Small Molecule *in situ* Resin Capture - an Organism Independent Strategy for Natural Product Discovery.

Authors:

Alexander Bogdanov [1], Mariam N. Salib [2], Alexander B. Chase [1,3], Heinz Hammerlindl [4], Mitchell N. Muskat [1], Stefanie Luedtke [5], Anthony J. O'Donoghue [5], Lani F. Wu [4], Steven J. Altschuler [4], Tadeusz F. Molinski [2], Paul R. Jensen [1].

Affiliations:

[1] Scripps Institution of Oceanography University of California, San Diego, La Jolla, CA 92093, USA

[2] Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA

[3] Department of Earth Sciences, Southern Methodist University, Dallas, TX 75205, USA

[4] Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94158, USA

[5] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093

ABSTRACT

Microbial natural products remain a crucial source of lead compounds for drug discovery. Yet, commonly employed discovery techniques are plagued by the rediscovery of known compounds, a narrow taxonomic focus on the relatively few microbes that can be cultured, and laboratory growth conditions that do not support biosynthetic gene expression. Here we exploit a culture independent approach to microbial natural product discovery that we call the Small Molecule in situ Resin Capture (SMIRC) technique. SMIRC represents a new approach to mine natural products from underexplored 'chemical space'. We demonstrate the application of this approach with the discovery of numerous new compounds including cabrillospiral, which possesses a new carbon skeleton with a functional group not previously observed among natural products. In contrast to traditional culture-based methods, this compound first approach can capture structurally complex small molecules across all domains of life in a single deployment while relying on Nature to provide the complex and poorly understood environmental cues needed to elicit biosynthetic gene expression. We target marine habitats given the potential for chemical novelty, yet the SMIRC method is applicable across other environments. We demonstrate that SMIRC recovers novel compounds with sufficient yields for NMR-based structure assignment and employ a variety of approaches including expanded deployments, in situ cultivation, and metagenomics to address compound re-supply. This approach has the potential to provide unprecedented access to new natural product chemotypes with broad implications for drug discovery.

INTRODUCTION

Microbial natural products account for many of today's essential medicines including most antibiotics [1, 2]. These compounds are traditionally discovered using a 'microbe first' approach,

where individual strains are isolated from Nature, cultured in the lab, and screened for biological activity with bioassays used to guide the isolation of active compounds from culture extracts. While once productive, the limitations to this approach are now well documented [3] and include i) the re-isolation of known compounds [4]; ii) the recognition that only a small percentage of bacterial diversity has been cultured [5]; and iii) observations that most biosynthetic gene clusters (BGCs) are silent under laboratory cultivation conditions [6]. Attempts to address these limitations include efforts to improve cultivation efficiency [7], co-cultivation or the addition of elicitors to activate silent BGCs [8, 9], and genome mining [10, 11], where BGCs are activated via genetic manipulation [12] or heterologous expression [13]. While genome mining has shown promise [14], it is time consuming, limited to cultured strains, and targets one BGC at a time with no a priori evidence that the products will be chemically novel or have medicinal relevance [15]. Recent evidence of unique biosynthetic potential in yet-to-be cultured bacteria [16], as well as meta-analyses of genomic [17] and metagenomic [18] data all suggest that our best discovery efforts to date have failed to realize the natural product biosynthetic potential encoded in microbial genomes.

Here we report a new approach for microbial natural product discovery that bypasses the initial need for laboratory cultivation. Rather than relying on cultured strains to drive discoveries, we instead isolate natural products directly from the environment in which they are produced. This culture independent approach, which we call the Small Molecule In Situ Resin Capture (SMIRC) technique, inverts traditional discovery paradigms by targeting compounds without initial consideration for the producing organism. SMIRC builds upon targeted, solid phase methods developed for toxin monitoring [19] and was first employed to detect natural products and their producers in marine sediments [20]. The technique is agnostic to the biological source of the compounds and thus can capture small molecules originating from bacteria to phytoplankton and, as we show, some that were previously reported from plants and invertebrates. As such, it requires no knowledge of the cultivation conditions required to induce biosynthetic gene expression; two major obstacles to current discovery strategies. Instead, our approach relies on the compendium of naturally occurring chemical cues and environmental conditions found in complex microbial communities to trigger natural product biosynthesis. It is designed to capture secreted compounds across a wide range of polarities while avoiding the complex mixtures obtained from direct sample extraction. Here we demonstrate SMIRC's effectiveness with the isolation and characterization of three new natural products from a single deployment site including one with a new carbon skeleton and another with promising biological activity.

RESULTS

SMIRC deployment and compound identification

Having demonstrated that our small molecule in situ resin capture (SMIRC) technique could be used to detect known microbial products in marine sediment [20], we revised our methods to determine if it could be used to isolate sufficient amounts of compound for NMR-based characterization. These changes included increasing the amount of HP-20 resin to 100g/replicate and using Nitex mesh enclosures along with other modifications that reduce chemical contaminants (Methods). With these modifications, we deployed SMIRC for seven days in a well-protected *Zostera marina* seagrass meadow (<2m) in Mission Bay, San Diego (Fig. 1A) and obtained on average 54.2 ±6.1 mg (n=4) of salt-free, dark brown extract per replicate. A microfractionation antibiotic assay (Methods) targeting an outer membrane deficient *E. coli* lptD4123 [21] revealed a large active peak with a UV maximum at 360 nm (Fig. 1B). High-

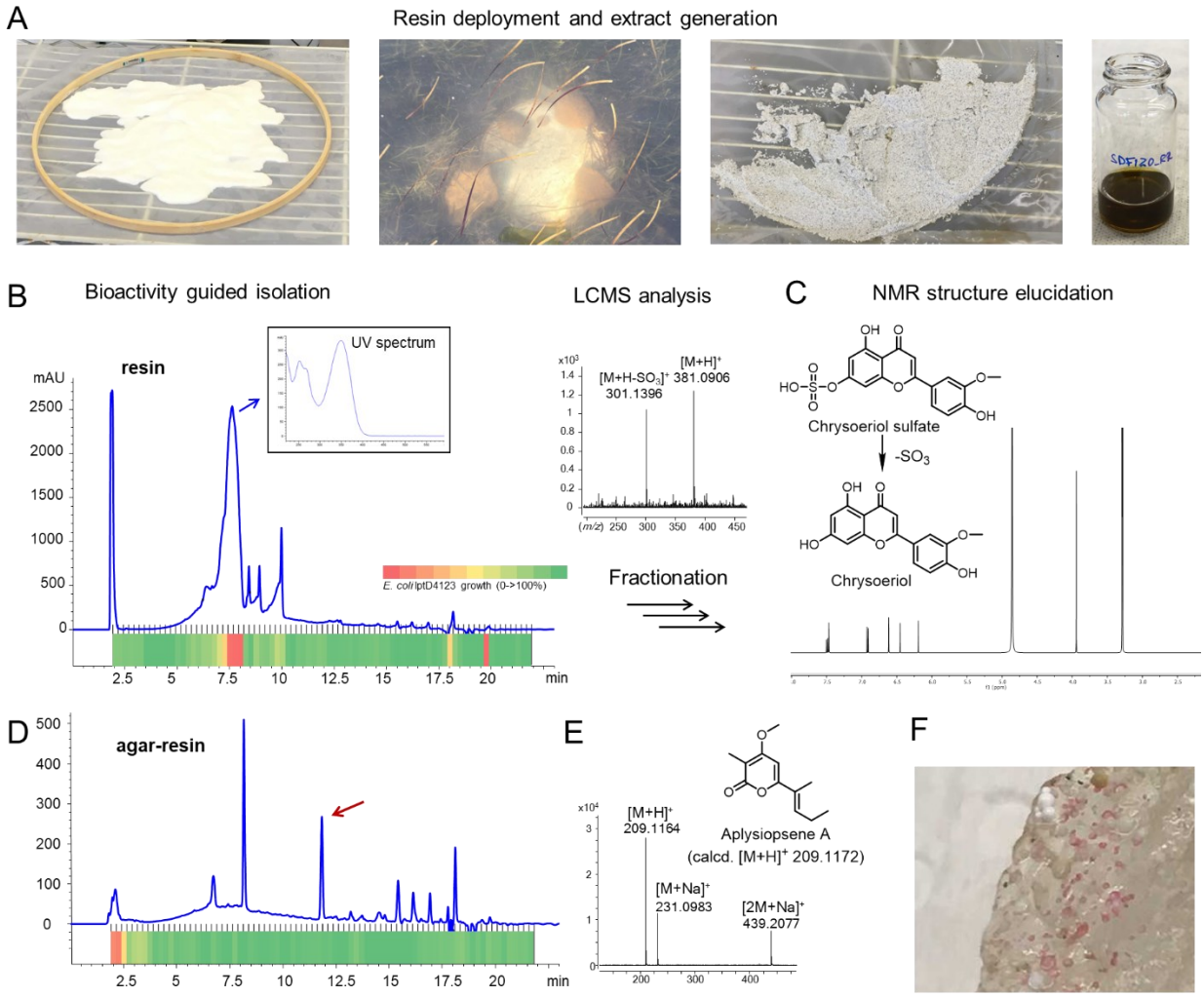

Figure 1. SMIRC deployments in *Zostera marina* meadow (Fiesta Island, Mission Bay, San Diego). A. From left to right: activated HP-20 resin before enclosing in Nitex, deployed resin, recovered resin before extraction, concentrated extract in MeOH. B. Bioactivity guided isolation of a sulfonated flavonoid: UV_{360} chromatogram associated with the microfractionation assay (80 fractions collected over 20 minutes). Heatmap shows E. coli LptD4123 growth inhibition (red = no growth). MS spectrum of the active compound shows loss of a sulfate moiety. C. ^{1}H NMR (500 MHz) and structure of the isolated flavonoid chrysoeriol. D. UV_{360} chromatogram associated with the microfractionation assay of modified (agar addition for in situ cultivation) SMIRC extract obtained a few meters apart at the same site reveal an additional peak (red arrow). E. Structure and MS spectrum of isolated aplysiopsene A, structure elucidated with 500 MHz NMR (^{1}H, COSY and HSQC). F. Pink colonies growing on agar/resin matrix that yielded aplysiopsene A.

