Published in final edited form as:

ACS Chem Biol. 2018 April 20; 13(4): 1029–1037. doi:10.1021/acschembio.7b01089.

Discovery of the Tyrobetaine Natural Products and Their Biosynthetic Gene Cluster *via* Metabologenomics

Elizabeth I. Parkinson[†], **James H. Tryon**[‡], **Anthony W. Goering**[‡,⊥], **Kou-San Ju**[†,#], **Ryan A. McClure**[‡,∇], **Jeremy D. Kemball**[†], **Sara Zhukovsky**[†], **David P. Labeda**[§], **Regan J. Thomson**[‡], **Neil L. Kelleher**[‡], and **William W. Metcalf**[*,†,||]

[†]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[‡]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

[§]Mycotoxin Prevention and Applied Microbiology Research Unit, USDA-ARS National Center for Agricultural Utilization Research, Peoria, Illinois 61604, United States

[||]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801 United States

Abstract

Natural products (NPs) are a rich source of medicines, but traditional discovery methods are often unsuccessful due to high rates of rediscovery. Genetic approaches for NP discovery are promising, but progress has been slow due to the difficulty of identifying unique biosynthetic gene clusters (BGCs) and poor gene expression. We previously developed the metabologenomics method, which combines genomic and metabolomic data to discover new NPs and their BGCs. Here, we utilize metabologenomics in combination with molecular networking to discover a novel class of NPs, the tyrobetaines: nonribosomal peptides with an unusual trimethylammonium tyrosine residue. The BGC for this unusual class of compounds was identified using metabologenomics and computational structure prediction data. Heterologous expression confirmed the BGC and suggests an unusual mechanism for trimethylammonium formation. Overall, the discovery of the

[*]Corresponding Author: metcalf@life.illinois.edu.
[⊥]Microbial Pharmaceuticals, Evanston, Illinois 60208, United States
[#]Department of Microbiology and the Division of Medicinal Chemistry and Pharmacognosy, The Ohio State University, Columbus, OH 43210, United States
[∇]AbbVie, Chicago, IL 60064, United States

ORCID

Kou-San Ju: 0000-0003-0438-6740
Ryan A. McClure: 0000-0002-0163-9290
Regan J. Thomson: 0000-0001-5546-4038
Neil L. Kelleher: 0000-0002-8815-3372
William W. Metcalf: 0000-0002-0182-0671

Notes

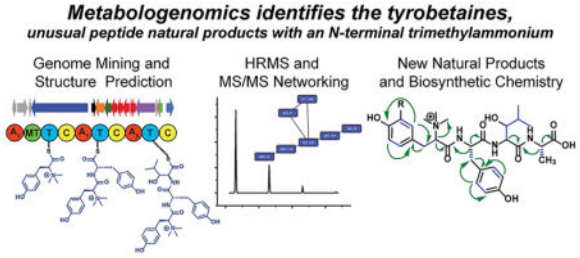
The authors declare the following competing financial interest(s): Anthony W. Goering, Regan J. Thomson, Neil L. Kelleher, and William W. Metcalf have a financial interest in metablogenomics via Microbial Pharmaceuticals. These potential conflicts of interests are managed through the respective university's conflict of interest policy.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschem-bio.7b01089. Detailed Materials and Methods, structure elucidation information, supplementary figures, supplementary tables, and NMR spectra (PDF)

tyrobetaines shows the great potential of metabologenomics combined with molecular networking and computational structure prediction for identifying interesting biosynthetic reactions and novel NPs.

Graphical Abstract



Historically, natural products (NPs) have been our best source of medicines with 64% of FDA-approved small molecule drugs being inspired by NPs.[1–3] Many of these bioactive NPs are secondary metabolites produced by actinomycetes: soil dwelling Gram-positive bacteria. Discovery of novel NPs from actinomycetes *via* traditional means (*i.e.*, screening crude fermentation broths for bioactivity)[4,5] is no longer viewed as a viable strategy due to the high rates of rediscovery.[6,7]

Improvements in sequencing technologies allowed whole genome sequencing of bacteria including *Streptomyces* (the largest genus within the Actinobacteria) and resulted in the birth of genome mining, which uses bioinformatics analysis of biosynthetic gene clusters (BGCs) to expedite discovery of NPs. These studies revealed that approximately 70% of putative BGCs differ significantly from known BGCs,[8–10] suggesting that numerous chemically diverse NPs have yet to be found.[11] New NPs have been discovered *via* genome mining.[12–18] However, large scale application has been slow to be realized.[18] This is in part due to bioinformatics challenges in identifying BGC families (GCFs) that produce the same molecule or close structural analogues.[11,18] Additionally, many clusters are silent (*i.e.*, present in the genome of a bacterium but transcriptionally inactive),[19,20] further complicating identification of novel NPs.

A few computational methods that utilize different parameters have been developed to identify GCFs.[11,18,21,22] The method that we previously developed analyzes the number of homologous genes shared, the proportion of nucleotides involved in pairwise alignment, and the amino acid sequence of the BGC proteins to group BGCs into GCFs.[11] This method correctly grouped 103 known BGCs into 41 GCFs that produce structurally related NPs. Additionally, it identified 4045 GCFs that have no characterized members.

While the discovery of new GCFs suggests new NPs, confirmation requires identification of the molecules. Without advanced knowledge of structures or bioactivities, detection of novel NPs is very difficult. To address this, we developed *metabologenomics*, an automated, untargeted method for identifying NPs based on binary correlation between a BGC and a molecule identified by liquid chromatography high-resolution mass spectrometry (LC-HRMS).[11] This method has two main advantages: (1) It requires no prior knowledge of the

molecular structure or bioactivity, and (2) it requires physical detection of the NPs by LC-HRMS, thus avoiding silent BGCs. In our preliminary report, analysis of 178 strains grown under four growth conditions allowed identification of 2521 unique compounds, 110 of which were known NPs. Of the known NPs, 27 had experimentally established BGCs, allowing for development of a scoring metric that identifies correlations between a NP and a GCF, with a score of 300 or better being highly predictive.[11] Metabologenomics has already allowed the identification of several novel NPs, including tambromycin[23] and the rimosamides.[24]

We hypothesized that metabologenomics could be improved by incorporation of molecular networking. Molecular networking is a pattern-based clustering of small molecules based on mass and fragmentation (MS^2) patterns.[25–29] It is a powerful method of dereplication because it allows determination of whether compounds with unique masses are structurally related to known NPs. Others have used molecular networking in combination with genome mining to identify novel NPs.[18,27,30–32] On the basis of these successes, we expected that incorporation of molecular networking into metabologenomics would allow identification of families of NPs with interesting chemical structures.

