmitz, S, Sekurova, ON, Zotchev, SB, Süssmuth, RD. Noursamycins, chlorinated cyclohexapeptides identified from molecular networking of *Streptomyces noursei* NTR-SR4. J. Nat. Prod. 2019; 82: 1478–1486. doi:10.1021/acs.jnatprod.8b00967

Nakajima, S, Niiyama, K, Ihara, M, Kojiri, K, Suda, H. Endothelin-binding inhibitors, BE-18257A and BE-18257B II. Structure Determination. J. Antibiot. 1991; 44: 1348–1356.

Naman, CB, Rattan, R, Nikoulina, SE, Lee, J, Miller, BW, Moss, NA, Armstrong, L, Boudreau, PD, Debonsi, HM, Valeriote, FA, Dorrestein, PC, & Gerwick, WH. Integrating molecular networking and biological assays to target the isolation of a cytotoxic cyclic octapeptide, samoamide A, from an American Samoan marine cyanobacterium. J. Nat. Prod. 2017; 80: 625–633. doi: 10.1021/acs.jnatprod.6b00907

Nett, M, Ikeda, H, Moore, BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat. Prod. Rep. 2009; 26: 1362–1384. doi: 10.1039/b817069j.

Ortiz-López, FJ, Carretero-Molina, D, Sánchez-Hidalgo, M, Martín, J, González, I, Román-Hurtado, F, de la Cruz, M, García-Fernández, S, Reyes, F, Deisinger, JP, Müller, A, Schneider, T, Genilloud, O. Cacaoidin, first member of the new lanthidin RiPP family. Angew. Chemie - International Edition, 2020; 59: 12654–12658. doi: 10.1002/anie.202005187.

Ouchene, R, Stien, D, Segret, J, Kecha, M, Rodrigues, AMS, Veckerlé, C, Suzuki, MT. Integrated metabolomic, molecular networking, and genome mining analyses uncover novel angucyclines from *Streptomyces* sp. RO-S4 strain isolated from Bejaia Bay, Algeria. Front. Microbiol. 2022; 13: 1–17. doi: 10.3389/fmicb.2022.906161.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014; 24: 206-214. doi: 10.1093/nar/gkt1226.

Palazzotto, E, Weber, T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. Curr. Opin. Microbiol. 2018; 45: 109–116. doi: 10.1016/j.mib.2018.03.004

Parveen, N, Cornell, KA. Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. Mol. Microbiol. 2011; 79: 7–20. doi: 10.1111/j.1365-2958.2010.07455.x

Pluskal, T, Castillo, S, Villar-Briones, A, Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 2010; 11: 1-11. doi: 10.1186/1471-2105-11-395

Prjibelski, A, Antipov, D, Meleshko, D, Lapidus, A, Korobeynikov, A. Using SPAdes de novo assembler. Curr. Protoc. Bioinformatics. 2020; 70: 1-29. doi: 10.1002/cpbi.102.

Reddy, N, Ramakrishna, D, Rajagopal, S. Optimization of culture conditions of *Streptomyces rochei* (MTCC 10109) for the production of antimicrobial metabolites. Egypt. J. Biol. 2011; 13: 21–29. doi: 10.4314/ejb.v13i1.4.

Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2016; 32: 929–931. doi: 10.1093/bioinformatics/btv681.

Roman-Hurtado, F, Sanchez-Hidalgo, M, Martin, J, Ortiz-Lopez, FJ, Genilloud, O. Biosynthesis and heterologous expression cacaoidin, the first member of the lanthidin family of RiPPs. Antibiotics. 2021. 10: 1-16. doi: 10.3390/antibiotics10040403.

Roman-Hurtado, F, Sanchez-Hidalgo, M, Martin, J, Ortiz-Lopez, FJ, Carretero-Molina, D, Reyes, F, Genilloud, O. One pathway, two cyclic non-ribosomal pentapeptides: Heterologous expression of BE-18257 antibiotics and pentaminomycins from *Streptomyces cacaoi* CA-170360. Microorganisms. 2021; 9: 1-13. doi: 10.3390/ microorganisms9010135.

Shannon, P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, Ramage, D, Amin, N, Schwikowski, B, Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498–2504. doi: 10.1101/gr.1239303.metabolite.

Skinnider, MA, Johnston, CW, Gunabalasingam, M, Merwin, NJ, Kieliszek, A., MacLellan, RJ, Li, H, Ranieri, MRM., Webster, ALH, Cao, MPT., Pfeifle, A, Spencer, N, To, QH, Wallace, DP, Dejong, CA, Magarvey, NA. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat. Commun. 2020; 11: 1-9. doi: 10.1038/s41467-020-19986-1

Soldatou, S, Eldjarn, GH, Huerta-Uribe, A, Rogers, S, Duncan, KR. Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. FEMS Microbiol. Lett. 2019; 366: 1–8. doi: 10.1093/femsle/fnz142.

Son, S, Jang, M, Lee, B, Hong, YS, Ko, SK, Jang, JH, Ahn, JS. Ulleungdin, a lasso peptide with cancer cell migration inhibitory activity discovered by the genome mining approach. J. Nat. Prod. 2018; 81: 2205–2211. doi: 10.1021/acs.jnatprod.8b00449.

Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011; 27: 1009-10. doi: 10.1093/bioinformatics/btr039.

Susana, SR, Salvador-Reyes, LA. Anti-inflammatory activity of monosubstituted xestoquinone analogues from the marine sponge *Neopetrosia compacta*. Antioxidants. 2022; 11: 607. doi:/10.3390/antiox11040607.

Teta, R, Della Sala, G, Glukhov, E, Gerwick, L, Gerwick, WH, Mangoni, A, Costantino, V. Combined LC-MS/MS and molecular networking approach reveals new cyanotoxins from the 2014 cyanobacterial bloom in Green Lake, Seattle. Environ. Sci. Technol. 2015; 49: 14301–14310. doi: 10.1021/acs.est.5b04415

Tian X, Zhang Z, Yang T, Chen M, Li J, Chen F, Yang J, Li W, Zhang B, Zhang Z, Wu J, Zhang C, Long L and Xiao J. Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. Front. Microbiol. 2016; 7: 1-16. doi: 10.3389/fmicb.2016.00998

Via, CW, Glukhov, E, Costa, S, Zimba, PV, Moeller, PDR, Gerwick, WH, Bertin, MJ. The metabolome of a cyanobacterial bloom visualized by MS/MS-based molecular networking reveals new neurotoxic smenamide analogs (C, D, and E). Front. Chem. 2018; 6: 1–9. doi: 10.3389/fchem.2018.00316

Wang, H, Yi, X, Zhou, Z, Yang, J, Pei, Y, Shi, S, Gao, C, Tian, X, Ju, J, Li, Q. Metabolic blockade-based genome mining of *Streptomyces cacaoi* SCSIO 68063: Isolation and identification of BE-18257 and pentaminomycin analogues. Tetrahedron. 2023; 130: 1-9. doi: 10.1016/j.tet.2022.133148.