resolution mass spectrometry detected compounds in the active fraction with *m/z* 381.0906 and 301.1396. The mass difference of 79.951 Da was attributed to the loss of SO_3 (calcd. 79.9568 Da). Bioassay-guided SPE C_{18} fractionation from one SMIRC replicate extract followed by HPLC purification yielded 2 mg of the major compound (*m/z* 301.1396), which was identified as the flavonoid chrysoeriol based on the comparison of MS and NMR data with published reports (Fig. 1C). Sulfonated flavonoids with antibacterial activity have been reported from *Z. marina* [22], providing a candidate producer for these compounds. While our goal was to capture novel microbial products, this finding raises interesting questions about the role of seagrass metabolites with antibiotic activity in structuring seagrass meadow microbiomes. We determined, by direct seawater extraction and quantitative HPLC analysis (Methods), that the concentration of chrysoeriol and related flavonoids at the deployment site was <1 µg/L, which is considerably lower

than the milligram quantities isolated from 100 g of HP-20 resin. Thus, our SMIRC modifications facilitated the capture and concentration of a natural product at levels that supported NMR-based structure elucidation and were unattainable using direct extraction given the trace amounts present in the environment.

In a follow-up deployment at the same site, we further modified the SMIRC technique by embedding the resin in an agar matrix with the aim of facilitating in situ microbial growth and increasing the yields of microbial-derived compounds. The addition of agar led to a different antibiotic profile (activity in most polar fractions) and to the detection of a major new peak in the UV chromatogram (Fig. 1D). We isolated this compound (0.4 mg) and identified it as aplysiopsene A (Fig. 1E) using NMR and tandem MS/MS (Figs. S48-S51). This α-pyrone was originally reported from the Atlantic Ocean sacoglossan mollusk *Aplysiopsis formosa* [23], while structurally related compounds have been reported from a sea fan derived fungus [24] and the actinobacterium *Nocardiopsis dassonvillei* [25]. Given that no sacoglossans were observed at the deployment site and that microbes produce closely related compounds, it can be suggested that aplysiopsene A represents another example in which microbes are the source of a compound originally reported from a marine invertebrate [26, 27], in this case a sacoglossan. We also observed pink colonies on the agar-resin matrix that yielded an aplysiopsene A enriched extract. However, we have not succeeded in bringing this microorganism into culture. The modified SMIRC technique represents a unique approach to in situ microbial cultivation and new opportunities to explore the microbial origins of natural products. These results provided a second example in which the SMIRC technique yielded sufficient compound for NMR-based structural elucidation and we asked if this technique could be further expanded to discover novel natural products directly from the environment.

We next deployed SMIRC in the Cabrillo State Marine Reserve (CSMR, Fig. 2A), a relatively pristine habitat comprised of sand patches and rocky reef dominated by the seagrass *Phyllospadix* sp. and diverse macroalgae and invertebrates. Deployments (without agar), which spanned 2-8 days depending on environmental conditions (i.e., tides and swells), yielded extract masses averaging 177.1 mg (±87.6 mg, N=17) and, in some cases, exceeded 300 mg per 100 g resin. While no activity was detected in our microfractionation antibiotic screening, the LCMS analysis revealed extensive chemical complexity (Fig 2B), a finding supported by the clusters of congested peaks likely composed of thousands of captured metabolites in the UV_{254} and total ion count (TIC) chromatograms (Fig. S9). To facilitate the interpretation of the mass spectrometry data, we utilized molecular networking associated with the GNPS platform [28]. Known natural compounds and environmental pollutants including lipids, carotenoids, flavonoids, bile acids, and melamine derivatives along with the microbial natural products okadaic acid and apratoxin A, were identified by comparison with over 580,000 reference MS/MS spectra in the GNPS library (Figs. S10-11). The detection of apratoxin A expands the geographic distribution of this compound family, that was initially reported from the tropical cyanobacterium "*Lyngbya majuscula*" [29] and is associated with dermatotoxic reactions [30], to temperate waters. We also identified pectenotoxin-2 by manual curation of the MS/MS spectra (Fig. S12) and comparison with published LCMS data [31]. Although pectenotoxin-2 is represented in the GNPS library, the compound was not annotated (even with relaxed precursor and fragment mass tolerance settings up to 1.0 Da and 0.5 Da respectively) highlighting the importance of manual analyses. We also detected trace amounts of aplysiopsene A in several of the CSMR extracts, further supporting our suggestion of a potential microbial origin. While additional marine natural products including callipeltin B, cyanopeptolin 880, several amphidinolides, kahalalide Y, and kabiramide D were

identified with the DEREPLICATOR+ algorithm [32], all of these annotations were determined spurious after manual analysis of the MS/MS spectra, further emphasizing that automated structure annotations should be interpreted with caution.

Fractionation of the crude extracts and subsequent molecular networking (GNPS) analysis of the LCMS data greatly increased the number of precursors that were selected for MS/MS fragmentation in the mass spectrometer. We compared the GNPS molecular networking outputs from the analyses of the LCMS data of 15 crude SMIRC extracts with the same 15 extracts and 15 SPE fractions from four of those extracts. After rigorous subtraction of controls, 1,231 ions (= nodes in the network) were detected in the analysis that included fractions versus 732 nodes detected in the crude extract analysis (68 % increase of molecules that can be potentially identified or targeted for isolation Fig. S13).

Natural product discovery using SMIRC

The small number of molecules that could be confidently identified in the SMIRC extracts suggests a high degree of chemical novelty and opportunities for natural product discovery. As such, we targeted a compound with a high-intensity molecular peak (*m/z* 328.2485 $[M+H]^+$ Fig. 2B) and calculated molecular formula of $C_{18}H_{33}NO_4$ (calcd $[M+H]^+$ 328.2483, 0.61 ppm) with three degrees of unsaturation that was detected in all replicate extracts from the CSMR deployments. This molecular formula matched only three compounds in the Dictionary of Natural Products (https://dnp.chemnetbase.com): i.e., 3-hydroxy-C_{14}-homoserine lactone and curvularides B and E, initially reported from various bacteria and a terrestrial endophytic fungus [33], respectively. Believing these were not good matches, we subjected two extracts with the highest estimated abundance of the compound to LCMS guided isolation (see Methods) and augmented UV detection with evaporative light-scattering (ELSD) detection given the absence of a strong chromophore. Repeated rounds of HPLC purification yielded ~40 μg of a pure compound as a white solid. An initial ^{1}H NMR spectrum (600 MHz, CD_3OD, 1.7 mm micro cryoprobe) revealed three doublet resonances attributable to three methyl groups (δ_H 0.90, J = 6.5 Hz; 0.96, J = 6.7; 1.20, J = 6.3 Hz) positioned on methine carbons (Fig. 2C), thus indicating that our targeted compound is neither 3-hydroxy-C_{14}-homoserine lactone, which possess only one methyl group, nor one of the curvularides, which have five methyl groups. The complete structure elucidation of this compound (see Supplementary), which we have called cabrillostatin (**1**, Fig. 3), was accomplished with 2D NMR experiments (^{1}H-^{1}H COSY, HSQC, HMBC). Cabrillostatin is a new natural product comprising the rare non-proteinogenic amino acid statine fused with a 9-hydroxydecanoic acid to form a 15-membered macrocycle containing both ester and amide functionalities. The characterization of this compound by NMR validates the application of SMIRC for natural product discovery.