Here, we used metabologenomics in combination with molecular networking to discover a new class of NPs, the tyrobetaines, and their BGC. The tyrobetaines are non-ribosomal peptides (NRPs) that have an unusual N-terminal trimethylammonium. The predicted BGC was confirmed *via* heterologous expression. Interestingly, the enzymatic activity responsible for the trimethylation of the N-terminal amine is probably localized to a domain within a nonribosomal peptide synthetase (NRPS), consistent with a novel type of NRPS *N*-methyltransferase that installs multiple methyl moieties. Phylogenetic analysis of the *N*-methyltransferase further supports its unusual reactivity. The discovery of the tyrobetaines and their BGC demonstrates the power of combining metabologenomics with molecular networking for the discovery of new NPs and interesting biosynthetic chemistries.

RESULTS AND DISCUSSION

Discovery of a New Family of NPs, the Tyrobetaines

Metabologenomics is a powerful technique for discovering a NP and its BGC, but it is less adept at identifying families of molecules. Using metabologenomics in combination with molecular networking, we identified a new family of NPs that we named the tyrobetaines. This family of NPs includes tyrobetaine (m/z = 587.3073), chlorotyrobetaine (m/z = 621.2667), tyrobetaine-2 (m/z = 387.1917), chlorotyrobetaine-2 (m/z = 421.1525), dichlorotyrobetaine (m/z = 655.2281), and dichlorotyrobetaine-2 (m/z = 455.1129, Figure S1A–F). Initially, tyrobetaine and chlorotyrobetaine were both discovered using metabologenomics. However, it was unclear from the original data whether these molecules were structurally related. MS^2 analysis and molecular networking revealed that tyrobetaine and chlorotyrobetaine clustered together, suggesting that they are structurally related and allowing us to hypothesize that tyrobetaine and chlorotyrobetaine were part of the same family of molecules. Additionally, the MS^2 analysis and molecular networking revealed other masses of interest that turned out to be the four additional tyrobetaine analogues mentioned above (Figure S1G–I).

Tyrobetaine is produced at detectable levels in 32 of the 269 strains examined (Table S1). Chlorotyrobetaine was detected less frequently (26 times) and was observed at lower levels than tyrobetaine. Tyrobetaine-2 is produced by 29 of the 32 tyrobetaine producers and often has a similar ion intensity to tyrobetaine. Chlorotyrobetaine-2 is produced less often than tyrobetaine-2 (26 strains) and is usually produced at lower levels than tyrobetaine-2. Chlorotyrobetaine and chlorotyr-obetaine-2 are produced by the same 26 strains.

To identify the tyrobetaine GCF, the metabologenomics score for each ion was analyzed (Table S2). GCFs are named by the type of NP followed by an underscore, the letters GCF, a period, and an identifying number (*e.g.*, NRPS_GCF.432) and can be accessed on our previously developed Web site (http://www.igb.illinois.edu/labs/metcalf/gcf/).[11] Several GCFs had high correlation scores (>300) to the tyrobetaines. This is probably because many of the tyrobetaine-producing strains are close relatives and, thus, share many of their BGCs (Figure S2A,B).[11]

The tyrobetaine GCF was identified by examining the predicted GCFs for tyrobetaine, chlorotyrobetaine, tyrobetaine-2, and chlorotyrobetaine-2. The four molecules share nine of their top 10 scoring GCFs (Table S2). Analysis of the top 10 tyrobetaine-producing strains revealed three GCFs were common to these strains (NRPS_GCF.424, NRPS_GCF.83, and NRPS_GCF.432, Table 1). AntiSMASH[33] and PRISM[34] prediction software were used to predict the size of the NPs produced by each BGC. NRPS_GCF.424 has one adenylation domain and is predicted to produce an NP that is smaller than tyrobetaine (~200–300 Da). NRPS_GCF.83 has eight adenylation domains and is predicted to produce an NP that is larger than tyrobetaine (~1100–1300 Da). NRPS_GCF.432 has four adenylation domains and is predicted to produce a molecule that is approximately the size of tyrobetaine (~400–500 Da, Figure S1J,K). NRPS_GCF.432 is thus the best candidate GCF for the tyrobetaines.

Distribution of NRPS_GCF.432

A blastp search[35] for proteins encoded by NRPS_GCF.432 revealed that 51 *Streptomyces* strains in the NCBI database have NRPS_GCF.432 (Table S1). We examined their distribution using a *Streptomyces* phylogeny that we generated. All strains with NRPS_GCF.432 fell into a single clade (Figure S2A,B). However, some strains (*e.g.*, *Streptomyces rimosus subsp. pseudoverticillatus* NRRL B-3698) produce tyrobetaine but appear to lack NRPS_GCF.432 (Table S1). The genome assemblies of these strains are of poor quality, leaving open the possibility that the BGC is present, but not observed. We found that NRPS_GCF.432 was present in *Streptomyces rimosus subsp. pseudoverticillatus* NRRL B-3698, based on a PCR assay for the TybD gene (Figure S2C).

Structure Elucidation of the Tyrobetaines

Tyrobetaine (12 mg) and chlorotyrobetaine (2 mg) were isolated from 25 L of spent media from *Streptomyces sp.* NRRL WC-3773. Tyrobetaine has an unusual N-terminal trimethylammonium along with the unnatural amino acid 3-hydroxyleucine (Figure 1a). Full structure elucidation can be found in the Supporting Methods, Figures S3 and S4, supporting NMR files, and Tables S3 and S4. MS^2 analysis suggests that tyrobetaine consists of a trimethylammonium tyrosine, tyrosine, hydroxylated leucine, and alanine (Figure 1b). NMR

analyses confirmed this. The second tyrosine was determined to be L-tyrosine *via* Marfey's analysis (Figure S3A). The alanine was primarily L-alanine with some D-alanine also observed (Figure S3B). It is unclear whether this center was racemized during purification or if both L- and D-alanine are added to the growing peptide. The hydroxyleucine appears to be a single stereoisomer (Figure S3C), but a lack of standards prevented its stereochemical elucidation. The stereochemical assignment for the first tyrosine was not determined due to the stability of the trimethylammonium moiety to acid hydrolysis.