Wang, M, Carver, JJ, Phelan, VV, Sanchez, LM, Garg, N, Peng, Y, Nguyen, DD, Watrous, J, Kapono, CA, Luzzatto-Knaan, T, Porto, C, Bouslimani, A, Melnik, AV, Meehan, MJ, Liu, WT, Crüsemann, M, Boudreau, PD, Esquenazi, E, Sandoval-Calderón, M, … Bandeira, N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 2016. 34: 828–837 doi: 10.1038/nbt.3597.

Wang, X, Subko, K, Kildgaard, S, Frisvad, JC, Larsen, TO. Mass Spectrometry-based network analysis reveals new insights into the chemodiversity of 28 species in *Aspergillus* section *Flavi*. Front. Fungal Biol. 2021; 2: 1–13. doi: 10.3389/ffunb.2021.719420.

Zeng, Y, Kulkarni, A, Yang, Z, Patil, PB, Zhou, W, Chi, X, Van Lanen, S, Chen, S. Biosynthesis of albomycin δ2 provides a template for assembling siderophore and aminoacyl-tRNA synthetase inhibitor conjugates. ACS Chem. Biol. 2012; 7: 1565–1575. doi: 10.1021/cb300173x

SUPPLEMENTARY INFORMATION

Table S1: Top 5 matches of extracted 16S rRNA sequence from the genome assembly against the 16S rRNA BLAST database.

Accession No.	Description	% Identity	Query Coverage	E-value	Bitscore
NR_041061.1	*Streptomyces cacaoi* strain NBRC 12748 16S ribosomal RNA, partial sequence	100.00	97	0	1949
NR_029087.1	*Streptomyces ferralitis* strain SFOp68 16S ribosomal RNA, partial sequence	98.42	99	0	1893
NR_180104.1	*Streptomyces alkaliterrae* strain OF1 16S ribosomal RNA, complete sequence	98.33	99	0	1890
NR_025615.1	*Streptomyces albus* strain DSM 40313 16S ribosomal RNA gene, partial sequence	98.33	99	0	1890
NR_114917.1	*Streptomyces erumpens* strain DSM 40941 16S ribosomal RNA, partial sequence	98.23	99	0	1882

Table S2: Top 5 matches in JSpecies whole-genome tetranucleotide correlation search.

Description	Z-score*
Streptomyces cacaoi NRRL B-1220	0.99993
Streptomyces cacaoi NBRC 12748	0.99992
Streptomyces cacaoi DSM 40057	0.99983
Streptomyces albus NRRL B-1685	0.97850
Streptomyces albus NRRL B-1335	0.97828

* Z-score >0.999 indicate reliable classification up to species level

Table S3: Genome Assembly and Annotation Metrics for *S. cacaoi* R2A-843A

Assembled Size (bp)	8,551,568
GC (%)	73.28
Contigs	861
Largest Contig Size (bp)	90,360
N50 / L50	17,925 / 140
Ave. Coverage Depth	101
Complete BUSCO (%)	93.92

Table S4: BLAST alignment details of mapped contigs to *S. cacaoi* CA-170360 *cpp* BGC (Genbank Accession No. MW038823.1; 48,011 bp)

Contig Name	Query From	Query To	Subject Length (bp)	Identity (%)	Align Cov (%)	E-value	Bit Score
NODE_95_length_22269_cov_52.377478	1	7,381	22,269	98.38	33.21	0	12,979
NODE_583_length_3291_cov_59.864655	7,884	11,144	3,291	97.96	100.00	0	5,679
NODE_860_length_512_cov_41.579310	11,135	11,649	512	96.89	100.00	0	859
NODE_153_length_16878_cov_54.935659	12,088	28,981	16,878	99.38	99.99	0	30,600
NODE_464_length_5337_cov_55.757605	29,035	34,368	5,337	98.95	100.00	0	9,542
NODE_140_length_17925_cov_50.650829	34,901	42,344	17,925	98.47	41.51	0	13,112

Table S5: Comparison of *cpp* adenylation domains between *S. cacaoi* CA-170360 (Roman-Hurtado et al., 2021) and R2A-843A

Module	AA Identity (%)	CA-170360 Substrate	R2A-843A Predicted Substrate	
			AntiSMASH	PRISM (Score)
BE-18275 A-C (*cppB*)				
L	99.0	Leucine	**Leucine**	Valine (408.0) Phenylalanine (403.3) Hyrdoxyphenylglycine (393.2) Tryptophan (360.3) Threonine (355.2) **Leucine (354.4)** Aspartate (349.7) Lysine (331.5) 3,5-dihydroxyphenylglycine (326.2) Glycine (109.7)
M1	97.32	Tryptophan	?	**Tryptophan (527.1)** Phenylalanine (488.3) Lysine (456.8) Serine (441.2) Tyrosine (439.6) Ornithine (434.3) Glycine (424.5) Hydroxyacetyl-glycine (420.5) Alanine (418.6) Beta-alanine (112.8)
M2	98.99	Leucine/Serine	?	**Serine (675.7)** Glutamate (526.4) Alanine (522.0) Glycine (505.9) Phenylalanine (502.5) Asparatate (487.7) Lysine (486.9) Arginine (479.9) Threonine (454.7) β-aminoisobutyric acid (f) (114.5)
M3	100.00	Alanine	**Alanine**	**Alanine (495.2)** Aspartate (431.3) Tryptophan (400.2) Asparagine (397.9) Arginine (395.7) Piperazic acid (395.1) Hydroxyacetyl-ornithine (383.0) Histidine (347.8) Modified tyrosine (329.1) Pipecolic acid (109.0)

