

Electronic Supplementary Information

Chemical profiling of two congeneric sea mat corals along the Brazilian coast: adaptive and functional patterns

L. V. Costa-Lotufo,^{a*} F. Carnevale-Neto,^b A. E. Trindade-Silva,^c R. R. Silva,^b G. G.
Z. Silva,^d D. V. Wilke,^c F. C. L. Pinto,^e B. D. Sahm,^a P. C. Jimenez,^f J. N. Mendonça,^b
T M. C. Lotufo,^g O. D. L. Pessoa,^e and N. P. Lopes^{b*}

* Corresponding author

L. V. Costa-Lotufo; costalotufo@gmail.com, N. P. Lopes; npelopes@fcfrp.usp.br

Material and Methods

Sampling

Sampling was conducted at Paracuru (“Pedra Rachada”, 3°23’54.94”S, 39°0’50.76”O) and Taíba (3°30’18.90”S, 38°53’32.43”O) beaches in Ceará State (CE), at Arvoredo Island (Rancho Norte’s” region 27°16’44.08”S, 48°22’29.79”O, Santa Catarina State (SC), and Arraial do Cabo (“Abobrinha” 22°59’11.47”S, 41°59’32.35”O) and Búzios (“Tartaruga’s” beach 22°59’11.47”S, 41°54’5.39”O) in Rio de Janeiro State (RJ), between April and August 2013. Healthy zoanths *P. caribaeorum* and *P. variabilis* were hand collected in Ceará State and Búzios during low tide and by scuba dive at all other sites. In Arraial do Cabo, RJ, only *P. caribaeorum* was found. Three colonies of both species were randomly chosen and hand collected, each one completing a 15mL Falcon tube for extracts preparation for metabolomics analysis. For metagenomics, the collected samples were superficially sterilized *in situ* by spraying polyps with ethanol (70%), sequentially immersed in sterile seawater to remove ethanol excess, and immediately frozen in liquid nitrogen (N₂) for transportation to the laboratory. These samples were stored at –70°C until DNA purification procedures.

Chemical Extraction procedures

Fresh samples, ca 5.0 g, of *P. caribaeorum* and *P. variabilis*, were cut in to small pieces and subjected to extraction with MeOH (15 mL) under sonication for 15 min (three time). The MeOH extracts were partitioned with *n*-hexane to yield the corresponding fractions, which were dried on Na₂SO₄ and evaporated under reduced pressure (30 °C) to subsequent metabolomic analysis.

Metabolomic analysis

CG-MS

Gas chromatography/Mass spectrometry analyses were performed in a Shimadzu GC-MS QP2010. DB5MS column (30 m x 0.25 mm x 0.25 µm film) was used with an oven temperature program of 100 °C (hold for 3 min), then heated to 220 °C at 10 °C/min (hold for

10 min), and finally, heated to 300 °C at 15 °C/min (hold for 20 min). Injection temperature was set to 250 °C. Helium was used as a carrier gas with a linear velocity of 40.8 cm/s and a split of 1:5. The mass spectrometer operated in electron ionization (EI) mode at 70 eV. The interface and ion source temperatures were 290 °C and 250 °C, respectively. Mass spectra were acquired from 40 to 600 *m/z*.

LC-DAD-IT and LC-DAD-TOF

LC-DAD-IT and LC-DAD- TOF analysis were performed using an UFLC (Shimadzu, Japan) consisting of two LC20AD solvent pumps, a SIL20A_{HT} auto sampler, a CTO20A column oven, a CBM20A system controller and a diode array detector (SPD-M20AV, Shimadzu). LC-DAD-MS/MS was acquired using the UFLC apparatus coupled with an Ion Trap Mass Spectrometer (amaZon SL, Billa Rica – US), while LC-DAD- TOF was conducted by UFLC equipped with an UltrOTOF (Bruker Daltonics, Billa Rica – US) mass spectrometer. The mobile phase (flow 1.0 mL min⁻¹) consisted of water (A) and acetonitrile (B) in the following gradient: 0.0 – 30.0 min (10 - 100% B); 3.0 – 40.0 min (100% B); 40.0 – 45.0 min – (100 - 10% B); 45.0 – 60.0 min (10% B). A C₁₈ - Luna (Phenomonex® – 250 mm x 4.6 mm x 5 µm) column was used and temperature was adjusted to 35 °C. IT acquisition parameters were as follows: capillary 3.5 kV, end plate offset 500 V, nebulizer 40 psi, dry gas (N₂) with flow of 8 L/min and temperature of 300 °C. CID fragmentation experiments were performed in autoMSⁿ mode using Enhanced resolution mode for MS and UltraScan mode for MS/MS acquisition. Experimental details are given in Table S1.

The high resolution MS experiments were conducted as follows: capillary voltage of 30 V; capillary temperature at 200 °C; source voltage of 4.5 kV; source current of 80 µA; nitrogen was used as the sheath gas; drying gas flow at 5.0 L/min; drying gas temperature at 180 °C; nebulizer gas pressure of 4 bar; positive and negative ESI modes. Spectra (*m/z* 50–1000) were recorded every 2 s. Accurate masses were obtained by using TFA-Na⁺ (sodiated trifluoroacetic acid) as the internal standard.

MALDI

A MALDI-TOF/TOF mass spectrometer (ultrafleXtreme, Bruker-Daltonics, Bremen, DE) equipped with a 1kH smartbeam-II laser (Nd:YAG – 355nm) was used to acquire the mass spectra. The MALDI plates used were all of the type MTP 384 GroundSteel (BrukerDaltonics, Bremen, DE) and the instrument was operated in the positive reflector ion mode. Spectra were acquired in the mass range between m/z 1000-5000 with the following instrumental conditions: ion source 1 at 25.00 kV, ion source 2 at 22.55 kV, lens at 8.30 kV, reflector 1 at 26.6 kV, reflector 2 at 13.35 kV. The parameter that were held constant for all of the experiments included the pulsed ion extraction (PIE), which was set to 120 ns, the laser attenuation, with an offset of 57 %, a range of 15 %, and the focus set to large, and the number of shots, which was set to 400 shots at a laser frequency of 1000 Hz. A solution of 20 mg/mL of matrix substance DHB (2,5-dihydroxybenzoic acid) was prepared using 0,1% of trifluoroacetic acid and acetonitrile 7:3 (v/v). Matrix solution was then mixed with 2 μ L of the hydroalcoholic extract of sample at 1 :1 (v/v) and 2 μ L of this mixture were spotted on 2 sample spots on the MALDI plate. Instrument calibration was achieved using a mixture of peptides (peptide calibration standard – Bruker).

Identification of non-polar metabolites

Metabolic identification by GC-MS was performed by two independent parameters: Kovats Index (KI) and fragmentation pattern. The KI values were calculated from the linear regression between the retention times obtained experimentally with values from literature. The equation was based on retention times of a mixture of linear hydrocarbons (C_9H_{20} - $C_{40}H_{82}$) injected under the same experimental conditions as samples. The fragmentation patterns were obtained by chromatographic deconvolution performed by AMDIS software,²⁴ followed by comparison of the mass spectra with Wiley 7 and NIST 08 databases.^{25,26} The metabolic detection took into consideration only molecules that contained KI errors \leq 1.5% and fragmentation pattern with similarity \geq 70 (determined by AMDIS match factor - MF).

Identification of polar metabolites

Dereplication initiated with calculation of molecular formulae using TOF-MS data (accurate molecular weights < 5ppm). The resulting formulae were compared with different metabolic database (METLIN, MarinLit, DMNP), considering chemotaxonomic information, and the possible putative structures were proposed by MS/MS-based fragmentation pattern. In case of LC peaks, retention time (shift tolerance of ± 0.05 min) and UV spectra were used as orthogonal information on peak annotation. Additionally, the dereplication workflow were assisted by molecular networking, i.e., clustering of MS/MS spectra by cosine similarity.^{27,28}

Molecular Networking

Molecular networking workflows organized large data sets of tandem mass spectra based on the similarity between fragmentation patterns of different, but related, precursor ions.¹¹ To perform the molecular networking, the LC-DAD-IT MS/MS data were converted to mzXML format directly from Bruker DataAnalysis 4.2. The resulting files contained scan number, precursor m/z and the m/z each ion observed in both ionization modes. In order to perform the molecular network with distinct ionization modes (ESI positive or negative), each mzXML data were converted in two new mzXML files containing one or another ionization mode, by an in-house R-algorithm. Once all LC-IT files were in text format and split in mzXML ESI positive or negative modes, data were subjected to Spectral Networks, which includes MS-Cluster, followed by visualization in Cytoscape 2.8.3.²⁹

Briefly, MS/MS spectra were converted into unit vectors and compared by cosine similarity.¹¹ This comparison was conducted between pairs of spectra that have at least six ions that match, cosine scores over 0.8, and at least two nodes to be in the top 10 cosine scores (K parameter) in both directions for an edge to connect them in Cytoscape. To avoid clustering spectra from same LC-peak, MS-Cluster algorithm was applied and combined with repeatedly acquired spectra from the same molecules into cluster-consensus spectra with a higher signal-to-noise ratio, in this work, with similar parent masses, within 0.5 Da and cosine score higher than 0.95 for each pair.¹¹ Then, Spectral Networks compared all possible vector pairs from

consensus MS/MS spectra, considering mass tolerance for fragment peaks (0.5 Da), parent mass tolerance (1.0 Da), the minimum percentage of overlapping masses between two spectra (set at 45%), the minimum number of matched peaks per spectral alignment,¹¹ the minimum percentage of matched peaks in a spectral alignment (40%), and a minimum cosine score of 0.8. The higher the cosine score between two spectra, the more similar the MS/MS spectra and, by extension, the more similar the corresponding molecules.¹¹

After organizing the spectra based on fragmentation similarity, the data were imported into Cytoscape and displayed as a network of nodes and edges.¹¹ To avoid misinterpretation of LC-contaminants or noise, blank injections (mobile phase) were input to Spectral Networks as a distinct sample group and identified on Cytoscape as white-color nodes. The network was organized with the organic layout; node colors were mapped based on the source files of the MS/MS and the edge thickness attribute was defined to reflect cosine similarity scores, with thicker lines indicating higher similarity.¹¹

Statistical analyses of metabolomics measurements

Format conversion: The Shimadzu .qgd format was converted to NetCDF format with OPENChrom software. The Bruker .d format was converted to mzXML directly from Bruker DataAnalysis 4.2. The Bruker MALDI-MS raw data was exported from flexAnalysis as a text file after external calibration (see data acquisition section).