Chlorotyrobetaine has a nearly identical structure to tyrobetaine but with a chlorine on its N-terminal tyrosine residue (Figure 1a). MS^2 and NMR data support this assignment (Figure S4A). The position of the chlorine is further supported by comparison with synthetic *N*,*N*,*N*-trimethyl-3-chlorotyrosine (Figure S4B).

While dichlorotyrobetaine, tyrobetaine-2, chlorotyrobetaine-2, and dichlorotyrobetaine-2 were not isolated, we deduced likely structures based on their HRMS and MS^2 data (Figures S1C–F and S4C–F). Dichlorotyrobetaine appears to be monochlorinated on each of the tyrosine residues (Figure S4E). The other molecules appear to contain only the two tyrosine residues and are likely biosynthetic intermediates or degradation products of tyrobetaine, chlorotyrobetaine, and dichlorotyrobetaine, respectively.

NRPS_GCF.432 Is the Tyrobetaine BGC

The genes in NRPS_GCF.432 from *Streptomyces* sp. WC-3703 were analyzed using BLAST[35] to assign their potential functions (Figure 2a and Table S5). The open reading frames for this BGC were named *tyb*. Boundaries for the BGC were estimated by comparing the BGCs from multiple tyrobetaine producing strains. NRPS_GCF.432 includes a NRPS (TybD) that is predicted to generate a tripeptide consisting of two tyrosine residues and an unknown amino acid (Figure 2b and Table S6). Additionally, TybD contains an *N*-methyltransferase domain that likely is responsible for the methylation of the N-terminal tyrosine. TybD ends with a condensation domain, which likely catalyzes the addition of the alanine. NRPS_GCF.432 also contains a phosphopantetheinyl transferase (TybE) and an amino acid adenylation domain (TybF) that are predicted to load the alanine (Table S6). TybP is predicted to encode a P450 monooxygenase that has 74% identity to Tem23, an enzyme that catalyzes the conversion of leucine to 3-hydroxyleucine in telomycin biosynthesis,[36] suggesting that TybP hydroxylates the leucine in tyrobetaine. Interestingly, this cluster lacks a halogenase to chlorinate chlorotyrobetaine. Also, no thioesterase domain is present to cleave the peptide from the NRPS.

NRPS_GCF.432 also contains genes with homology to ones that encode enzymes known to produce methoxymalonate (TybH-L) and a polyketide synthase (PKS, TybM). These genes are highly conserved, being present in all analyzed tyrobetaine producing strains. Thus, it is possible that a larger precursor is being produced that incorporates tyrobetaine and methoxymalonate. This could explain the lack of thioesterase observed in the NRPS. TybM contains a C-terminal thioester reductase domain (Table S5) that could be responsible for off-loading. However, no larger masses consistent with such a structure were observed in any of our mass-spectrometric analyses.

To experimentally demonstrate that NRPS_GCF.432 is the tyrobetaine BGC, the BGC from *Streptomyces rimosus subsp. rimosus* NRRL B-2659 was heterologously expressed in *Streptomyces lividans* 66 (*S. lividans* 66 pRAM6). This strain showed production of tyrobetaine and tyrobetaine-2 (see Figures 2c and S5A,B), while the parent *S. lividans* 66 that lacks NRPS_GCF.432 showed no production. Production of tyrobetaine by the heterologous expression strain is comparable to that of many of the native producing strains with an LC-HRMS intensity of 1.1E7 (see Tables 1 and S1). Co-spiking with purified tyrobetaine further confirmed that NRPS_GCF.432 is responsible for production of tyrobetaine. No chlorinated analogues were observed in the extracts from *S. lividans* 66 pRAM6.

The lack of a halogenase in the BGC combined with the inability of the heterologous expression strain to make chlorotyrobetaine or chlorotyrobetaine-2 implies that the halogenase is not found on NRPS_GCF.432. Analysis of the genomes of chlorotyrobetaine producing strains revealed only one halogenase that is known to chlorinate a product similar to the tyrobetaines: the halogenase present in the complestatin BGC. The complestatin GCF is NRPS_GCF.83, a cluster with high correlation scores for both chlorotyrobetaine and chlorotyrobetaine-2, suggesting that it could be the halogenase responsible for chlorination of these compounds (see Tables 1 and S1 and S2). This halogenase typically transfers two chlorines to hydroxyphenylglycine residues. However, some halogenases such as Hpg13 from the biosynthesis of enduracidin are known to install two chlorine atoms on its native substrate but only one when acting on a non-native substrate such as ramoplanin,[37] suggesting that the site and number of chlorines transferred may be substrate dependent. If the complestatin halogenase is indeed responsible for chlorination of chlorotyrobetaine, it suggests that it is a relatively promiscuous enzyme that might be used for chlorination of other tyrosine containing NRPs, a highly sought after reaction in the pharmaceutical industry.[37]

Feeding Experiments to Study Tyrobetaine Biosynthesis

To further investigate the biosynthesis of the tyrobetaines, the native producer *Streptomyces sp.* NRRL WC-3703 was grown with stable isotope labeled L-tyrosine, L-leucine, L-alanine, or methionine (Figure 3 and Figure S6). Growth on D_4-labeled L-tyrosine resulted in incorporation of four or eight deuterons consistent with two tyrosine residues and either three, four, or seven deuterons in chlorotyrobetaine, consistent with one tyrosine and one chlorinated tyrosine. Growth on D_{10}-labeled L-leucine resulted in the incorporation of eight or nine deuterons in tyrobetaine and chlorotyrobetaine, which corroborates the incorporation of hydroxyleucine. The greater incorporation of eight compared to nine deuterons suggests that the hydroxyleucine may be epimerized to the D-amino acid prior to incorporation. However, no epimerase exists in the cluster to support this hypothesis. Growth on D_4-labeled L-alanine resulted in incorporation of four deuterons providing evidence that L-alanine is present. Growth on methionine (methyl-$^{13}CD_3$) resulted in the incorporation of three $^{13}CD_3$ methyl groups, suggesting that methylation of the nitrogen is *via S*-adenosyl methionine (SAM).

