LC-HRMS-Database Screening Metrics for Rapid Prioritisation of Samples to Accelerate the Discovery of Structurally New Natural Products

Jioji N. Tabudravu,[†,‡] *Léonie Pellissier,*[‡] *Alan James Smith,*[‡] *Karolina Subko,*[‡] *Caroline Autréau,*[‡] *Klaus Feussner,*[§] *David Hardy,*[⊥] *Daniel Butler,*[O] *Richard Kidd,*[▽] *Edward J. Milton,*[O] *Hai Deng,*[‡] *Rainer Ebel,*[‡] *Marika Salonna*[◿] *Carmela Gissi,*[◿Ω] *Federica Montesanto,*[#] *Sharon M. Kelly,*[□] *Bruce F. Milne,*[◊] *Gabriela Cimpan,*[O] *Marcel Jaspars**[‡]

[†]*School of Forensic and Applied Sciences, Faculty of Science & Technology,University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK.*

[‡]*Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, AB24 3UE, Scotland, UK.*

[§]*Institute of Applied Sciences, Faculty of Science, Technology and Environment, University of the South Pacific, Laucala Campus, Private Mail Bag, Suva, Fiji Islands.*

[⊥]*Thermo Fisher Scientific, Altrincham Business Park, 1 St George's Court, Altrincham WA14 5TP, UK.*

[▽]*Publisher, Data & Databases, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK.*

[O]*Advanced Chemistry Development, UK Ltd. Venture House, Arlington Square, Downshire Way, Bracknell, Berks. RG12 1WA, UK.*

[◊]*CFisUC, Department of Physics, University of Coimbra, Rua Larga, 3004-516, Coimbra, Portugal*

[◿]*Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari "A. Moro", Via Orabona 4, 70125 Bari, Italy.*

[Ω]*IBIOM, Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari, CNR, Via Amendola 165/A, 70126 Bari, Italy.*

[#]*Department of Biology - LRU CoNISMa, University of Bari, Via Orabona 4, 70125, Bari, Italy*

[□]*Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, G128QQ, UK.*

ABSTRACT

In order to accelerate the isolation and characterisation of structurally new or novel secondary metabolites, it is crucial to develop efficient strategies that prioritise samples with greatest promise early in the workflow so that resources can be utilised in a more efficient and cost-effective manner. We have developed a metrics-based prioritisation approach using exact LC--HRMS which uses data for 24,618 marine natural products held in the PharmaSea database. Each sample was evaluated and allocated a metric score by a software algorithm based on the ratio of new masses over the total (sample novelty), ratio of known masses over the total (chemical novelty), number of peaks above a defined peak area threshold (sample complexity), and peak area (sample diversity). Samples were then ranked and prioritized based on these metric scores. To validate the approach, 8 marine sponges and 6 tunicate samples collected from the Fiji Islands were analysed, metric scores calculated and samples targeted for isolation and characterisation of new compounds. Structures of new compounds were elucidated by spectroscopic techniques, including 1D and 2D NMR, MS and MS/MS. Structures were confirmed by Computer Assisted Structure Elucidation methods (CASE) using the ACD/Structure Elucidator Suite.

Natural products have been our most productive and valuable source of new drugs to date.[1,2] For example, 70% of small molecule therapeutics to treat infectious disease, and 77% of small molecules to treat cancer can trace their origin back to natural products.[1] The number is expected to increase as drug discovery from new ecological niches and microbial sources,[3–7] and genome mining[8] is now being realised. Natural products are much more attractive as drug leads, as they are known to occupy a much wider chemical space[9,10] and are more drug-like than compounds derived from combinatorial chemistry.[11,12] However, finding new natural products is far from trivial and one of the reasons for this, is the large number of known natural products that now stands at more than 300,000.[13] This, unfortunately leads to higher rates of compound redundancy if chemical dereplication which is the process of compound identification is not performed early on in the project.[14,15] Historically, natural product sources were usually chosen either randomly or on the basis of their ecology/geography and/or taxonomy. This meant determination of chemical diversity, which usually occurred after extraction, isolation and purification had been performed, resulted in high redundancy rates. To overcome this problem, several dereplication methods have been developed.[13,16] One example is the use of genetic information of the organism[17] particularly in microorganisms.[8,13,18,19] However, challenges remain as it has been shown that even strains with similar 16S gene sequences do not necessarily produce the same chemistry[20,21] indicating that the induction and/or activation of biosynthetic machinery for natural product production is far more complex than originally thought, and relies on a complex interaction of environmental, chemical, biochemical and biological stimuli.[8,22] Even if the organism's metabolome is successfully stimulated, the next challenge is detection and identification of new secondary metabolites that are often produced in small quantities, in a background of known metabolites present in higher quantities. It is therefore imperative to devise new strategies that rapidly identify known compounds early on in the project, so that resources can be concentrated only on the discovery of structurally new or novel ones. At the forefront of this chemical dereplication process is the use of LC hyphenated with detectors like MS,[23,24] photodiode array (DAD),[23] evaporative light scattering (ELSD),[25] and NMR.[26] Recently, we proposed a new strategy of identifying known microbial natural products utilising a combination of HRMS and predicted LC retention time of 5,098 compounds from *Streptomyces*.[27] Another approach used a predicted ^{13}C NMR chemical shifts database to screen for similar compounds in an extract.[28]

The goal of this study was to develop a LC-HRMS-database-software-integrated tool that can process, screen, and prioritise samples based on three metrics: novelty, complexity and

diversity. To demonstrate the effectiveness of this approach we analysed 14 marine sponge and tunicate extracts, ranked them based on software-derived metrics followed by isolation and structure elucidation of compounds using HPLC, NMR and MS. The use of HRMS,[20,23,29,30] and NMR[13,31,32] have been well documented as chemical dereplication tools in the literature, but the use of a software algorithm approach that provides numerical scores on key compound indicators like novelty, diversity and complexity has yet to be explored. This approach has the potential to vastly improve the productivity of natural products drug discovery programmes; particularly those involving large sample numbers.

Results and Discussion

An LCMS profile of a bacterial culture extract for example, provides *m/z* values corresponding to the metabolome of the strain under defined experimental conditions. This concept has been successfully utilised to characterise, identify, and discriminate samples.[15,27,33] In this context, the data processing algorithm, ACD/IntelliXtract (IX), part of the ACD/MS Workbook Suite[34] was scripted to provide metrics on sample novelty (number of unidentified *m/z* ions over the total number of detectable ions), chemical novelty (total number of identified masses in the sample over the total number of ions), sample complexity (number of peaks above a defined threshold), and sample diversity (abundance) taking into consideration peak areas and heights. The script has been written for execution at the tail end of the normal LCMS data processing sequence which normally occurs in a series of steps starting with file conversion, followed by feature detection, normalisation, and finally linking with external databases for compound identification.[35,36] A couple of software packages are currently available that can successfully perform these tasks. For example MZmine,[37,38] XCMS,[39,40] MAVEN,[41] and the ACD/MS Workbook Suite.[34] We preferred the commercially available software package provided by ACD/Labs, as it offered more flexibility and versatility allowing database creation and search capabilities (SpectrusDB), NMR data processing, NMR chemical shift prediction and structure elucidation (Structure Elucidator), retention time prediction (ChromGenius), and others that use the same interface.[34] Two databases were created for identifying the masses (*m/z*) generated by IntelliXtract. PharmaSeaDB (Created using the ACD/Spectrus Enterprise platform) is a database of 24,618 compounds derived from MarinLit[42] containing structures, molecular formulas, $[M+H]^+$, $[M-H]^-$, M^+ and M^- (Figure S28). The second database, MatrixDB containing more than one thousand compound masses (*m/z*) derived from culture media, solvent blanks and other contaminants is stored within the Postgre-SQL server.[34] There is

flexibility in the use of other matrix databases depending on application, for example, use of fungal culture media instead of bacteria. Dereplication starts by inputting the LCMS data file name (Figure S23) and the data file processed by the scripted IX (IX.mcr) resulting in a table of indexed masses (Table S24), which are then deconvoluted in terms of retention time and percentage total ion chromatogram (%TIC, Table S25). These indexed masses are then screened against the two databases (MatrixDB and PharmaSeaDB (Figure S28)), yielding a table (Table S26) of new masses (masses not identified in either database). All this information is used to calculate the prioritisation metrics (Table S27). The 'media fraction' metric provides an indication of how clean the sample is from external contaminants. It is calculated by dividing the number of matrix ions found in the sample over the total number of ions detected. The 'sample novelty' metric is the ratio of new masses to the total masses after the exclusion of matrix components. Samples with high sample novelty indices (Table S27) indicate low 'hit' rates in PharmaSeaDB. 'Sample complexity' is based on the number of new peaks above a set peak area threshold, within the retention time window. Samples with high complexity indices suggest large number of new masses either close together or spread out within the chromatogram. The 'sample diversity' metric indicates the intensity of peaks based on the sum of peak heights, multiplied by the logarithm of the peak areas of the new masses.