Data preprocessing: The LCMS iontrap data and GCMS data were preprocessed with XCMS R package.³⁰ The MALDI-MS text spectra were preprocessed with MALDIquant R package.³¹

Data processing: The peak tables obtained from each technique were autoscaled to equal variable importance. After sample scaling, datasets from each technique were scaled by total variation using the Frobenius norm to normalize their influence on the joint dataset.³² The metabolomic profiles acquired from multiple analytical sources (LCMS, GCMS and MALDI-MS) were analyzed by integrated unsupervised methods Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) with Manhattan distance and Ward grouping

method. The joint dataset produced by data fusion was further subject to Partial Least Square - Discriminant Analysis (PLS-DA) classification model in order to rank the individual ion importance to the classification model.³³

Isolation procedures for major metabolite validation

P. caribaeorum (2.0 kg) and *P. variabilis* (1.8 kg), both collected at Paracuru beach, Ceará State (“Pedra Rachada”, 3°23’54.94”S, 39°0’50.76”O), were extracted with MeOH (3 × 1000 mL) under sonication for 15 minutes. The MeOH suspensions were filtered and evaporated under reduced pressure to approximately 1/3 of the total volume and then partitioned with *n*-hexane and EtOAc (3x 200 mL of each solvent), from which were obtained the corresponding fractions: *P. caribaeorum* (*n*-hexane, 3.9 g; EtOAc, 0.5 g; hydroalcoholic, 60.0 g) and *P. variabilis* (*n*-hexane, 11.3 g; EtOAc, 0.2 g; hydroalcoholic, 40.5 g).

The *n*-hexane fraction from *P. caribaeorum* was subjected to a silica gel chromatography column and eluted with *n*-hexane, *n*-hexane:EtOAc 2:1 and 1:1, EtOAc, EtOAc:MeOH 2:1 and 1:1, followed by MeOH. The MeOH fractions were subjected to HPLC using a semi-preparative C₁₈ column, MeOH as solvent and a flow rate of 4.7 mL min⁻¹, from which the palyosulfonoceramides A (30.1 mg) and B (12.5 mg) were isolated.³⁴

The EtOAc fraction from *P. caribaeorum* was fractionated through a SPE cartridge using a gradient of MeOH/H₂O (2:8, 4:6, 6:4, 8:2 and 10:0) as eluent. Fractions H₂O/MeOH 6:4 (22.0 mg) and H₂O/MeOH 4:6 (14.0 mg) were further purified by HPLC using a semi-preparative C₁₈ column, a solvent gradient of H₂O (0.1% TFA)/CH₃CN (9.5:0.5 → 1:9.0), at a flow rate of 4.7 mL min⁻¹ to afford the pure following compounds: 20-hydroxyecdysone (4.6 mg, *t_R* 7.4 min), 3-*O*-acetyl-20-hydroxyecdysone (2.6 mg, *t_R* 11.7 min) and 2-*O*-acetyl-20-hydroxyecdysone (1.4 mg, *t_R* 13.3 min).³⁵

Applying the described procedures to the corresponding fractions obtained from *P. variabilis*, palyosulfonoceramides A (40.2 mg) and B (18.8 mg) were isolated from the *n*-hexane fraction, while 20-hydroxyecdysone (3.6 mg), a mixture of 3-*O*-acetyl-20-

hydroxyecdysone (0.5 mg), and zoanthusterone (1.2 mg, t_R 14.5 min) were isolated from EtOAc fraction.³⁴

Metagenomic DNA extraction

Metagenomic DNA purification was performed following a protocol previously developed specifically for coral tissues.³⁶ Briefly, 1 g of tissue was sliced from thawed samples (colonies' polyps) and macerated until powdered in liquid nitrogen (N_2) using crucible and pestle. Approximately 150 mg of powdered tissues was then transferred to 2 mL microtubes containing 1 ml of a *Cetyltrimethyl Ammonium Bromide* (CTAB) lysis buffer [2%(m/v) CTAB (Sigma Aldrich), 1.4 M NaCl, 20 mM EDTA, 100mM Tris-HCl (pH 8.0), with freshly added 5 μ g proteinase K (v/v; Invitrogen) and 1% 2-mercaptoethanol (Sigma Aldrich)] and submitted to five freeze-thawing cycles (-80 °C to 65 °C). Bulk DNA extraction proceeded with two phenol:chloroform:isoamyl alcohol (25:24:1) washes followed by one chloroform wash. DNA purification was obtained by precipitation with isopropanol and 5 M ammonium acetate, washing with 70% ethanol and elution with TE buffer (10 mM Tris-HCl e 1 mM de EDTA). Metagenomic DNA purity and integrity was respectively evaluated in a NANODROP spectrophotometer (Thermo Fisher Scientific Inc.), and with electrophoretic run in 1% agarose gel. Accurate quantifications of double strain DNA were then performed with Qubit 2.0 fluorometer (Thermo Fisher Scientific Inc.) following manufacturer's instructions and up to 4 ng of each sample was used for construction of the metagenomic libraries.

Metagenomes sequencing and annotation

Zoanthids shotgun metagenomic libraries were prepared with the *Nextera XT DNA Sample Preparation Kit* (Illumina) following the manufacturer's instructions. Libraries were sequenced by the MiSeq Desktop Sequencer using the 500-cycle (250 bp paired-end runs) *MiSeq Reagent Kits v2* chemistry (Illumina). Metagenomic mate FASTQ paired-end reads were merged and annotated in MG-RAST server (version 3.6)³⁷ using the defined filtering pipeline

to de-replicate, removed of human (*Homo sapiens*, NCBI v 36) sequences and dynamic trimming for quality (> 15 phred).

Automated taxonomic classifications were performed by MG-RAST considering Best Hit matches under default settings (cutoff *e*-value of $1e^{-5}$, minimal identity of 60% and minimal alignment lengths of 15aa), and using M5NR (Gen Bank) as the referential databases. For functional annotations, Real Times Metagenomics (RTMg)³⁸ was applied, since it uses a *k*-mer-based approach, which is known to provide more robust annotation than other tools based on homology.³⁹ RTMg annotations were performed at default settings and *k*-mers matches combined under two level hierarchies of the subsystems ontology.⁴⁰