Feeding studies with 3-chlorotyrosine were performed with the heterologous expression strain *S. lividans* 66 pRAM6. We hypothesized that if the tyrosine was chlorinated before being loaded onto the NRPS, we should be able to see production of chlorinated analogues. A similar technique was used by Müller and co-workers to show that the chlorinated tyrosine found in chondrochlorens A and B was made *via* loading of 3-chlorotyrosine.[38] Growth of *S. lividans* 66 pRAM6 on various concentrations of 3-chlorotyrosine (1 μM to 1 mM) resulted in no detectable chlorotyrobetaine or chlorotyrobetaine-2. These results suggest that chlorination occurs either on the PCP-bound tyrosine or on the product after release from the NRPS. Halogenation of a PCP-bound amino acid has recently been shown to be the mechanism utilized in glycopeptide biosynthesis,[39] further supporting our hypothesis that the halogenase from the complestatin GCF may be responsible for the halogenation of these molecules.

Phylogenetic Analysis of the NRPS *N*-Methyltransferase

The mechanism of trimethylammonium installation appears to be different from other trimethylammonium containing compounds. Trimethylation of amines is typically performed by stand-alone SAM-dependent methyltransferases. For example, in some bacteria, phosphatidylcholine is formed after three methylation events catalyzed by phosphatidylethanolamine *N*-methyltransferase (PmtA).[40] Although heterologous expression shows that the genes needed for trimethylation lie within NRPS_GCF.432, no stand-alone *N*-methyltransferases are present in the BGC. The isotope labeling studies suggest that the methyl groups are from SAM, and there is a SAM-dependent *N*-methyltransferase domain in the NRPS. However, to the best of our knowledge, NRPS encoded *N*-methyltransferase domains have only been reported to transfer a single methyl group to the nitrogen of PCP-linked amino acids.[41,42] The ability of the *N*-methyltransferase domain in TybD to transfer three methyl groups suggests it is a highly unusual NRPS *N*-methyltransferase domain.

A phylogenetic tree comparing the *N*-methyltransferase of TybD to proteins from the UniProtKB/Swiss-Prot database was generated (Figure 4a and S7A). Twenty-three methyltransferases with known products were identified to have significant similarity (E value < 10, identity > 25%) to the *N*-methyltransferase domain of TybD. Additionally, two methyltransferases known to trimethylate their substrates (Dph5 and EgtD) were included. Interestingly, the methyltransferase domain of TybD fell between the clade containing the NRPS *N*-methyltransferase domains, which transfer a single methyl group, and PmtA, which transfers three methyl groups.[40] The other trimethylating enzymes (Dph5 and EgtD) were distantly related. Dph5 catalyzes the trimethylation of a modified histidine on translation elongation factor 2 and is only found in eukaryotes and archaea, potentially explaining the distance.[43] EgtD is commonly found in bacteria but is known to very selectively modify histidine.[44] Comparison of crystal structures of EgtD and PmtA reveal similar SAM binding regions but drastically different substrate binding regions, providing a potential explanation for the difference.[44] A phylogenetic tree comparing the *N*-methyltransferase of TybD to nonredundant protein sequences in NCBI was also generated (Figure 4b and S7B). The TybD *N*-methyltransferases from tyrobetaine producing strains were all organized into a single clade. A few *N*-methyltranferases known to transfer a single

methyl group were found (*e.g.*, those in the biosynthesis of lyngbyatoxin, pristinamycin, or virginiamycin), but all were distantly related. No known trimethylating enzymes were found on this tree, which includes the 500 homologues most similar to the methyltransferase domain of TybD. Interestingly, the *N*-methyltransferases most closely related to TybD are annotated as hypothetical NRPS. These homologues also have their *N*-methyltransferases located "first" (i.e., before the first peptide carrier protein), suggesting they could possibly transfer three methyl groups to the free N-terminus of their putative NRPS products. The rest of the *N*-methyltransferases on the tree are also annotated as "hypothetical NRPS" with a mixture of the *N*-methyltransferases being "first" and "not first" (*i.e.*, unable to transfer three methyl groups).

Bioactivity Analysis of the Tyrobetaines

Compounds containing trimethylammonium moieties (*e.g.*, glycine betaine, proline betaine, and carnitine) are commonly used as osmolytes in all domains of life,[45–47] but they are extraordinarily rare in peptide NPs. A SciFinder search for compounds containing trimethylammonium tyrosine revealed only one peptide NP, 4862F, an HIV-1 protease inhibitor produced by *Streptomyces* I03A-04862.[48] The few other matches were free trimethylammonium tyrosine; derivatives of free trimethylammonium tyrosine from lichen,[49] marine sponges,[50–53] plants,[54] beetles,[55] and human blood;[56] or synthetic derivatives of peptide NPs.[57–60] These results indicate that tyrobetaine is a structurally interesting NP.

The single agent antibiotic and anticancer activity of tyrobetaine and chlorotyrobetaine was tested (Table 2). They showed no detectable activity at 100 μM against either the ESKAPE pathogens (***E****nterococcus faecalis* ATCC 19433, ***S****taphylococcus aureus* ATCC 29213, ***K****lebsiella pneumoniae* ATCC 27736, ***A****cinetobacter baumannii* ATCC 19606, ***P****seudomonas aeruginosa* PAO1, and ***E****scherichia coli* ATCC 25922) or the human lung cancer cell line A549. We hypothesized that tyrobetaine might synergize with oxytetracycline given its co-occurrence in the *S. rimosus* strains (Figure S2B). However, no synergy was observed when *E. coli* ATCC 25922 was cotreated with the two compounds. It was also proposed that tyrobetaine might be a protease inhibitor given its structural similarity to the angiotensin converting enzyme inhibitor K-26[61,62] and the HIV-1 protease inhibitor 4862F.[48] Tyrobetaine showed no activity against these proteases (Table 2). We expect that tyrobetaine does have a biological function that we have yet to discover.