M4	100.00	Valine/Leucine	**Valine**	**Valine (493.9)** Phenylalanine (490.4) Hydroxyphenylglycine (471.4) Threonine (461.8) Lysine (453.5) Isoleucine (452.4) Arginine (441.6) Aspartate (441.2) 6-chloro-4-hydroxyindole-3-carboxylic acid (111.1)
Pentaminomycin (*cppM*)				
L	99.31	Valine/Leucine/ Phenylalanine	**Leucine**	Hydroxyphenylglycine (350.7) **Valine (327.0)** **Phenylalanine (313.7)** Threonine (304.5) Tryptophan (304.0) Glycine (290.1) Lysine (288.5) Serine (283.1) Alanine (275.2) N5-hydroxyornithine (109.0)
M1	98.79	Valine	**Valine**	Phenylalanine (488.3) **Valine (470.9)** Hydroxyphenylglycine (446.6) Isoleucine (446.0) Tryptophan (443.4) Ornithine (440.8) Serine (438.0) Lysine (433.8) Leucine (433.5) Hydroxyvaline (110.2)
M2	99.97	Tryptophan	**Tryptophan**	Serine (429.2) Glycine (420.5) **Tryptophan (409.7)** Ornithine (371.8) Phenylalanine (368.9) Proline (364.1) Tyrosine (359.9) Hydroxyphenylglycine (351.2) Lysine (330.8) α-ketoisovalerate (108.8)
M3	98.75	Arginine	**Arginine**	Ornithine (490.5) **Arginine (468.1)** Lysine (440.7) Aspartate (438.6) Hydroxyacetyl-ornithine (418.3) Phenylalanine (416.1) Glutamate (414.5) Glutamine (385.3) β-hydroxy-aspartic acid (352.0) Anthranilic acid (110.1)
M4	99.74	Leucine/ Phenylalanine	?	**Phenylalanine (503.9)** Tryptophan (494.3) Tyrosine (460.9) Alanine (459.2) Serine (443.5) Lysine (434.1) Hydroxyacetyl-ornithine (429.8) **Leucine (424.5)** Glycine (421.0) Anthranilic acid (111.8)

Table S6: BLAST alignment details of mapped contigs to *S. cacaoi* CA-170360 cacaoidin BGC (Genbank Accession No. MT210103.1; 29,760 bp)

Contig Name	Query From	Query To	Subject Length (bp)	Identity (%)	Align Cov (%)	E-value	Bit Score
NODE_374_length_7407_cov_44.319100	1	2,828	7,407	96.78	38.15	0	4,711
NODE_191_length_14452_cov_45.536278	2,858	17,345	14,452	98.97	100.00	0	25,896
NODE_38_length_35596_cov_47.836144	17,363	26,034	35,596	96.89	24.36	0	15,633

Table S7: Bioactivities of compound 1 and 2.

	% cell viability			
	HCT116		MCF-7	
	10 µg/mL	1 µg/mL	10 µg/mL	1 µg/mL
BE-18257A (**1**)	102.82 ± 2.79	131.21 ± 7.76	58.05 ± 8.38	110.92 ± 4.57
BE-18257B (**2**)	103.10 ± 2.03	128.02 ± 4.71	23.79 ± 4.87	92.47 ± 1.37
Taxol (100 nM)	20.80 ± 1.22		65.04 ± 0.96	
Taxol (1 nM)	88.88 ± 9.05		80.60 ± 4.13	

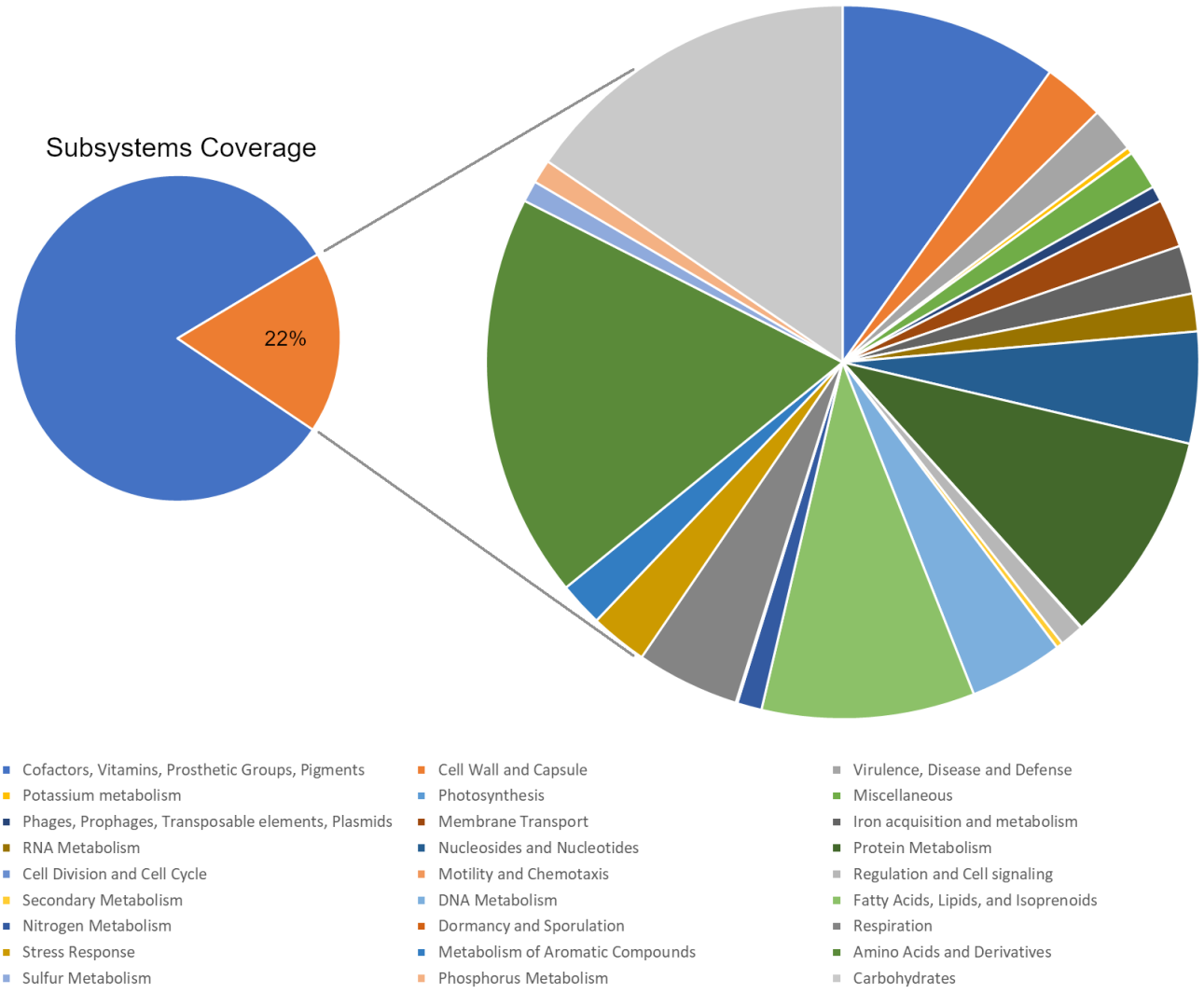


Figure S1: Functional annotation of *S. cacaoi* R2A-843A using RAST Subsystems.