Supplementary Figures

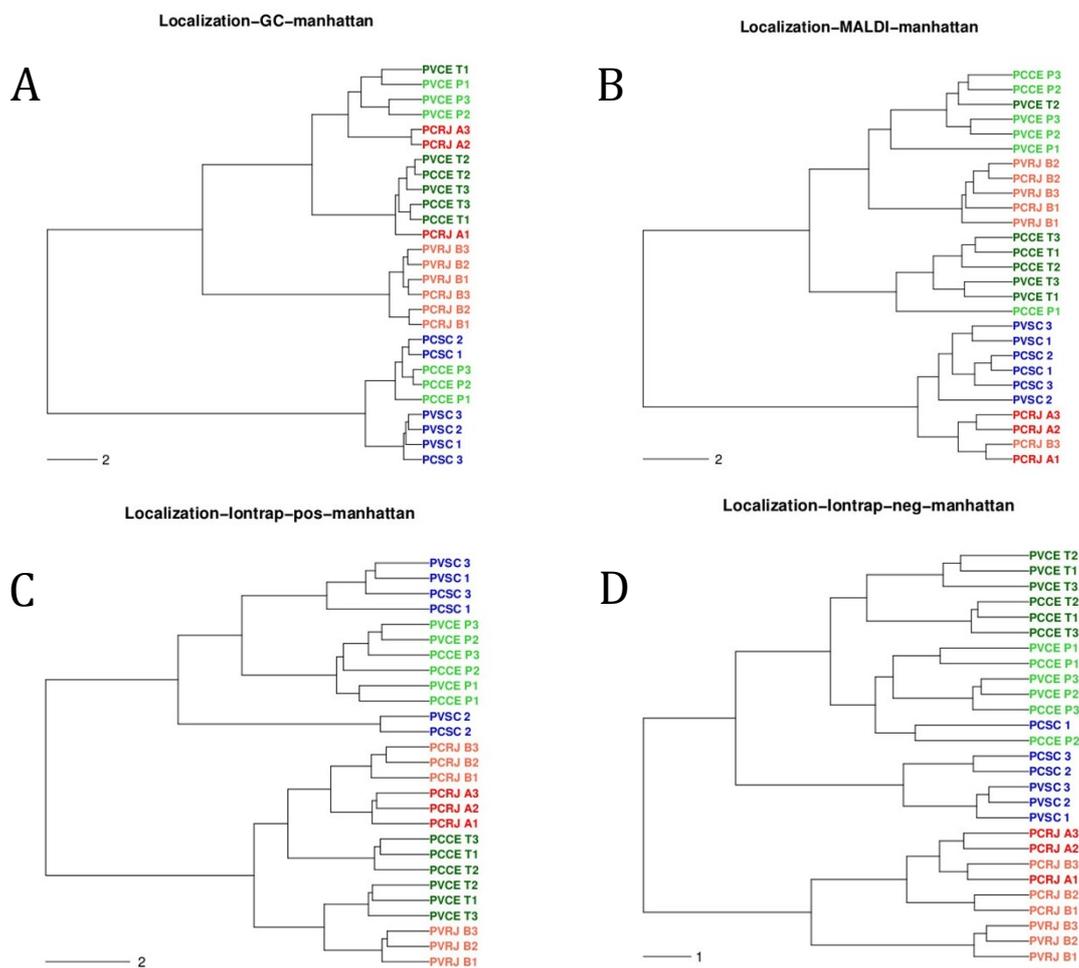


Fig. S1. Unsupervised hierarchical clusterization of *Palythoa* species considering the metabolite profiles acquired in A) GC-MS, B) MALDI-TOF MS, C) LC-MS (ESI+) and D) LC-MS (ESI-). Colors indicated sampling location: Red for Rio de Janeiro State, green for Ceará State and yellow for Santa Catarina State.

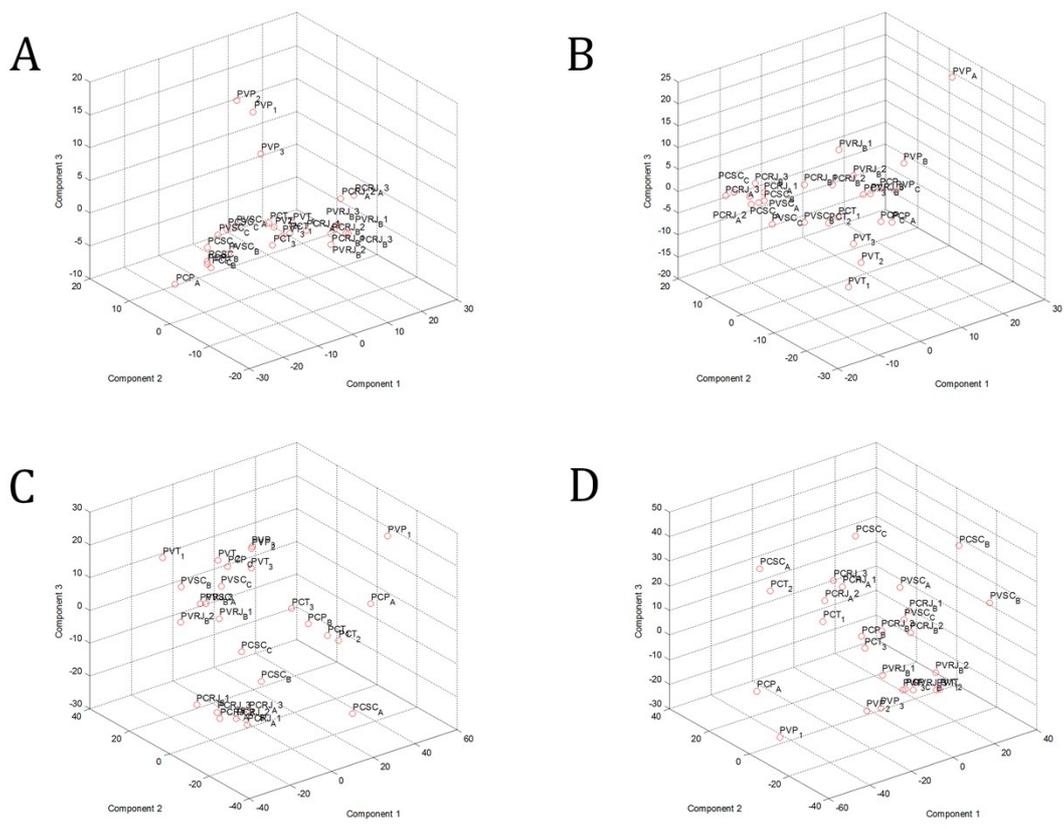


Fig. S2. Principal Component Analysis (PCA) PC1-PC3 score plot of *Palythoa* species considering the metabolite profiles acquired in A) GC-MS, B) MALDI-TOF MS, C) LC-MS (ESI +) and D) LC-MS (ESI -).

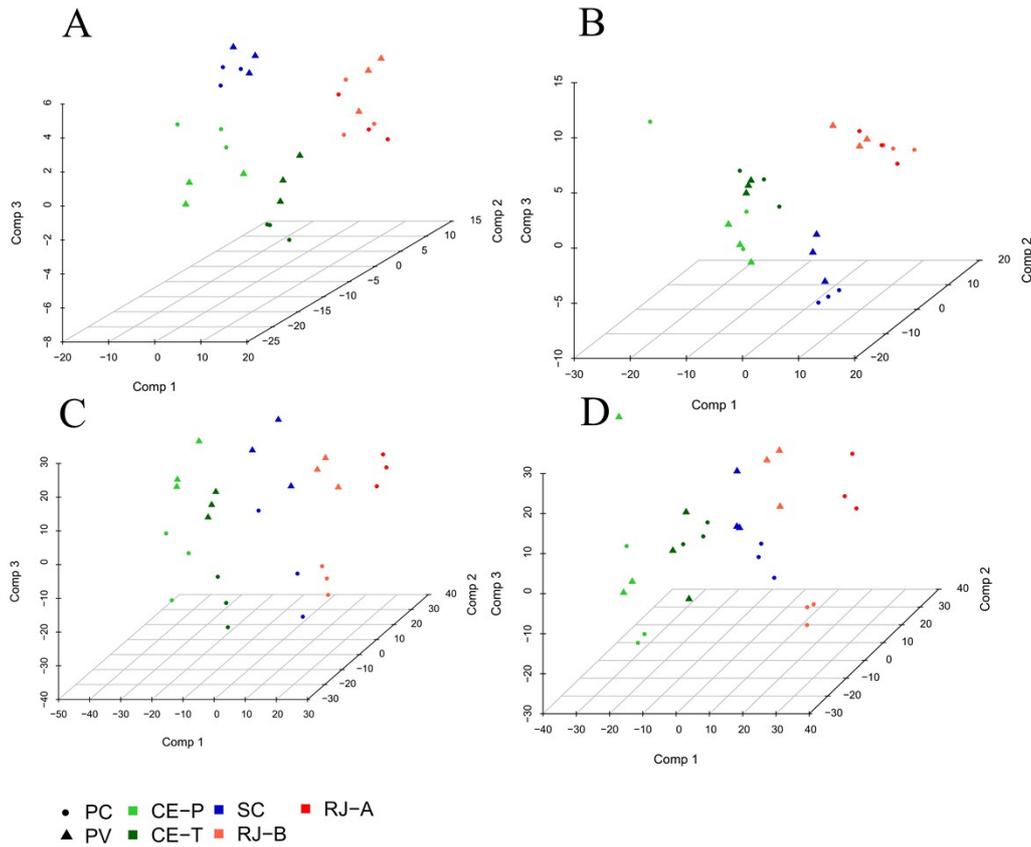


Fig. S3. Partial least square discriminant analysis (PLS-DA) LV1-LV3 score plot of *Palythoa* species considering the metabolite profiles acquired in A) GC-MS, B) MALDI-TOF MS, C) LC-MS (ESI +) and D) LC-MS (ESI -). PC = *Palythoa cariboreum*, PV = *P. variabilis*, CE-P = sample from Paracuru beach in Ceara State, CE-T = sample from Taiba beach in Ceara State, SC = sample from Arvoredo beach in Santa Catarina State, RJ-B = sample from Buzios beach in Rio de Janeiro State, RJ-A = sample from Arraial do Cabo beach in Rio de Janeiro State.

