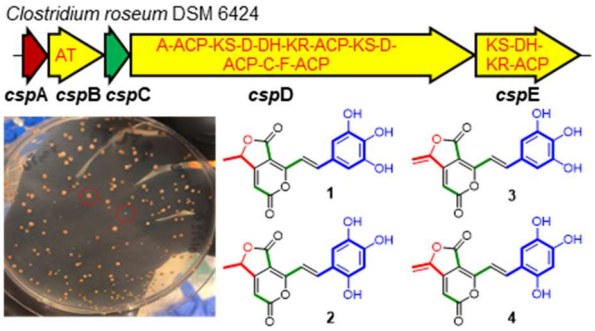
Discovery and biosynthesis of clostyrylpyrones from the obligate anaerobe *Clostridium roseum*

Jeffrey S Li[†], Yongle Du[†], Di Gu[‡], Wenlong Cai[†], Allison Green[†], Samuel Ng[†], Alexander Leung[†], Wenjun Zhang*[†§]

[†]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States

[‡]Department of Chemistry, University of California, Berkeley, California 94720, United States

[§] Chan Zuckerberg Biohub, San Francisco, California 94158, United States



ABSTRACT: Anaerobic bacteria are a promising new source for natural product discovery. Examination of extracts from the obligate anaerobe *Clostridium roseum* led to discovery of a new family of natural products, the clostyrylpyrones. The polyketide synthase-based biosynthetic mechanism of clostyrylpyrones is further proposed based on bioinformatic, gene knockout and heterologous expression studies.

Natural products (NPs) are important sources of privileged chemical scaffolds for development of drugs and other bioactive compounds.[1] Whereas natural product discovery has been historically biased toward more easily culturable aerobic microbes, anaerobes are now recognized as an underutilized reservoir for the discovery of NPs, based on genomic analyses demonstrating the widespread occurrence of biosynthetic gene clusters (BGCs) encoding major classes of NPs such as polyketides and non-ribosomal peptides.[2] Increasingly available chemical data for anaerobe-derived NPs speak to the structural and functional diversity which can be accessed by exploring this resource.[3-8]

The Class Clostridia is a subset of the anaerobes which are particularly enriched in biosynthetic potential for polyketides based on genomic analyses.[9] Polyketides are biosynthesized by characteristic multidomain assembly line enzymes which join malonyl-CoA monomers by decarboxylative Claisen condensation, often in iterative or modular processing.[10] While polyketide synthase (PKS) genes have been detected in many Clostridia, very few examples of polyketides have been reported (**Figure 1**), all of which are associated with iterative PKSs. For example, clostrienose is a product of an iterative type I single-module PKS from *Clostridium acetobutylicum.*[11] Clostrubins are aromatic polyketides biosynthesized by a type II PKS in both *Clostridium beijerinckii* and *Clostridium puniceum*.[12,13] Through heterologous expression of a type II PKS BGC of *Blautia wexlerae* (belong to the class Clostridia) in a *Bacillus subtilis* host, a new aromatic polyketide, wexrubicin, was also discovered.[8] These compounds exhibit structural and functional diversity and complexity, underscoring the value of studying polyketides from Clostridia.

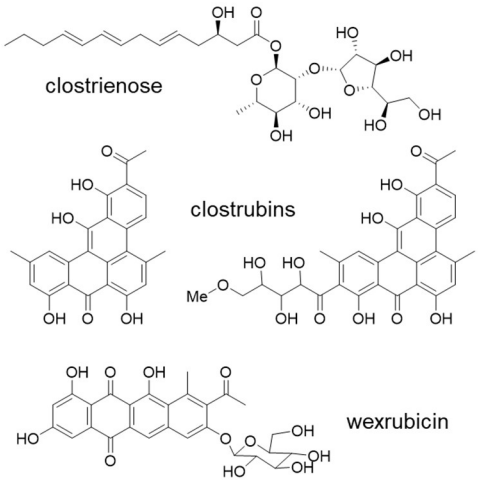


Figure 1. Polyketide NPs associated with Clostridia.