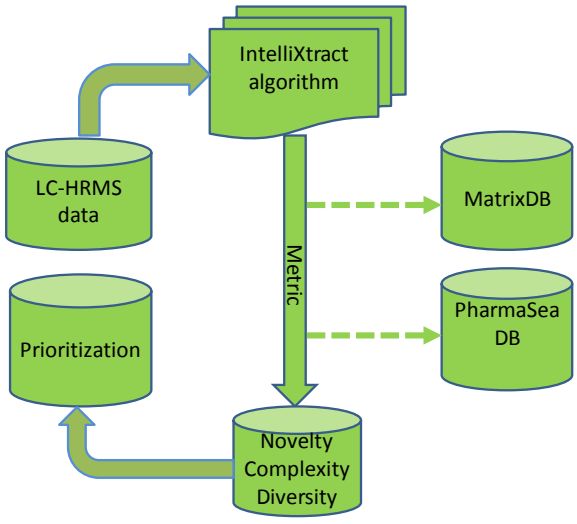


Figure 1. Metric score calculation work-flow. HR-LCMS data is loaded and processed by ACD/IntelliXtract (IX.mcr) that interrogated both MatrixDB and PharmaSeaDB yielding data for calculation of novelty, complexity, diversity and media score metrics. All metric data and ion chromatograms are stored in the PharmaSeaDB (Figure 2). Other information such as total detected peaks (*m/z)*, new peaks and associated retention times, peak areas, and peak heights are automatically saved as text files.

Screening of compounds in MatrixDB and PharmaSeaDB was performed based on the principle of mass-matching[43,44] with an error setting of ± 0.0025 Da between experimental HRMS data (accurate masses) and exact masses of known compounds held in the databases. Data processing settings used by ACD/IntelliXtract (IX.mcr) were optimised using LC-HRMS data of bacteria and marine sponge extracts, that had been pre-cleaned using solid phase extraction (SPE) to minimise ion matrix interference within the ion source of the mass spectrometer[45,46] (Figure S35). Optimization of processing settings is important, as it affects the peak-picking, alignment, and data accuracy of the applied algorithm, particularly in complex biological samples.[47,48] Each of the 14 samples (Table S38) used in this study was fractionated by SPE to yield four fractions (56 fractions in total), analysed by LC-HRMS, processed, and metric calculations performed. Samples were then ranked based on these metrics for isolation of compounds for structure elucidation.

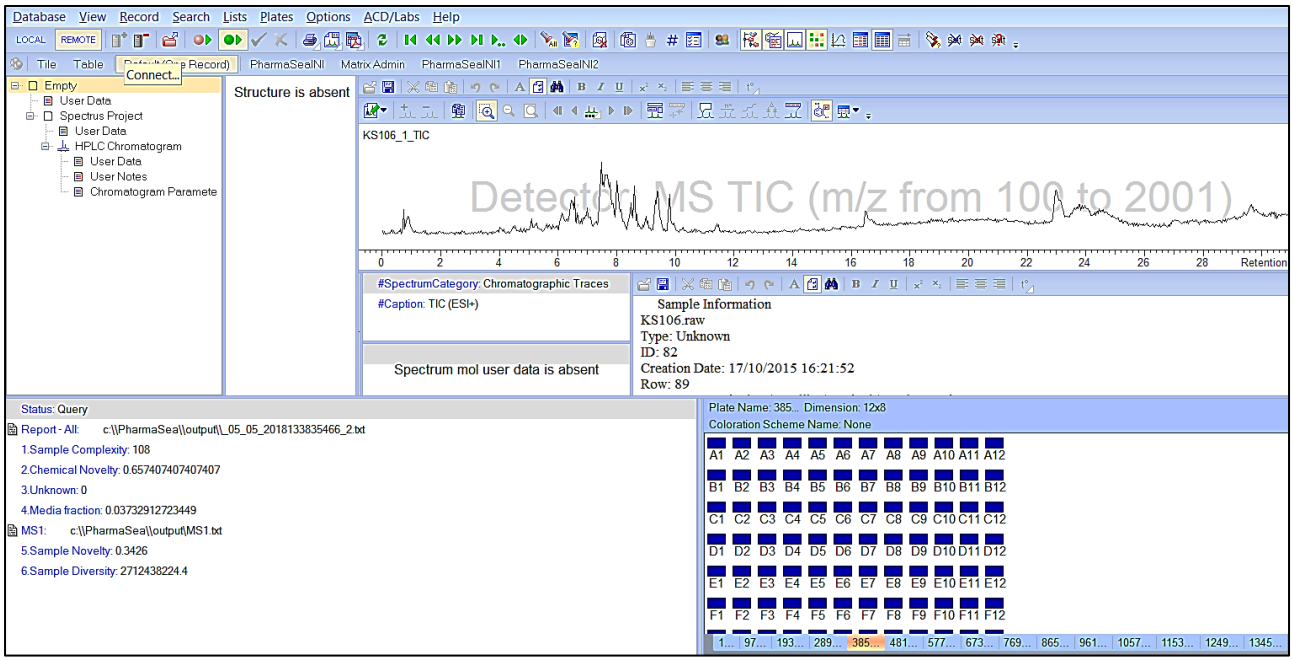


Figure 2. An example of a typical output profile showing total ion chromatogram (TIC), sample information, and metric data (complexity, novelty, diversity, and media) stored in ACD/Spectrus DB - PharmaSeaDB. A sample novelty metric score of 0.34 for example indicates that 34% of the compounds in the sample are not found in PharmaSeaDB. A media fraction metric score of 0.037 indicates 3.7% of peaks in the sample originate from the matrix.

Sample Prioritisation. Detailed analysis was carried out to determine the accuracy of the metric scores. Eden-6-50, the second highest ranked sample (Table S27) based on sample novelty showed 97 out of 109 detected masses (Tables S30, S31) were not found in

PharmaSeaDB, yielding a sample novelty index of 0.89 (Figure S29) or approximately 89% of the compounds in this sample were not found in PharmaSeaDB. In comparison, Tavarua-2-SPE-50 (Tables S32, S33) ranked 56th overall, showed 11 masses out of 74 were not found in PharmaSeaDB (Sample novelty = 0.15, Figure S34) indicating about 15% of the compounds in this sample were not in the database. It needs to be pointed out that some masses could be annotated more than once depending on the charged adducts that have been formed including $[M + Na]^+$, $[M + NH_4]^+$, $[2M + H]^+$ which could distort the results. Adduct formation varied between samples in this study, where some had none to some with about 10% of masses occurring as other adducts. Adduct formations are very difficult to predict as they are highly dependent on MS ion-source settings, solvents used, and analyte concentration.[23,49] In addition some molecular ions masses may be lost due to in source fragmentations.[50]

Sample novelty metric scores were ranked in decreasing order from Tavarua-8-100 to Tavarua-2-50 (Figure 3, Table S27). Further analysis of this data tends to suggest a possible association of sample novelty with compound polarity, where most of the new compounds were found in the medium to less polar fractions (Figure 4). Sample complexity and diversity metrics are shown in Tables S36 and S37 respectively. Analysis of matrix contamination (media fraction metric) indicate it is insignificant for these marine extracts (<10% of total masses, Table S27). Overall, an ideal sample is one that scores high in sample novelty, complexity and diversity, but low in chemical novelty (indication of known compounds), and media components. However, this is often not straightforward as exemplified by Tavarua-8-100 which was ranked 1st on sample novelty (Table S27), 25th on complexity (Table S36), and 15th on diversity (Table S37). This makes the task of sample prioritisation challenging, but because isolating structurally new or novel compounds was the main goal of this study, the sample novelty metric was given more weight than the others. To determine if the predictions by these metrics would yield new compounds, two samples were chosen, one from either end of the sample novelty metric scale for isolation and structure elucidation work.