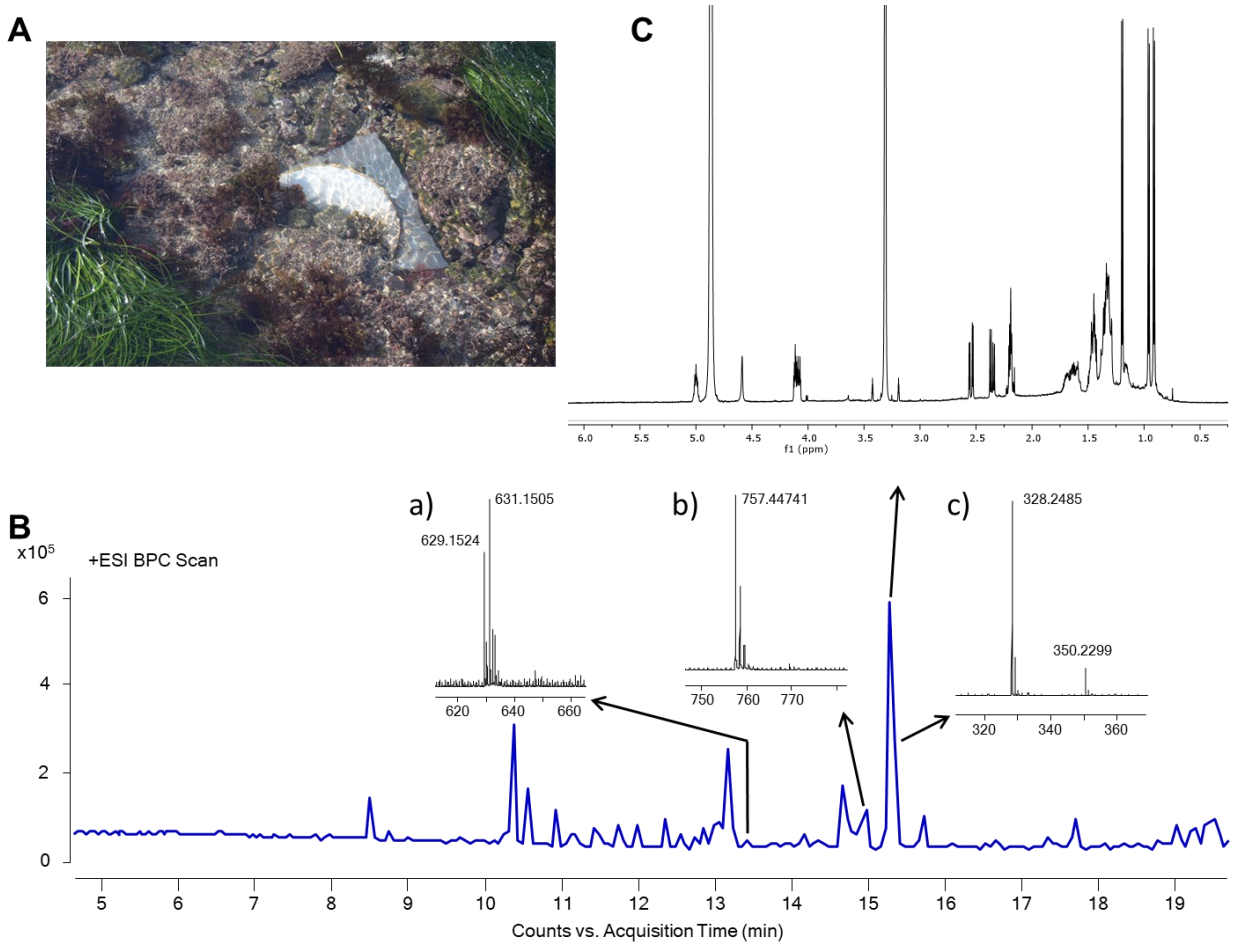

Figure 2. SMIRC deployment at Cabrillo State Marine Reserve (CSMR). A. In situ SMIRC package deployed on rocky substrate. B. ^{1}H NMR spectrum of isolated compound *m/z* 328.2485, cabrillostatin (**1**), recorded with 40 μg (600 MHz, CD_3OD, 1.7 mm cryoprobe). C. BPC chromatogram of a SMIRC extract and mass spectra (a, b, and c) of compounds targeted for isolation and characterization.

In the course of isolating cabrillostatin (**1**), we detected a second compound of interest based on the MS spectrum, which displayed an isotopic pattern indicative of a dihalogenated molecule bearing one Cl and one Br (Fig. 2B). The best-fit molecular formula $C_{29}H_{40}BrClO_9$, with nine degrees of unsaturation, was calculated from the highest intensity isotopic group at *m/z* 629.1524 (calcd for *m/z* $[M\text{-}H_2O+H]^+$ 629.1512, -1.9 ppm). This formula received no matches in database searches (DNP) suggesting it represented a second new natural product. Multiple HPLC purification steps (Methods) yielded 50 μg of pure compound from a 100 mg SMIRC extract. A series of NMR experiments (^{1}H, ^{1}H-^{1}H COSY, HSQC, HMBC, NOESY, 600 MHz, 1.7 mm microcryoprobe, CD_3CN) conducted over long acquisition times (>72 h for HMBC) left several important correlations missing in the HMBC spectrum. Pooling samples from two SMIRC extracts yielded ~100 μg of compound while NMR experiments (^{1}H, COSY and HSQC) repeated in CD_3OD improved peak shapes and resolved signals and *S/N*. An LR-HSQMBC (optimized for J = 8 Hz) experiment revealed additional correlations not observed in CD_3CN [34]. Extensive analysis of the NMR data recorded in both solvents (see Supplementary) allowed us to establish the structure as a new bromo-chloro-polyketide that we have called cabrillospiral A (**2**) (Fig. 3). This compound exhibits rare features including an α-halogenated 6,5-spiroketal bicycle and an α-β unsaturated

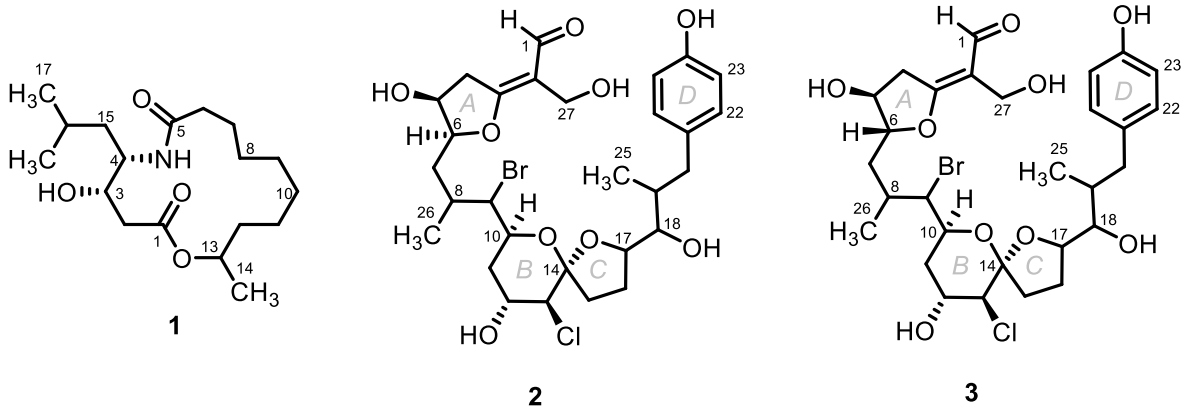

Figure 3. Structures of cabrillostatin (**1**) and cabrillospirals A (**2**) and B (**3**).

aldehyde fused within a furan ring. The latter, a vinylogous formate ester, is unprecedented among both the natural product and synthetic literature (SciFinder). Cabrillospiral A (**2**) has 11 chiral centers distributed across four independent stereogenic substructures making stereochemical elucidation a challenge. Interpretation of the NOESY data and *J-J* coupling constants provided key insights into the stereochemistry of **2** and led to the assignment of the $\Delta^{2,3}$ double bond geometry and of the relative stereochemistry six chiral carbons (see Supplementary). The full assignment of the relative and absolute stereochemistry is ongoing. Cabrillospiral A (**2**) eluted with two minor compounds (Fig. S4B) that shared the same chromophore and were indistinguishable from **2** by MS/MS. After repeated rounds of HPLC purification from several extracts, we obtained one analog in sufficient amount (27 µg) to record a ^{1}H NMR, HSQC, and COSY experiments and elucidated the structure as cabrillospiral B (**3**), which we propose is the C-6 epimer of **2**.

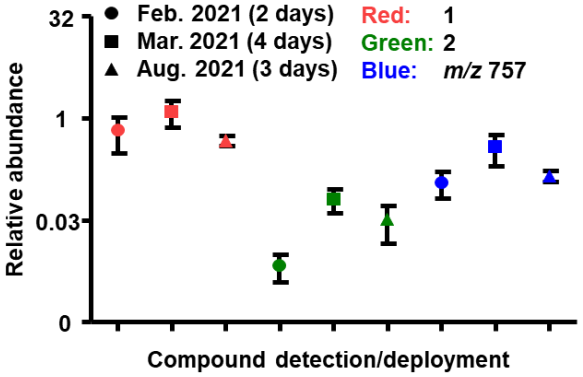

Figure 4. Reproducibility of three target compounds at the CSMR site. In all cases, compound yields were highest for the longest deployments.

In addition to compounds **1**-**3**, we targeted and isolated several other potentially new natural products from the CSMR deployments (Table 1 and supplementary Figs. S19-S29). In total, from a single deployment site, we have identified compounds belonging to two new compound classes (**1** and **2**-**3**) and at least 11 others that are likely new. Our ability to consistently detect target compounds **1**, **2**, and *m/z* 757.4474 from three deployments spanning seven months suggests that subsequent deployments represent a method to obtain additional quantities of lead compounds. Evidence that the greatest yields were obtained from the longest deployments (Fig. 4) justifies additional time course studies.

