

ARTICLE

Identification of a tambjamine gene cluster in *Streptomyces* reveals convergent evolution of the biosynthetic pathway†

Neil L. Grenade, Dragos S. Chiriac, Graeme W. Howe and Avena C. Ross*

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Bacterial natural products are an immensely valuable source of therapeutics. As modern DNA sequencing efforts provide increasing numbers of microbial genomes, it is clear that the molecules produced by most natural product biosynthetic gene clusters (BGCs) remain unknown. Genome mining makes use of bioinformatic techniques to elucidate the natural products produced by these “orphan” BGCs. Here, we report the use of sequence similarity networks (SSNs) and genome neighborhood networks (GNNs) to identify an orphan BGC that is responsible for the production of the antitumor tambjamine BE-18591 in *Streptomyces albus* NRRL B-2362. Although BE-18591 is a close structural analogue of tambjamine YP1 produced by *Pseudoalteromonas tunicata*, the biosynthetic routes to produce these molecules differ significantly. Notably, the C₁₂-alkylamine tail that is appended onto the bipyrrole core of tambjamine YP1 is derived from fatty acids siphoned from the primary metabolism of the pseudoalteromonad, whilst the *S. albus* NRRL B-2362 BGC encodes a dedicated system for the *de novo* biosynthesis of the alkylamine portion of tambjamine BE-18591. These remarkably different biosynthetic strategies represent a striking example of convergent BGC evolution, with selective pressure for the production of tambjamins seemingly leading to the emergence of separate biosynthetic pathways in pseudoalteromonads and streptomycetes that ultimately produce closely related compounds.

Introduction

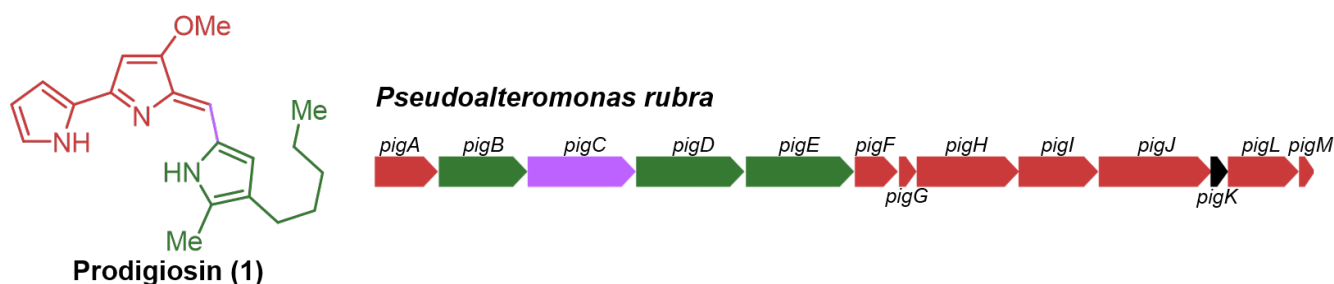
Bacterial natural products (NPs) have a central role in the discovery and development of clinically relevant medicines.¹ The discovery of bacterial NPs has traditionally relied on a bioactivity assay-guided fractionation whereby complex fermentation extracts are partially separated and the resulting fractions are tested for bioactivity.² While initially effective, this technique quickly plateaued in efficacy when the pool of readily accessible NPs was exhausted and the ubiquity of such microbial NPs led to high rates of rediscovery.³ In the post-genomic age, bioinformatic tools have been developed that circumvent the issue of NP rediscovery by facilitating the identification and prioritization of biosynthetic gene clusters (BGCs) responsible for the production of novel NPs.^{4, 5} The dereplication of known bacterial NPs through bioinformatics, collectively known as genome mining, provides an opportunity to probe the true biosynthetic capacity of bacteria.⁶ In one particularly illustrative example, genome mining of 75 *Salinospora* strains revealed 124 BGCs, only 11 of which could be linked to a characterized molecule.⁷ Furthermore, nearly 60% of the identified BGCs were present in only one or two strains, suggesting that most are acquired by horizontal gene transfer (HGT) and investigation of new strains will reveal new molecules.⁷ The potential of bioprospecting to discover new

therapeutic NPs has driven the development of many genome mining tools including AntiSMASH,⁸ PRISM,⁹ BiG-SCAPE,¹⁰ GNPS,¹¹ EvoMining,¹² ARTS,¹³ and the Enzyme Function Initiative - Enzyme Similarity Tools (EFI-EST).^{14, 15} While these tools use different algorithms to analyse and functionally annotate microbial genomes, each is designed with a common goal: the rapid and precise identification of BGCs in microbial genomes that are responsible for the production of novel NPs.