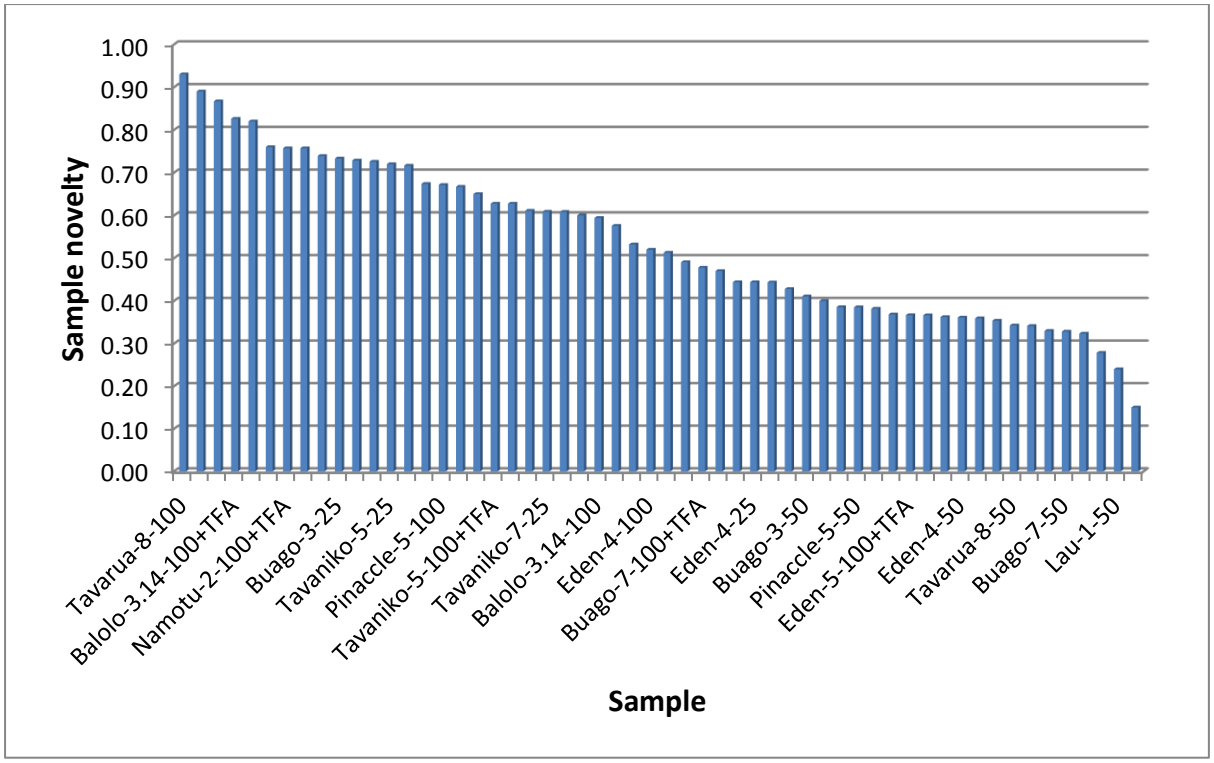


Figure 3. Sample novelty metric scores of 56 fractions in decreasing order calculated as the ratio of compounds not found in PharmaSeaDB against the total number of masses found per sample. Samples with higher novelty metric scores suggest presence of a large number of compounds not found in PharmaSeaDB.

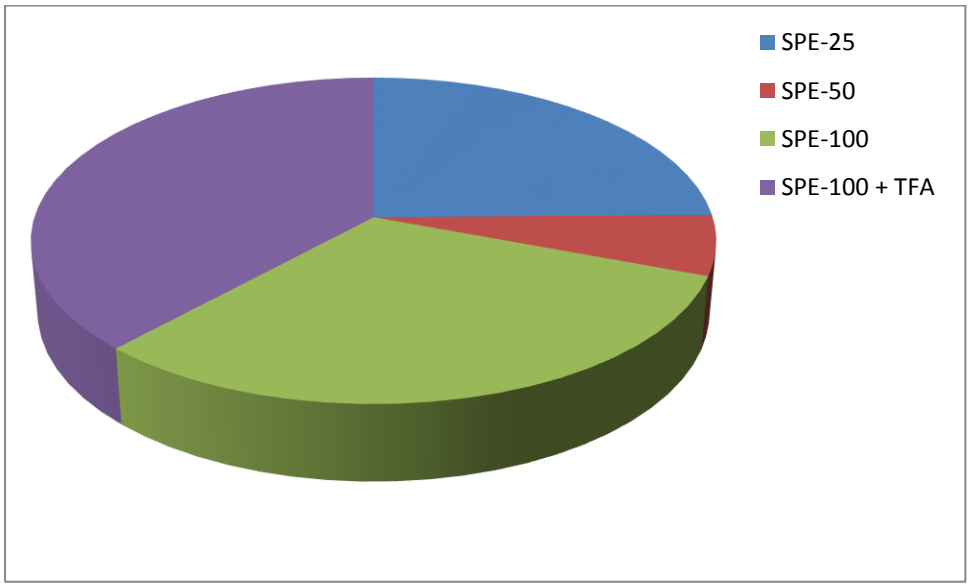
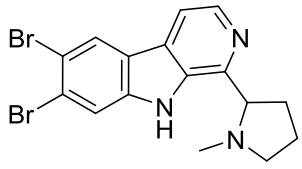
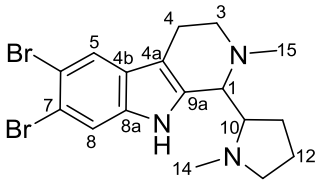


Figure 4. Association of new compounds (not found in PharmaSeaDB) with fraction-types based on analysis of sample novelty scores greater than 0.50 where about 50% of the compounds are expected to be new.

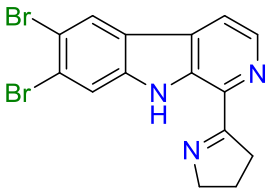
Case Study 1 – Tavarua-8. This case study describes the isolation of two new compounds from the highest ranked fraction, Tavarua-8-100 based on the sample novelty metric. Tavarua-8 was collected from a reef on the island of Tavarua (17.850° S, 177.183° E), Fiji Islands, and identified to genus level as either *Eudistoma* (Polycitoridae) *or Pseudodistoma* (Pseudodistoma) based on detailed morphological and molecular analysis (Supporting Information Part C). A voucher specimen (USP12339) has been deposited at the South Pacific Regional Herbarium, University of the South Pacific, Suva, Fiji. This sample was fractionated by SPE using 100% MeOH to yield the fraction Tavarua-8-100. The high sample novelty metric (0.93, Table S27) indicated a significant number (93%) of compounds in this fraction were not found in PharmaSeaDB and were potentially new (Figure S41, Table S42). Targeted purification by reversed-phase HPLC based on UV profile and LC retention time of the unknown compounds yielded two new eudistomin derivatives **1** (1.4 mg), and **2** (0.9 mg. Yields of other unknown compounds were below the detection limit for NMR spectroscopy and their structures remain undetermined.



Eudistomin Z1 (**1**)



Eudistomin Z2 (**2**)



Eudistomin V (**3**)

Table 1. NMR Spectroscopic Data (600/150 MHz, CD_3OD) for Compounds 1 and 2

	1				2			
Pos.	δ_C[a], type	δ_H (*J* in Hz)	COSY	HMBC $^1H \rightarrow ^{13}C$	δ_C[a], type	H (*J* in Hz)	COSY	HMBC $^1H \rightarrow ^{13}C$
1	138.1, C				60.3, CH	3.30, ovlp		
3	139.1, CH	8.47, d (5.4)	4	1, 4a, 4b, 9a	45.3, CH_2	3.30, ovlp 3.09, dd (15.4, 4.2)	4	4a
4	116.5, CH	8.15, d (5.4)	3	3, 4a, 4b, 9a	16.1, CH_2	2.60, dd (15.4, 4.2) 2.99, ovlp	3	9a

4a	130.0, C				107.9, C			
4b	123.1, C				127.5, C			
5	127.2, CH	8.60, s		4a, 6, 7, 8a	123.1, CH	7.79, s		4a, 6, 7, 8a
6	115.8, C				116.4, C			
7	125.1, C				113.7, C			
8	117.7, CH	7.99, s		4b, 6, 7, 8a	116.6, CH	7.69, s		4b, 6, 7
8a	141.9, C				136.4, C			
9a	134.9, C				130.8, C			
10	69.8, CH	5.07, ovlp			70.4, CH	3.80, m		14
11	31.9, CH_2	2.89, m	12		28.7, CH_2	2.28, m	12	
		2.37, ovlp	11, 13				11, 13	
12	23.5, CH_2	2.20, ovlp			23.5, CH_2	2.20, m		
		3.97, m	12				12	14
13	57.0, CH_2	3.44, m			57.0, CH_2	3.66, m		
14	40.5, CH_3	2.99, S		10	42.1, CH_3	3.00, s		10, 13
15					40.1, CH_3	2.55, s		1, 3

Ovlp = overlap

[a]Carbons extracted from 2D NMR (HSQC and HMBC) data

Compound **1** (eudistomin Z1) showed a HRESIMS ion at *m/z* 407.9702 $[M+H]^+$ (Δ 0.8 ppm) (Figure S15) for the expected molecular formula of $C_{16}H_{15}Br_2N_3$, and requiring 10° of unsaturation[51], and indicated the presence of two bromine atoms.[52] Interpretation of the ^{13}C NMR spectrum of **1** showed similarities to the NMR data for eudistomin V(**3**).[53] Careful analysis of 1D and 2D NMR data (Table 1, Figure 5) suggested **1** contained the sub-unit 1-methylpyrrolidine instead of 3,4-dihydro-2*H*-pyrrole found in eudistomin V(**3**) to yield eudistomin Z1(**1**) a new 6,7-dibromo-1-(1-methylpyrrolidin-2-yl)-9*H*-β-carboline. A good correlation between experimental and predicted ^{13}C NMR data (R^2= 0.9991) provides additional evidence that the proposed structure is correct (Figure S16).[54] The absolute configuration of Z1 has yet to be determined.