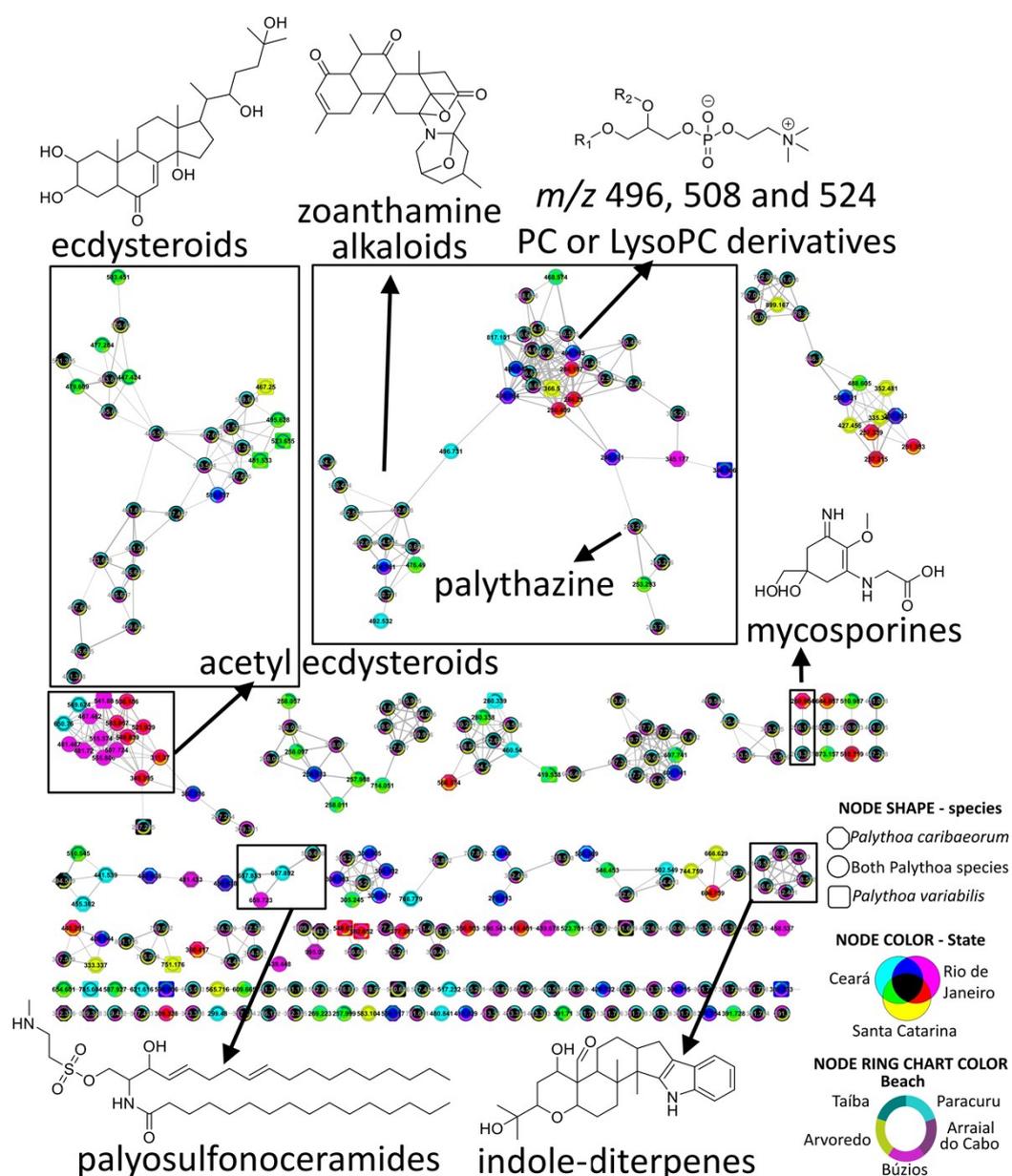


Fig. S4. *Palythoa* holobiont molecular network based on LC-MS/MS (ESI+) similarities patterns. Node color reflects the State where samples were collected. Node chart shows the numbers of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

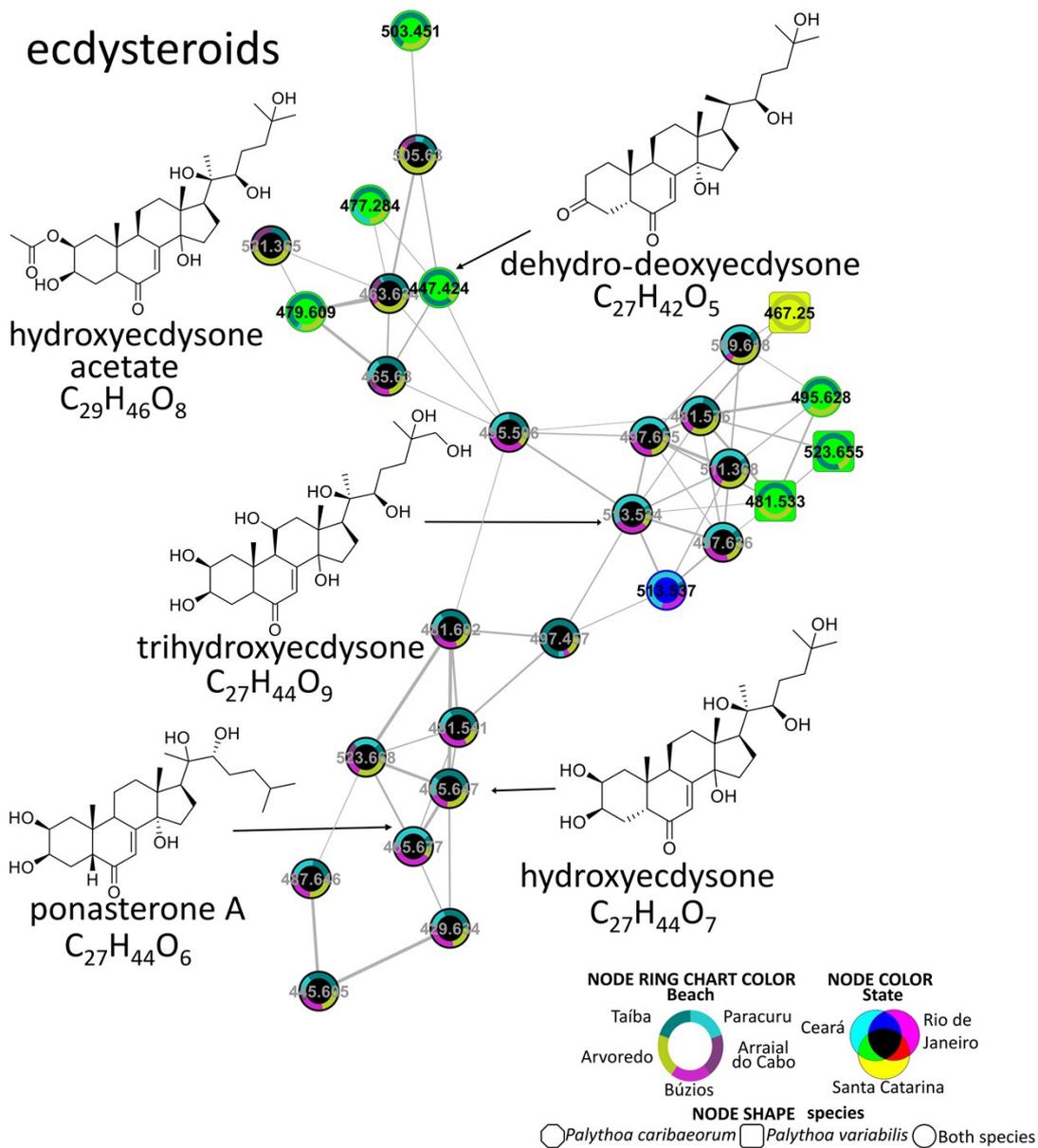


Fig. S5. Zoom of Fig. S4 showing a cluster of nodes identified as ecdysteroids. Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

acetyl ecdysteroids

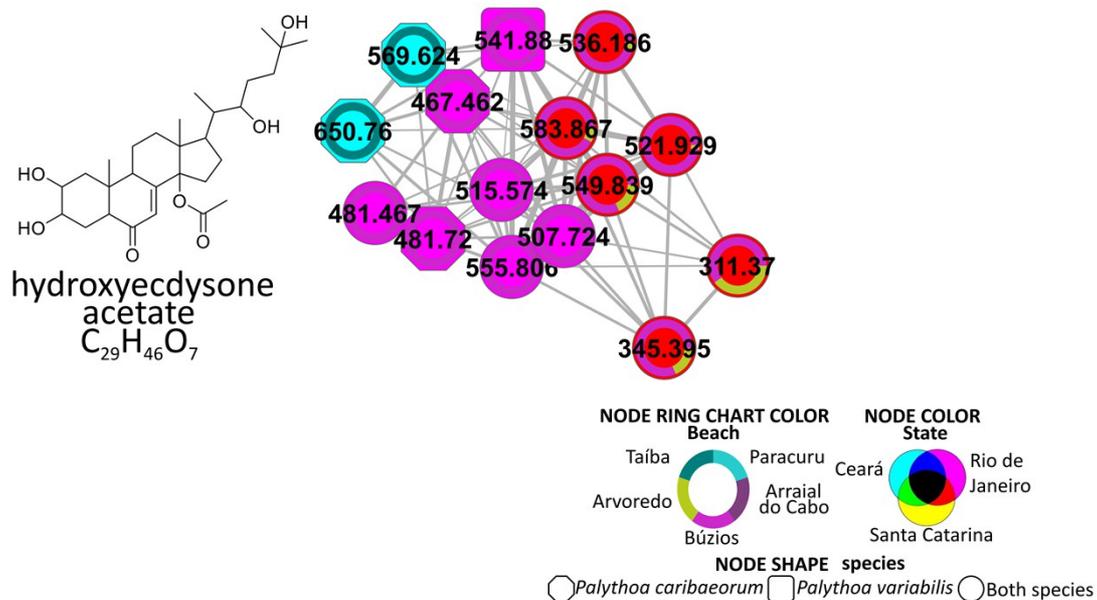


Fig. S6. Zoom of Fig. S4 showing a cluster of nodes composed with acetyl ecdysteroids.

Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

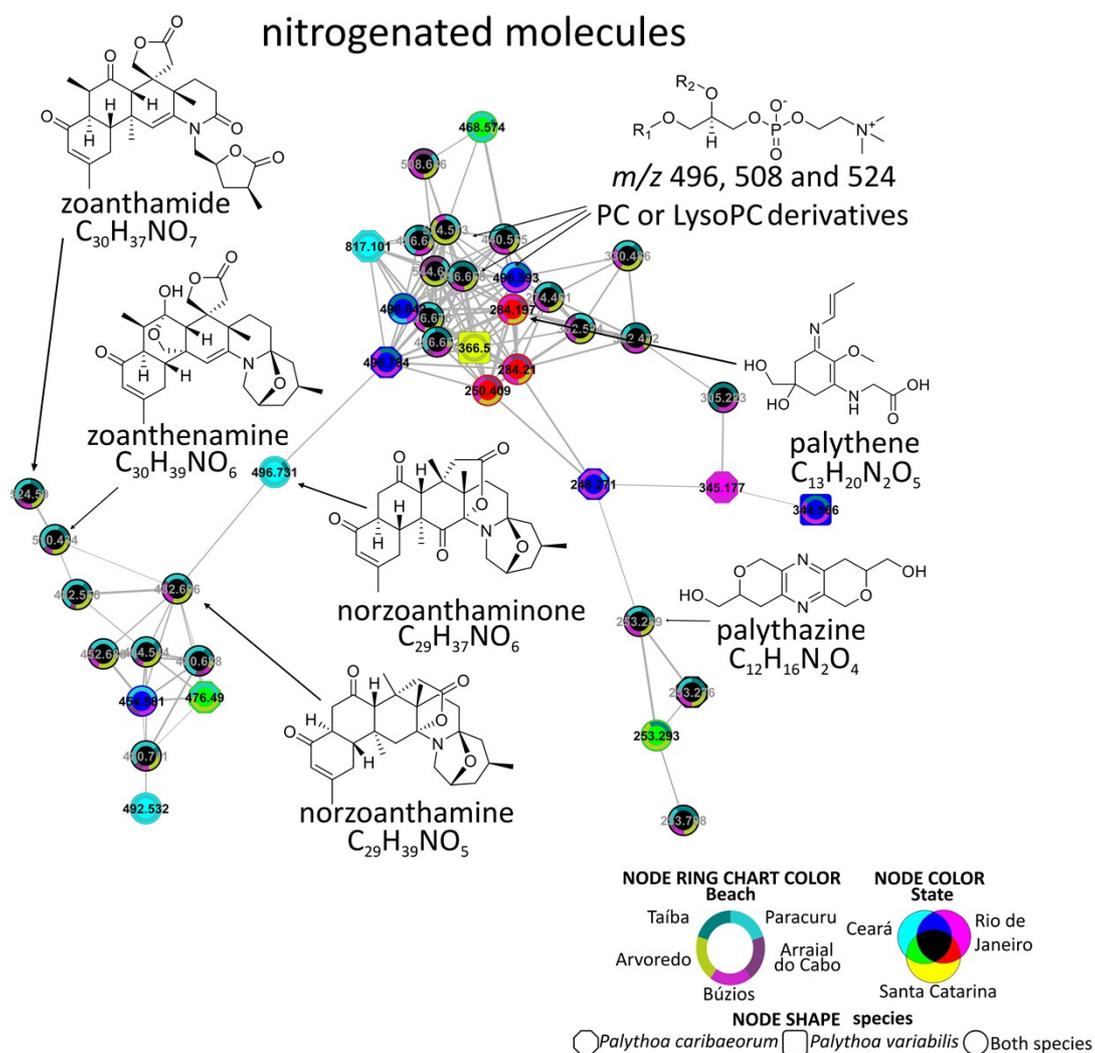


Fig. S7. Zoom of Fig. S4 showing a cluster of nodes composed with nitrogenated metabolites (zoanthamine alkaloids, phosphatidylcholine derivatives and pyrazines) Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

indole-diterpenes

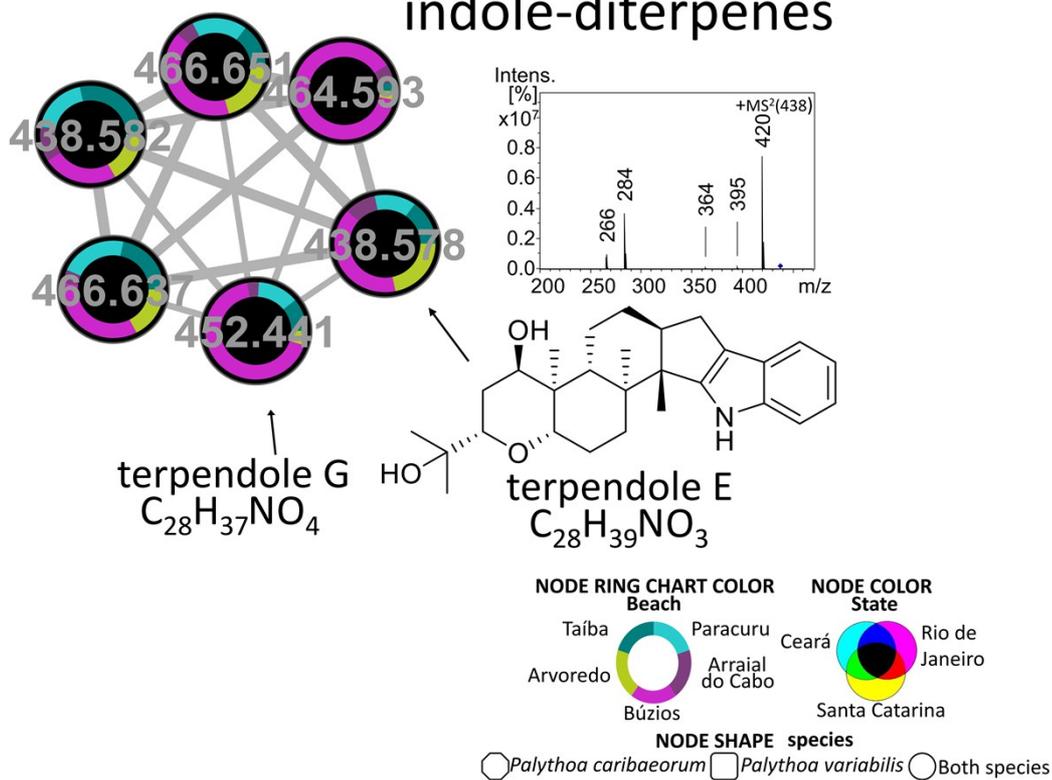


Fig. S8. Zoom of Fig. S4 showing a cluster of nodes identified as indole diterpene. Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

mycosporine

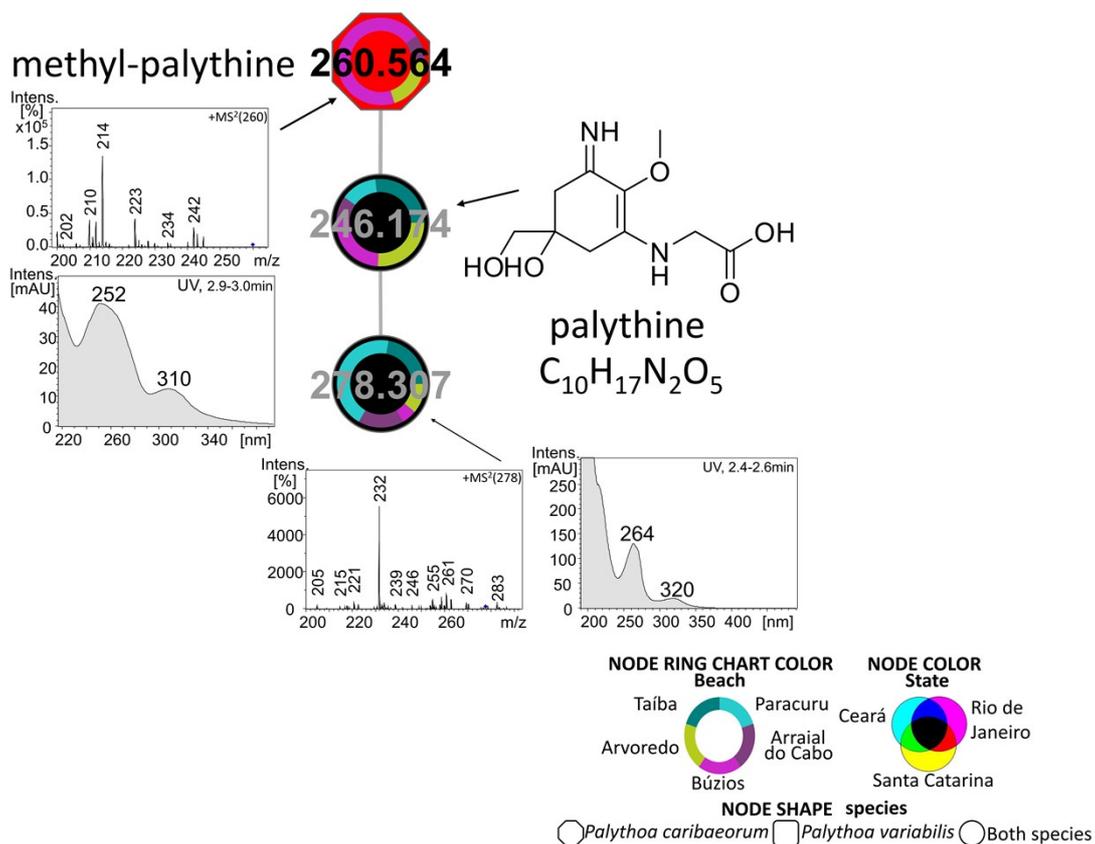


Fig. S9. Zoom of Fig. S4 showing a cluster of nodes identified as mycosporine. Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