In drug development, medicinal chemists are frequently challenged with modifying the core scaffold of molecules derived from NPs to improve bioavailability while retaining therapeutic efficacy.¹ Nature, however, has been utilising evolution to generate structural diversity in secondary metabolites for countless years. Firn's ‘Screening Hypothesis’ suggests that bioactive compounds are inherently rare and organisms that develop the capacity to generate molecular diversity around a central bioactive scaffold have an evolutionary advantage.¹⁶ Traits such as enzyme promiscuity, branched, and horizontally transferred pathways confer this advantage to organisms by allowing structural diversity to be sampled at minimal energetic cost.¹⁶ As Nature has likely sampled structural analogues of evolutionarily advantageous NPs, we hypothesized that orphan BGCs that are similar, but not identical, to gene clusters for known NPs would produce related and potentially useful bioactive agents.

Department of Chemistry, Queen's University, Kingston, ON K7L 3N6, Canada.
E-mail: avena.ross@queensu.ca

†Electronic Supplementary Information (ESI) available: Bioinformatics, NMR and MS data. See DOI: 10.1039/x0xx00000x



Prodigiosin (1)

Figure 1. The structure of prodigiosin and the BGC responsible for its biosynthesis in *P. rubra*. Genes are coloured according to their biosynthetic role in NP production. Genes are colored according to their biosynthetic role in the NP production (Red: MBC core biosynthesis; Green: MAP biosynthesis; Purple: condensation enzyme; Black: unknown function).

Prodigiosin (**1**, Figure 1), the archetypal prodiginine, is produced by a number of *Serratia* spp. and marine bacteria including *Hahella chejuensis* and *Pseudoalteromonas rubra*.^{17, 18} This NP contains a tripyrrolic structure that is ultimately assembled by the PigC-catalyzed condensation of 4-methoxy-2,2'-bipyrrole-5-carboxaldehyde (MBC) and methylamylpyrrole (MAP) (Figure S1).^{19, 20} A considerable number of naturally occurring analogues of this tripyrrolic NP exist including the *Streptomyces coelicolor* metabolites undecylprodigiosin (**2**, Figure 2) and streptorubin B (**3**, Figure S2).^{21, 22} A second group of bipyrrolic analogues are the tambjamines exemplified by tambjamine YP1 (**4**, Figure 2) from *Pseudoalteromonas tunicata* and produced predominantly by marine microorganisms.^{23–25} Although **4** also contains the common MBC moiety, the final NP incorporates an unsaturated alkylamine tail in place of the third pyrrole moiety found in prodiginines. These pigmented compounds display a wide array of bioactivities, including antibacterial,^{26, 27} antimalarial,^{28, 29} immunosuppressive,^{30, 31} and antitumor activities.^{32, 33}

Here, we report the use of sequence similarity networks (SSNs) and genome neighborhood networks (GNNs) to explore the structural diversity of the BGCs and corresponding molecules in the tambjamine and prodiginine families of NPs. Genes encoding the biosynthesis of **1** in *Serratia marcescens* were used to construct SSNs at varying alignment scores (ASs). These networks, in combination with GNNs, suggested the role of an orphan BGC (*tab*) in the biosynthesis of tambjamine BE-18591 (**5**) in a number of streptomycetes. Extracts of candidate producers were screened using LC-MS, and BE-18591 (**5**) was identified conclusively in extracts of *Streptomyces albus* NRRL B-2362. While **5** was previously isolated from an unidentified streptomycete,^{34, 35} the work reported here describes the first tambjamine BGC in *Streptomyces* and suggests the co-evolution of tambjamines in Actinobacteria and Proteobacteria.

Results and Discussion

In combination with the EFI-EST and EFI-GNT, the prodigiosin (**1**) BGC is a valuable starting point for the identification of new analogues or producers of prodiginines and tambjamines. The biosynthesis of prodigiosin (**1**) is well understood, and the specific enzymes responsible for the synthesis of the common MBC core have been characterised.³⁶ We anticipated homologues of these enzymes (PigA, PigF–J, and PigL–M) to be

involved in all prodiginine and tambjamine biosynthetic pathways and, therefore, variants of the condensation enzyme PigC should be present in the BGCs for all prodiginines and tambjamines. As these PigC variants must also recognize the nucleophilic partner (pyrrole or amine), sequence variation between these condensation enzymes might also provide some insight into the identity of the nucleophile coupled to the bipyrrole core. As such, our search for prodiginine and tambjamine-related BGCs began with the construction of SSNs based on PigC from *Serratia marcescens* (UniProt ID: Q5W252).