Table 1. Additional isolated compounds from CSMR deployment site.

m/z	Molecular formula	Purity	Known natural products with the same mass or formula
757.4474 $[M+H]^+$	$C_{43}H_{64}O_{11}$	semi pure	gambierol, amphidinolide C2, glycosylated steroids
456.1317 $[M+Na]^+$	$C_{20}H_{29}Cl_2NO_5$	4 isomers in a HPLC fraction	none
1211.6231 $[M+Na]^+$	$C_{63}H_{96}O_{21}$	HPLC fraction	altohyrtin C, spongistatin 2, steroids
1086.6190 $[M+NH_4]^+$	$C_{52}H_{80}N_{10}O_{14}$	semi pure	none
836.4025 $[M+H]^+$	$C_{42}H_{72}BrCl_2NO_6$	HPLC fraction	none
237.1492 $[M+H]^+$	$C_{14}H_{20}O_3$	HPLC fraction	>30
488.257 $[M+H]^+$	$C_{23}H_{34}ClNO_8$	HPLC fraction	none
606.3572 $[M+H]^+$	$C_{34}H_{52}ClNO_6$	HPLC fraction	none
644.3936 $[M+H]^+$	$C_{34}H_{58}ClNO_8$	semi pure	none
730.2286 $[M+Na]^+$	$C_{33}H_{48}Cl_3NO_9$	semi pure	none
913.4561 $[M+Na]^+$	$C_{47}H_{70}O_{16}$	semi pure	glycosylated steroids

*Bioactivity of cabrillostatin (**1**), cabrillospiral A (**2**) and B (**3**).*

Given that the amino acid statine is a known pharmacophore of protease inhibitors [35], cabrillostatin (**1**) was tested for activity against the aspartic protease cathepsin D. Surprisingly, **1** potentiated the activity of this protease in a concentration dependent manner from 1.5-fold at 100 nM to almost 2-fold at 10 µM (Fig. S14). Testing of **1** in an expanded panel of proteases did not reveal additional activities. Multiple SMIRC deployments at the CSMR site allowed us to isolate sufficient amounts of cabrillostatin (additional 26 µg) and cabrillospirals A (75 µg) and B (27 µg) for testing in a panel of nine diverse cancer cell lines (Supplementary Table 1) using cell painting [36] and high content imaging. We used our customized image analysis pipeline to generate phenotypic profiles that represent an in-depth quantification of cell morphology and biomarker staining in comparison to 28 reference compounds across six diverse drug categories with known distinguishable phenotypes [37] (Fig. S15A, Supplementary Table 2). Principal component analysis (PCA) clearly separated cabrillostatin from the DMSO controls in most cell lines (Fig. 5A). Quantifying the separation showed that cabrillostatin was significantly bioactive ($p<10^{-6}$ at 10 µM) in six of the nine cell lines (Fig. 5B, Supplementary Table 3). The phenotypic profiles of cabrillostatin relative to the reference compounds were unique (confidence score >0.1 [37], Supplementary Table 4) indicating a different mechanism of action. We next tested a decreasing dose series (**1**-**3** at 10 µM, 1 µM, 0.33 µM) in A549 cells. While this independent dose-response experiment validated the bioactivity of the 10 µM concentration, we could not detect bioactivity of the lower doses (Fig. S15B). In contrast to **1**, cabrillospirals **2**-**3** induced no significant phenotypic changes that can be detected using cell painting and were deemed non-bioactive in this assay.

We additionally tested cabrillostatin (**1**) and cabrillospiral A (**2**) in induced pluripotent stem cell derived cardiomyocytes (IPSC-CM), which allow the quantification of Ca^{2+} transients that correspond to cardiomyocyte beating [38]. The IPSC-CM model can sensitively detect changes of cardiomyocyte beating induced by anti-arrhythmic or anti-cancer drugs (Fig. S16A and B, Supplementary Table 5) which could unveil effects of the natural products that influence Ca^{2+} handling in IPSC-CMs. In concordance with the bioactivity detected in cancer cells using phenotypic profiling, **1**, but not **2** and **3**, showed a clear alteration of IPSC-CM beating (Fig. 5C, Fig. S16C). Cabrillostatin increased the peak amplitude with a small decrease of peak frequency (Fig. 5D), which was validated in an independent batch of IPSC-derived cardiomyocytes (Fig. S16D). The observed changes in the IPSC-CM Ca^{2+} transients suggest a potential positive

inotropic effect of cabrillostatin and warrants further investigations. Taken together, our SMIRC approach allowed us to isolate novel natural products at quantities sufficient for testing in mammalian cell models.

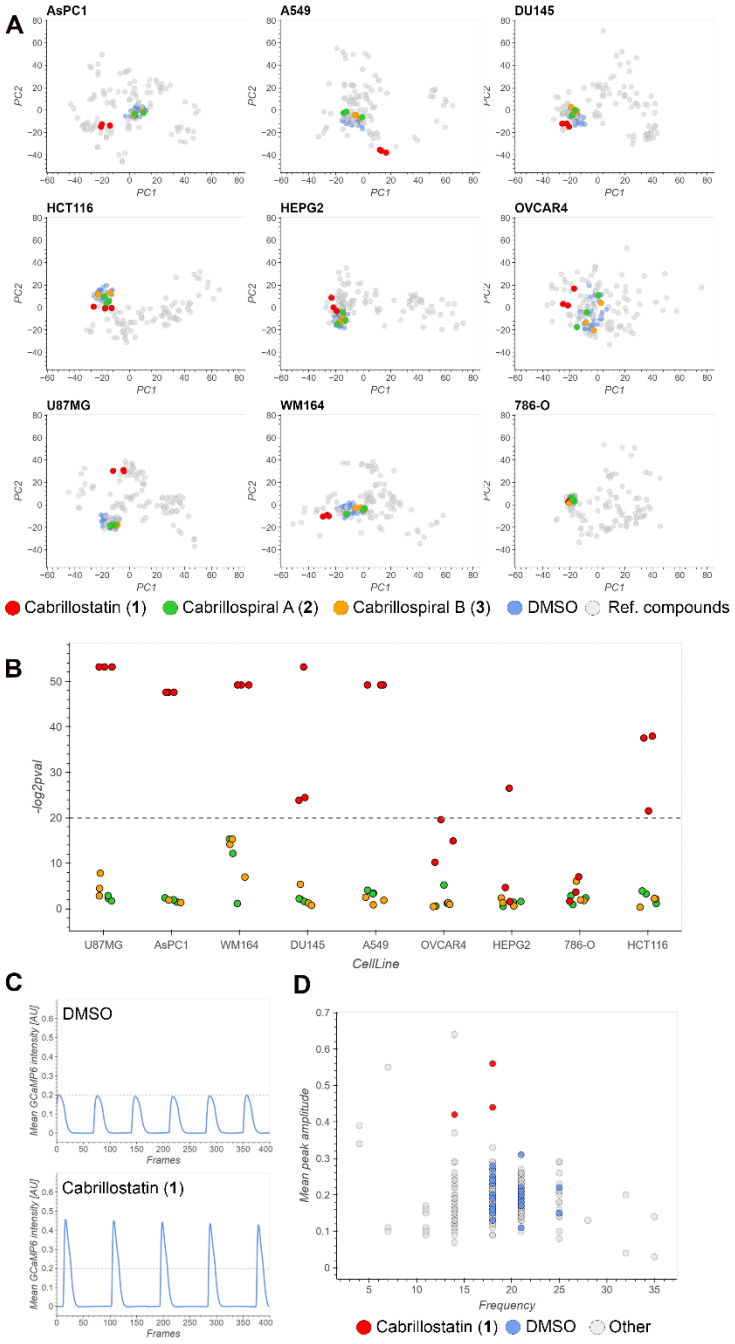

Figure 5. Bioactivity results for cabrillostatin (**1**), and cabrillospirals A (**2**) and B (**3**). A. Principal component analysis of high dimensional phenotypic profiles of **1**-**3** (red, green, yellow, respectively), reference compounds (grey), and DMSO (blue) in nine diverse cell lines. B. Bioactivity expressed as significance (-log2pval) across all cell lines. C. Representative Ca^{2+} transient traces of DMSO (top) and **1** (bottom). D. Scatter plot showing beating frequency and mean peak amplitude of **1** (red), DMSO (blue) and anti-arrhythmic and anti-cancer drugs (grey).