Compound **2** (eudistomin Z2) showed a HRESIMS ion at *m/z* 426.0174, $[M+H]^+$ (Δ 0.2 ppm) (Figure S21) for the expected molecular mass for $C_{17}H_{21}Br_2N_3$, and requiring 8° of unsaturation, and indicated the presence of two bromine atoms as found in **1**. Closer inspection of ^{13}C NMR data indicated similarities to compound **1** (eudistomin Z1), and eudistomins V (**3**).[53] Careful analysis of 1D and 2D NMR (Table 1, Figures 5) data suggested **2** had lost unsaturation between N-2 and C-3 as found in **1**, but N-2 had gained a methyl group yielding eudistomin Z2 (**2**), a new 6,7-dibromo-1-methylpyrrolidin-2-yl)-2,3,4,9-tetrahydro-1*H*-β-carboline. A good correlation between experimental and predicted ^{13}C NMR data (R^2=

0.9974)[54] provides additional evidence that the proposed structure is correct (Figure S22). The absolute configuration of Z2 has yet to be determined.

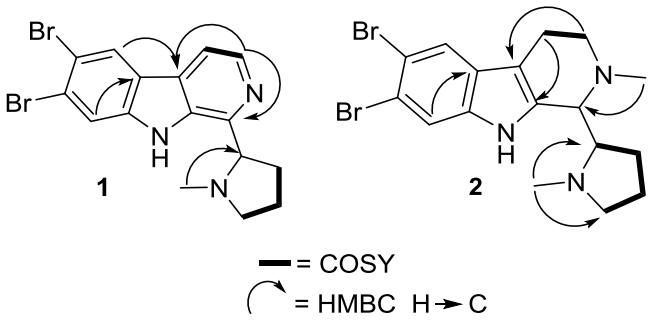
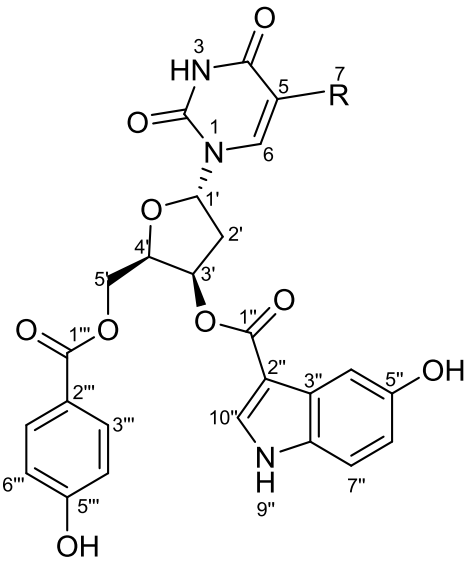


Figure 5. COSY and HMBC correlations for compounds **1** and **2**

Case Study 2 - Tavarua-2. This case study describes the isolation and characterisation of two novel nucleosides, that were predicted to be known based on matching masses in PharmaSeaDB. This perhaps underlines one of the limitations of this technique, where matching masses in the database could eliminate compounds with the same mass or molecular formula but possess different chemical structures. Tavarua-2, a pink mottled tunicate, (Figure S61) was collected at a depth of 20 m from the island of Tavarua (17.863°S, 177.192°E), Fiji Islands (Table S38). Tavarua-2 has been assigned to the family Didemnidae based on morphological and 18S rDNA sequence analysis (Supporting Information Part C). A voucher specimen (USP12338) has been deposited at the South Pacific Regional Herbarium, University of the South Pacific, Suva, Fiji. Tavarua-2-50, an SPE fraction of Tavarua-2 (eluted with 50% MeOH-H_2O; sample novelty = 0.15), contained 11 potentially new compounds (Table S32). The identity of the dominant compound (*m/z* = 522, Figure S45) was of interest to us, as the proton NMR profile (Figure S1) did not match any of the predicted[34] ^{1}H NMR profiles of the MS-matched compounds in PharmaSeaDB. A recalculation of the metric data using a narrower mass range (0.0001 Da (0.2 ppm at *m/z* of 500)) showed the mass was not in PharmaSeaDB (Figure S43, Table S44) warranting full structural investigation. HPLC purifications of this fraction yielded compounds **4** and **5**. The molecular formula of **4** (Tavarua deoxyriboside A) was established by HRESIMS (Figure S5) as $C_{26}H_{24}N_3O_9$, and suggested 17° of unsaturation. The ^{13}C data extracted from HSQC[55] and HMBC[32,56] NMR spectra indicated the presence of 26 carbons (Table 2) in the form of one methyl group (δ_C 11.8), 2 sp^3 methylenes (δ_C 37.1,

63.5), 9 sp^2 methines (δ_C 131.7, 131.7, 115.8, 115.8, 136.7, 105.8, 112.5, 112.3, 133.0), 3 sp^3 methines (δ_C 85.6, 83.1, 75.1), 7 sp^2 non-protonated carbons (δ_C 120.1, 162.5, 104.6, 126.8, 131.4, 152.6, 110.7), two amide carbonyl (δ_C 164.7, 150.9), and two ester carbonyl (δ_C 164.7, 165.9) groups.



R = Me, Tavarua deoxyriboside A (**4**)
R = H, Tavarua deoxyriboside B (**5**)

Use of 1D and 2D NMR data (Figures S1-S4) enabled the construction of four substructures (Figure 6). The presence of a 2'-deoxy ribose unit was defined by COSY correlations: H-1' and the non-equivalent CH_2 protons (H-2'A and H-2'B); H-2B' and H-3' and between H-3' and H-4' (Substructure B, Figure S3). ^{13}C chemical shifts of this pentose moiety are very similar to those found for 3-acetyl-5-methyl-2'deoxyuridine previously isolated from a marine sponge derived culture of *Streptomyces microflavus*.[57] Long range HMBC correlations (Figure S4) between the *sp*^{2} methine at δ_H 7.39 ppm to the two carbonyls at δ_C 164.7 and 150.9 ppm, between the methyl group at δ_H 1.42 ppm and the amide carbonyl at δ_C 164.7 ppm suggested the presence of a thymine moiety in the structure of **4** (substructure A).

Table 2. NMR Spectroscopic Data (600/150 MHz, CD_3OD) for Tavarua Deoxyriboside A (**4**)

pos.	δ_C[a], type	δ_H (*J* in Hz)	COSY	HMBC $^1H\rightarrow{}^{13}C$
2	150.9, C			
4	164.7, C			

5	110.8, C			
6	136.7, CH	7.39, d (1.2)		1’, 2, 4
7	11.8, CH_3	1.42, s		4, 5, 6
1'	85.6, CH	6.40, dd (8.7, 5.8)	2’a, 2’b	2’
2'	37.1, CH_2	a: 2.62, m	1’, 2’b, 3’	1', 3'
		b: 2.49, m	2’a	
3'	75.1, CH	5.72, bm	4', 2’b	1', 2', 1’’
4'	83.1, CH	4.51, m	3'	
	63.5, CH_2	a: 4.91, ovlp	5’b	
5’		b: 4.52, dd (12.1, 3.6)	5’a	
1''	164.7, C			
2''	104.6, C			
3''	126.8, C			
4''	105.8, CH	7.51, d (2.4)	6’’	5'', 10’’
5''	152.6, C			
6''	112.5, CH	6.78, dd (8.8, 2.4)	7'', 4’’	5'', 2’’
7''	112.3, CH	7.30, d (8.8)	6''	5''
8''	131.4, C			
9''				
10''	133.0, CH	7.98, s		2'', 3''
1'''	165.9, C			
2'''	120.1, C			
3'''/7’’’	131.7, CH	7.95, d (8.8)	4'''	1''', 5’’’
4'''/6’’’	115.8, CH	6.86, d (8.8)	3'''	5''', 2’’’
5'''	162.5, C			