Using the InterProScan tool,³⁷ the *S. marcescens* PigC was identified as a member of Pfams 01326 and 00391. These Pfams correspond to mobile domains that utilize a phosphorylated histidine residue to facilitate phosphoryl group transfer between remote reactive sites.^{38, 39} The EFI-EST was used to construct SSNs for both Pfams (Figure S3–S4). At an AS of 120, the query sequence partitioned into a small cluster of 114 unique sequences (Figure S5). This cluster also contained PigC variants from many other prodiginine/tambjamine producers (eg. *Serratia* sp. ATCC 39006, *Pseudoalteromonas rubra*, *Streptomyces coelicolor* A3(2)) and very likely represents an isofunctional cluster of condensation enzymes. In a new network created from this cluster, the AS was increased until clusters of proteins from known BGCs began to fractionate. Condensation enzymes for **1**, **2**, **4**, tambjamine MYP1 (**6**, Figure S2), and roseophilin (**7**, Figure S2) biosynthesis broke off into separate small clusters or singletons (Figure S6).

A small cluster of proteins from Streptomycetes, including the condensation enzyme responsible for the biosynthesis of marineosin (**8**, Figure S2), was targeted for further analysis using the genome neighbourhood tool (GNT) hosted by the EFI. This tool allows the genetic context of target proteins to be explored as a function of the SSN clustering. The resulting GNNs for the targeted Streptomycetes cluster showed significant variation in the genes surrounding the PigC homologues (Figure S7). We were particularly intrigued by the genetic context of the PigC variant in *Streptomyces albus*. While this BGC contains all genes necessary for MBC biosynthesis, similar to the known Streptomycete metabolite undecylprodigiosin,¹⁸ genes responsible for the synthesis of the third pyrrole involved in

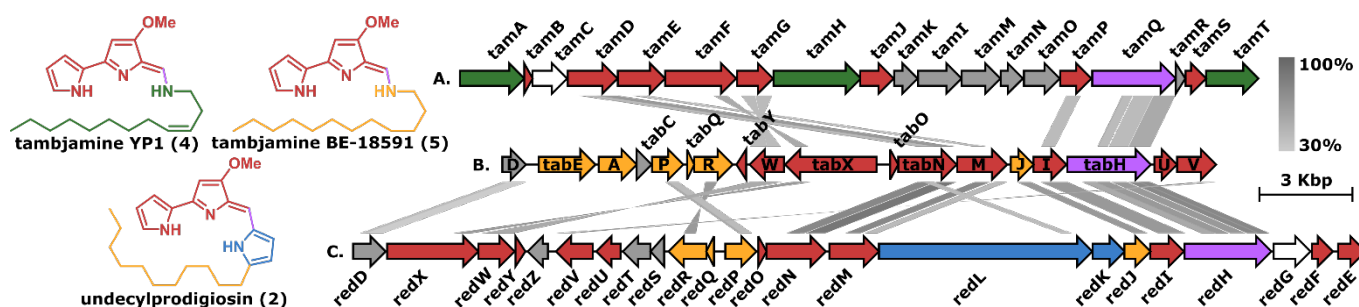


Figure 2. A comparison of the BGCs responsible for the production of (A) tambjamine YP1 (4) from *P. tunicata*, and (C) undecylprodigiosin (2) from *S. coelicolor* with (B) the novel *tab* BGC identified in *S. albus* responsible for BE-18591 (5). Pairwise tblastx hits between BGCs are represented by shaded gray bars reflecting the relative similarity between two regions (30–100%). Genes are coloured according to putative biosynthetic roles (Red: MBC biosynthesis, Green: dodecanoic processing enzymes, Purple: Condensation enzymes, Orange: fatty acid synthesis, Blue: synthesis of the pyrrole moiety of 2-undecylpyrrole, White: macrocyclization; Gray: unknown function).

Table 1. The putative functions and biosynthetic roles of the gene products of the *tab*, *red*, *tam*, and *pig* BGCs.

Tab Proteins	Red Proteins	Tam Proteins	Pig Proteins	Putative Function	Biosynthetic role
TabD	RedD	-	-	No assigned function	-
TabE	-	-	-	Aldehyde dehydrogenase	DDA
TabA	-	TamH	-	Thioester reductase & Aminotransferase	DDA/DDEA
TabC	-	-	-	No assigned function	-
TabP	RedP	-	-	B-ketoacyl-ACP synthase	2-UP/DDA
TabQ	RedQ	-	-	Acyl-carrier protein	2-UP/DDA
TabR	RedR	-	-	B-ketoacyl-ACP synthase	2-UP/DDA
TabY	RedY	TamR	PigK	No assigned function	-
TabW	RedW	TamG	PigA	L-prolyl-PCP dehydrogenase	MBC
TabX	RedX	TamF	PigJ	B-ketoacyl synthase	MBC
TabO	RedO	TamB	PigG	Petidyl-carier protein	MBC
TabN	RedN	TamD	PigH	Aminotransferase	MBC
TabM	RedM	TamE	PigI	L-prolyl-AMP ligase	MBC
TabJ	RedJ	-	-	Thioesterase	2-UP/DDA
TabI	RedI	TamP	PigF	SAM dependent O-methyl transferase	MBC
TabH	RedH	TamQ	PigC	Condensation enzyme	Condensation
TabU	RedU	TamS	PigL	Phosphopantetheinyl transferase	MBC
TabV	RedV	TamJ	PigM	Oxidase/dehydrogenase	MBC
-	-	TamA	-	Adenylation-ACP	DDEA
-	-	TamT	-	Dehydrogenase	DDEA