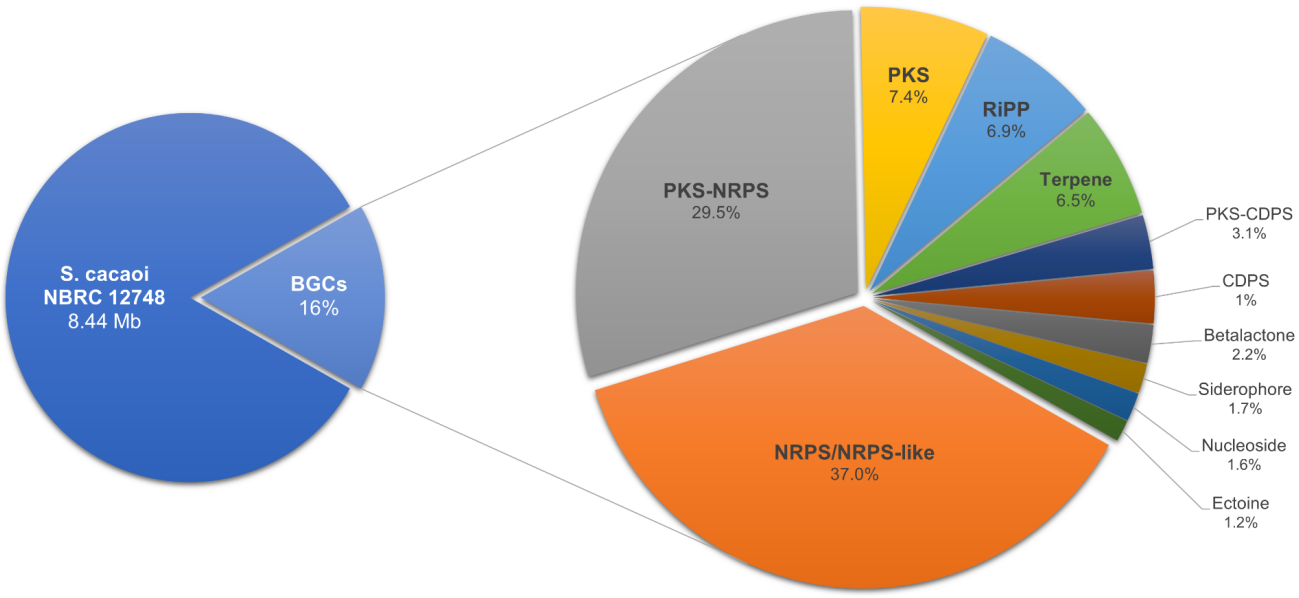


Figure S2: Genome fraction of *S. cacaoi* NBRC 12748 and predicted secondary metabolites encoded in the genome.

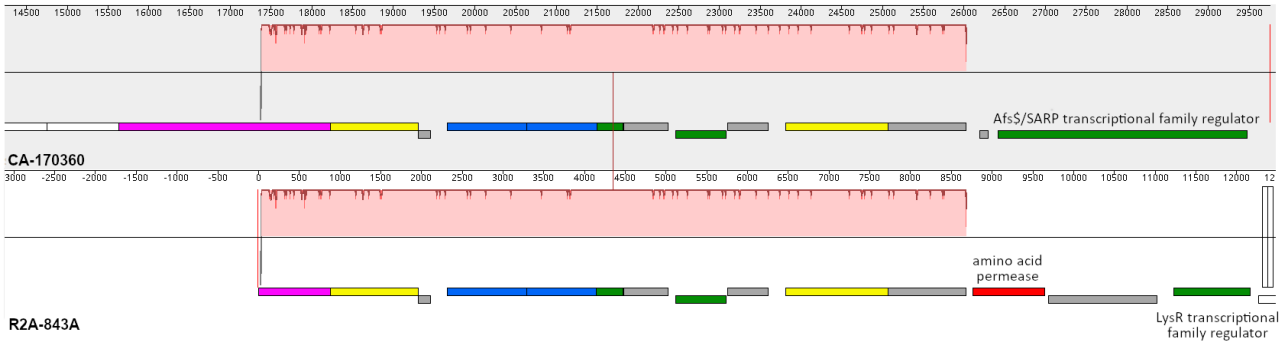


Figure S3: Alignment of *S. cacaoi* CA-170360 and R2A-843A showing difference in gene content at the end of the cacaoidin BGC.