Ovlp = overlap

[a]Carbons extracted from 2D NMR (HSQC and HMBC) data

Chemical shifts and integration of the coupled aromatic sp^2 methines (δ_H 7.95, 6.86 ppm, J = 8.8 Hz) suggested the presence of a para-hydroxy-substituted benzene ring (substructure C).[58] COSY correlations between the sp^2 methine at δ_H 7.30, 6.78, and 7.51 ppm; HMBC correlations between the sp^2 carbon at δ_C 152.6 ppm and the methine at δ_H 7.30 ppm confirmed the presence of a 5-hydroxy-1*H*-indole-3-carboxylic acid moiety (substructure D). The four substructures were assembled based on long range HMBC correlations. A HMBC correlation between the sp^2 carbon at 136.7 ppm (C-6) and the methine at 6.40 ppm (H-1’) established the link between substructures A and B. Thymidine-2’-deoxyriboside (substructures A and B) is well established in marine natural products.[57,59,60]

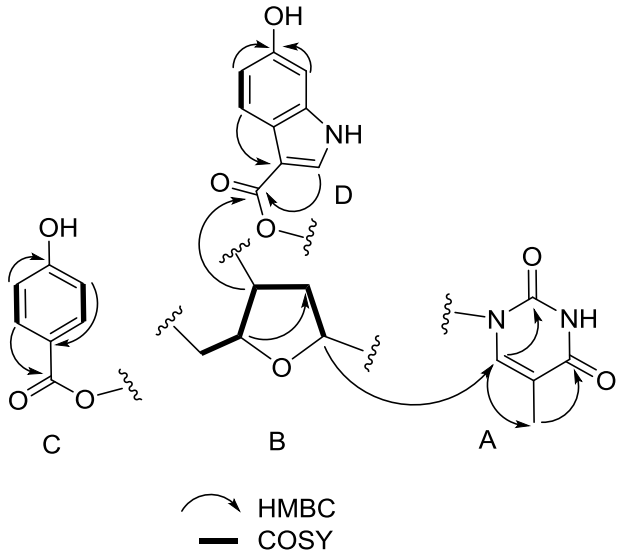


Figure 6. Substructures A-D of Tavarua deoxyriboside A (**4**) with COSY and HMBC correlations.

An HMBC correlation between C-1'' and the proton at H-3' established the link between substructures B and D. Establishing HMBC correlations between fragment B and C proved difficult, as HMBC correlations in fragment C were all internal. Nevertheless, the link was established by MS/MS fragmentation data as shown in (Figure S6), providing further evidence of a new thymidine-2'-deoxyriboside nucleoside. Additional proof for the structure of Tavarua deoxyriboside A was performed by comparing predicted and experimental ^{13}C NMR data. Predicted ^{13}C NMR data was performed using the ACD/Structure Elucidator Suite utilising the HOSE algorithm.[61] This strategy has been shown to be effective in predicting correct structures of natural products.[32,54] A good correlation (R^2 = 0.9991) was obtained suggesting that the proposed planar structure for Tavarua deoxyriboside A was correct (Figure 7).

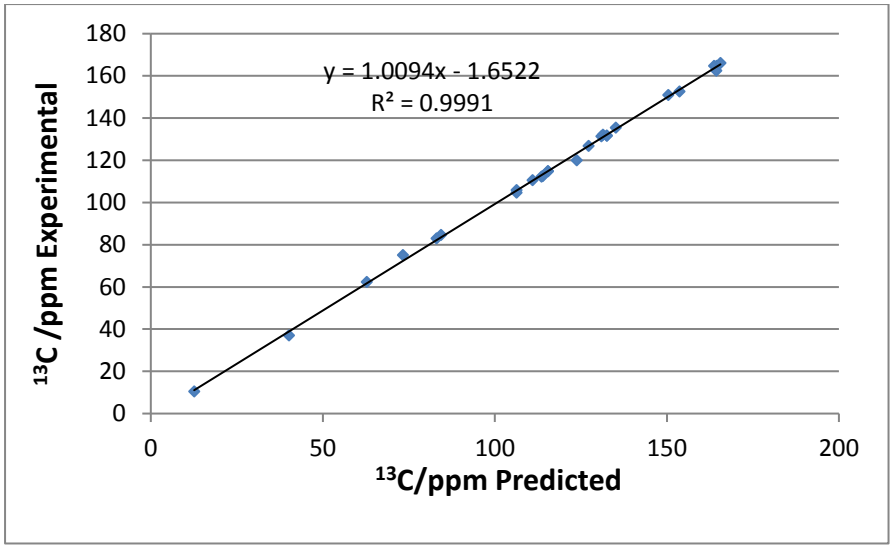


Figure 7. Experimental vs predicted ^{13}C NMR data for Tavarua deoxyriboside A (**4**).

The relative configurations at positions C-1', C-3' and C-4' in **4** have been assigned based on observed 2D ROESY correlations (Figure S7) between H-2b' (δ_H 2.49) and H-6 (δ_H 7.39), H-1' (δ_H 6.40) and H3' (δ_H 5.72), and between H-3' and H-4'(δ_H 4.51). Weak correlations were observed between H2b' and both H-1' and H-3'. The data is consistent with a deoxyribose linked α to N-1. Figure S8 shows the calculated global minimum energy conformation for compound **4** using ChemBio3D Ultra 12.0.[62]

The absolute configuration of the deoxyribose unit in **4** for the positions C-1', C-3'and C-4' has been determined to be *S,R,R* based on comparison of experimental electronic circular dichroism (ECD) spectrum with that computationally derived using the Avogadro molecular modelling software version 1.1.1 and energy minimized with the MMFF94 force field.[63,64] Calculation of all eight stereoisomers yielded *S,R,R* as the best fit (Supporting Information Part B).

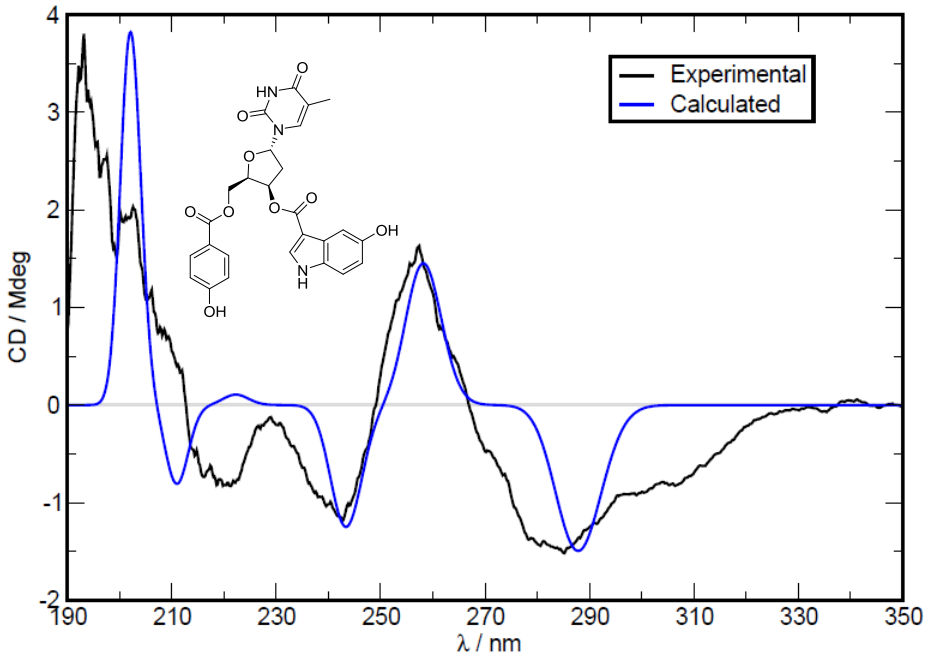


Figure 8. Overlay of calculated (blue) over experimental (black) ECD spectra for Tavarua deoxyribose (**4**) indicating the *S,R,R* configuration at positions C-1',C-3',C-4'.

Tavarua deoxyriboside B (**5**) showed a molecular formula of $C_{25}H_{22}N_3O_9$ suggesting 17 degrees of unsaturation. The LCMS data suggested the loss of a methyl group from the structure of **4** (Figure S9). This was confirmed by inspection of the ^{1}H NMR spectrum (Figure S10) of **5** which had lost the proton singlet at δ_H 1.40 ppm in **4** and replaced by a proton at δ_H 5.45 ppm (J =8.1 Hz) yielding a new pyrimidine-2'-deoxyribose nucleoside (**5**).