prodigine formation are absent suggesting this is a tambjamine BGC (Figure S8). Additionally, the remaining genes share similarity with the fatty acid synthase in the *red* cluster that generates the twelve-carbon tail of undecylprodigiosin in *S. coelicolor* (Figure 2C and S13). While there are currently no published tambjamine BGCs from Streptomyces, these observations led us to obtain a series of *S. albus* strains containing the putative tambjamine BGC from the NRRL Culture Collection. After completing cultivation experiments to test for production of tambjamins, gratifyingly, one of the organisms harbouring a BGC from this cluster, *Streptomyces albus* NRRL B-2362, was found to produce a molecule with *m/z* 358.3 and absorbance at 405 nm (Figure S9) matching data originally reported for tambjamine BE-18591 (5, Figure 2).^{34, 35}

Subsequent rounds of culturing, extraction, and purification described in the methods led to the isolation of 0.8 mg of pure 5, which was analyzed by HR-MS and 1D ¹H and COSY NMR. The observed *m/z* and chemical shifts are in excellent agreement with previously reported values³⁵ (Figures S10–S12, Table S2) and confirm that *S. albus* NRRL B-2362 produces tambjamine BE-18591. As there is only one PigC variant in the *S. albus* NRRL B-2362 genome with any significant similarity with other prodigine/tambjamine condensation enzymes, we have confidently assigned the associated genes as the *tab* cluster: the first explicit tambjamine BGC in Streptomyces.

A comparison of the *tab* gene cluster in contrast to the *red* and *tam* BGCs can be seen in Figure 2 with the associated gene annotations given in Table 1. The *tab* BGC contains the genes required for the production of MBC, which can also be found in the *pig*, *red*, and *tam* gene clusters that produce **1**, **2**, and **4** respectively (Figure 2, Figure S8). The *tab* gene cluster also contains several genes (*tabJ*, *P*, *Q*, and *R*) that are homologues of those involved in 2-undecylpyrrole (2-UP) biosynthesis in the production of **2** (Figure S13). The homologous genes are likely required for production of dodecanoic acid in tambjamine **5** biosynthesis (Figure 3A), as in the 2-UP BGC. The *tab* cluster, however, lacks a homologue of the PKS-NRPS, *redL* (Figure 2A), which is thought to introduce an additional malonate and a glycine onto the fatty acid chain of 2-UP that ultimately undergoes cyclization to produce the third alkylated pyrrole (Figure S13). The similarity of the *red* and *tab* BGCs, despite producing different products is perhaps less surprising when considering that both are found in *Streptomyces*. What is more intriguing is the finding that the BGCs responsible for the biosynthesis of tambjamines **4** and **5** are significantly different, despite producing structurally complex NPs that only differ by a single degree of unsaturation in the dodecylamine (DDA) tail.

The *tam* cluster has no homologues of *red/tabJ*, *P*, *Q*, *R* with which to construct the fatty acid tail that is appended onto MBC in the biosynthesis of **4**. Instead, the pathway appears to siphon dodecanoic acid from primary metabolism³⁶ and then loads it onto TamA (Figure 3B) for further processing.⁴⁰ The origin of dodecanoic acid is not the only difference between the *tam* and *tab* clusters. In the biosynthesis of tambjamine **4**, TamT is proposed to introduce the alkene in the fatty tail. Then TamH, a bifunctional enzyme, is thought to reduce the dodecanoic acid-TamA thioester adduct and carry out reductive amination of the resultant *cis*-dodec-3-en-1-al to give the *cis*-dodec-3-en-1-amine (DDEA), which is ultimately condensed with MBC.⁴⁰ In contrast, the *Tab* pathway is missing a TamT homologue, consistent with the presence of a fully saturated fatty acyl tail in **5** (Figure 3A). Bioinformatic analysis of TabA (the closest TamH homologue) suggests that this protein has an aminotransferase domain but does not have a reductase domain (Figure S14). As such, TabA will complete the aminotransfer step but not the reductive offloading from the carrier protein. It is currently unclear what the substrate for TabA is and how it is generated. TabE is found only in the *Tab* pathway and is predicted to be an aldehyde dehydrogenase but may favour the reverse reaction and reduce dodecanoic acid to dodecanal to produce a substrate for TabA. However, TabE's exact biosynthetic role remains unclear and further experiments will be needed to confirm what role if any it plays in the biosynthesis of **5**.