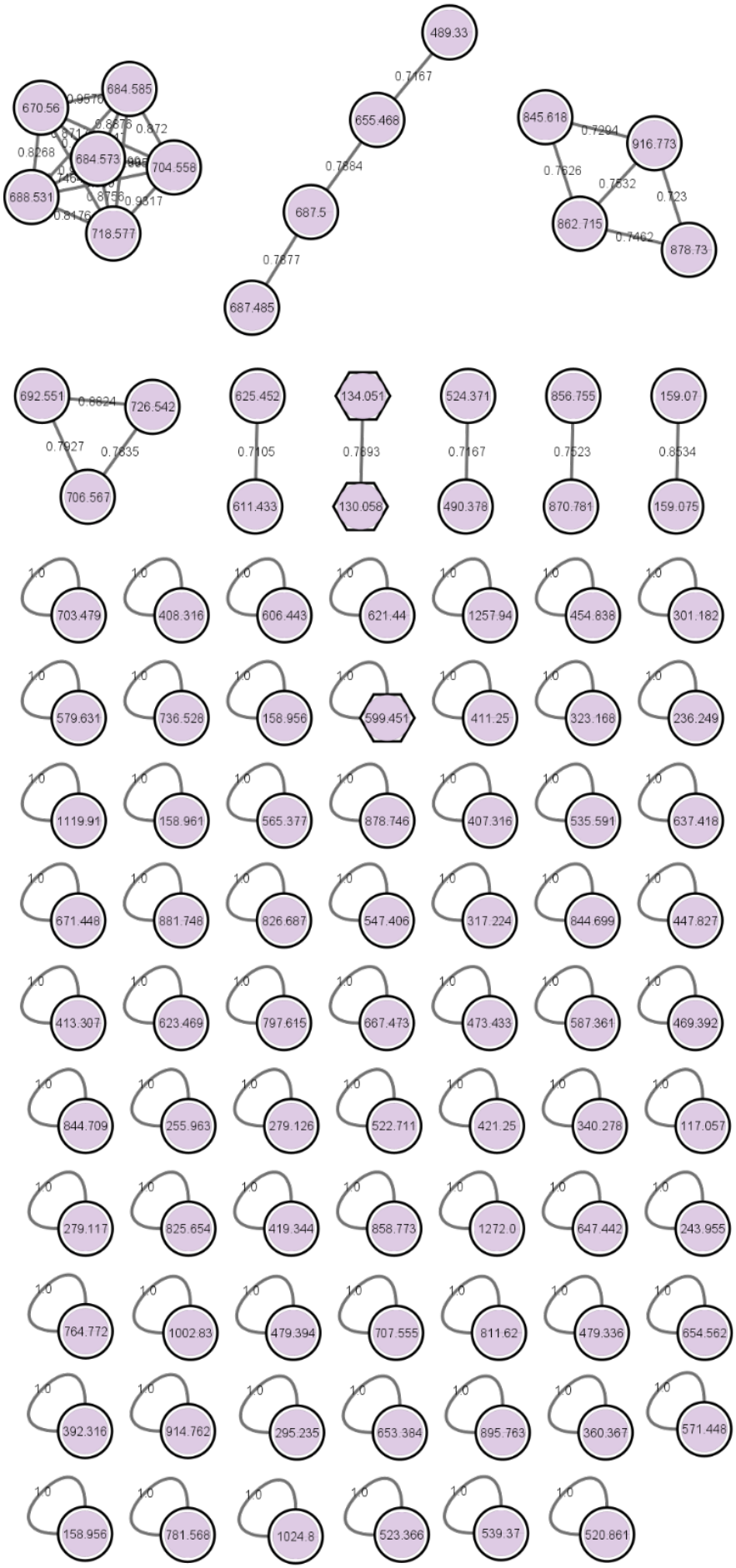


Figure S4: Full molecular network generated from the HRMS/MS spectra of the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A using the Global Natural Products Social Molecular Networking (GNPS) platform.

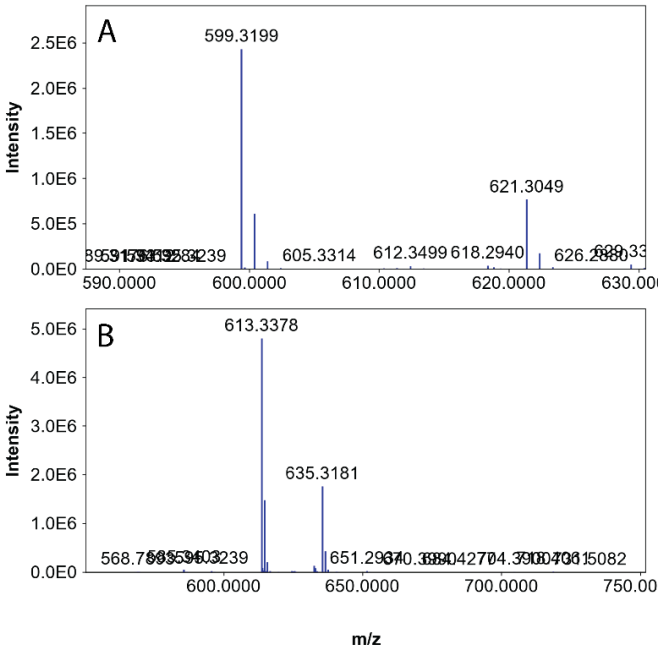


Figure S5: HRESIMS spectra of 1 (A) and 2 (B) from the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A.