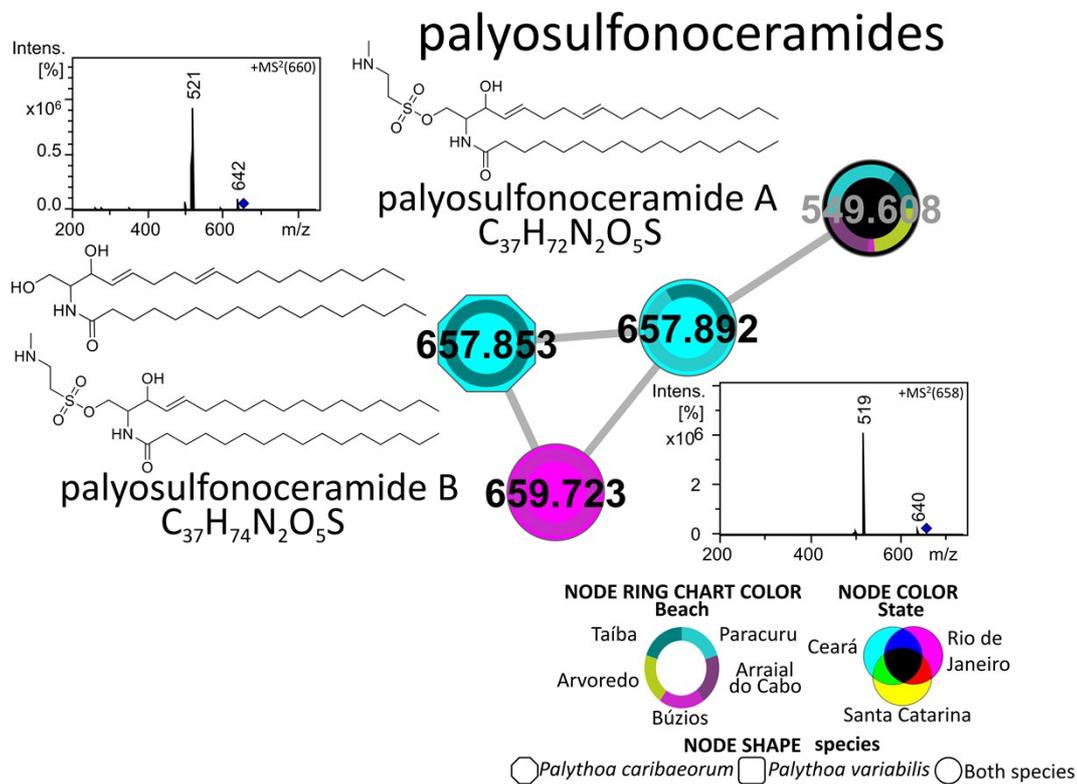


Fig. S10. Zoom of Fig. S4 showing a cluster of nodes identified as sulphoceramide. Node color reflects the State sample was collected, node chart shows the number of spectra for each beach sample was collected. Node shape represents the distribution of the parent ion among the two species.

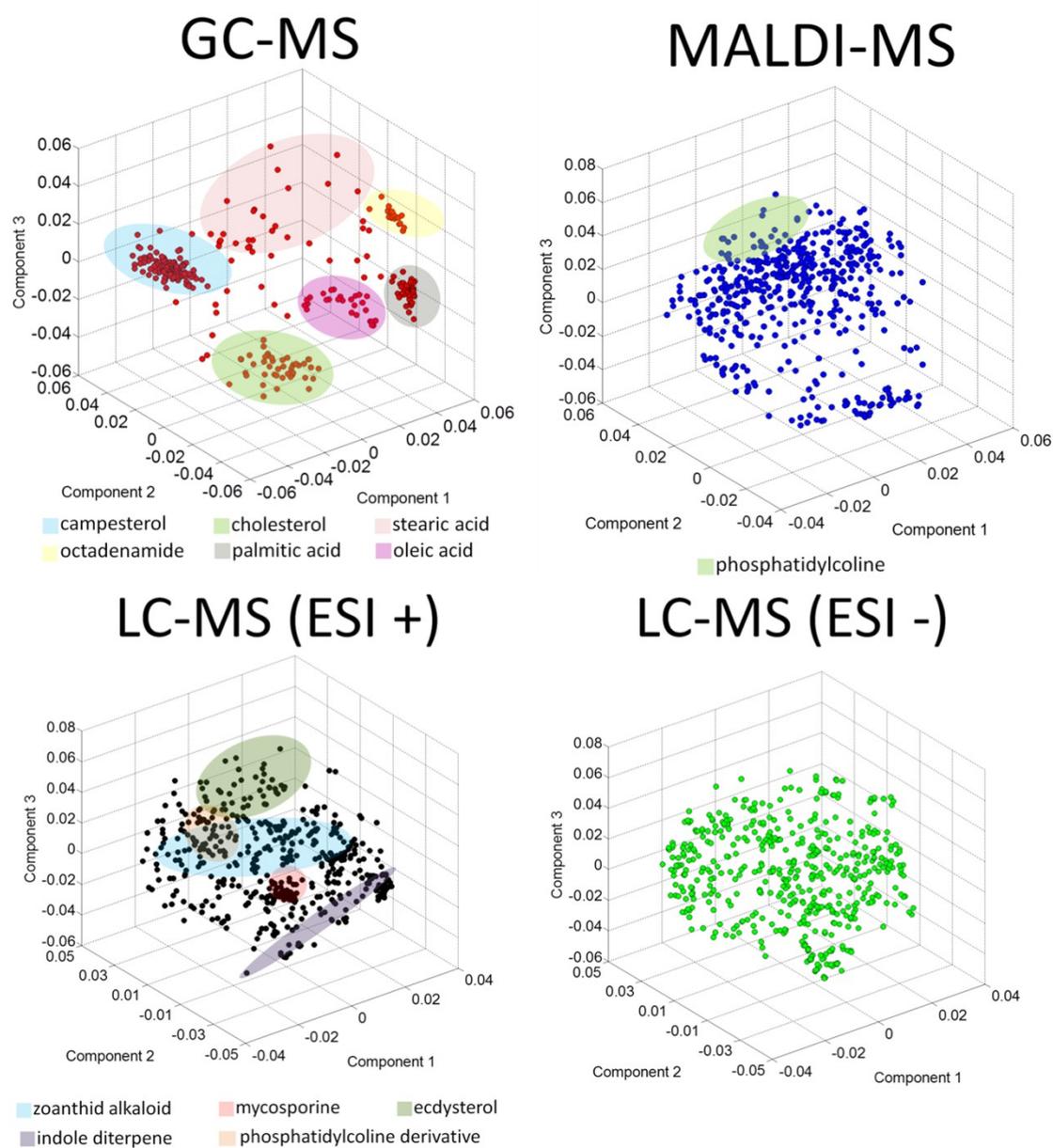


Fig. S11. Principal Component Analysis (PCA) PC1-PC3 loading plots representing metabolite profiles obtained by GC-MS, MALDI-TOF MS, ESI positive LC-MS and ESI negative LC-MS analyses showing parent ion distribution.

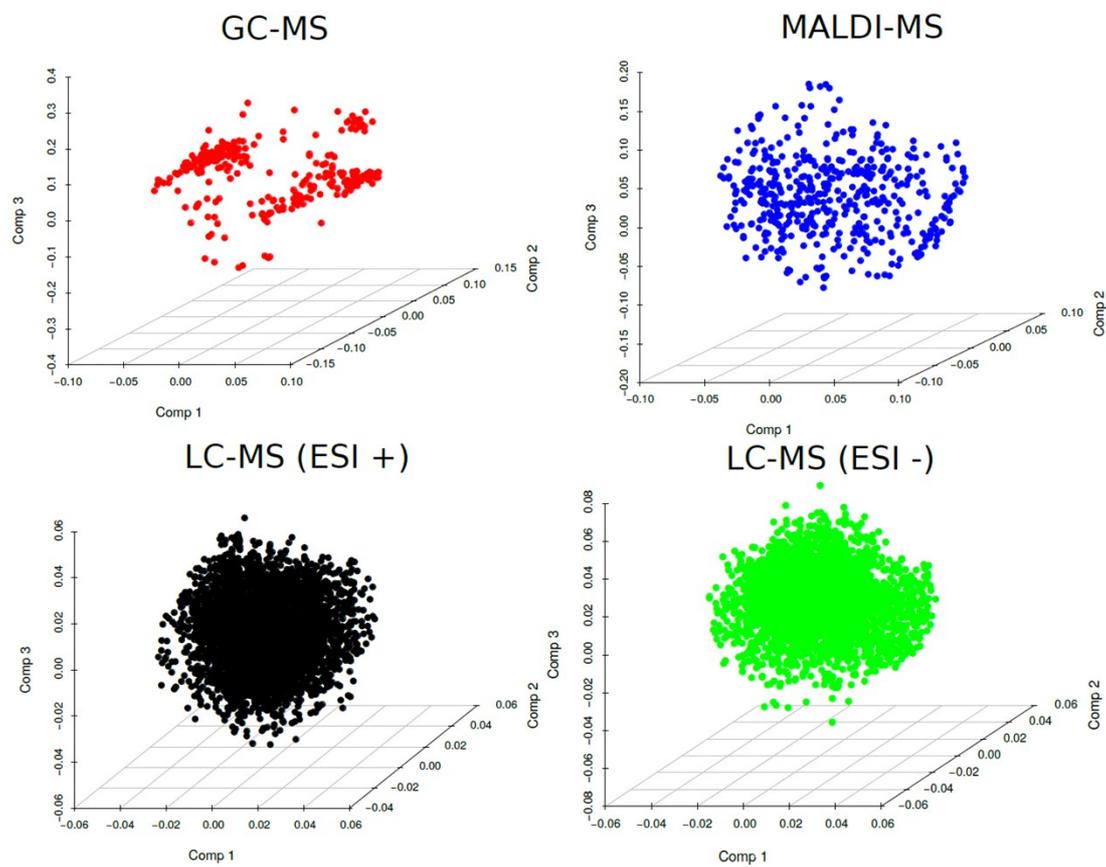


Fig. S12. Partial least square (PLS) discriminant analysis (PLS-DA) LV1-LV3 loading plots representing metabolite profiles obtained by GC-MS, MALDI-TOF MS, ESI positive LC-MS and ESI negative LC-MS analyses showing parent ion distribution.