The genetic differences between the *tab* and *tam* BGCs suggest that despite the similarity of the fatty amine tails of tambjamines **4** and **5**, their biosynthetic origins are quite different and most likely emerged through HGT and convergent evolution. In an analogous manner, prodiginines **1** and **2** bear a striking resemblance to each other. However, despite

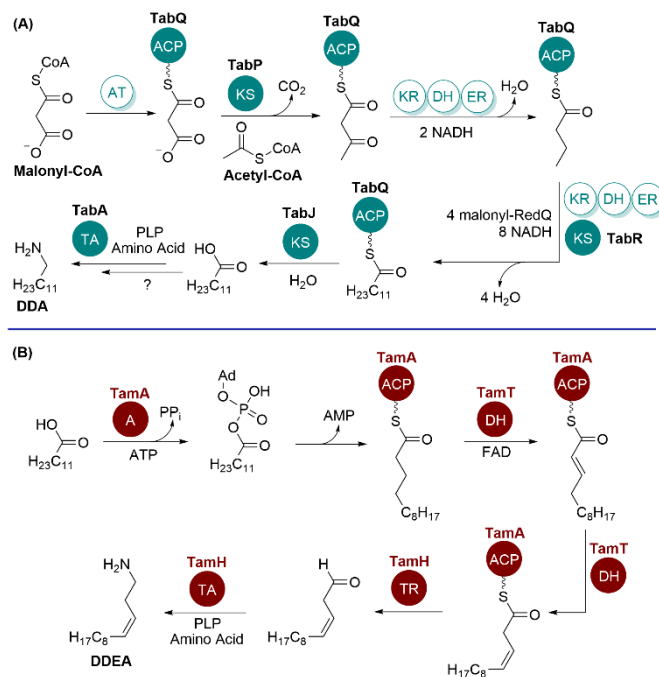


Figure 3. Proposed biosynthetic schemes for (A) the production of DDA by the *Tab* pathway and (B) DDEA by the *Tam* pathway.

conservation of the MBC biosynthetic enzymes, there are also marked differences in how the *Pig* and *Red* pathways synthesize the third pyrrole moiety. While the *Pig* pathway constructs MAP by co-opting an 8-carbon fatty acid from primary fatty acid metabolism (much like in the *Tam* pathway), the *red* pathway builds up 2-UP entirely *in situ* from multiple malonyl-CoA units (as proposed in *Tab* pathway).³⁶ Therefore, we now have two examples of streptomycete pathways building fatty acids and pseudoalteromonads co-opting them from primary metabolism, suggesting distinct evolutionary origins for these seemingly related pathways (Figure 4). This evolutionary insight into the emergence of the *pig*, *tam*, *red*, and *tab* gene clusters in *Streptomyces* and *Pseudoalteromonas* was made possible by the holistic comparison of the genetic differences in the BGCs and the structural relationship between the corresponding prodiginine/tambjamine NPs.

Conclusions

Bioinformatic analysis of homologues of key prodiginine and tambjamine biosynthetic enzymes has led to the identification of an orphaned BGC (*tab*) for the known tambjamine BE-18591 (**5**). Putative functions were assigned to genes within the *Tab* pathway and compared to homologous genes in other prodiginine/tambjamine pathways. Given the evidence for convergent evolution of the pathways of prodiginines **1** and **2**, similarities between the biosynthesis of MAP and DDEA, and homology between *Tab/RedJ*, *P*, *Q* and *R*, a parallel conclusion about evolutionary convergence can be made for the biosynthesis of tambjamines **4** and **5**. Proteobacteria recruit primary metabolites into shunt pathways for the synthesis of these NPs, while actinomycetes rely on dedicated systems within the BGCs. The result in either case is the production of