Clostridium roseum DSM 6424 (*Cro*), first isolated from German maize,[14] is an important anaerobe with broad biotechnological significance for production of organic solvents and biofuels.[15] Recently, *Cro* has been highlighted as one promising organism for NP discovery due to the presence of large thio-templated BGCs in its genome.[3] *Cro* is noteworthy among anaerobes in its production of pigments, including yellow-orange compounds which are secreted into the culture media. We thus firstly targeted these compounds for structural elucidation.

HPLC examination of ethyl acetate extracts of *Cro* culture revealed the presence of several peaks with UV/vis maxima around 450 nm (**Figures 2** and **S1**). HR-ESI-MS analyses of the three major peaks (I-III) in positive ion mode enabled prediction of their associated chemical formulae: I, $C_{16}H_{10}O_7$ (calculated for $C_{16}H_{11}O_7^+$: m/z 315.0499; found: m/z 315.0503); II, $C_{16}H_{12}O_7$ (calculated for $C_{16}H_{13}O_7^+$: m/z 317.0656; found: m/z 317.0663); III, $C_{16}H_{12}O_8$ (calculated for $C_{16}H_{13}O_8^+$: m/z 333.0605; found: m/z 333.0607) (**Figure S1**). The predicted molecular formulae and similarities in the UV spectra suggested these peaks were congeners. For purification and structural elucidation, a four-liter batch culture of *Cro* was extracted and purified by size-exclusion chromatography with a Sephadex LH-20 column packing, followed by reverse-phase C18 HPLC purification. Peak I was isolated as a brownish amorphous solid with a yield of 0.4 mg; Peak II as an orange amorphous solid with 0.56 mg. Peak I material exhibited remarkable thermochromic and solvatochromic properties after isolation (**Figure S2**). The yield of Peak III was too low to be further characterized by NMR spectroscopic analysis.

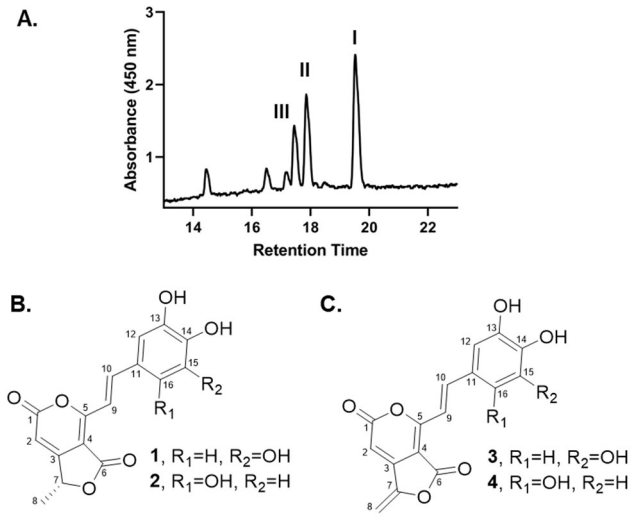


Figure 2. Discovery of clostyrylpyrones. (A) UV traces (450 nm) during LC-MS analysis of *Cro* extracts. (B) Structures of compounds isolated from Peak II. (C) Structures of compounds isolated from Peak I.

A detailed analysis of the 1D and 2D NMR spectra (^{1}H, ^{13}C, gHSQC, and gHMBC) of Peak I and II led to the discovery of four chemical structures, named the clostyrylpyrones A-D (**1-4**) (**Figures 2, S3** and **S4, Table S1**). In particular, Peak II contained a mixture of two structural isomers, **1** (74%) and **2,** which were inseparable during HPLC purification; similarly, Peak I contained compounds **3** (53%) and **4**. These deductions are supported by the following evidence: The obtained ^{1}H NMR and HSQC spectra displayed signals for four hydroxyl groups (δ_H 9.34, 13/15-OH of **1**; δ_H 9.11, 13-OH of **2**; δ_H 9.08, 14-OH of **2**), two trans-disubstituted olefins (δ_H 7.62, H-10 of **1**; δ_H 7.55, H-10 of **1**; δ_H 7.30, H-9 of **1**; δ_H 7.27, H-9 of **2**), six aromatic or olefinic protons (δ_H 6.73, H-12 and H-16 of **1**; δ_H 6.70, H-12 of **2**; δ_H 6.58, H-15 of **2**; δ_H 6.30, H-2 of **1**; δ_H 6.23, H-2 of **2**), two oxygen bearing methines (δ_H 5.53, H-7 of **1/2**), and two methyl groups (δ_H 1.53, H_3-8 of **1/2**). Furthermore, the ^{13}C NMR and HSQC spectra showed signals for eighteen quaternary sp^2 carbons, ten aromatic or olefinic methines, two oxygen bearing alkyl methines, and two methyl groups. The presence of each dilactone fragment was proposed through their proton and carbon chemical shifts and supported by their ^3J-HMBC cross-peaks from H-2 to C-4, H-7 to C-2 and H-8 to C-3, as well as ^2J-HMBC from H-2 to C-1 and H-7 to C-3. They were independently connected to a double bond based on ^3J-HMBC correlations from H-9 to C-4 and H-10 to C-5. In addition, the other side of each double bond was linked to pyrogallol or hydroxyquinol by the observations from HMBC correlation. The absolute configuration of compounds **1/2** was assigned by the ECD spectrum (**Figure S5**). The n → π* transition bands at 243 nm showed a negative Cotton effect and 264 nm showed a positive Cotton effect, defining the absolute configuration at C-7 as R.[16,17]