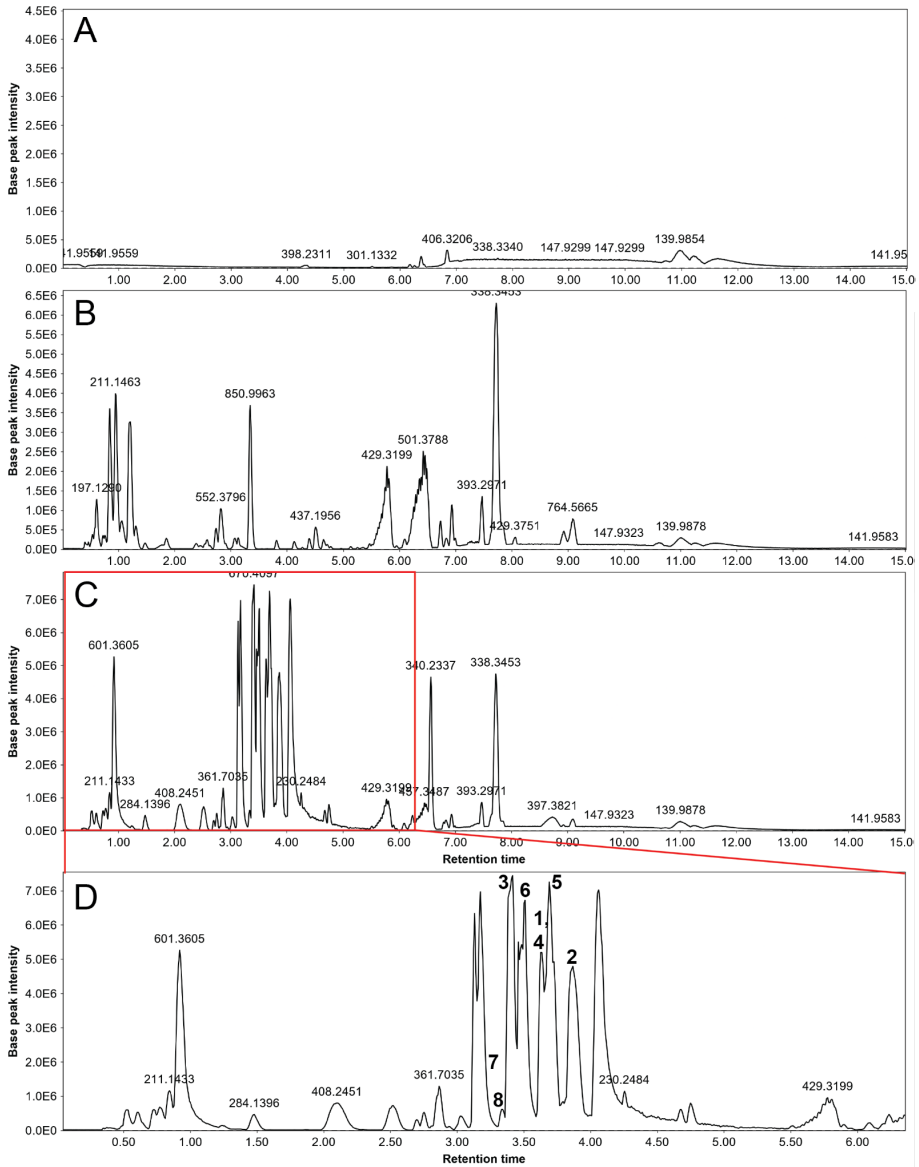


Figure S6: HRESIMS chromatograms of solvent blank (A), crude 100% MeOH Diaion extracts of R2A media only (B) and of *S. cacaoi* strain R2A-843A (C). An enlarged portion of t_R = 0.0-6.0 min is shown in D and labeled with the identified compounds in this study.

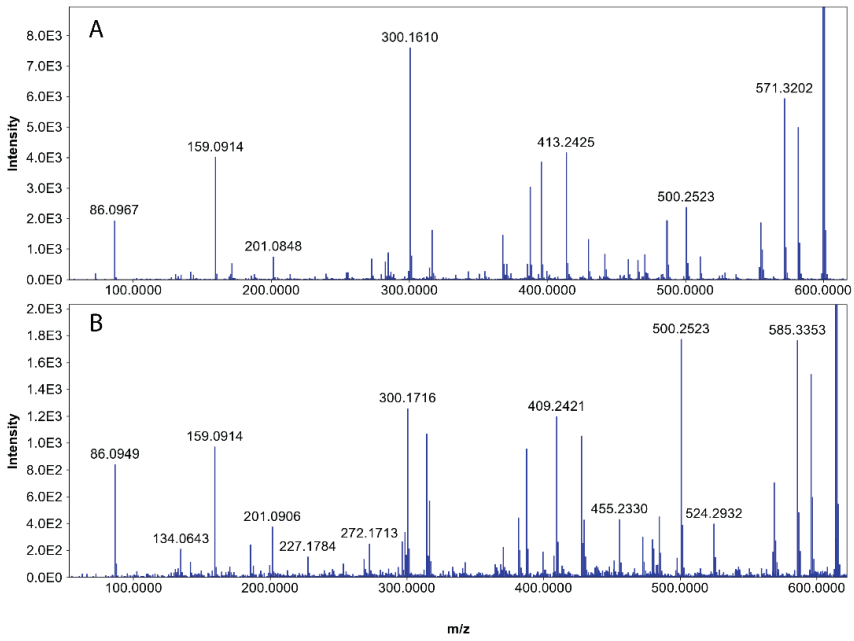


Figure S7: HRESIMS/MS spectra of 1 (A) and 2 (B) from the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A using ramp collision energy 20-40 eV.

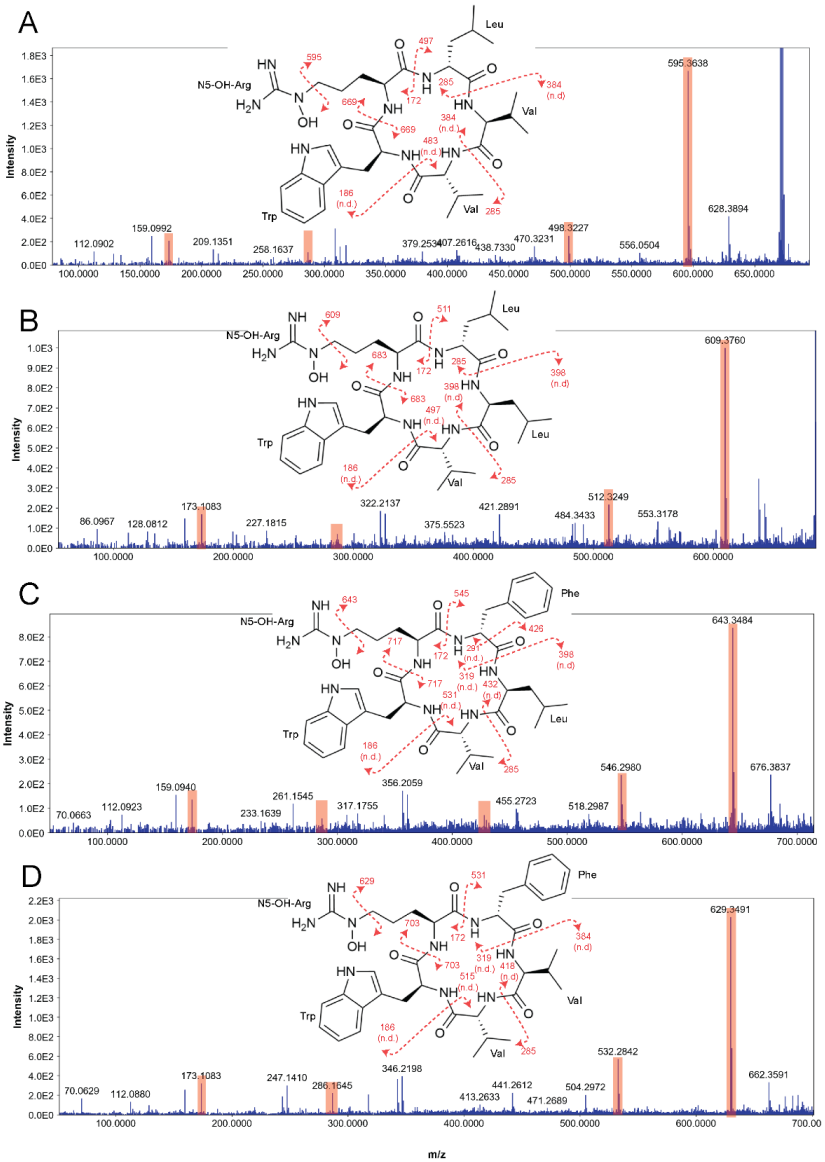


Figure S8: HRESIMS/MS spectra and structures of 3 to 6 (A to D) from the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A using ramp collision energy 20-40 eV. Highlighted in red are the detected fragment ions. The dashed lines through the structures indicate the "y" and "b" fragments and the described numbers indicate the corresponding m/z value (n.d.: not detected).