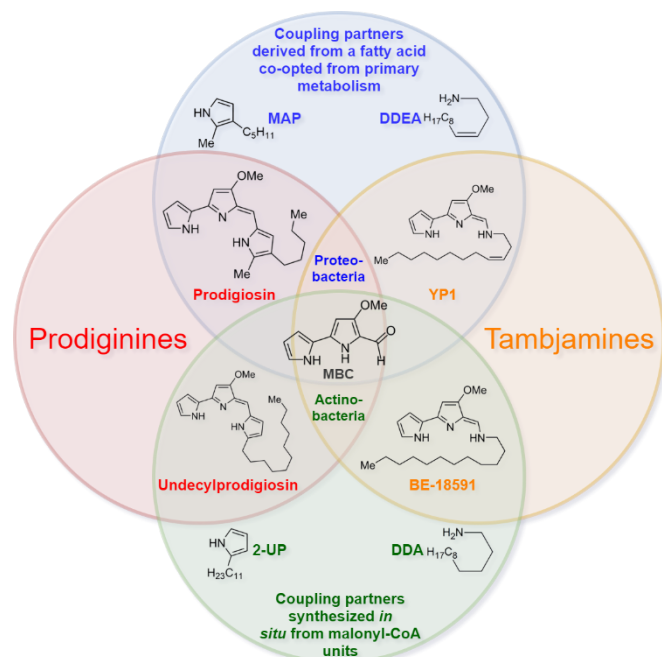


Figure 4. Comparing the source of the nucleophilic pyrrole/amine that is appended to the MBC core in the biosynthesis of prodiginine and tambjamine NPs.

structurally similar compounds that confer advantages to their respective hosts.

The SSNs and GNNs reported here also provide several promising potential sources of tambjamines from a wide variety of genera that may provide additional examples of HGT and convergent evolution of BGCs between bacterial strains.

Additional experimental work to isolate key enzymes within the Tab and Tam pathways to test activity and substrate scope will paint a more complete picture of the evolution that emerged between these disparate bacteria. Given the therapeutic potential of this class of natural products, uncovering these mechanisms may contribute enormously to understanding how to engineer these pathways to yield useful bioactives.

Materials and methods

General experimental protocols

All NMR spectra were recorded on a Bruker Ascend 700 MHz instrument in CDCl_3 . Crude extracts and preparative fractions were separated on a Waters H-Class UPLC with a Phenyl-Hexyl column (Waters CSH Phenyl-Hexyl; 2.1mm x 50mm; 1.7 μm particle size) using a linear mobile phase gradient from 5% to 95% methanol buffered with 0.1 % formic acid flowing at 300 $\mu\text{L}/\text{min}$ over 20 minutes. Analytes were detected by diode array absorbance from 200 – 600 nm and electrosprayed into a single quadrupole mass spectrometer at a 2.5 kV electrospray voltage scanning 200 – 1200 m/z . High-resolution mass spectra were collected on a Thermo Orbitrap Velos Pro by direct infusion at 10 $\mu\text{L}/\text{min}$ with an electrospray voltage of 5.0 kV scanning 200 – 2000 m/z . MS2 fragmentation spectra were collected by collision induced dissociation at a normalized collision energy of 35%.

Bioinformatics

The UniProt entries for known PigC variants from *S. marcescens* (UniProt ID Q5W252) and *S. sp.* ATCC 39006 (Q5W269) are functionally annotated with the SwissProt description “Prodigiosin synthesizing transferase PigC.” The amino acid sequence of PigC from *S. marcescens* was retrieved from UniProt (Version 2020_03) and analyzed with the InterProScan tool.³⁷ This enzyme was annotated as a member of Pfams PF00391 (designated as “PEP-utilizing enzyme, mobile domain”) and PF01326 (designated as “Pyruvate phosphate dikinase, AMP/ATP-binding domain”). An all-by-all BLAST was performed for the UniRef 90 versions of each of these Pfams using the EFI-EST.^{14, 15, 41} All resulting sequence similarity networks (SSNs) were visualized and manipulated in Cytoscape 3.8.0. Due to the considerable size of the PEP-utilizing Pfams, 50% identity RepNode networks were used throughout the subsequent analysis. Initial SSNs were constructed at alignment scores (ASs) of 80 (%ID > 22%) and 100 (%ID > 24%) for Pfams 01326 and 00391, respectively. ASs were increased incrementally until the PigC nodes separated into putative isofunctional clusters (AS 120 with the both networks). For both networks, the emergent isofunctional clusters contained the same sequences. As such, subsequent analysis was performed with the PF01326 network. A daughter network was created from this cluster, and the AS was increased incrementally to evaluate the degree of relatedness between each putative PigC variant. GNNs were created for the putative PigC isofunctional clusters at various ASs using the default parameters of the EFI-GNT webtool. The resulting BGCs were manually curated to evaluate their similarity and biosynthetic potential for the production of prodiginines and tambjamines. Pairwise BGC comparisons were performed using tblastx as implemented in EasyFig.⁴² For the sake of clarity, a cut-off was used to display only those tblastx hits with %ID > 30% and an alignment length > 100bp.