The difference of 2 Da between the masses of **1/2** and **3/4** suggested that **3/4** might be dehydrogenated products of **1/2**. Dehydrogenation of the double bond of C-7 to C-8 in **1/2** was confirmed by the deshielding of the chemical shifts of C-7 and C-8 from δ_C 76.1 (CH) and 19.3 (CH_3) in **1/2** to δ_C 150.7 (C) and 94.8 (CH_2) in **3/4,** respectively. The NMR spectroscopic data of **3** and **4** were similar to those of **1** and **2** except as noted above and were fully consistent with its structural assignment as dehydrogenated products of **1** and **2** at C-7 and C-8, as confirmed by the HMBC cross-peaks (**Figure S4**).

We have thus discovered a group of new compounds sharing an unprecedented fused five- and six-membered dilactone scaffold with a trihydroxystyrene substituent at the C-5 position. It is notable that the styrene-substituted six-membered pyrone is a motif commonly found from eukaryotic phenylpropanoid NPs (**Figure S6**). These known styrylpyrones are often biosynthesized from aromatic amino acids which have undergone ammonia elimination and extension by type III PKS in plants[18,19] or iterative type I PKS in fungi.[20] We then reasoned that **1-4** could be biosynthesized by a similar logic, via a PKS promoted extension of an aromatic starter unit, although the enzymatic machinery for the assembly of the unique dilactone scaffold remained elusive.

To probe the biosynthesis of clostyrylpyrones, we scanned the *Cro* genome using antiSMASH[21] for PKS-containing BGCs and identified a candidate locus for clostyrylpyrone biosynthesis, here designated as *cspA-E* (**Figure 3a**). Bioinformatic analyses of the five genes revealed a *trans*-AT five-module PKS complex, encoded by *cspB, cspD* and *cspE*, which is in agreement with the predicted five PKS building monomers of clostyrylpyrones (one aromatic starter unit and four extender units). *CspA* and *cspC* encode two modification enzymes, 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) synthase and a dehydrogenase, which might be involved in the aromatic starter unit biosynthesis and PKS chain modification, respectively. To confirm the requirement of *csp* for clostyrylpyrone biosynthesis, a *cspD* inactivation mutant (Δ*cspD*) was constructed in *Cro* using a suicide vector through double crossover, and the genotype of the resulting mutant was verified by PCR (**Figure S7**). The production of clostyrylpyrones was completely abolished in Δ*cspD* (**Figure 4a**), confirming that the identified *csp* cluster is involved in the biosynthesis of clostyrylpyrones. No homologous gene clusters were identified from public databases based on multi-Blast analysis, consistent with the unique chemical structures of clostyrylpyrones.