*Environmental distributions of compounds **1** and **2**:*

The GNPS platform was recently expanded to include MASST (Mass Spectrometry Search Tool) [39], a web-enabled search engine that is the equivalent of NCBI BLAST for MS/MS datasets in the MassIVE repository. Using MASST, we parsed ~2,700 untargeted, small-molecule datasets with the aim of identifying the environmental distributions and potential biotic origins of cabrillostatin (**1**), cabrillospiral (**2**), aplysiopsene A, and 11 other yet to be characterized compounds (Tables S5-S6) identified in the SMIRC extracts. We confidently identified **1** in 11 datasets (collected from 2019-2021) that were largely associated with the study of dissolved organic matter in nearshore (Imperial Beach, Scripps Pier CA) and deep ocean (California Current Ecosystem) seawater samples. Interestingly, a match from the metabolomic analysis of a dinoflagellate culture (*Ostreopsis* sp.) led us to suspect this organism was a potential source of cabrillostatin, however it was also detected in controls from the growth medium that were prepared using natural seawater [40], thus weakening any potential connection between source and producer. We extracted 1 L of seawater from the CSMR deployment site with 20 g HP-20 resin and confirmed the presence of cabrillostatin (**1**) with LCMS analysis (Fig. S17). Cabrillostatin appears to be widely distributed in Pacific Ocean seawater and is likely produced by a planktonic organism. Resolving the structure of **1** adds to

our understanding of the composition of dissolved organic matter (DOM) [41, 42] and raises questions about the ecological roles of what appears to be a widespread and biologically active natural product. Interestingly, cabrillospirals were not detected in any of the MassIVE datasets, suggesting that some compounds are site specific and justifying broader SMIRC deployments to gauge the extent of chemical diversity that can be captured.

Metagenomic analyses:

To link compounds identified from the CSMR site to the producing organisms, we generated untargeted shotgun metagenomes from marine sediment and bulk water samples (average 76.4±12M paired end reads per sample). Both the sediment and seawater communities were dominated by Proteobacteria and Bacteroidetes, with higher relative abundances of Actinobacteria and Planctomycetes in the sediment samples (Fig S18A). We recovered 1,843 BGCs (N50=18 kbp) and many of the 'larger" BGCs (52%, >10 genes) were predicted to be complete (i.e., not on a contig edge). BGCs were largely represented by nonribosomal peptide synthetases (NRPS; 28% of all BGCs), terpenes (25.7%), and ribosomally synthesized and post-translationally modified peptides (RiPPs; 17.7%). Notably, these BGCs could be clustered into 631 gene cluster families (GCFs; Figure S18B), of which only four (those encoding flexirubin, anabaenopeptin, ectoine, and carotenoids like zeaxanthin) shared similarity to BGCs in the MiBIG database [43], emphasizing both the biosynthetic complexity and potential novelty of these environments.

Based on the structure of cabrillostatin (**1**), we expected a PKS-NRPS hybrid BGC that would include a leucine starter unit for the NRPS, a hybrid KS domain to elongate the statine amino acid moiety followed by a highly reducing KS to account for the saturated acyl component. However, among the 43 PKS-NRPS hybrid BGCs recovered across all samples, none encoded the predicted enzymes. Interestingly, highly-reducing PKSs domains are traditionally of fungal origin [44], while the amino acid statine is primarily associated with bacteria [45]. Given the possibility that a low abundant organism (thus, not assembled) could account for production, we extracted all KS domains from the unassembled reads along with experimentally characterized fungal highly reducing [46-48] and statine-related KS domains (i.e., didemnin B and burkholdac A) from published literature [27, 49], and characterized them with NaPDoS2 [50]. Unfortunately, none clustered with these reference domains in a KS phylogeny, leaving the biosynthetic producer of **1** unresolved and demonstrating the limitations of metagenomics when applied to complex communities.

In the case of cabrillospiral A (**2**), biosynthetic hooks include a 10+ module type I PKS (T1PKS), halogenases, and methyltransferases to account for the C-25 and C-26 methyl, and C-27 hydroxymethyl groups. However, all of the 68 T1PKS BGCs detected are likely incomplete due to the assembly challenges associated with their repetitive, modular organization [51]. Yet, our analyses yielded two candidates (Fig. S18C, both of which binned into a single high-quality planctomycete MAG (4.96 Mbp; 98.2% complete, 3.6% contamination). Planctomycetes represent a promising and largely unexplored source of specialized metabolites [52]. In an attempt to resolve the T1PKS modules, we searched the unassembled data and found >10 highly similar KS domains (96-98% amino acid similarity) that may comprise the unassembled T1PKS BGC associated with the biosynthesis of **2**. Unfortunately, long-read Nanopore sequencing on the samples enriched in this MAG failed to cover this BGC, again likely due to the complexity of marine sediment communities. While inconclusive, the planctomycete MAG remains our top candidate for the production of compound **2**.

DISCUSSION/CONCLUSION

Explorations of biotic diversity have proven key to the discovery of biologically active natural products including many of today's most beneficial medicines. The discovery process has invariably started with the collection or cultivation of an organism followed by extraction and either chemical or bioassay-guided compound isolation. While these approaches proved effective for decades, it has become increasingly difficult to find structurally unique compounds that can be incorporated into screening programs. In the case of microbial natural products, the realization that genome sequences contain significant, unrealized biosynthetic potential has launched the field of genome mining and driven extensive efforts to activate silent biosynthetic gene clusters through co-cultivation, elicitor addition, or genetic manipulation [13]. Despite these efforts, current approaches to natural product discovery, including genome mining, have failed to capture the vast extent of chemical diversity predicted to exist in nature.

Here we report the Small Molecule In Situ Resin Capture (SMIRC) technique, a compound first approach to natural product discovery in which adsorbent resins are used to capture compounds directly from nature. In this study, we elucidated the structures of three novel natural products – a hybrid NRPS-PKS derived cabrillostatin (**1**) and the halogenated polyketides cabrillospirals A and B (**2**-**3**) from a single deployment site (CSMR). The latter possesses a new carbon skeleton and a functional group (vinylogous formate ester) that is unprecedented in the natural product or synthetic literature. The low Tanimoto similarity coefficient (0.47) calculated for cabrillospiral A demonstrates its uniqueness among natural products. Cabrillostatin (**1**) has intriguing effects on cardiomyocyte beating frequency and amplitude suggesting positive inotropic activity. In cancer cell lines **1** induces phenotypic changes that are unique among the reference compounds and indicates a different mode of action. Furthermore, cabrillostatin (**1**) seems to be widespread in the Pacific Ocean water and might represent an intriguing molecule for ecological studies. While our efforts to link these compounds to their biological origins remain ongoing, we speculate they are microbial products and have presented an in situ cultivation method that changed the profile of the captured metabolome and led to the isolation of aplysiopsene A, a compound previously reported from a marine sacoglossan [23]. In addition to these compounds, we isolated several other potentially new compounds from the same deployment site.

While further refinements to the SMIRC technique are needed, we show that compounds can be recovered in sufficient quantities for NMR-based structure elucidation. In some cases, structures could be solved using conventional 5 mm room temperature probes. We further show that some compounds can be detected from the same site over time, suggesting that scale-up deployments represent a potential mechanism to address expected supply issues. Alternatively, compounds such as cabrillostatin represent feasible synthetic targets while others will require investment from the broader synthetic community. Metagenomic and cultivation approaches provide additional opportunities to address low compound yields. Ultimately, supply issues will need to be addressed on a compound-by-compound basis and provide opportunities to further advance the approaches used to exploit chemical diversity.

In summary, SMIRC provides a unique approach to access chemical space that has proven otherwise inaccessible using current methods. It is agnostic to the producing organism and benefits from the complex and poorly understood environmental cues required to stimulate compound production, thus circumventing many of the current bottlenecks associated with genome mining and microbial cultivation. By incorporating up front bioassays, this approach can be used to directly target bioactive compounds without the need to build large culture collections.

SMIRC is relatively simple to assemble and deploy and requires analytical instrumentation that is readily available at most universities. Moving forward, this approach presents opportunities to gain new insights into the ecological roles of natural products and their applications for drug discovery.

MATERIALS AND METHODS

General Experimental Procedures.

1D and 2D NMR spectra were measured at 23 °C on a JEOL ECZ spectrometer (500 MHz), equipped with a 5 mm $^{1}H\{^{13}C\}$ room temperature probe, or on a Bruker Avance III (600 MHz) NMR spectrometer with a 1.7 mm $^{1}H\{^{13}C/^{15}N\}$ microcryoprobe. The NMR spectra were referenced to the solvent signals (CHD_2OD, δ_H 3.31, CD_3OD, δ_C 49.00 ppm; CHD_2CN, δ_H 1.94, CD_3CN, δ_C 1.74 ppm). LC-HRMS was performed on an Agilent 6530 Accurate-Mass QToF with ESI-source. The mass spectrometer was coupled with an Agilent 1260 Infinity HPLC equipped with a degasser, binary pump, autosampler, DAD detector and a 150x4.6 mm Kinetex C_{18} 5 µm column (Phenomenex, USA). The instrument was calibrated using the Agilent Reference Calibration Mix. Preparative, semipreparative, and analytical HPLC were completed on a system consisting of an Agilent 1100 G1312A binary pump, Agilent 1100 G1315A DAD UV/Vis detector, Agilent 1100 G1313A autosampler and an Agilent 1100 G1322A degasser, coupled to a Shimadzu low temperature evaporative light scattering detector (ELSD) through a Hewlett Packard 35900E interface. ECD spectra were measured on a JASCO J-810 spectropolarimeter at 23 °C in quartz cells of 5 mm path length.

HP-20 resin preparation, deployment, extraction.