Microbial strain culturing

Streptomyces albus NRRL B-2362, *Streptomyces sp.* NRRL B-1437, *Streptomyces sp.* NRRL F-5639, *Streptomyces sp.* NRRL F-6602, *Streptomyces sp. albus* NRRL S-1521, and *Streptomyces sp. NRRL WC-3753* were obtained from the Agricultural Research Services (ARS) Culture Collection. The strains were maintained on MS Agar at 28 °C. MS agar contains 10g/L mannitol, 10g/L Soya Flower (Red Mills), and 15g/L agar with pH adjusted to 7.2. Spores were resuspended with 8 ml of sterile water and 100 μL was used to inoculate 50 ml of liquid MS media (10g/L mannitol and 10g/L Soya Flower (Red Mills); pH = 7.2) in a 250 ml Erlenmeyer flask with aeration springs. Cultures were grown at 28 °C with shaking at 150 RPM for a total of 7 days. Frozen stocks were prepared by combining 0.5 mL of culture supernatant with 0.5 mL of a sterile 50% glycerol solution and flash-freezing the mixture in liquid N_2 . Glycerol stocks were stored at -80 °C.

Production and isolation of BE-18591

Streptomyces sp. NRRL B-2362 was grown on MS agar for 4 days. Spores were resuspended with 8 ml of sterile water and 100 μ l was used to inoculate 10 x 50 ml of MS media. The cultures were grown at 28 °C and 150 RPM for 7 days. The mycelia were collected by centrifugation at 12,000 RPM for 10 minutes. The supernatant was decanted, and the mycelia were extracted with 400 ml of EtOAc and stirred for over 2 hours at room temperature. The mycelia were removed by gravity filtration, the filtrate was dried with MgSO_4 . The EtOAc was removed under reduced pressure and the remaining residue was redissolved in 2 mL MeOH. The extract was filtered through a 0.22-micron filter (Millipore) and purified by semipreparative scale HPLC using a Waters C8 column (250 x 10 mm) with dH_2O and MeOH containing 0.1% formic acid as the mobile phase at a flow rate of 3 mL/min. A linear gradient between 60% and 90% MeOH was used during the first 10 minutes followed by isocratic flow of 90% MeOH for 5 minutes. BE-18591 eluted at 10 minutes. The fractions containing BE-18591 were pooled and evaporated under reduced pressure to yield approximately 1.6 mg /L of compound. HRMS and ^1H NMR matched the literature.³⁵

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We gratefully acknowledge the Agricultural Research Service Culture Collection (NRRL) for access to bacterial strains and the Natural Sciences and Engineering Research Council of Canada for financial support (Discovery Grants to GWH and ACR).

Notes and references

- D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629.
- M. G. Weller, *Sensors*, 2012, **12**, 9181.
- M. Tulp and L. Bohlin, *Bioorg. Med. Chem.*, 2005, **13**, 5274.
- H. Brötz-Oesterhelt and P. Sass, *Future Microbiol.*, 2010, **5**, 1553.
- S. P. Gaudencio and F. Pereira, *Nat. Prod. Rep.*, 2015, **32**, 779.
- B. O. Bachmann, S. G. Van Lanen and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 175.
- N. Ziemert, A. Lechner, M. Wietz, N. Millan-Aguinaga, K. L. Chavarria and P. R. Jensen, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, E1130.
- K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de Los Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2017, **45**, W36.
- M. A. Skinnider, N. J. Merwin, C. W. Johnston and N. A. Magarvey, *Nucleic Acids Res.*, 2017, **45**, W49.
- J. C. Navarro-Munoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. D. Cappelini, A. W. Goering, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *Nat. Chem. Biol.*, 2020, **16**, 60.
- M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapon, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crusemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrew, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. B. P. D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutierrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828.
- P. Cruz-Morales, J. F. Kopp, C. Martinez-Guerrero, L. A. Yanez-Guerra, N. Selem-Mojica, H. Ramos-Aboites, J. Feldmann and F. Barona-Gomez, *Genome Biol. Evol.*, 2016, **8**, 1906.
- M. D. Mungan, M. Alanjary, K. Blin, T. Weber, M. H. Medema and N. Ziemert, *Nucleic Acids Res.*, 2020, **48**, W546.
- R. Zallot, N. O. Oberg and J. A. Gerlt, *Curr. Opin. Chem. Biol.*, 2018, **47**, 77.
- R. Zallot, N. Oberg and J. A. Gerlt, *Biochemistry*, 2019, **58**, 4169.
- R. D. Firn and C. G. Jones, *Nat. Prod. Rep.*, 2003, **20**, 382.
- F. Wrede and O. Hettche, *Ber. Dtsch. Chem. Ges.*, 1929, **62**, 2678.
- N. R. Williamson, P. C. Fineran, F. J. Leeper and G. P. Salmond, *Nat. Rev. Microbiol.*, 2006, **4**, 887.
- A. S. Klein, A. Domrose, P. Bongen, H. U. C. Brass, T. Classen, A. Loeschcke, T. Drepper, L. Laraia, S. Sievers, K. E. Jaeger and J. Pietruszka, *ACS Synth. Biol.*, 2017, **6**, 1757.
- K. J. Picott, J. A. Deichert, E. M. deKemp, V. Snieckus and A. C. Ross, *ChemBioChem*, 2020, **21**, 1036.
- S. Tsao, B. A. M. Rudd, X. He, C. Chang and H. G. Floss, *J. Antibiot.*, 1985, **38**, 128.
- M. Redenbach, H. M. Kieser, D. Denapate, A. Eichner, J. Cullum, H. Kinashi and D. A. Hopwood, *Mol. Microbiol.*, 1996, **21**, 77.
- A. Franks, P. Haywood, C. Holmström, S. Egan, S. Kjelleberg and N. Kumar, *Molecules*, 2005, **10**, 1286.
- F. Ballestrero, T. Thomas, C. Burke, S. Egan and S. Kjelleberg, *Appl. Environ. Microbiol.*, 2010, **76**, 5710.