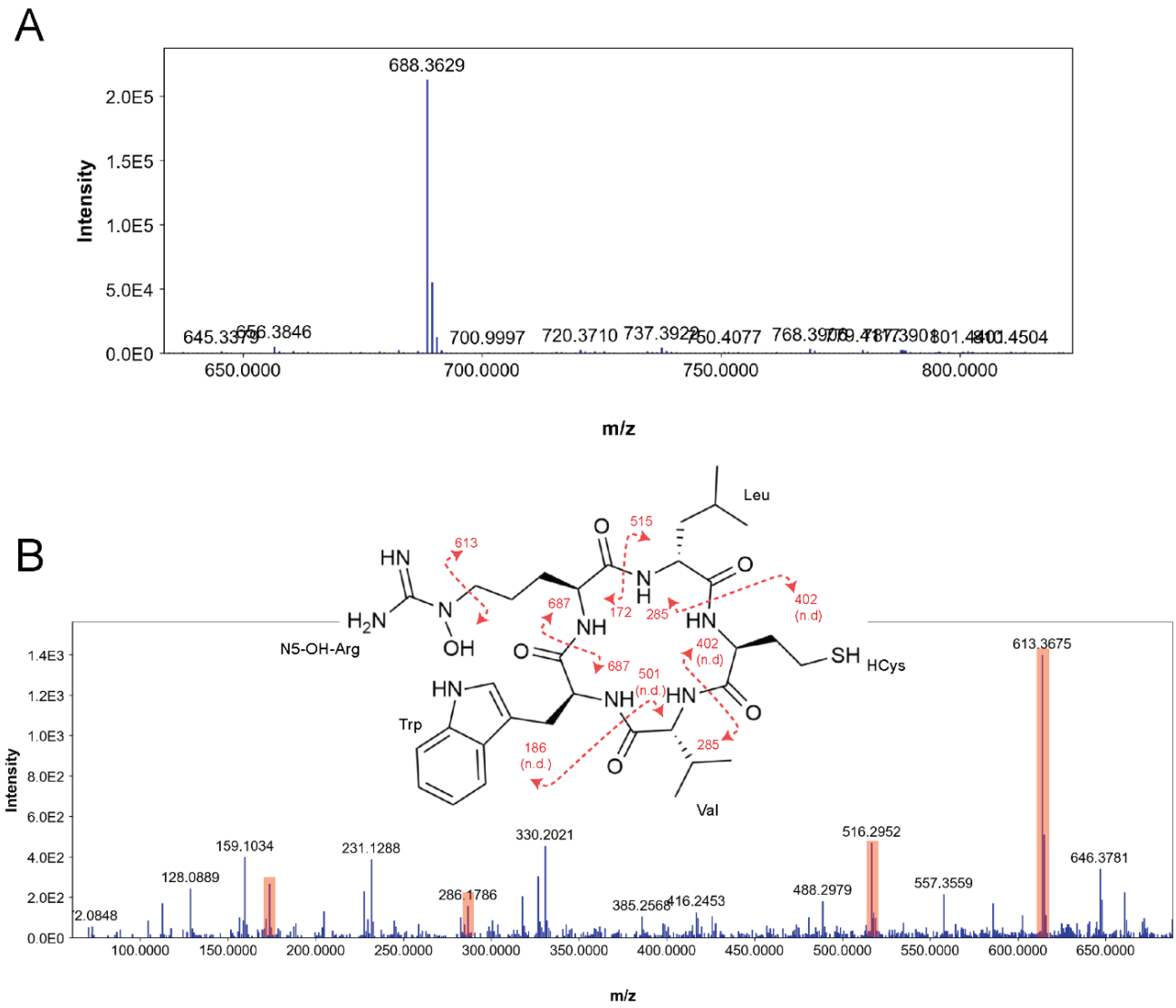


Figure S9: HRESIMS (A), HRESIMS/MS spectra and proposed structure (B) of 7 from the crude 100% MeOH Diaion extract of *S. cacaoi* strain R2A-843A using ramp collision energy 20-40 eV. Highlighted in red are the detected fragment ions. The dashed lines through the structure indicate the "y" and "b" fragments and the described numbers indicate the corresponding m/z value (n.d.: not detected).