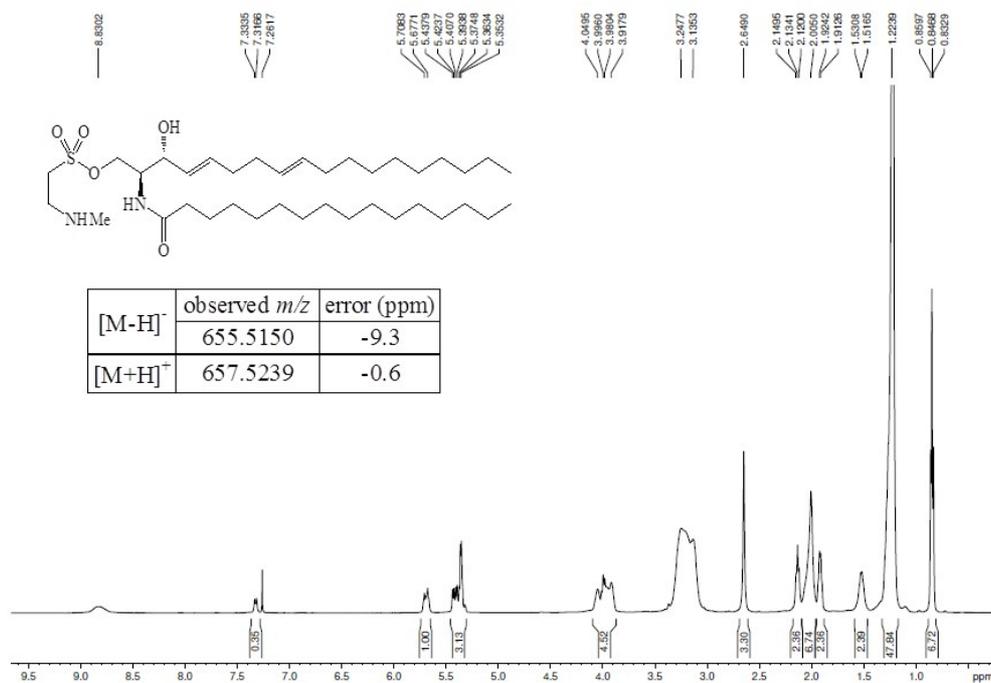


Fig. S13. ¹H NMR (CDCl₃/CD₃OD 4:1, 500 MHz) and HRMS (ESI + and -) spectra of the *palyosulfonoceramide A*.

¹H NMR data *palyosulfonoceramide A*: (500 MHz, CDCl₃/MeOD 4:1): δ 8.83 (1H, s br, NH), 7.32 (1H, d, *J* = 8.4 Hz, NH), 5.70-5.67 (1H, m, H-5), 5.40 (1H, dd, *J* = 15.4, 7.1 Hz, H-4), 5.39-5.35 (2H, m, H-8, H-9), 4.04-3.91 (4H, m, 2H-1, H-2, H-3), 3.13 (2H, s br, H-2''), 2.65 (3H, s, N-Me), 2.13 (1H, t, *J* = 7.4 Hz, H-2'), 2.02-2.00 (6H, m, H-6, H-7, H-1''), 1.92-1.91 (2H, m, H-10), 1.53-1.51 (2H, m, H-3'), 1.22 (38, s br, H-11/17, H-4'/15'), 0.84 (6H, t, *J* = 6.4 Hz, H-18, H-16').

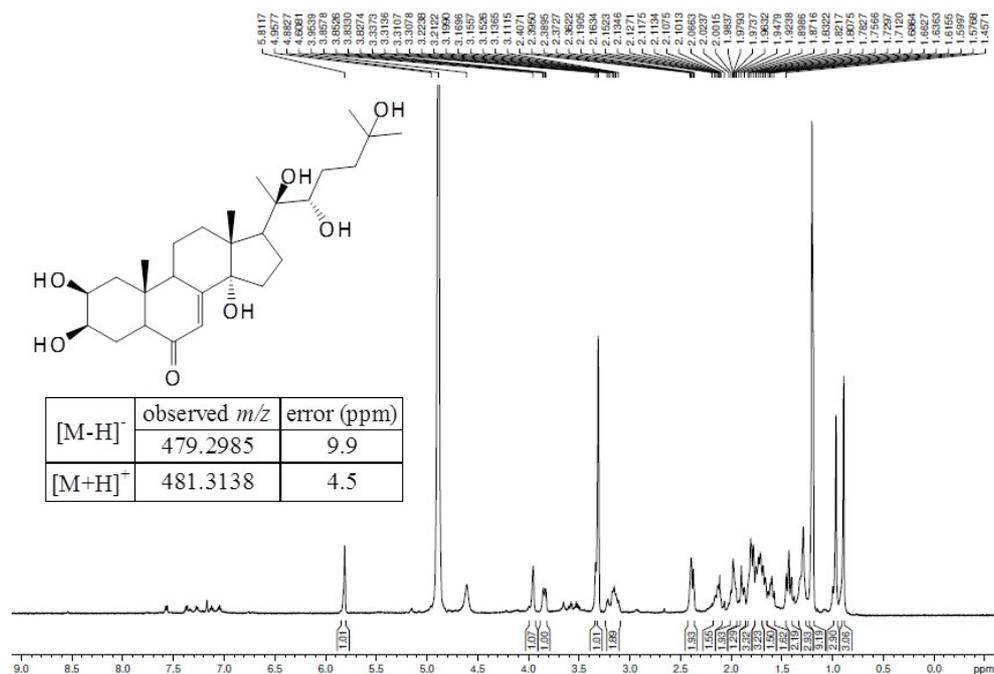


Fig. S15. ¹H NMR (CD₃OD, 500 MHz) and HRMS (ESI + and -) spectra of the 20-hydroxyecdysone.

¹H NMR data of 20-hydroxyecdysone: (500 MHz, MeOD): δ 5.82 (1H, s br, H-7), 3.95 (1H, s br, H-3), 3.84 (1H, td, *J* = 11.8, 4.0 Hz, H-2), 3.31* (1H, H-22), 3.15 (1H, m, H-9), 2.40-2.36 (2H, m, H-5, H-17), 2.13 (1H, m, H-12a), 2.00-1.97 (2H, m, H-15a, H-16a), 1.89 (1H, m, H-12b), 1.80 (3H, m, H-1a, H-11a, H-24a), 1.75 (1H, m, H-16b), 1.72-1.68 (3H, m, 2H-4, H-11b), 1.63 (1H, m, H-23a), 1.57 (1H, m, H-15b), 1.43-1.39 (2H, m, H-1b, H-24b), 1.29 (1H, m, H-23b), 1.20 (3H, s, H-27), 1.19 (6H, s, H-21, H-26), 0.96 (3H, s, H-19), 0.88 (3H, s, H-18).

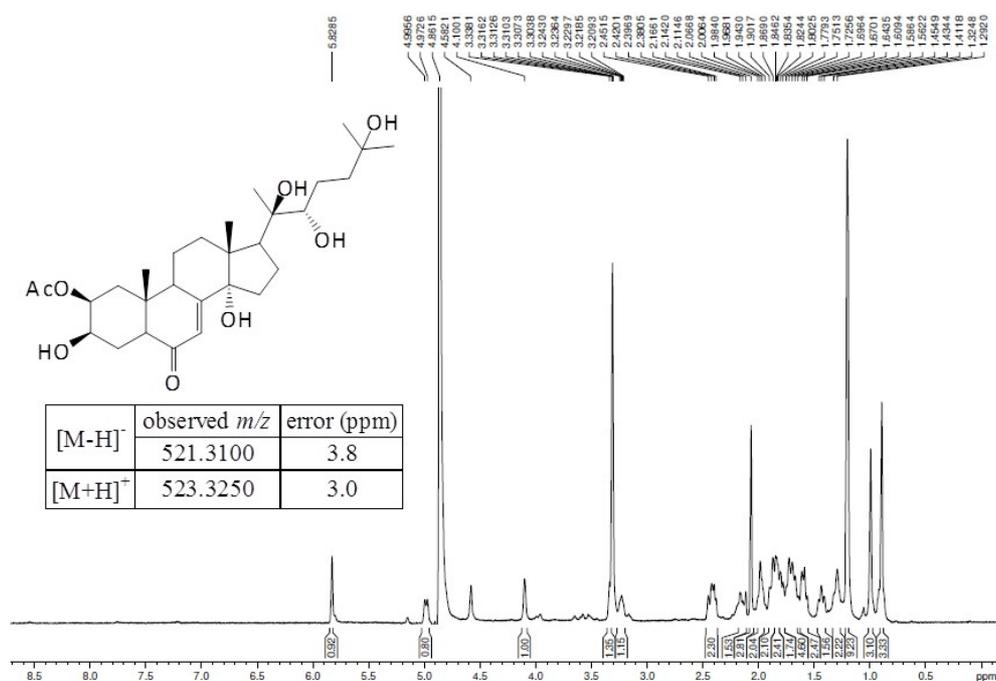


Fig. S16. ¹H NMR (CD₃OD, 500 MHz) and HRMS (ESI + and -) spectra of the 2-*O*-acetyl-20-hydroxyecdysone.