CONCLUSIONS

In this study, metabologenomics and molecular networking were used to discover the tyrobetaines, a new class of NPs with an unusual N-terminal trimethylammonium. Metabologenomics combined with computational structure prediction allowed identification of the tyrobetaine BGC, which was confirmed using heterologous expression. Intriguingly, this discovery revealed a previously undescribed activity for a NRPS *N*-methyltransferase, trimethylammonium formation. Additionally, this study demonstrated the power of combining metabologenomics with molecular networking and computational structure prediction for the discovery of novel NPs and their biosynthetic enzymes.

METHODS

Details of experimental procedures are provided in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Agricultural Research Service of the United States Department of Agriculture for providing the actinobacterial strains. We thank P. Hergenrother (UIUC) for providing the human cell lines and the ESKAPE pathogens. This work was funded by the National Institutes of Health (GM P01 GM077596 awarded to W.W.M. and R01AT009143 awarded to R.J.T. and N.L.K.). NMR spectra were recorded on a 600 MHz instrument purchased with support from National Institutes of Health grant S10 RR028833. E.I.P. was supported by National Institutes of Health Ruth L. Kirchstein National Service Research Award F32GM120999. Additional funding was provided by a proof-of-concept award from the Institute of Genomic Biology, University of Illinois at Urbana-Champaign.

References

1. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod. 2012; 75:311–335. [PubMed: 22316239]
2. Cragg GM, Newman DJ. Natural products: a continuing source of novel drug leads. Biochim Biophys Acta, Gen Subj. 2013; 1830:3670–3695.
3. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs from 1981 to 2014. J Nat Prod. 2016; 79:629–661. [PubMed: 26852623]
4. Cragg GM, Grothaus PG, Newman DJ. Impact of Natural Products on Developing New Anti-Cancer Agents. Chem Rev. 2009; 109:3012–3043. [PubMed: 19422222]
5. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discovery. 2015; 14:111–129. [PubMed: 25614221]
6. Baltz RH. Renaissance in antibacterial discovery from actinomycetes. Curr Opin Pharmacol. 2008; 8:557–563. [PubMed: 18524678]
7. Forseth RR, Fox EM, Chung D, Howlett BJ, Keller NP, Schroeder FC. Identification of cryptic products of the gliotoxin gene cluster using NMR-based comparative metabolomics and a model for gliotoxin biosynthesis. J Am Chem Soc. 2011; 133:9678–9681. [PubMed: 21612254]
8. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T, Kikuchi H, Shiba T, Sakaki Y, Hattori M. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. Proc Natl Acad Sci U S A. 2001; 98:12215–12220. [PubMed: 11572948]
9. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature. 2002; 417:141–147. [PubMed: 12000953]
10. Sidebottom AM, Carlson EE. A reinvigorated era of bacterial secondary metabolite discovery. Curr Opin Chem Biol. 2015; 24:104–111. [PubMed: 25461728]
11. 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. [PubMed: 25262415]
12. Lautru S, Deeth RJ, Bailey LM, Challis GL. Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. Nat Chem Biol. 2005; 1:265–269. [PubMed: 16408055]

13. Pidot S, Ishida K, Cyrulies M, Hertweck C. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. Angew Chem, Int Ed. 2014; 53:7856–7859.

14. Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BVB, Montiel D, Brady SF. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. Proc Natl Acad Sci U S A. 2015; 112:4221–4226. [PubMed: 25831524]

15. Kang HS, Brady SF. Mining Soil Metagenomes to Better Understand the Evolution of Natural Product Structural Diversity: Pentangular Polyphenols as a Case Study. J Am Chem Soc. 2014; 136:18111–18119. [PubMed: 25521786]

16. Liu WT, Kersten RD, Yang YL, Moore BS, Dorrestein PC. Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in Streptomyces roseosporus. J Am Chem Soc. 2011; 133:18010–18013. [PubMed: 21999343]

17. Ju KS, Gao J, Doroghazi JR, Wang K-Ka, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, Lee J, Cioni JP, Evans BS, Hirota R, Labeda DP, van der Donk Wa, Metcalf WW. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. Proc Natl Acad Sci U S A. 2015; 112:12175–12180. [PubMed: 26324907]

18. Medema MH, Fischbach MA. Computational approaches to natural product discovery. Nat Chem Biol. 2015; 11:639–648. [PubMed: 26284671]

19. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. Nat Commun. 2013; 4:2894. [PubMed: 24305602]

20. Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol. 2015; 13:509–523. [PubMed: 26119570]

21. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, Birren BW, Takano E, Sali A, Linington RG, Fischbach MA. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. Cell. 2014; 158:412–421. [PubMed: 25036635]

22. Ziemert N, Lechner A, Wietz M, Millán-Aguiñaga N, Chavarria KL, Jensen PR. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. Proc Natl Acad Sci U S A. 2014; 111:E1130–E1139. [PubMed: 24616526]

23. 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. [PubMed: 27163034]

24. McClure RA, Goering AW, Ju KS, Baccile JA, Schroeder FC, Metcalf WW, Thomson RJ, Kelleher NL. Elucidating the Rimosamide-Detoxin Natural Product Families and Their Biosynthesis Using Metabolite/Gene Cluster Correlations. ACS Chem Biol. 2016; 11:3452–3460. [PubMed: 27809474]

25. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC. Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A. 2012; 109:E1743–E1752. [PubMed: 22586093]

26. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, Glukhov E, Wodtke A, de Felicio R, Fenner A, Wong WR, Linington RG, Zhang L, Debonsi HM, Gerwick WH, Dorrestein PC. Molecular networking as a dereplication strategy. J Nat Prod. 2013; 76:1686–1699. [PubMed: 24025162]

27. Nguyen DD, Wu CH, Moree WJ, Lamsa A, Medema MH, Zhao X, Gavilan RG, Aparicio M, Atencio L, Jackson C, Ballesteros J, Sanchez J, Watrous JD, Phelan VV, van de Wiel C, Kersten RD, Mehnaz S, De Mot R, Shank EA, Charusanti P, Nagarajan H, Duggan BM, Moore BS, Bandeira N, Palsson BØ, Pogliano K, Gutiérrez M, Dorrestein PC. MS/MS networking guided analysis of molecule and gene cluster families. Proc Natl Acad Sci U S A. 2013; 110:E2611–E2620. [PubMed: 23798442]