25. K. J. Picott, J. A. Deichert, E. M. deKemp, V. Snieckus and A. C. Ross, *ChemBioChem*, 2019, **20**, 1.
26. T. Danevcic, M. Boric Vezjak, M. Zorec and D. Stopar, *PLoS One*, 2016, **11**, e0162412.
27. F. E. Sakai-Kawada, C. G. Ip, K. A. Hagiwara and J. D. Awaya, *Front. Microbiol.*, 2019, **10**, 1715.
28. A. J. Castro, *Nature*, 1967, **213**, 903.
29. M. Isaka, A. Jaturapat, J. Kramyu, M. Tanticharoen and Y. Thebtaranonth, *Antimicrob. Agents Chemother.*, 2002, **46**, 1112.
30. S. Songia, A. Mortellaro, S. Taverna, C. Fornasiero, E. A. Scheiber, E. Erba, F. Colotta, A. Mantovani, A. M. Isetta and J. Golay, *J. Immunol.*, 1997, **158**, 3987.
31. R. D'Alessio, A. Bargiotti, O. Carlini, F. Colotta, M. Ferrari, P. Gnocchi, A. Isetta, N. Mongelli, P. Motta, A. Rossi, M. Tibolla and E. Vanotti, *J. Med. Chem.*, 2000, **43**, 2557.
32. R. Francisco, R. Perez-Tomas, P. Gimenez-Bonafe, V. Soto-Cerrato, P. Gimenez-Xavier and S. Ambrosio, *Eur. J. Pharmacol.*, 2007, **572**, 111.
33. T. F. Ho, C. J. Ma, C. H. Lu, Y. T. Tsai, Y. H. Wei, J. S. Chang, J. K. Lai, P. J. Cheuh, C. T. Yeh, P. C. Tang, J. Tsai Chang, J. L. Ko, F. S. Liu, H. E. Yen and C. C. Chang, *Toxicol. Appl. Pharmacol.*, 2007, **225**, 318.
34. K. Kojiri, S. Nakajima, H. Suzuki, A. Okura and H. Suda, *J. Antibiot.*, 1993, **46**, 1799.
35. S. Nakajima, K. Kojiri and H. Suda, *J. Antibiot.*, 1993, **46**, 1894.
36. D. X. Hu, D. M. Withall, G. L. Challis and R. J. Thomson, *Chem. Rev.*, 2016, **116**, 7818.
37. P. Jones, D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez and S. Hunter, *Bioinformatics*, 2014, **30**, 1236.
38. O. Herzberg, C. C. H. Chen, G. Kapadia, M. McGuire, L. J. Carroll, S. J. Noh and D. Dunaway-Mariano, *Proc. Natl. Acad. Sci. U.S.A.*, 1996, **93**, 2652.
39. K. Lim, R. J. Read, C. C. H. Chen, A. Tempczyk, M. Wei, D. Ye, C. Wu, D. Dunaway-Mariano and O. Herzberg, *Biochemistry*, 2007, **46**, 14845.
40. P. M. Marchetti, V. Kelly, J. P. Simpson, M. Ward and D. J. Campopiano, *Org. Biomol. Chem.*, 2018, **16**, 2735.
41. J. A. Gerlt, J. T. Bouvier, D. B. Davidson, H. J. Imker, B. Sadkhin, D. R. Slater and K. L. Whalen, *Biochim. Biophys. Acta, Proteins Proteomics*, 2015, **1854**, 1019.
42. M. J. Sullivan, N. K. Petty and S. A. Beatson, *Bioinformatics*, 2011, **27**, 1009.