Compounds containing the indole-3-carbonyl motif in natural products are not common and its presence commonly can be found in a small group of natural products, such as indolyl-3-carbonyl-alpha-L-rhamnopyranoside from *Streptomyces* sp. GT 061150[65] and 1H-Indole-3-carbothioic acid *S*-methyl ester from the marine bacterium *Oceanibulbus indolifex* Hel 45.[66] Nucleosides containing benzoyl ester motifs have been previously isolated from the marine fungus *Aspergillus versicolor* derived from the gorgonian *Dichotella gemmacea*[67] and related compounds from the marine ascidian *Atriolum robustum*.[68]

CONCLUSION

The study has shown that a simple, sample prioritisation approach, based on calculated metric scores can successfully identify samples that contain potentially new or novel compounds as it

resulted in the isolation and structure elucidation of two new eudistomin-analogues and two new nucleosides. The study has also highlighted a general limitation of the method, where new or novel compounds could be missed if they share the same HRMS data.[69] We are currently working on improving the accuracy of this strategy by including MS/MS, and predicted HPLC retention time data,[27] to provide better and reliable identification of known compounds. Molecular networking[70] based on MS/MS has been used successfully to identify compound families that can be helpful as an additional prioritization filter. For compounds that are not in the integrated database statistical analysis like use of PCA can help identify compounds that are most interesting.[71] Nevertheless the strategy discussed in this paper has demonstrated the importance of sample prioritisation for rapid discovery of new or novel natural products, particularly when working with larger sample sets.

EXPERIMENTAL SECTION

General Experimental Procedures. The optical rotation was recorded using a Bellingham & Stanley, Model ADP410 Polarimeter at 589 nm at 25 °C. UV spectra were recorded on a photodiode array (DAD)-HPLC system. ECD spectra were obtained using a Jasco J-810 spectropolarimeter[72] at 20 °C. The sample was measured at a concentration of 0.2 mg/mL dissolved in spectral grade methanol using a 0.5 mm pathlength quartz cuvette. IR spectra were recorded on a Perkin Elmer UATR Two, Model L1600300. NMR data, both 1D and 2D were recorded on a Bruker AVANCE III HD Prodigy TCI Cryoprobe at 600 and 150 MHz for ^{1}H and ^{13}C respectively. This instrument was optimized for ^{1}H observation with pulsing/decoupling of ^{13}C and ^{15}N with 2H lock channels equipped with shielded z-gradients and cooled preamplifiers for ^{1}H and ^{13}C. The ^{1}H and ^{13}C chemical shifts were referenced to the solvent signals (δ_H 3.31 and δ_C 49.00 in CD_3OD). High resolution mass spectrometry data were measured using a Thermo Fisher Scientific LTQXL-Discovery Orbitrap[73] coupled to an Accela UPLC-DAD system. The following conditions were used for mass spectrometric analysis: capillary voltage 45 V, capillary temperature 320 °C, auxiliary gas flow rate 10 -20 arbitrary units, sheath gas flow rate 40-50arbitrary units, spray voltage 4.5 kV, mass range 100-2000 amu, resolution 30,000 for MS and 7,500 for MS/MS. Solid phase extractions (SPE) were performed using Phenomenex[74] C18 cartridges (Strata C18-E, 55um, 70Å). Semi-preparative HPLC purifications were performed using an Agilent[75] 1100 HPLC system consisting of a

binary pump, degasser, photodiode array detector (DAD), and a preparative fraction collector. All solvents were of HPLC grade.

LCMS Analysis. Each fraction (56 in total) was prepared in 100% MeOH to produce a 0.5 mg/mL solution with 4 μL injected by HPLC, and then post-column split (3:4 (v:v), MS:DAD). Analysis by LCMS was carried out using polarity switch mode where only the positive mode was under high resolution. Both the MS/MS and negative mode were measured at 7,500 peak resolution respectively. HPLC separations were carried out on an Agilent Technologies[75] Poroshell 120, EC-C18, 2.1 x 100 mm at 0.4 mL/min flow rate. The solvents used: A (100% CH_3CN + 0.1% formic acid), B (100% H_2O + 0.1% formic acid) on a solvent gradient system from 0 to 100% CH_3CN in 25 min and flushed with 100% B for another 5 min, followed by a 5 min equilibration time before the next injection.

Data Analysis and Database Searching. Each LCMS data set (RAW file format) was loaded into ACD/Spectrus DB Enterprise (Figure S23), and a query was created through ACD/IntelliXtract (IX.mcr) using optimized data processing parameters. This was followed by screening of the matrix database (MatrixDB) to identify matrix contaminants, followed by screening of the 24,618 compounds in PharmaSeaDB, and subsequent calculations performed to generate metrics output files.

Sample Collection. A collection of 8 marine sponges and 6 tunicates was collected from the west coast of Viti Levu, Fiji Islands (-17.839 S, 177.199 E) in March of 2009 (Table S38). Subsamples for DNA analysis were prepared and stored in RNAlater stabilization solutions.[76] Samples were stored in plastic containers, preserved in 100% MeOH and shipped to the University of Aberdeen where they were stored at 4 °C before processing for analysis. Taxonomic identifications of Tavarua8 (Tava8) and Tavarua2 (Tava2) were carried out using morphological and molecular DNA approaches (Supplementary Information C). Sequences for Tava8 and Tava2 have been deposited at the European Nucleotide Archive (ENA)[77] with AC numbers LR136919-20, LR136924, and LR136942 (Tables: S64, S65).

Extraction and Isolation. Samples were extracted with MeOH (3x) followed by CH_2Cl_2 (3x), placed in glass vials and dried under the flow of nitrogen gas at 39 °C using a Microlab Aarhus A/S Supertherm mini-oven evaporator.[78] The dried samples were then fractionated into four fractions based on polarity using C18 solid phase extraction (SPE) cartridges. After column conditioning as recommended by Phenomenex[74] each sample was loaded to the SPE cartridge and then flushed with 100% H_2O to remove salts and highly polar compounds.[46] This was

followed by flushing with 25% MeOH in H_2O (SPE-25), followed by 50% MeOH/H_2O (SPE-50), then by 100% MeOH (SPE-100). Finally, the column was flushed with 100% MeOH containing 0.05% TFA (SPE-100 + TFA). All fractions were dried as before and stored at 4 °C awaiting analysis.

Case Study 1. The fraction SPE-100% MeOH (Tavarua-8) was purified using an ACE 5 C18 HL, 250 x 100 mm column[79] and a solvent gradient system from 95% H_2O in MeOH to 100% MeOH from 0-30 min at a flow rate of 2.0 mL/min yielding the new compounds: **1** (1.4 mg), **2 (**0.9 mg).

Case Study 2. The SPE-50 fraction of Tavarua-2 was purified using an ACE 5 C18 HL, 250 x 100 mm HPLC column and a solvent gradient system from 0 to 100% CH_3CN containing 0.05% TFA in 30 min at a flow rate of 1.5 mL/min to yield **4** (1.2 mg) and **5** (0.2 mg).

*Eudistomin Z1 (**1**):* Light brownish oil; UV (MeOH) λ_{max} 400, 370, 300, 250 nm (Figure S50); IR (MeOH) ν_{max} 3300, 2870, 1580, 1430, 1155, 1310, 1150 cm^{-1} ; ^1H and ^{13}C NMR data (CD_3OD, 600 and 150 MHz, respectively), Table 1; HRESIMS *m/z* 407.9702 $[M+H]^+$ (calculated for $C_{16}H_{15}Br_2N_3$, 407.9711, Δ = 0.8 ppm) (Figure S15).

*Eudistomin Z2 (**2**):* Light brownish oil; UV (MeOH) λ_{max} 300, 250 nm (Figure S51); IR (MeOH) ν_{max} 3400, 2900, 1600, 1535, 1470, 1295, 1090 cm^{-1}; ^1H and ^{13}C NMR data (CD_3OD, 600 and 150 MHz, respectively), Table 1; HRESIMS *m/z* 426.0174, $[M+H]^+$ (calculated for $C_{17}H_{21}Br_2N_3$, 426.0181, Δ = 0.2 ppm) (Figure S21).

*Tavarua deoxyriboside A (**4**)*: Colourless oil; $[\alpha]_D^{20}$ -110° (c 0.11 g/100 mL, MeOH); UV (MeOH-H_2O) λ_{max} 216, 260, 300 (Figure S46); ECD (0.0004 M, MeOH) λ ($\Delta\varepsilon$) 220 (-0.8), 230 (-0.2), 243 (-1.2), 257 (1.6), 285 (-1.5) nm (Figure S49); IR (MeOH) ν_{max} 3257, 2925, 2854, 1686, 1608, 1516, 1441, 1374, 1354,1269, 1203, 1165, 1099, 1052, 1099, 853, 772 (Figure S48); ^1H and ^{13}C NMR data (CD_3OD, 600 and 150 MHz, respectively), see Table 2; HRESIMS *m/z* 522.1508 $[M+H]^+$ (calculated for $C_{26}H_{24}N_3O_9$, 522.1507, Δ = 0.47 ppm) (Figure S5).