28. Henke MT, Soukup AA, Goering AW, McClure RA, Thomson RJ, Keller NP, Kelleher NL. New Aspercryptins, Lipopeptide Natural Products, Revealed by HDAC Inhibition in *Aspergillus nidulans ACS*. Chem Biol. 2016; 11:2117–2123.
29. Covington BC, McLean JA, Bachmann BO. Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. Nat Prod Rep. 2017; 34:6–24. [PubMed: 27604382]
30. Duncan KR, Crüsemann M, Lechner A, Sarkar A, Li J, Ziemert N, Wang M, Bandeira N, Moore BS, Dorrestein PC, Jensen PR. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from salinispora species. Chem Biol. 2015; 22:460–471. [PubMed: 25865308]
31. Kleigrewe K, Almaliti J, Tian IY, Kinnel RB, Korobeynikov A, Monroe EA, Duggan BM, Di Marzo V, Sherman DH, Dorrestein PC, Gerwick L, Gerwick WH. Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. J Nat Prod. 2015; 78:1671–1682. [PubMed: 26149623]
32. Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MRM, Yang L, Zechel DL, Ma B, Magarvey NA. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. Nat Commun. 2015; 6:8421. [PubMed: 26412281]
33. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011; 39:W339–W346. [PubMed: 21672958]
34. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, Wyatt MA, Magarvey NA. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res. 2015; 43:9645–9662. [PubMed: 26442528]
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
36. Fu C, Keller L, Bauer A, Brönstrup M, Froidbise A, Hammann P, Herrmann J, Mondesert G, Kurz M, Schiell M, Schummer D, Toti L, Wink J, Müller R. Biosynthetic Studies of Telomycin Reveal New Lipopeptides with Enhanced Activity. J Am Chem Soc. 2015; 137:7692–7705. [PubMed: 26043159]
37. Latham J, Brandenburger E, Shepherd SA, Menon BRK, Micklefield J. Development of Halogenase Enzymes for Use in Synthesis. Chem Rev. 2018; 118:232–269. [PubMed: 28466644]
38. Rachid S, Scharfe M, Blöcker H, Weissman KJ, Müller R. Unusual Chemistry in the Biosynthesis of the Antibiotic Chondrochlorens. Chem Biol. 2009; 16:70–81. [PubMed: 19171307]
39. Kittilä T, Kittel C, Tailhades J, Butz D, Schoppet M, Büttner A, Goode RJA, Schittenhelm RB, van Pee KH, Süssmuth RD, Wohlleben W, Cryle MJ, Stegmann E. Halogenation of glycopeptide antibiotics occurs at the amino acid level during non-ribosomal peptide synthesis. Chem Sci. 2017; 8:5992–6004. [PubMed: 28989629]
40. Geiger O, López-Lara IM, Sohlenkamp C. Phosphatidylcholine biosynthesis and function in bacteria. Biochim Biophys Acta, Mol Cell Biol Lipids. 2013; 1831:503–513.
41. Ansari M, Sharma J, Gokhale RS, Mohanty D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. BMC Bioinf. 2008; 9:454.
42. Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr Opin Chem Biol. 2001; 5:525–534. [PubMed: 11578925]
43. Zhu X, Kim J, Su X, Lin H. Reconstitution of diphthine synthase activity in vitro. Biochemistry. 2010; 49:9649–9657. [PubMed: 20873788]
44. Vit A, Misson L, Blankenfeldt W, Seebeck FP. Ergothioneine Biosynthetic Methyltransferase EgtD Reveals the Structural Basis of Aromatic Amino Acid Betaine Biosynthesis. ChemBioChem. 2015; 16:119–125. [PubMed: 25404173]
45. Burg MB, Ferraris JD. Intracellular organic osmolytes: function and regulation. J Biol Chem. 2008; 283:7309–7313. [PubMed: 18256030]

46. Naresh Chary V, Dinesh Kumar C, Vairamani M, Prabhakar S. Characterization of amino acid-derived betaines by electrospray ionization tandem mass spectrometry. J Mass Spectrom. 2012; 47:79–88. [PubMed: 22282093]
47. Rhodes D, Hanson AD. Quaternary Ammonium and Tertiary Sulfonium Compounds in Higher Plants. Annu Rev Plant Physiol Plant Mol Biol. 1993; 44:357–384.
48. Liu X, Gan M, Dong B, Zhang T, Li Y, Zhang Y, Fan X, Wu Y, Bai S, Chen M, Yu L, Tao P, Jiang W, Si S. 4862F, a New Inhibitor of HIV-1 Protease, from the Culture of Streptomyces I03A-04862. Molecules. 2013; 18:236–243.
49. Bernard T, Goas G. Biosynthese de la sticticine chez le lichen Lobaria laetevirens. Physiol Plant. 1981; 53:71–75.
50. Shen S, Liu D, Wei C, Proksch P, Lin W. Purpuroines A-J, halogenated alkaloids from the sponge Iotrochota purpurea with antibiotic activity and regulation of tyrosine kinases. Bioorg Med Chem. 2012; 20:6924–6928. [PubMed: 23131412]
51. Galeano E, Thomas OP, Robledo S, Munoz D, Martinez A. Antiparasitic bromotyrosine derivatives from the marine sponge Verongula rigida. Mar Drugs. 2011; 9:1902–1913. [PubMed: 22073002]
52. Tsuda M, Endo T, Watanabe K, Fromont J, Kobayashi J. Nakirodin A, a bromotyrosine alkaloid from a Verongid sponge. J Nat Prod. 2002; 65:1670–1671. [PubMed: 12444697]
53. Ciminiello P, Dell'Aversano C, Fattorusso E, Magno S, Pansini M. Chemistry of verongida sponges. 10 Secondary metabolite composition of the caribbean sponge Verongula gigantea. J Nat Prod. 2000; 63:263–266. [PubMed: 10691724]
54. Shrestha T, Bisset NG. Quaternary nitrogen compounds from south American moraceae. Phytochemistry. 1991; 30:3285–3287.
55. Eggenberger F, Daloze D, Pasteels JM, Rowell-Rahier M. Identification and seasonal quantification of defensive secretion components of Oreina gloriosa (Coleoptera: Chrysomelidae). Experientia. 1992; 48:1173–1179.
56. Chaleckis R, Ebe M, Pluskal T, Murakami I, Kondoh H, Yanagida M. Unexpected similarities between the Schizosaccharomyces and human blood metabolomes, and novel human metabolites. Mol BioSyst. 2014; 10:2538–2551. [PubMed: 25010571]
57. Sasaki Y, Watanabe Y, Ambo A, Suzuki K. Synthesis and biological properties of quaternized N-methylation analogs of D-Arg-2-dermorphin tetrapeptide. Bioorg Med Chem Lett. 1994; 4:2049–2054.
58. Thorne GC, Ballard KD, Gaskell SJ. Metastable decomposition of peptide [M + H]+ ions via rearrangement involving loss of the C-terminal amino acid residue. J Am Soc Mass Spectrom. 1990; 1:249–257.
59. Kawai M, Fukuta N, Ito N, Kagai T, Butsugan Y, Maruyama M, Kudo Y. Preparation and opioid activities of N-methylated analogs of [D-Ala2,Leu5]enkephalin. Int J Pept Protein Res. 1990; 35:452–459. [PubMed: 2165469]
60. Kawai M, Fukuta N, Ito N, Ohya M, Butsugan Y, Maruyama M, Kudo Y. Preparation and properties of trimethylammonium group-containing analogues of [D-Ala2, Leu5]-enkephalin and gramicidin S. J Chem Soc, Chem Commun. 1986; 0:955–956.
61. Yamato M, Koguchi T, Okachi R, Yamada K, Nakayama K, Kase H, Karasawa A, Shuto K. K-26, a novel inhibitor of angiotensin I converting enzyme produced by an actinomycete K-26. J Antibiot. 1986; 39:44–52. [PubMed: 3005218]
62. Ntai I, Bachmann BO. Identification of ACE pharmacophore in the phosphonopeptide metabolite K-26. Bioorg Med Chem Lett. 2008; 18:3068–3071. [PubMed: 18096385]