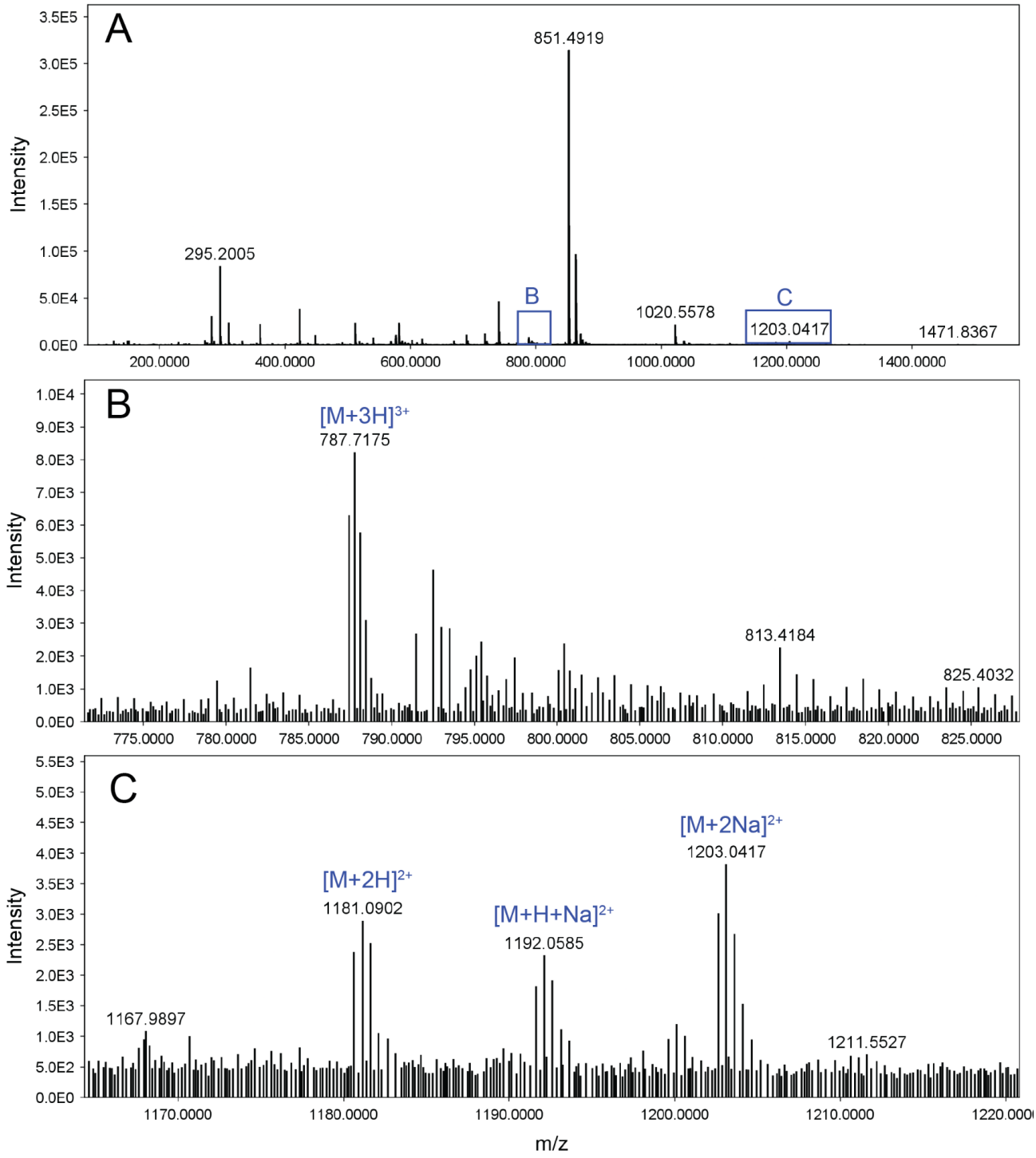
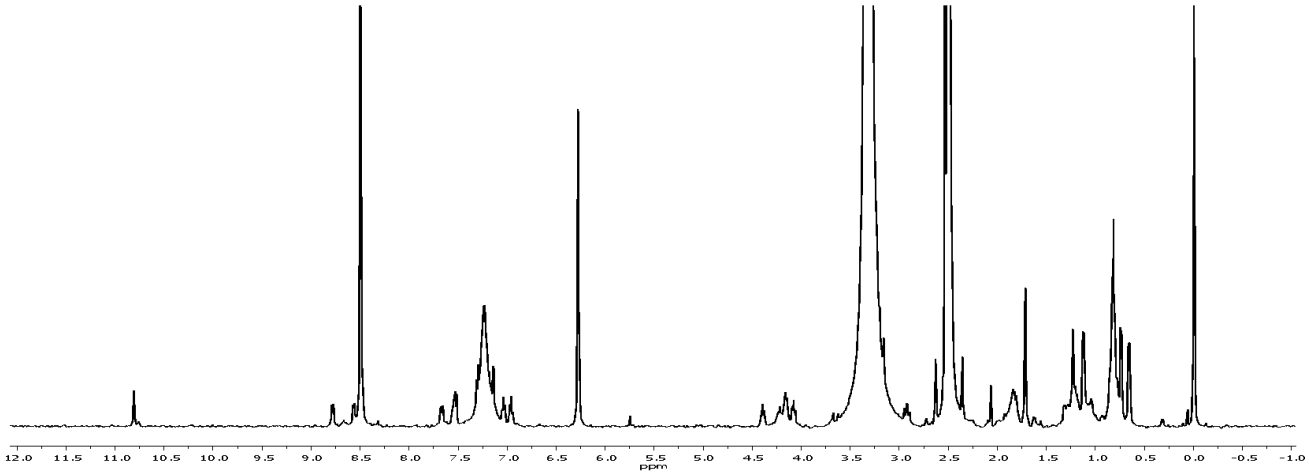


Figure S10: HRESIMS evidence of cacaoidin (8). Mass spectrum of R2A-843A at t_R=3.32 min (A) where the triply charged adduct $[M+3H]^{3+}$ (B), and doubly charged adducts $[M+2H]^{2+}$, $[M+H+Na]^{2+}$, and $[M+2Na]^{2+}$ (C) for 8 were observed.

A



B

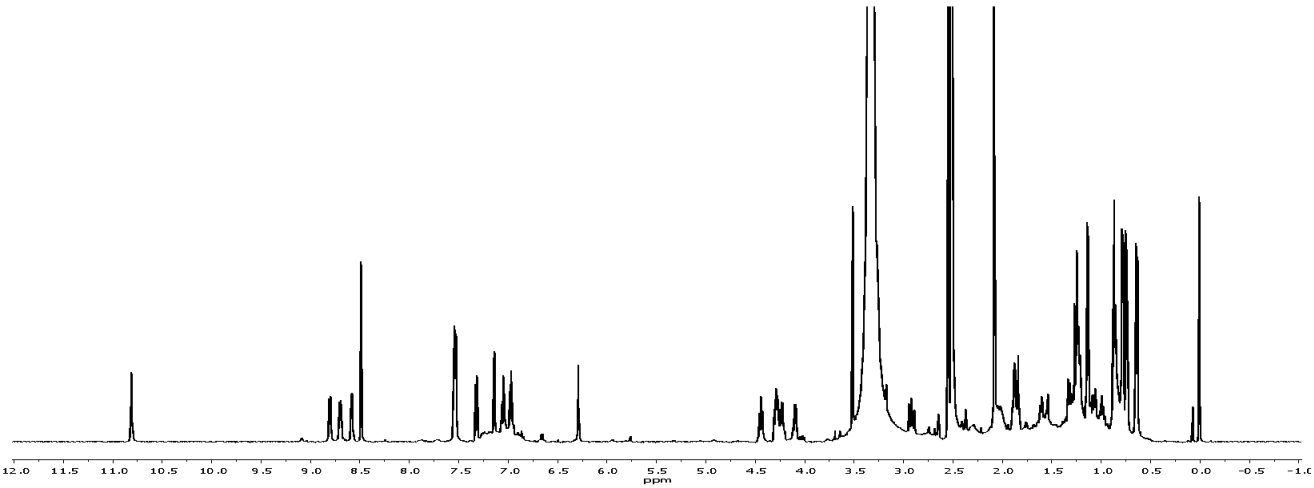


Figure S11: ^{1}H NMR spectrum of 1 (A) and 2 (B) in DMSO-d_6 (500 MHz).