*Tavarua deoxyriboside B (**5**)*: Colourless oil; UV (MeOH)/H_20 λ_{max} 216, 260, 351 (Figure S47). ^1H data (CD_3OD, 600 MHz) δ 7.93 (2H, d, *J* = 8.8 Hz, H3’’’/H7’’’), 7.93 (1H, s, H-10’’), 7.70 (1H, d, *J* = 8.1 Hz, H-6), 7.50 (1H, d, *J* =2.3 Hz, H-4’’), 7.29 (1H, d, *J* = 8.8 Hz, H-7’’), 6.84 (1H, d, *J* = 8.8 Hz, H-4’’’/H6’’’), 6.78 (1H, dd, *J* = 8.5, 2.5 Hz, H-6’’), 6.34 (1H, dd, *J* = 8.1,

6.3 Hz, H-1'), 5.66 (1H, bm, H-3'), 5.45 (1H, d, *J* = 8.1 Hz, H-5), 4.77 (1H, dd, *J* = 12.1, 3.7, H-5a'), 4.64 (1H, dd, *J* = 12.1, 3.7, H5b'), 4.52 (1H, overlap, H-4'), 2.64 (1H, m, H -2a'), 2.49 (1H, m, H-2b'), (Figure S10); HRESIMS *m/z* 508.1355 $[M+H]^+$ (calculated for $C_{25}H_{22}N_3O_9$, 508.1351, Δ = -0.82 ppm) (Figure S9).

ASSOCIATED CONTENT

Supporting Information

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

UV, ECD, IR, ^1H, ^{13}C, COSY, ROESY, HSQC, HMBC, HRESIMS ES^+ fragmentation data, metric data.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel: +44 1772 893489. E-mail: jtabudravu@uclan.ac.uk.; Tel: +44 1224 272895. E-mail: m.jaspars@abdn.ac.uk.

ORCID

Jioji N. Tabudravu: 0000-0002-6930-6572

Notes

The authors declare the following competing financial interest(s): G.C., E.J.M. are full-time employees at ACD/Labs. D.H. and D.B. are former employees of ACD/Labs. The other authors declare no competing financial interest.

ACKNOWLEDGMENTS

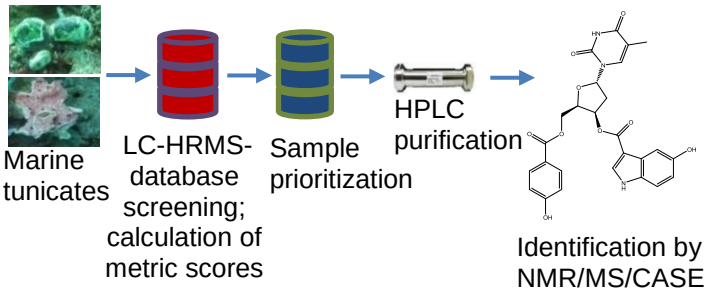
The research leading to these results received funding from the European Union's Seventh Framework Programme (FP7/2007-2013 under grant agreement no. 312184 "PharmaSea" to M.J., R.E., J.T., H.D., R.K. and G.C. J.T. wishes to thank V. Paget and the ACD/Labs Software Development Team for software assistance and G. McGibbon of ACD/Labs for constructive discussions. M.S., C.G, F.M. wish to thank Francesco Mastrototaro, Department of Biology - LRU CoNISMa, University of Bari, Via Orabona 4, 70125, Bari, Italy for help with ascidian identification. J.T. and M.J. wish to thank R. Gray of the Marine Biodiscovery Centre, University of Aberdeen for 2D NMR spectroscopic data.

REFERENCES

(1) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016**, *79*, 629–661.
(2) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2007**, *70*, 461–477.
(3) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2004**, *67*, 1216–1238.
(4) Montaser, R.; Luesch, H. *Future Med. Chem.* **2011**, *3*, 1475–1489.
(5) Blunt, J. W.; Copp, B. R.; Keyzers, R. A.; Munro, M. H. G.; Prinsep, M. R. *Nat Prod Rep* **2015**, *32*, 116–211.
(6) Blunt, J. W.; Copp, B. R.; Keyzers, R. A.; Munro, M. H. G.; Prinsep, M. R. *Nat Prod Rep* **2016**, *33*, 382–431.
(7) Tedesco, P.; Maida, I.; Palma Esposito, F.; Tortorella, E.; Subko, K.; Ezeofor, C.; Zhang, Y.; Tabudravu, J.; Jaspars, M.; Fani, R.; et al. *Mar. Drugs* **2016**, *14*, 83.
(8) Walsh, C. T.; Fischbach, M. A. *J. Am. Chem. Soc.* **2010**, *132*, 2469–2493.
(9) Rosén, J.; Gottfries, J.; Muresan, S.; Backlund, A.; Oprea, T. I. *J. Med. Chem.* **2009**, *52*, 1953–1962.
(10) Larsson, J.; Gottfries, J.; Bohlin, L.; Backlund, A. *J. Nat. Prod.* **2005**, *68*, 985–991.
(11) Harvey, A. L. *Curr. Opin. Chem. Biol.* **2007**, *11*, 480–484.
(12) Kennedy, J. P.; Williams, L.; Bridges, T. M.; Daniels, R. N.; Weaver, D.; Lindsley, C. W. *J. Comb. Chem.* **2008**, *10*, 345–354.
(13) Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. *Phytochem. Rev.* **2017**, *16*, 55–95.
(14) Ito, T.; Odake, T.; Katoh, H.; Yamaguchi, Y.; Aoki, M. *J. Nat. Prod.* **2011**, *74*, 983–988.
(15) Ito, T.; Masubuchi, M. *J. Antibiot. (Tokyo)* **2014**, *67*, 353–360.
(16) Gaudêncio, S. P.; Pereira, F. *Nat Prod Rep* **2015**, *32*, 779–810.
(17) Emerson, D.; Agulto, L.; Liu, H.; Liu, L. *BioScience* **2008**, *58*, 925.
(18) Ross, A. C.; Gulland, L. E. S.; Dorrestein, P. C.; Moore, B. S. *ACS Synth. Biol.* **2015**, *4*, 414–420.
(19) Tawfike, A.; Tawfik, N.; Edrada-Ebel, R. *Planta Med.* **2015**, *81*.
(20) Hou, Y.; Braun, D. R.; Michel, C. R.; Klassen, J. L.; Adnani, N.; Wyche, T. P.; Bugni, T. S. *Anal. Chem.* **2012**, *84*, 4277–4283.
(21) Freel, K. C.; Nam, S.-J.; Fenical, W.; Jensen, P. R. *Appl. Environ. Microbiol.* **2011**, *77*, 7261–7270.
(22) Derewacz, D. K.; Covington, B. C.; McLean, J. A.; Bachmann, B. O. *ACS Chem. Biol.* **2015**, *10*, 1998–2006.