The HP-20 resin was activated by gently stirring in methanol for one hour on a shaker at 40 rpm. The resin was then filtered through a Büchner funnel (200 µm pores) under reduced pressure and washed three times with methanol (HPLC grade) to remove contaminants. Subsequently, the resin was washed three times with high purity (HPLC or MIlliQ-grade) water. The activated resin could be stored in high purity water at room temperature (20˚C) for several weeks. Before deployment, 100 g activated resin was packaged between two layers of Nitex nylon mesh (120 µm, Genesee Scientific) within a 30 cm wooden embroidery hoop (Fig. 1A).

Packaged resin was secured in the field using aluminum tent stakes or rocks. Upon recovery, SMIRC replicates were processed individually. The package surface was carefully washed in the laboratory with deionized water to displace sea water and remove debris from the outside before transferring the resin into a Büchner funnel (200 µm). There, the resin was washed once with high purity water (MilliQ or HPLC grade) and eluted three times with 200 mL HPLC grade methanol under reduced pressure. The solvent was evaporated on a rotary evaporator to yield salt-free crude extracts. Contamination was minimized by removing the resin from the Nitex prior to extraction, using HPLC-grade solvents, avoiding contact with plastic ware, washing glassware with conc. HCl and pre-rinsing immediately before use with high-purity solvents. Background contaminants included caprolactam (from Nitex mesh), phthalate plasticizers and widely used poly(ethyleneglycol) (PEG) of variable chain lengths. (MW ~500-1000 Da).

LCMS analyses.

Aliquots (5 µL) of the methanolic extracts (1 mg/mL) were injected onto the LC-HRMS and eluted (1 mL/min) using an isocratic mobile phase (20:80 acetonitrile (ACN):H_2O and 0.1 % formic acid

(FA) for 2 min followed by a gradient elution to 5:95 ACN:H_2O over 18 min, isocratic for 2 min, then increasing to 100% ACN over 1 min, and finishing with 100% ACN for 2 min. MS data were acquired over the range 135-1700 *m/z* in positive mode. All solvents were of LCMS grade. The data was manually analyzed using the Masshunter Qualitative Analysis Software B.05.00. Molecular networking and automated library searches were performed using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu) [53] and were visualized using Cytoscape 3.8.2 [54] and the GNPS in-browser network visualizer. Single spectra of interest were queried in MASST associated with the GNPS platform [39] and publicly available metabolomic datasets were searched with default parameters (https://masst.ucsd.edu/).

Isolation of chrysoeriol and aplysiopsene A

Crude extract generated at the Fiesta Island deployment site were fractionated using C_{18} (2 g) and a stepped-gradient of 20 mL of MeOH:H_2O (25:75, 50:50, 75:25, 90:10 and 100:0 MeOH:H_2O) and collected as fractions SPE1-5, respectively. Bioactivity guided, the active fraction SPE2 was submitted to HPLC separation equipped with a 10x250 mm C_8 column, isocratic MeOH:H_2O (69:31) mobile phase with 0.1% formic acid (FA) and 2.5 mL/min flow to yield 2 mg of chrysoeriol. Crude agar-resin extract was fractionated following the same procedure. SPE fractions were subsequently analyzed on UV HPLC and the targeted compound was detected in SPE3. This fraction was submitted to further separation on HPLC equipped with 4.6x150 mm C_{18} column, isocratic ACN:H_2O (50:50) mobile phase with 0.1% (FA) and 1 mL/min flow to yield 0.4 mg of aplysiopsene A.

*Isolation of cabrillostatin (**1**) and cabrillospirals A-B (**2**-**3**).*

Crude resin extracts generated at CSMR deployment site were fractionated in the same way as described above. SPE fractions were subsequently analyzed using LCMS and the targeted compounds were identified in fraction SPE3. The latter was submitted to further separation on HPLC (4.6x150 mm C_{18} column, isocratic ACN:H_2O (35:65) mobile phase with 0.05% FA, 1 mL/min flow). Target compounds were quantified by solvent ^{13}C satellites (QSCS) [55].

Quantification of flavonoids in seawater.

Seawater was collected from SMIRC deployment site at Fiesta Island, Mission Bay, San Diego just above the seagrass meadow. 1 L was extracted using Waters Oasis HLB SPE column and eluted with 3 mL of MeOH. The extract was analyzed on an HPLC with UV detection at 360 nm using chromatography conditions described above for antibiotic screening (microfractionation). Methanolic solution of previously isolated chrysoeriol (1 mg/mL) was used as a standard. The sea water extraction efficiency was determined with adding 50 µg of chrysoeriol to 1 L seawater, subsequent extraction and HPLC analysis.

Antibiotic screening.

Crude extracts were screened for antibiotic activity following HPLC microfractionation. In brief, 1 mg (*e.g.,* 33 µL at 30 mg/mL) of crude extract was injected into an analytical HPLC (reversed-phase 150x4.6 mm Kinetex C_{18} 5 µm column, flowrate 0.8 mL/min) using a gradient from 10-100% ACN (with 0.1 % FA) over 16 min then isocratic for 4 min. Fractions (200 µL) were collected every 15 seconds starting at 1.8 minutes (system dead time) directly into a flat bottom 96-well micro-titer plates using a Gilson FC204 fraction collector yielding a total of 80 fractions. The

fraction plates were dried under vacuum for 4 hours and the content of each well dissolved in 10 µL of DMSO and 190 µL of LB medium inoculated with *E. coli LptD*4213 [21] strain (outer membrane defect that facilitates antibiotic entry [56], grew overnight in LB) added to each well. Positive (10 µm chloramphenicol) and negative (10 µmol streptomycin and solvent/media) controls were tested for each plate. The plates were incubated for 16 hours on a shaker (120 rpm, 30°C) and turbidity (OD_{650}) measured before and after incubation on a Molecular Devices Emax Precision Plate Reader. Growths inhibition was calculated as the reduction in OD relative to solvent/media controls.

Cell lines

A549 (lung carcinoma), OVCAR4 (ovarian carcinoma), DU145 (prostatic carcinoma), 786-O (renal cell carcinoma) and AsPC1 (pancreas adenocarcinoma) cells were purchased from ATCC. HEPG2 (hepatocellular carcinoma), U87MG (glioblastoma) and HCT116 (colorectal carcinoma) were purchased from the UCSF cell line core facility (UCSF CCF). WM164 was a gift by Dr. Helmut Schaider. A549, OVCAR4, 786-O, WM164 and AsPC1 were maintained in RPMI 1640 (Life Technologies). DU145, HEPG2, U87MG and HCT116 were maintained in DMEM with 4.5 g/dL glucose (Life Technologies). Both media were supplemented with 10% FBS (Gemini) and 1% penicillin/streptomycin (Thermo Fisher Scientific). Cell lines were maintained in culture in a humidified 37 ˚C incubator with 5% CO2 for a maximum of 6 weeks. Frozen cell lines were stored in liquid nitrogen with 10% DMSO. All cell lines are routinely tested for mycoplasma as described previously [57, 58].

Compound library

All compounds were purchased from Selleck and/or MedChemExpress and were tested in 2 doses (1:1000 and 1:10000 dilutions). The reference compounds showed a high classification accuracy (mean ~94% with each cell lines >90%), confirming high quality of phenotypic profiling for each cell line (Supplementary Table 1). Compounds and stock concentrations are summarized in Supplementary Table 2 and 5.

Drug addition

Cells were seeding into 384 well Cell Carrier Ultra plates (Perkin Elmer) at an empirically determined density in 75 µL of the respective media per well. Compounds were added directly to the well using the ECHO 650 liquid handling system (Beckman Coulter) integrated into a Perkin Elmer EXPLORER G3 WORKSTATION. Transfers and plate handling was set up using the PerkinElmer's plate::works™ (v6.2) software.

Cell painting

Staining with cell painting dyes was adjusted from Bray et al. (2016) [36]. Briefly, 10 µL of a 8x Mitotracker deep red (Life Technologies) master mix (in DMEM at a concentration of 5 µM) was added to each well, and incubated for 30 min. Cells were then fixed by adding 30 µL 16% paraformaldehyde (PFA) (final concentration of 4% PFA) for 30 min at RT. Plates were washed once with 1x HBSS (Life Technologies) and permeabilized with 1x HBSS+ 0.5% (vol/vol) Triton X-100 solution for 30 min. Next, cells were washed twice with 1x HBSS and the 30µl of a staining master mix in 1x HBSS +1% (wt/vol) BSA solution (Table S1) was added and incubated in the dark at RT for 30 min. All dyes were purchased from Life Technologies. Finally, cells were washed three times with 1x HBSS, covered with aluminum foil seals, and light-protected until imaging. All

pipetting steps were automated using the Perkin Elmer EXPLORER G3 WORKSTATION and performed by MultiFloFX and 405 TS washer (both BioTek).

High-content imaging and data analysis.