A. *Clostridium roseum* DSM 6424 genome, *csp* locus

Gene	Size (aa)	Accession	Deduced Role and Annotation	Closest protein homolog found in:	Homolog Accession	Identity/ Similarity (%/%)
cspA	342	OOM01707.1	DAHP synthase	*Candidatus Dichloromethanomonas elyunquensis*	RNC29451.1	63/80
cspB	1083	OOM01708.1	acyltransferase	*Hungateiclostridium alkalicellulosi*	WP_148889894.1	52/70
cspC	372	OOM01709.1	dehydrogenase	*Methylomusa anaerophila*	WP_126309542.1	67/84
cspD	4807	OOM01710.1	polyketide synthase	*Paenibacillus polymyxa*	WP_049828323.1	39/58
cspE	1210	OOM01711.1	polyketide synthase	*Pyxidicoccus* sp. MCy9557	ASA76633.1	42/61
ORF 1	534	OOM01712.1	transporter	*Bacillus cereus*	WP_098796821.1	42/64
ORF 2	90	OOM01713.1	hypothetical protein	Desulfobacteraceae sp.	TET87557.1	59/77
ORF 3	575	OOM01714.1	transporter	*Clostridium felsineum*	WP_077893534.1	100/100
ORF 4	279	OOM01704.1	transcription regulator	*Clostridium felsineum*	WP_077893537.1	100/100
ORF 5	472	OOM01705.1	peptidase-like superfamily	*Clostridium felsineum*	WP_077893536.1	99/100
ORF 6	183	OOM01706.1	membrane lipoprotein	*Clostridium felsineum*	WP_077893535.1	98/98

B.

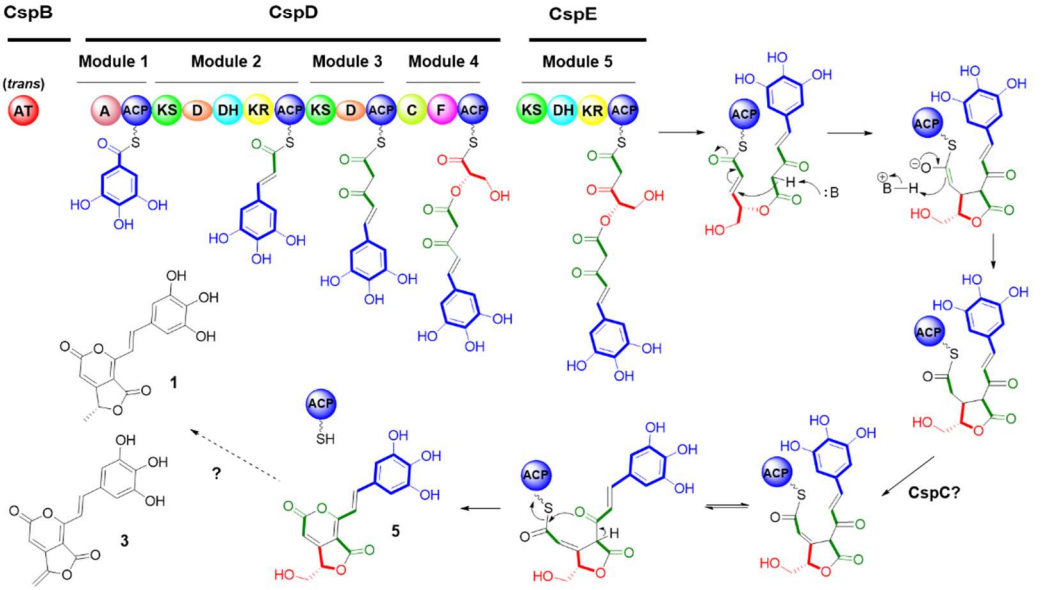


Figure 3. Map of *csp* gene neighborhood and proposed biosynthetic pathway. (A) Organization of *csp* genes, nearby open reading frames (ORFs), and their deduced roles based on sequence homology. Genes are colored by function: light blue, PKS related genes; dark blue, other biosynthetic genes; black, transcriptional regulator; grey, other; green, transporter. Domain abbreviations: AT, acyltransferase; A, adenylation; ACP, acyl carrier protein; KS, ketosynthase; D, *trans*-AT docking; DH, dehydratase; KR, ketoreductase; F, FkbH-like. (B) Proposed biosynthesis of clostyrylpyrones. For simplicity, gallic acid is depicted as a representative starting monomer.