(23) Nielsen, K. F.; Månsson, M.; Rank, C.; Frisvad, J. C.; Larsen, T. O. *J. Nat. Prod.* **2011**, *74*, 2338–2348.
(24) Klitgaard, A.; Iversen, A.; Andersen, M. R.; Larsen, T. O.; Frisvad, J. C.; Nielsen, K. F. *Anal. Bioanal. Chem.* **2014**, *406*, 1933–1943.
(25) Cremin, P. A.; Zeng, L. *Anal. Chem.* **2002**, *74*, 5492–5500.
(26) Hubert, J.; Nuzillard, J.-M.; Purson, S.; Hamzaoui, M.; Borie, N.; Reynaud, R.; Renault, J.-H. *Anal. Chem.* **2014**, *86*, 2955–2962.
(27) Chervin, J.; Stierhof, M.; Tong, M. H.; Peace, D.; Hansen, K. Ø.; Urgast, D. S.; Andersen, J. H.; Yu, Y.; Ebel, R.; Kyeremeh, K.; et al. *J. Nat. Prod.* **2017**, *80*, 1370–1377.
(28) Bakiri, A.; Hubert, J.; Reynaud, R.; Lanthony, S.; Harakat, D.; Renault, J.-H.; Nuzillard, J.-M. *J. Nat. Prod.* **2017**, *80*, 1387–1396.
(29) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. *Anal. Chem.* **2014**, *86*, 10780–10788.
(30) El-Elimat, T.; Figueroa, M.; Ehrmann, B. M.; Cech, N. B.; Pearce, C. J.; Oberlies, N. H. High- *J. Nat. Prod.* **2013**, *76*, 1709–1716.
(31) Pierens, G. K.; Mobli, M.; Vegh, V. *Anal. Chem.* **2009**, *81*, 9329–9335.
(32) Rateb, M. E.; Tabudravu, J.; Ebel, R. NMR Characterisation of Natural Products Derived from Under-Explored Microorganisms. In *Nuclear Magnetic Resonance*; Ramesh, V., Ed.; Royal Society of Chemistry: Cambridge, 2016; Vol. 45, pp 240–268.
(33) Hur, M.; Campbell, A. A.; Almeida-de-Macedo, M.; Li, L.; Ransom, N.; Jose, A.; Crispin, M.; Nikolau, B. J.; Wurtele, E. S. *Nat. Prod. Rep.* **2013**, *30*, 565.
(34) ACD/Labs.com :: Your Partner in Chemistry Software for Analytical and Chemical Knowledge Management, Chemical Nomenclature, and In-Silico PhysChem and ADME-Tox http://www.acdlabs.com/ (accessed Nov 15, 2016).
(35) Sugimoto, M.; Kawakami, M.; Robert, M.; Soga, T.; Tomita, M. *Curr. Bioinforma.* **2012**, *7*, 96–108.
(36) Katajamaa, M.; Orešič, M. *J. Chromatogr. A* **2007**, *1158*, 318–328..
(37) MZmine 2 http://mzmine.github.io/ (accessed Jun 8, 2017).
(38) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*, 234.
(39) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
(40) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035–5039.
(41) Clasquin, M. F.; Melamud, E.; Rabinowitz, J. D. LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine. In *Current Protocols in Bioinformatics*; Baxevanis, A. D., Petsko, G. A., Stein, L. D., Stormo, G. D., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2012.
(42) MarinLit - A database of the marine natural products literature http://pubs.rsc.org/marinlit/ (accessed Jun 3, 2017).
(43) Ausloos, P.; Clifton, C. L.; Lias, S. G.; Mikaya, A. I.; Stein, S. E.; Tchekhovskoi, D. V.; Sparkman, O. D.; Zaikin, V.; Zhu, D. T *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 287–299.
(44) Brown, M.; Dunn, W. B.; Dobson, P.; Patel, Y.; Winder, C. L.; Francis-McIntyre, S.; Begley, P.; Carroll, K.; Broadhurst, D.; Tseng, A.; et al. *The Analyst* **2009**, *134*, 1322.
(45) Panuwet, P.; Hunter, R. E.; D'Souza, P. E.; Chen, X.; Radford, S. A.; Cohen, J. R.; Marder, M. E.; Kartavenka, K.; Ryan, P. B.; Barr, D. B. *Crit. Rev. Anal. Chem.* **2016**, *46*, 93–105.
(46) Månsson, M.; Phipps, R. K.; Gram, L.; Munro, M. H. G.; Larsen, T. O.; Nielsen, K. F. *J. Nat. Prod.* **2010**, *73*, 1126–1132.
(47) Brodsky, L.; Moussaieff, A.; Shahaf, N.; Aharoni, A.; Rogachev, I. *Anal. Chem.* **2010**, *82*, 9177–9187.
(48) Gürdeniz, G.; Kristensen, M.; Skov, T.; Dragsted, L. O. *Metabolites* **2012**, *2*, 77–99.
(49) Kind, T.; Fiehn, O. A *Bioanal. Rev.* **2010**, *2*, 23–60.
(50) Xu, Y.-F.; Lu, W.; Rabinowitz, J. D. *Anal. Chem.* **2015**, *87*, 2273–2281.
(51) Fred W. McLafferty. Interpretation of mass spectra, third edition. University science books, Mill valley, California, 1980. pp. xvii + 303 - White V - 1982 - Biological Mass Spectrometry -

Wiley Online Library http://onlinelibrary.wiley.com/doi/10.1002/bms.1200090610/abstract (accessed Nov 16, 2016).
(52) Pretsch, E.; Bühlmann, P.; Badertscher, M. *Structure Determination of Organic Compounds: Tables of Spectral Data*, 4th ed.; Springer-Verlag: Berlin Heidelberg, 2009.
(53) Davis, R. A.; Carroll, A. R.; Quinn, R. J. *J. Nat. Prod.* **1998**, *61*, 959–960.
(54) Elyashberg, M.; Williams, A. J.; Blinov, K. *Nat. Prod. Rep.* **2010**, *27*, 1296.
(55) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
(56) Martin, G. E.; Crouch, R. C. I *J. Nat. Prod.* **1991**, *54*, 1–70.
(57) Li, K.; Li, Q.-L.; Ji, N.-Y.; Liu, B.; Zhang, W.; Cao, X.-P. *Mar. Drugs* **2011**, *9*, 690–695.
(58) Stierhof, M.; Hansen, K. Ø.; Sharma, M.; Feussner, K.; Subko, K.; Díaz-Rullo, F. F.; Isaksson, J.; Pérez-Victoria, I.; Clarke, D.; Hansen, E.; et al. *Tetrahedron* **2016**, *72*, 6929–6934.
(59) Searle, P. A.; Molinski, T. F. *J. Nat. Prod.* **1994**, *57*, 1452–1454.
(60) Kondo, K.; Shigemori, H.; Ishibashi, M.; Kobayashi, J. *Tetrahedron* **1992**, *48*, 7145–7148.
(61) Bremser, W. *Anal. Chim. Acta* **1978**, *103*, 355–365.
(62) Corporation, C. ChemBio3D Ultra 14 Suite https://www.cambridgesoft.com/Ensemble_for_Chemistry/details/Default.aspx?fid=13&pid=668 (accessed Jan 7, 2019).
(63) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminformatics* **2012**, *4*, 17.
(64) Halgren, T. A. I. *J. Comput. Chem.* **1996**, *17*, 490–519.
(65) Hu, J. F.; Wunderlich, D.; Sattler, I.; Härtl, A.; Papastavrou, I.; Grond, S.; Grabley, S.; Feng, X. Z.; Thiericke, R. *J. Antibiot. (Tokyo)* **2000**, *53*, 944–953.
(66) Wagner-Dobler, I. Oceanibulbus Indolifex Gen. Nov., Sp. Nov., a North Sea Alphaproteobacterium That Produces Bioactive Metabolites. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 1177–1184.
(67) Chen, M.; Fu, X.-M.; Kong, C.-J.; Wang, C.-Y. *Nat. Prod. Res.* **2014**, *28*, 895–900.
(68) Kehraus, S.; Gorzalka, S.; Hallmen, C.; Iqbal, J.; Müller, C. E.; Wright, A. D.; Wiese, M.; König, G. M. *J. Med. Chem.* **2004**, *47*, 2243–2255.
(69) Hoang, T. P. T.; Roullier, C.; Boumard, M.-C.; Robiou du Pont, T.; Nazih, H.; Gallard, J.-F.; Pouchus, Y. F.; Beniddir, M. A.; Grovel, O. *J. Nat. Prod.* **2018**, *81*, 2501–2511.
(70) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; et al. *J. Nat. Prod.* **2013**, *76*, 1686–1699.
(71) Chanana, S.; Thomas, C.; Braun, D.; Hou, Y.; Wyche, T.; Bugni, T. *Metabolites* **2017**, *7*, 34.
(72) Jasco UK https://www.jasco.co.uk/ (accessed Dec 14, 2018).
(73) Mass Spectrometry https://www.thermofisher.com/uk/en/home/industrial/mass-spectrometry.html (accessed Jan 10, 2017).
(74) Solid Phase Extraction (SPE) Method Development Tool from Phenomenex https://www.phenomenex.com/Tools/SPEMethodDevelopment (accessed Jan 10, 2017).
(75) Agilent | LC Columns http://www.agilent.com/en-us/products/liquid-chromatography/lc-columns#0 (accessed Jan 10, 2017).
(76) Thermo Fisher Scientific - UK https://www.thermofisher.com/uk/en/home.html (accessed Dec 15, 2018).
(77) European Nucleotide Archive < EMBL-EBI https://www.ebi.ac.uk/ena (accessed Jan 10, 2019).
(78) Mikrolab Aarhus A/S Laboratorieudstyr http://www.mikrolab.dk/ (accessed Jan 10, 2017).
(79) Hichrom: the HichrOM(e) of chromatography http://www.hichrom.com/index.htm (accessed Jan 11, 2019).


Marine tunicates
LC-HRMS-database screening; calculation of metric scores
Sample prioritization
HPLC purification
Identification by NMR/MS/CASE