Imaging was performed on the Operetta CLS confocal spinning-disk high content analysis system (Perkin-Elmer) using a 20x water-immersion objective. Each well was imaged at 5 fields of view, in 5 channels (filters according to wavelength information in Table M1) in 3 z-planes. Cell segmentation and single cell feature extraction was performed using the Harmony™ software (Perkin-Elmer) from maximum projections of each image. In total ~1100 features were calculated for each cell including intensity, morphology, and texture features. Phenotypic profiles were calculated using the Kolmogorov-Smirnov (KS) statistic as previously described [37]. "Bioactivity", reference compound classification and novel compound class prediction were calculated on a well level based on the phenotypic profiles as previously described [37]. All computations were performed using Python (v3.9).

Table M1: cell painting biomarker set

Name	**Stock concentration**	**Final concentration [µl/ml]**	**Excitation wavelength [nm]**
Phalloidin	2 units	0.5	568
Concanavalin A	5mg/ml	20	488
SYTO 14 green	5 mM	0.6	490-515
Hoechst	10mg/ml	0.5	385
Wheat-germ agglutinin	1mg/ml	1.5	555
Mitotracker deep red	5uM	0.5uM	647

IPSC derived cardiomyocyte differentiation

IPSC derived cardiomyocytes are generated from human IPSCs (WTC background genetically modified to inducibly express dCas9-KRAB fusion protein and constitutively express GCaMP6 purchased from the Gladstone institute stem cell core facility) using previously published differentiation, expansion, and maturation protocols with modifications [38, 59, 60]. Briefly, hiPSCs are dissociated into a single cell suspension using TrypLE™ Express and seeded into Geltrex™ LDEV-Free Reduced Growth Factor Basement Membrane Matrix-coated 12 well-plates at 75k cells per well in E8 media supplemented with 10 µM Y-27632. Cells are fed daily with E8 media and differentiation was initiated once the hiPSCs reached 80-90% confluence (~48h after

seeding) by addition of RPMI-1640 supplemented with B27 without insulin and 7.5 μM CHIR99021. Exactly 48h later (±15 minutes), media is replaced with RPMI-1640 supplemented with B27 without insulin and 7.5 μM IWP2. After an additional 48h (±15 minutes), media is replaced with RPMI-1640 supplemented with B27-without insulin. Two days later the media is changed to RPMI-1640 supplemented with B27 containing insulin and 1% P/S that is replenished every 2-3 days. Spontaneous beating is generally observed 8-10 days after addition of CHIR99021. iCMs from wells that show spontaneous beating in >80% of the surface area are then exposed to lactate enrichment medium and are re-platted into 384 well Cell Carrier Ultra plates (Perkin Elmer) after 5 days of metabolic selection. Finally, cells are switched to maturation medium for 10 days before initiating the compound screen.

IPSC-derived cardiomyocyte Ca^{2+} transient imaging and analysis

IPSC derived cardiomyocyte Ca^{2+} transients were monitored using a EVOS M7000 Imaging System (Thermo Fisher Scientific) for 20 seconds at a frame rate of 30hz in the GFP channel (Excitation max 488 nm). Mean fluorescence intensities for each field of view were extracted using a custom image analysis pipeline in python. The signal was baseline corrected using the Python package BaselineRemoval (v0.1.3) and the mean signal intensity of the last 400 frames were plotted for each compound.

Metagenomic analyses.

Sediment (February and March 2021) and bulk water samples (August 2021) were collected in parallel to the SMIRC deployments. For the sediment samples (N=7), ~80g of sediment was collected in sterile WhirlPak bags and transferred on ice back to Scripps Institution of Oceanography for storage. For bulk water samples (N=3), 2L of seawater was collected and immediately filtered (0.2 μm Sterivex). All samples were stored (-20°C) until further processing. DNA was extracted using a phenol-chloroform protocol [61] and assessed using a Nanodrop for quality control and concentration estimates. When needed, low concentration extractions from the same sediment/water samples were pooled prior to sequencing. Libraries for all samples were prepared using the Nextera XT DNA library preparation kit (Illumina) for sequencing on an Illumina NovaSeq 6000 system with 150-bp paired end reads.

Raw reads were quality trimmed and adapters removed using the BBMap package [62]. For taxonomic and biosynthetic analyses, samples were processed using a read-based approach to limit biases introduced from assembled [63]. Briefly, single-copy phylogenetic marker genes were extracted with hidden Markov models (HMMs) [64] from each sample (62.6±13k amino acid merged reads) and classified against a reference genomic database using phylogenetic inferences with pplacer (v1.1.alpha17) [65]. To identify biosynthetic domains, translated reads were searched against the NaPDoS2 [50] database for ketosynthase (KS) and condensation (C) domains.

For the assembly-based approach, filtered reads were normalized with BBNorm (target=40 mindepth=5) and assembled with the IDBA-UD assembler using the pre-correction flag (mink=30 maxk=200 step=10) [66]. Assembled contigs (>2000bp) were searched for predicted biosynthetic gene clusters (BGCs) with antiSMASH (v5.1.2; both fungal and bacterial versions) [67] and were clustered into gene cluster families (GCFs) based on BGC relatedness with BiG-SCAPE (v1.1.2) [68]. Finally, all contigs were binned into metagenome assembled genomes (MAGs) by first calculating coverage profiles with bowtie2 [69] followed by an unsupervised binning approach with Metabat2 across individual samples [70]. All MAGs were checked for quality (i.e., completeness

and contamination) with checkM [71], dereplicated with dRep [72], and assigned taxonomy with GTDB-Tk [73].

ACKNOWLEDGMENTS

We thank Linh Anh Cat and Lauren Pandori (Cabrillo State Marine Reserve) for their assistance with field work and access to the restricted area of the park, Brendan M. Duggan (Skaggs School of Pharmacy, UC San Diego) for access to NMR equipment and expert guidance, Douglas A. Sweeney, Alyssa M. Demko and Leesa J. Klau for their contributions to the development of SMIRC. This research was supported by NIH grants R01GM085770 to P.R.J. and R21AT010493 to P.R.J. and T.F.M.

1. Berdy, J., *Bioactive microbial metabolites. A personal view.* J Antibiot, 2005. **58**(4): p. 1-26.
2. *World Health Organization Model List of Essential Medicines – 22nd List*. 2021, World Health Organization: Geneva. p. 1-66.
3. Baltz, R.H., *Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?* J Ind Microbiol Biotechnol, 2006. **33**(7): p. 507-13.
4. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery.* Nat. Rev. Drug Discov., 2005. **4**(3): p. 206-220.
5. Zengler, K., et al., *Cultivating the uncultured.* Proc. Natl. Acad. Sci., 2002. **99**(24): p. 15681-15686.
6. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-147.
7. Kaeberlein, T., K. Lewis, and S.S. Epstein, *Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment.* Science, 2002. **296**(5570): p. 1127-1129.
8. Seyedsayamdost, M.R., *High-throughput platform for the discovery of elicitors of silent bacterial gene clusters.* Proc Natl Acad Sci, 2014. **111**(20): p. 7266-7271.
9. Schroeckh, V., et al., *Intimate bacterial–fungal interaction triggers biosynthesis of archetypal polyketides in Aspergillus nidulans.* Proc Natl Acad Sci, 2009. **106**(34): p. 14558-14563.
10. Challis, G.L., *Mining microbial genomes for new natural products and biosynthetic pathways.* Microbiol, 2008. **154**(6): p. 1555-1569.
11. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making?* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 175-184.
12. Ochi, K. and T. Hosaka, *New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters.* Appl Microbiol Biotechnol, 2013. **97**(1): p. 87-98.
13. Rutledge, P.J. and G.L. Challis, *Discovery of microbial natural products by activation of silent biosynthetic gene clusters.* Nat Rev Microbiol, 2015. **13**(8): p. 509-523.
14. Medema, M.H., T. de Rond, and B.S. Moore, *Mining genomes to illuminate the specialized chemistry of life.* Nat Rev Genetics, 2021: p. 1-19.
15. Medema, M.H. and M.A. Fischbach, *Computational approaches to natural product discovery.* Nat Chem Biol, 2015. **11**(9): p. 639-648.
16. Crits-Christoph, A., et al., *Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis.* Nature, 2018: p. 1.
17. Gavriilidou, A., et al., *A global survey of specialized metabolic diversity encoded in bacterial genomes.* bioRxiv, 2021.
18. Nayfach, S., et al., *A genomic catalog of Earth's microbiomes.* Nat Biotechnol, 2020: p. 1-11.
19. Lane, J.Q., et al., *Application of Solid Phase Adsorption Toxin Tracking (SPATT) for field detection of the hydrophilic phycotoxins domoic acid and saxitoxin in coastal California.* Limnol Oceanography: Methods, 2010. **8**(11): p. 645-660.
20. Tuttle, R.N., et al., *Detection of Natural Products and Their Producers in Ocean Sediments.* Applied and Environmental Microbiology, 2019. **85**(8): p. e02830-18.
21. Nonejuie, P., et al., *Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules.* Proc Natl Acad Sci, 2013. **110**(40): p. 16169-16174.
22. Zidorn, C., *Secondary metabolites of seagrasses (Alismatales and Potamogetonales; Alismatidae): Chemical diversity, bioactivity, and ecological function.* Phytochemistry, 2016. **124**: p. 5-28.
23. Ciavatta, M.L., et al., *Aplysiopsenes: an additional example of marine polyketides with a mixed acetate/propionate pathway.* Tetrahedron Letters, 2009. **50**(5): p. 527-529.
24. Rukachaisirikul, V., et al., *Cyclohexene, diketopiperazine, lactone and phenol derivatives from the sea fan-derived fungi Nigrospora sp. PSU-F11 and PSU-F12.* Archives of Pharmacal Research, 2010. **33**(3): p. 375-380.