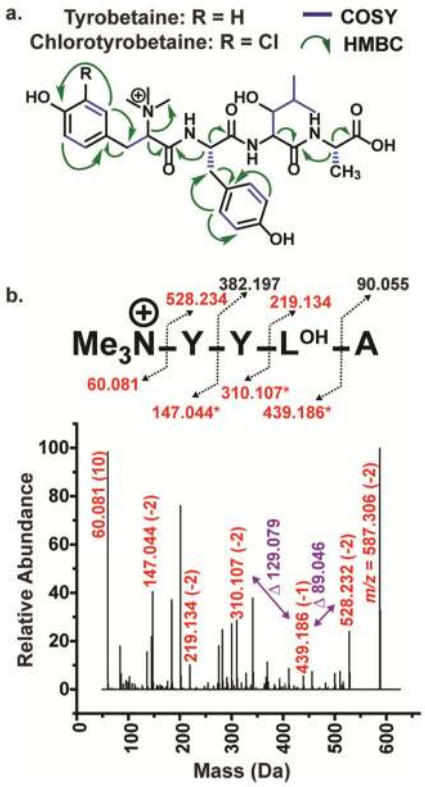


Figure 1.

Structure elucidation of tyrobetaine and chlorotyrobetaine. (a) The structure of tyrobetaine and chlorotyrobetaine with key NMR correlations indicated. (b) The MS^2 spectrum for tyrobetaine with key masses indicated. *Indicates the mass after removal of the trimethylammonium. L^{OH} = hydroxyleucine. Red masses indicate observed masses. Δppm values are indicated in parentheses after the key masses. Key mass losses are in purple.

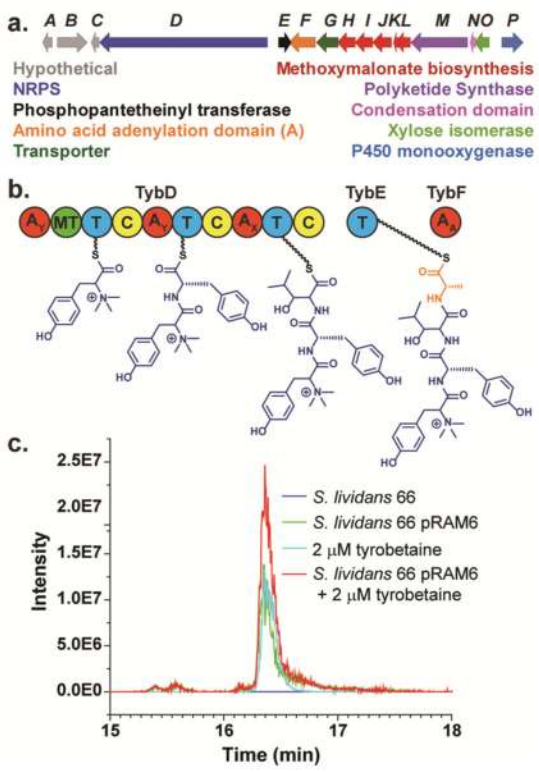


Figure 2.

Tyrobetaine BGC and heterologous expression. (a) The BGC for the tyrobetaines from *Streptomyces* sp. WC-3703. Arrows indicate genes. The color of the arrow corresponds to the type of gene (indicated below the arrows). See Table S5 for BLAST analysis. (b) AntiSMASH NRPS domain predictions for TybD, TybE, and TybF. A = adenylation domain with subscript indicating the predicted amino acid (Y = tyrosine, X = no consensus, A = alanine). MT = *N*-methyltransferase. T = thiolation domain. C = condensation domain. See Table S6 for NRPS adenylation domain predictions. (c) Extracted ion chromatogram for tyrobetaine (587.30–587.31) in wild type *S. lividans* 66, *S. lividans* 66 pRAM6 (heterologous expression of NRPS_GCF.432), purified tyrobetaine, and *S. lividans* 66 pRAM6 spiked with purified tyrobetaine.