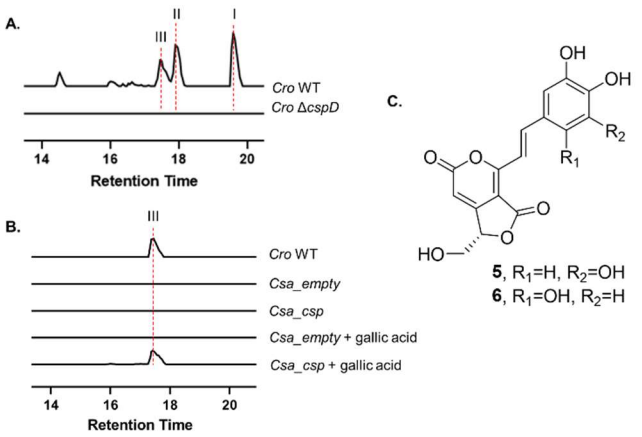


Figure 4. LC-HRMS analysis of metabolites. Extracted ion chromatograms are shown. The calculated masses (I: *m/z* = 315.0499 $[M+H]^+$; II: *m/z* = 317.0656 $[M+H]^+$; III: *m/z* = 333.0605 $[M+H]^+$) with a 10-p.p.m. mass error tolerance were used. (A) Deletion of *cspD* abolishes clostyrylpyrone production in *Cro*. (B) Heterologous expression of *cspA-E*. (C) Putative structures of compound **5**/**6** from peak III of *Cro*.

The identification of *csp* has allowed us to propose a putative enzymatic pathway for clostyrylpyrone biosynthesis (**Figure 3b**). The PKS assembly line starts with the activation of 3,4,5-trihydroxybenzoic acid (gallic acid) or 2,4,5-trihydroxybenzoic acid by the initial adenylation (A) domain of CspD. *In silico* sequence alignment and modeling of this A domain identified residues potentially interacting with the aromatic substrates and yielded an unprecedented specificity motif[22] (MPWGLVDGIK), consistent with incorporation of novel trihydroxybenzoic acid substrates (**Figure S8**). While the biogenesis of 2,4,5-trihydroxybenzoic acid remains unknown, gallic acid has been known to derive from the shikimic/chorismic acid pathways in both plant and bacterial model organisms.[23] We further identified the predicted enzymes *in silico*, including the dedicated 3-dehydroshikimate reductase (AroE), involved in gallic acid biosynthesis which are encoded elsewhere on the *Cro* genome (**Figure S9)**. In addition, *cspA* encodes an extra copy of DAHP synthase, which catalyzes the committed first step in the shikimic/chorismic acid pathways. CspA is thus proposed to further channel the primary metabolites to the trihydroxybenzoic acid starter unit synthesis in *Cro*.

After the starter unit activation and loading onto CspD, the chain is elongated twice using malonyl-CoA by the next two typical PKS modules of CspD, before the action of the fourth atypical module of CspD with domains organized into C-F-ACP (C: condensation, F: FkbH-like domain, ACP: acyl carrier protein) (**Figure 3b)**. The presence of an F domain is reminiscent of the biosynthetic mechanism of agglomerin and other tetronate NPs which feature exocyclic-olefin bearing lactones, as in **3** and **4**.[24] We thus propose that this module enables the incorporation of D-glycerate, derived from a primary metabolite such as D-1,3-bisphosphoglycerate, via a regiospecific ester bond formation into the PKS assembly line.[25,26] After the last chain extension promoted by CspE, a Michael addition reaction takes place to generate the five-membered lactone, accompanied by the chain release from the assembly line through the six-membered pyrone formation. The PKS complex lacks an obvious chain release domain, and the fused five- and six-membered dilactone cyclization is thus most likely spontaneous associated with product release. CspC, a predicted dehydrogenase, may catalyze the double bond formation between C-2 and C-3 on or post the assembly line to yield the proposed hydroxylated intermediates, **5** and **6**. It is notable that **5/6** have chemical formulae of $C_{16}H_{12}O_8$, consistent with the predicted formula associated with the Peak III minor product (**Figure 2**). **3/4** could form from **5/6** through acetylation–elimination as observed in tetronate antibiotics biosynthesis,[27] but the homologous acetyltransferase and eliminating enzyme were not identified from the genome of *Cro*, suggesting that the conversion of **5/6** to **3/4** and **1/2** may involve a different mechanism and the relevant biosynthetic genes are not clustered with *csp*.