25. Fu, P., et al., *α-Pyrones and Diketopiperazine Derivatives from the Marine-Derived Actinomycete Nocardiopsis dassonvillei HR10-5.* Journal of Natural Products, 2011. **74**(10): p. 2219-2223.
26. Zan, J., et al., *A microbial factory for defensive kahalalides in a tripartite marine symbiosis.* Science, 2019. **364**(6445): p. eaaw6732.
27. Xu, Y., et al., *Bacterial Biosynthesis and Maturation of the Didemnin Anti-cancer Agents.* Journal of the American Chemical Society, 2012. **134**(20): p. 8625-8632.
28. Wang, M., et al., *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.* Nat Biotechnol, 2016. **34**(8): p. 828-837.
29. Luesch, H., et al., *Total Structure Determination of Apratoxin A, a Potent Novel Cytotoxin from the Marine CyanobacteriumLyngbyamajuscula.* Journal of the American Chemical Society, 2001. **123**(23): p. 5418-5423.
30. Werner, K.A., L. Marquart, and S.A. Norton, *Lyngbya dermatitis (toxic seaweed dermatitis).* International Journal of Dermatology, 2012. **51**(1): p. 59-62.
31. Suzuki, T., et al., *Liquid chromatography–mass spectrometry of spiroketal stereoisomers of pectenotoxins and the analysis of novel pectenotoxin isomers in the toxic dinoflagellate Dinophysis acuta from New Zealand.* Journal of Chromatography A, 2003. **992**(1): p. 141-150.
32. Mohimani, H., et al., *Dereplication of microbial metabolites through database search of mass spectra.* Nature Communications, 2018. **9**(1).
33. Chomcheon, P., et al., *Curvularides A-E: Antifungal Hybrid Peptide-Polyketides from the Endophytic Fungus Curvularia geniculata.* Chemistry - A European Journal, 2010. **16**(36): p. 11178-11185.
34. Williamson, R.T., et al., *LR-HSQMBC: A Sensitive NMR Technique To Probe Very Long-Range Heteronuclear Coupling Pathways.* The Journal of Organic Chemistry, 2014. **79**(9): p. 3887-3894.
35. Nguyen, J.-T., et al., *Design of Potent Aspartic Protease Inhibitors to Treat Various Diseases.* Archiv der Pharmazie, 2008. **341**(9): p. 523-535.
36. Bray, M.-A., et al., *Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes.* Nature Protocols, 2016. **11**(9): p. 1757-1774.
37. Kang, J., et al., *Improving drug discovery with high-content phenotypic screens by systematic selection of reporter cell lines.* Nature Biotechnology, 2016. **34**(1): p. 70-77.
38. Mandegar, Mohammad A., et al., *CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs.* Cell Stem Cell, 2016. **18**(4): p. 541-553.
39. Wang, M., et al., *Mass spectrometry searches using MASST.* Nature Biotechnology, 2020. **38**(1): p. 23-26.
40. Ternon, E., et al., *On the Hunt for New Toxin Families Produced by a Mediterranean Strain of the Benthic Dinoflagellate Ostreopsis cf. ovata.* Toxins, 2022. **14**(4): p. 234.
41. Wegley Kelly, L., et al., *Molecular Commerce on Coral Reefs: Using Metabolomics to Reveal Biochemical Exchanges Underlying Holobiont Biology and the Ecology of Coastal Ecosystems.* Frontiers in Marine Science, 2021. **8**.
42. Pendergraft, M.A., et al., *Bacterial and Chemical Evidence of Coastal Water Pollution from the Tijuana River in Sea Spray Aerosol.* Environmental Science & Technology, 2023.
43. Kautsar, S.A., et al., *MIBiG 2.0: a repository for biosynthetic gene clusters of known function.* Nucleic Acids Research, 2020. **48**(D1): p. D454-D458.
44. Cox, R.J., *Curiouser and curiouser: progress in understanding the programming of iterative highly-reducing polyketide synthases.* Natural Product Reports, 2023. **40**(1): p. 9-27.
45. Walsh, C.T., R.V. O'Brien, and C. Khosla, *Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds.* Angewandte Chemie International Edition, 2013. **52**(28): p. 7098-7124.

46. Xu, Y., et al., *Characterization of the Biosynthetic Genes for 10,11-Dehydrocurvularin, a Heat Shock Response-Modulating Anticancer Fungal Polyketide from Aspergillus terreus.* Applied and Environmental Microbiology, 2013. **79**(6): p. 2038-2047.
47. Zhou, H., et al., *A polyketide macrolactone synthase from the filamentous fungus Gibberella zeae.* Proceedings of the National Academy of Sciences, 2008. **105**(17): p. 6249-6254.
48. Courtial, J., et al., *Characterization of NRPS and PKS genes involved in the biosynthesis of SMs in Alternaria dauci including the phytotoxic polyketide aldaulactone.* Scientific Reports, 2022. **12**(1).
49. Biggins, J.B., C.D. Gleber, and S.F. Brady, *Acyldepsipeptide HDAC Inhibitor Production Induced in Burkholderia thailandensis.* Organic Letters, 2011. **13**(6): p. 1536-1539.
50. Klau, L.J., et al., *The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function.* Journal of Biological Chemistry, 2022. **298**(10): p. 102480.
51. Udwary, D.W., et al., *Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica.* Proceedings of the National Academy of Sciences, 2007. **104**(25): p. 10376-10381.
52. Kallscheuer, N. and C. Jogler, *The bacterial phylum Planctomycetes as novel source for bioactive small molecules.* Biotechnology Advances, 2021. **53**: p. 107818.
53. Wang, M., et al., *Sharing and community curation of mass spectrometry data with GNPS.* Nat Biotechnol, 2016. **34**(8): p. 828.
54. Shannon, P., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Research, 2003. **13**(11): p. 2498-2504.
55. Dalisay, D.S. and T.F. Molinski, *NMR Quantitation of Natural Products at the Nanomole Scale.* Journal of Natural Products, 2009. **72**(4): p. 739-744.
56. Ruiz, N., et al., *Chemical conditionality: A geneticstrategy to probe organelle assembly.* Cell, 2005. **121**(2): p. 307-317.
57. Uphoff, C. and H. Drexler, *Detection of mycoplasma in leukemia–lymphoma cell lines using polymerase chain reaction.* Leukemia, 2002. **16**(2): p. 289-293.
58. Uphoff, C.C. and H.G. Drexler, *Detecting Mycoplasma Contamination in Cell Cultures by Polymerase Chain Reaction*. Humana Press. p. 319-326.
59. Feyen, D.A.M., et al., *Metabolic Maturation Media Improve Physiological Function of Human iPSC-Derived Cardiomyocytes.* Cell Reports, 2020. **32**(3): p. 107925.
60. Buikema, J.W., et al., *Wnt Activation and Reduced Cell-Cell Contact Synergistically Induce Massive Expansion of Functional Human iPSC-Derived Cardiomyocytes.* Cell Stem Cell, 2020. **27**(1): p. 50-63.e5.
61. Patin, N.V., et al., *Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates.* Microbial Ecology, 2013. **65**(3): p. 709-719.
62. Bushnell, B., *BBMap: A Fast, Accurate, Splice-Aware Aligner*. 2014, Lawrence Berkeley

National Lab (LBNL): Berkeley, CA.
63. Chase Alexander, B., et al., *Microdiversity of an Abundant Terrestrial Bacterium Encompasses Extensive Variation in Ecologically Relevant Traits.* mBio, 2017. **8**(6): p. e01809-17.
64. Johnson, L.S., S.R. Eddy, and E. Portugaly, *Hidden Markov model speed heuristic and iterative HMM search procedure.* BMC Bioinformatics, 2010. **11**(1): p. 431.
65. Matsen, F.A., R.B. Kodner, and E.V. Armbrust, *pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree.* BMC Bioinformatics, 2010. **11**(1): p. 538.
66. Peng, Y., et al., *IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.* Bioinformatics, 2012. **28**(11): p. 1420-1428.

67. Blin, K., et al., *antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.* Nucleic Acids Research, 2019. **47**(W1): p. W81-W87.
68. Navarro-Muñoz, J.C., et al., *A computational framework to explore large-scale biosynthetic diversity.* Nature Chemical Biology, 2020. **16**(1): p. 60-68.
69. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. **9**(4): p. 357-359.
70. Kang, D.D., et al., *MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies.* PeerJ, 2019. **7**: p. e7359.
71. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.* Genome Research, 2015. **25**(7): p. 1043-1055.
72. Olm, M.R., et al., *dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication.* The ISME Journal, 2017. **11**(12): p. 2864-2868.
73. Chaumeil, P.-A., et al., *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.* Bioinformatics, 2020. **36**(6): p. 1925-1927.