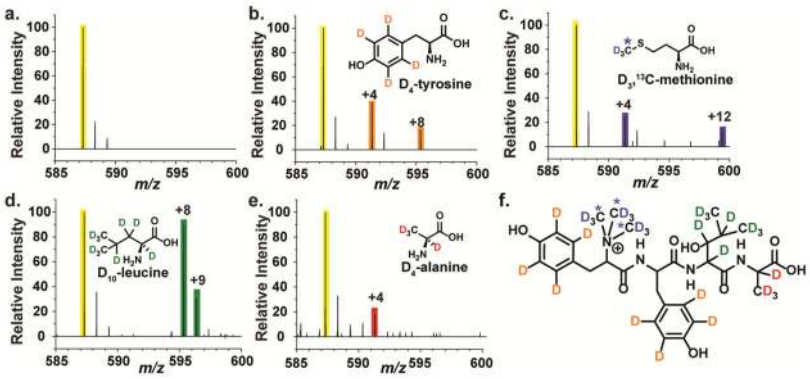


Figure 3.

Feeding experiments with stable isotope labeled amino acids. Mass spectrum from spent media from *Streptomyces sp.* NRRL WC-3703 grown on (a) medium alone or medium with (b) D_4-L-tyrosine, (c) D_3,^{13}C-methionine, (d) D_{10}-L-leucine, or (e) D_4-L-alanine. (f) Structure of tyrobetaine with likely location of stable isotopes. Yellow bar indicates the *m*/*z* for tyrobetaine. Colored bars indicate isotopically labeled tyrobetaine. See Figure S6 for chlorotyrobetaine.

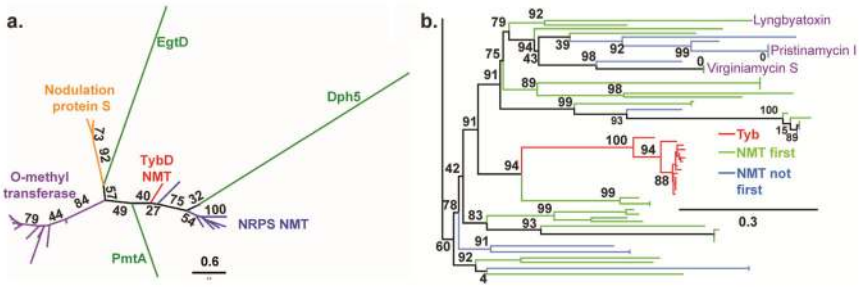


Figure 4.

Phylogenetic analysis of TybD *N*-methyltransferase. (a) Phylogenetic tree for the *N*-methyltransferase of TybD compared to that of known methyltransferases from the UniProtKB/Swiss-Prot database. NMT = *N*-methyltransferase. Functions of proteins are indicated by color and defined by text of the same color. Black numbers are bootstrap support percentages. (b) A portion of the phylogenetic tree for the *N*-methyltransferase of TybD compared to that of nonredundant protein sequences from NCBI. The products of known NRPS *N*-methyltransferases are indicated in purple text. Black numbers are FastTree support values. See Figure S7B for the full figure.

Table 1

Streptomyces Strains That Produce the Tyrobetaines and Potential Biosynthetic Gene Clusters[a]

Strain ID	HRMS Intensity				NRPS Gene Cluster Family[b]								
	Tyb	Cl-Tyb	Tyb-2	Cl-Tyb-2									
NRRL WC-3868	1.4E+09	7.8E+06	2.2E+09	1.3E+07	424	83	432	525	436	599	465	469	85
NRRL WC-3703	1.3E+09	8.9E+07	1.1E+09	4.1E+08	424	83	432	525	436	599	465	469	85
NRRL WC-3882	1.1E+09	5.1E+06	6.6E+08	2.0E+07	424	83	432	525	436		465	469	85
NRRL WC-3877	6.4E+08	7.2E+07	1.6E+09	2.9E+08	424	83	432	525	436	599	465	469	85
NRRL WC-3773	5.9E+08	6.8E+06	1.2E+09	1.2E+08	424	83	432			599	465		85
NRRL WC-3929	4.4E+08	4.1E+07	3.4E+08	4.8E+07	424	83	432	525	436	599	465	469	85
NRRL B-2659	3.2E+08	3.2E+07	3.0E+07	4.4E+08	424	83	432	525	436	599	465	469	
NRRL WC-3927	3.1E+08	7.5E+07	2.3E+08	4.0E+07	424	83	432	525	436	599	465	469	85
NRRL WC-3908	3.1E+08	1.8E+06	1.3E+09	6.7E+06	424	83	432			599			85
NRRL B-2661	1.8E+08	2.8E+07	6.1E+07	1.8E+07	424	83	432	525	436	599	465	469	85

[a]HRMS = high resolution mass spectrometry. Tyb = tyrobetaine. Cl-Tyb = chlorotyrobetaine. Tyb-2 = tyrobetaine-2. Cl-Tyb-2 = chlorotyrobetaine-2.

[b]Gene cluster families are named by TypeOfNaturalProduct_GCF.IdentifyingNumber (*e.g.*, NRPS_GCF.424). The numbers indicated here refer to the identifying number for an NRPS GCF. The GCFs can be found at http://www.igb.illinois.edu/labs/metcalf/gcf/.

Table 2

Antibiotic Activity, Anticancer Activity, and Protease Inhibition of Tyrobetaine and Chlorotyrobetaine[a]

organism/protease	tyrobetaine MIC[b] or IC_{50}[c] (μM)	chlorotyrobetaine MIC[b] or IC_{50}[c] (μM)	oxytetracycline MIC (μg/mL)	oxytetracycline + 32 μg/mL tyrobetaine MIC (μg/mL)
E. faecalis ATTC 19433	>100	>100	ND	ND
S. aureus ATCC 29213	>100	>100	ND	ND
K. pneumoniae ATCC 27736	>100	>100	ND	ND
A. baumannii ATCC 19606	>100	>100	ND	ND
P. aeruginosa PAO1	>100	>100	ND	ND
E. coli ATCC 25922	>100	>100	4	8
A549 human lung cancer	>100	>100	ND	ND
ACE	>10	ND	ND	ND
HIV-1 protease	>10	ND	ND	ND

[a] ND = not determined. ACE = angiotensin converting enzyme.

[b] Minimum inhibitory concentration (MIC) for bacteria determined based on CLSI guidelines.

[c] 50% inhibitory concentration (IC_{50})