To further confirm the role of *csp* in clostyrylpyrone biosynthesis, we cloned the 22 kb *cspA-E* into a plasmid for heterologous expression studies in *Clostridium saccharoperbutylacetonicum* N1-4 (*Csa*).[28] Based on the proposed biosynthetic activity of *cspA-E*, we proposed that the functional expression of *csp* would yield product **5/6**. As expected, a new product was detected in the engineered strain of *Csa_csp* when the culture was supplemented with gallic acid. This product demonstrated identical retention time and mass-to-charge ratio to that of the Peak III minor product (predicted to be **5/6**) from the wild-type *Cro*, and the product was thus predicted to be **5** as its production required gallic acid (**Figure 4b**). The heterologous expression strain of *Csa_csp* did not produce peaks corresponding to compounds **1-4**, which is consistent with the prediction that *csp* lacks genes responsible for subsequent maturation of the biosynthetic intermediates **5/6**. This work thus adds another example of non-clustered biosynthetic genes in Clostridia, raising concerns toward employing a heterologous expression strategy for revealing native natural products.

In summary, we here report the discovery of the clostyrylpyrones, a novel family of polyketides isolated from an obligate anaerobe. Structural elucidation is described for four compounds (**1** to **4**), all featuring an unprecedented scaffold of styrylpyrone fused to a tetronate-like heterocycle. We have further identified the biosynthetic gene cluster of clostyrylpyrones by genome mining, targeted gene deletion and heterologous expression, which sets the stage for deciphering the chemical logic and enzymatic machinery in clostyrylpyrone biosynthesis. A complex modular PKS with a *trans*-acting AT are employed in biosynthesis, and the unique dilactone scaffold is formed due to the enzymatic activity of FkbH-like domain containing PKS module as well as the non-enzymatic cyclization promoted by Michael addition and pyrone formation. The discovery of the clostyrylpyrones expands the limited chemical space of known anaerobe-derived polyketides and encourages future efforts in mining new chemical scaffolds from anaerobes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Experimental details, NMR data, and Supplementary Figures 1-9.

AUTHOR INFORMATION

Corresponding Author

* E-mail: wjzhang@berkeley.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank J. Pelton (UC Berkeley) for helping with NMR spectroscopic analysis. This work was supported by the National Institutes of Health (DP2AT009148) and the Chan Zuckerberg Biohub Investigator Program.

REFERENCES

1. Newman, D., & Cragg, G. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79(3), 629-661 (2016).

2. Letzel, A., Pidot, S. J., Hertweck, C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat. Prod. Rep.* 30(3), 392-428 (2013).

3. Li, J. S., Barber, C. C., Zhang, W. Natural products from anaerobes. *J. Ind. Microbiol. Biotechnol.* 46(3-4), 375-383 (2019).

4. Ishida, K., Shabuer, G., Schieferdecker et al. Oak-associated Negativicute equipped with ancestral aromatic polyketide synthase produces antimycobacterial dendrubins. *Chem. Eur. J.* Chem.202001939. doi: 10.1002/chem.202001939 (2020).

5. Schieferdecker, S., Shabuer, G., Letzel et al. Biosynthesis of diverse antimicrobial and antiproliferative acyloins in anaerobic bacteria. *ACS Chem. Biol.* 14(7), 1490-1497 (2019).

6. Schieferdecker, S., Shabuer, G., Knuepfer, U., Hertweck, C. Clostrindolin is an antimycobacterial pyrone alkaloid from *Clostridium beijerinckii*. *Org. Biomol. Chem.* 17(25), 6119-6121 (2019).

7. Li, J. S., Barber, C. C., Herman, N. A., Cai, W., Zafrir, E., Du, Y., Zhu, X., Skyrud, W., Zhang, W. Investigation of secondary metabolism in the industrial butanol hyper-producer *Clostridium saccharoperbutylacetonicum* N1-4. *J. Ind. Microbiol. Biotechnol.* 47, 319-328 (2020).

8. Sugimoto, Y., Camacho, F. R., Wang, S., Chankhamjon, P., Odabas, A., Biswas, A., Jeffrey, P. D., Donia, M. S. A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science*. 366, 6471 (2019).

9. Behnken, S. & Hertweck, C. Cryptic polyketide synthase genes in non-pathogenic *Clostridium* spp. *PloS One.* 7(1), e29609 (2012).

10. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* 48(26), 4688-4716 (2009).

11. Herman, N., Kim, S. J., Li, J. S., Cai, W., Koshino, H., Zhang, W. The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation. *Nat. Comm.* 8(1), 1514 (2017).

12. Pidot, S., Ishida, K., Cyrulies, M., Hertweck, C. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Ang. Chem. Int. Ed.* 53(30), 7856-7859 (2014).

13. Shabuer, G., Ishida, K., Pidot, S., Roth, M., Dahse, H. M., Hertweck, C. Plant pathogenic anaerobic bacteria use aromatic polyketides to access aerobic territory. *Science*. 350(6261), 670-674 (2015).

14. McCoy, E. & McClung, L. Studies on anaerobic bacteria. VI. The nature and systematic position of a new chromogenic *Clostridium. Arch. Mikrobiol.* 6, 230-238 (1935).

15. Poehlein, A., Solano, J. D. M., Flitsch, S. K., Krabben, P., Winzer, K., Reid, S. J., Jones, D. T., Green, E., Minton, N. P., Daniel, R., Dürre, P. Microbial solvent formation revisited by comparative genome analysis. *Biotechnol. Biofuels.* 10(1), 58 (2017).

16. Li, X., Ferreira, D., Ding, Y. Determination of absolute configuration of natural products: Theoretical calculation of electronic circular dichroism as a tool. *Curr. Org. Chem.* 14(16), 1678-1697 (2010).

17. Wen, C. N., Hu, D. B., Bai, X., Wang, F., Li, Z. H., Feng, T. Liu, J. K. Chemical constituents from fruiting bodies of Basidiomycete *Perenniporia subacida. Fitoterapia*. 109, 179-184 (2016).

18. Goh, S. H., Ee, G. C. L., Chuah, C. H., Wei, C. Styrylpyrone derivatives from *Goniothalamus dolichocarpus. Aus. J. Chem.* 48(2), 199-205 (1995).

19. Pluskal, T., Torrens-Spence, M. P., Fallon, T. R., De Abreu, A., Shi, C. H., Weng, J. K. The biosynthetic origin of psychoactive kavalactones in kava. 5(8) 867-878 (2019).

20. Lee, I. K. & Yun, B. S. Styrylpyrone-class compounds from medicinal fungi *Phellinus* and *Inonotus* spp., and their medicinal importance. *J. Antibiotics.* 64(5), 349-359 (2011).

21. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47(W1), W81-W87 (2019).

22. Stachelhaus, T., Mootz, H. D., Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6(8), 493-505 (1999).

23. Muir, R. M., Ibáñez, A. M., Uratsu, S. L., Ingham, Elizabeth S., Leslie, C. A., McGranahan, G. H., Batra, N., Goyal, S., Joseph, J., Jemmis, E. D., Dandekar, A. M. Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). *Plant Mol. Biol.* 75(6), 555-565 (2011).

24. Vieweg, L., Reichau, S., Schobert, R., Leadlay, P. F., Süssmuth, R. D. Recent advances in the field of bioactive tetronates. *Nat. Prod. Rep.* 31(11), 1554-1584 (2014).

25. Sun, Y., Hong, H., Gillies, F., Spencer, J. B., Leadlay, P. F. Glyceryl-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics. *ChemBioChem*. 9(1), 150-156 (2008).

26. Dorrestein, P. C., Van Lanen, S. G., Li, W., Zhao, C., Deng, Z., Shen, B., Kellener, N. L. The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *J. Am. Chem. Soc.* 128(32), 10386-10387 (2006).

27. Kanchanabanca, C., Tao, W., Hong, H., Liu, Y., Hahn, F., Samborskyy, M., Deng, Z., Sun, Y., Leadlay, P. F. Unusual acetylation-elimination in the formation of tetronate antibiotics. *Angew. Chem. Int. Ed.* 52(22), 5785-5788 (2013).

28. Herman, N., Li, J., Bedi, R., Turchi, B., Liu, X., Miller, M., Zhang, W. Development of a high-efficiency transformation method and implementation of rational metabolic engineering for the industrial butanol hyper-producer Clostridium saccharoperbutylacetonicum strain N1-4. *Appl. Environ. Microbiol*. 83(2), e02942-16 (2017).