¹H NMR data of 2-*O*-acetyl-20-hydroxyecdysone: (500 MHz, MeOD): δ 5.82 (1H, s br, H-7), 4.98 (1H, d, *J* = 11,5 Hz, H-2), 4.10 (1H, m, H-3), 3.30* (1H, H-22), 3.22 (1H, m, H-9), 2.43 (1H, dd, *J* = 8.7, 2.6 Hz, H-5), 2.39 (2H, dd, *J* = 8.5, 2.5 Hz, H-17), 2.15 (1H, m, H-12a), 2.06 (3H, s, 2-Ac), 2.00-1.97 (2H, m, H-15a, H-16a), 1.90-1.86 (3H, m, H-1a, H-12b, H-16b), 1.83 (1H, m, H-4a), 1.80-1.78 (2H, m, H-11a, H-24a), 1.72 (1H, m, H-4b), 1.67-1.64 (2H, m, H-11b, H-23a), 1.60-1.57 (2H, m, H-1a, H-15b), 1.43 (1H, m, H-24b), 1.29 (1H, m, H-23b), 1.22 (3H, s, H-27), 1.20 (3H, s, H-21), 1.19 (3H, s, H-26), 0.99 (3H, s, H-19), 0.89 (3H, s, H-18).

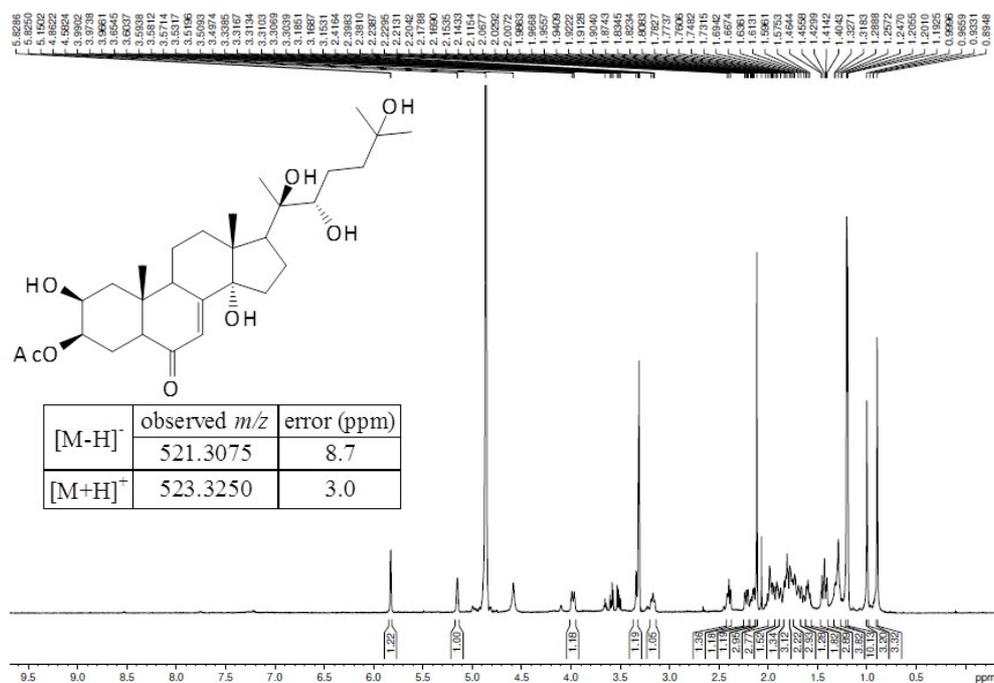


Fig. S17. ¹H NMR (CD₃OD, 500 MHz) and HRMS (ESI + and -) spectra of the 3-O-acetyl-20-hydroxyecdysone

¹H NMR data of 3-O-acetyl-20-hydroxyecdysone: (500 MHz, MeOD): δ 5.82 (1H, s br, H-7), 5.15 (1H, s br, H-3), 3.97 (1H, s br, H-2), 3.30* (1H, H-22), 3.16 (1H, m, H-9), 2.39 (2H, t, *J* = 9.0 Hz, H-17), 2.22 (1H, d, *J* = 11.4 Hz, H-5), 2.14 (1H, m, H-12a), 2.11 (3H, s, 3-Ac), 1.99-1.97 (2H, m, H-15a, H-16a), 1.91-1.89 (2H, m, H-4a, H-12b), 1.79-1.77 (2H, m, H-4a, H-24a), 1.72 (1H, m, H-11a), 1.67-1.65 (2H, m, H-11b, H-23a), 1.60 (1H, m, H-15b), 1.43-1.40 (2H, m, H-1b, H-24b), 1.31-1.28 (2H, m, H-4b, H-23b), 1.20 (6H, s, H-27 / H-21), 1.19 (3H, s, H-26), 0.99 (3H, s, H-19), 0.89 (3H, s, H-18).

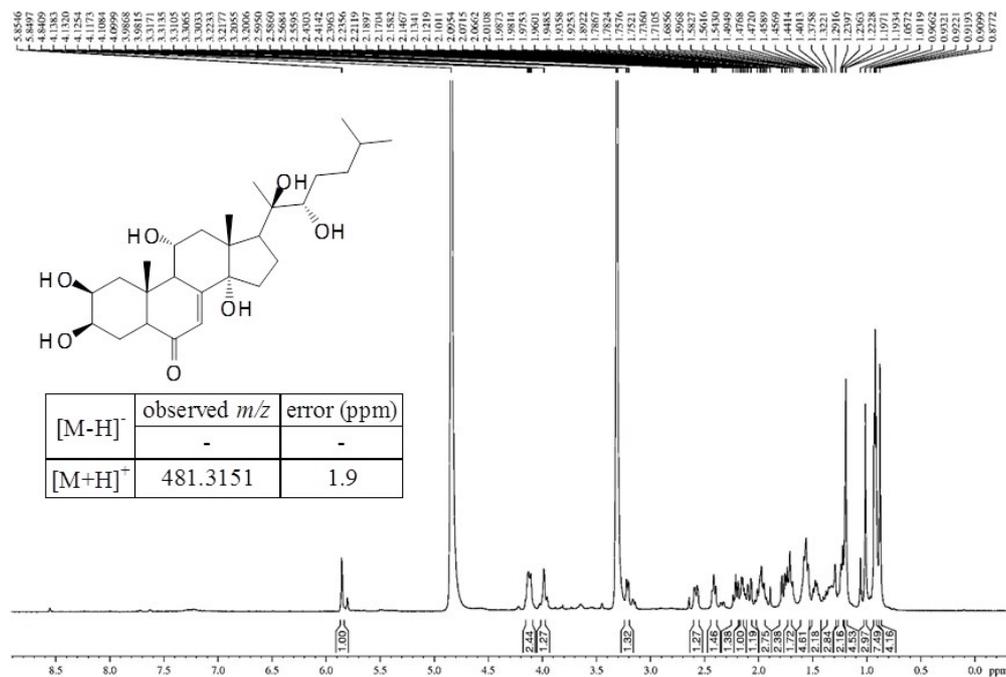


Fig. S18. ¹H NMR (CD₃OD, 500 MHz) HRMS-ESI+ spectra of the zoanthusterone

¹H NMR data of zoanthusterone: (500 MHz, MeOD): δ 5.84 (1H, d, *J* = 2.5 Hz, H-7), 4.13-4.09 (2H, m, H-11, H-3), 3.98 (1H, d, *J* = 2.7 Hz, H-2), 3.31* (1H, H-22) 3.20 (1H, dd, *J* = 8.9, 2.8 Hz, H-9), 2.57 (H, dd, *J* = 13.3, 4.5 Hz, H-5), 2.40 (1H, t, *J* = 8.9 Hz, H-17), 2.21 (1H, t, *J* = 11.1 Hz, H-12a), 2.17-2.12 (1H, m, H-1a), 2.08 (H, dd, *J* = 14.8, 2.8 Hz, H-16a), 1.98-1.92 (3H, m, H-12b, H-15a, H-4a), 1.76 (H, dd, *J* = 14.8, 2.1 Hz, H-16b), 1.73-1.68 (2H, m, H-15b, H-4b), 1.59-1.54 (3H, m, H-25, H-23a, H-24a), 1.49-1.44 (1H, m, H-1b), 1.40-1.37 (1H, m, H-24b), 1.22 (1H, t, *J* = 6.8 Hz, H-23b), 1.19 (3H, s, H-21), 1.01 (3H, s, H-19), 0.92 (3H, d, *J* = 5.0 Hz, H-26), 0.90 (3H, d, *J* = 5.0 Hz, H-27), 0.87 (3H, s, H-18).

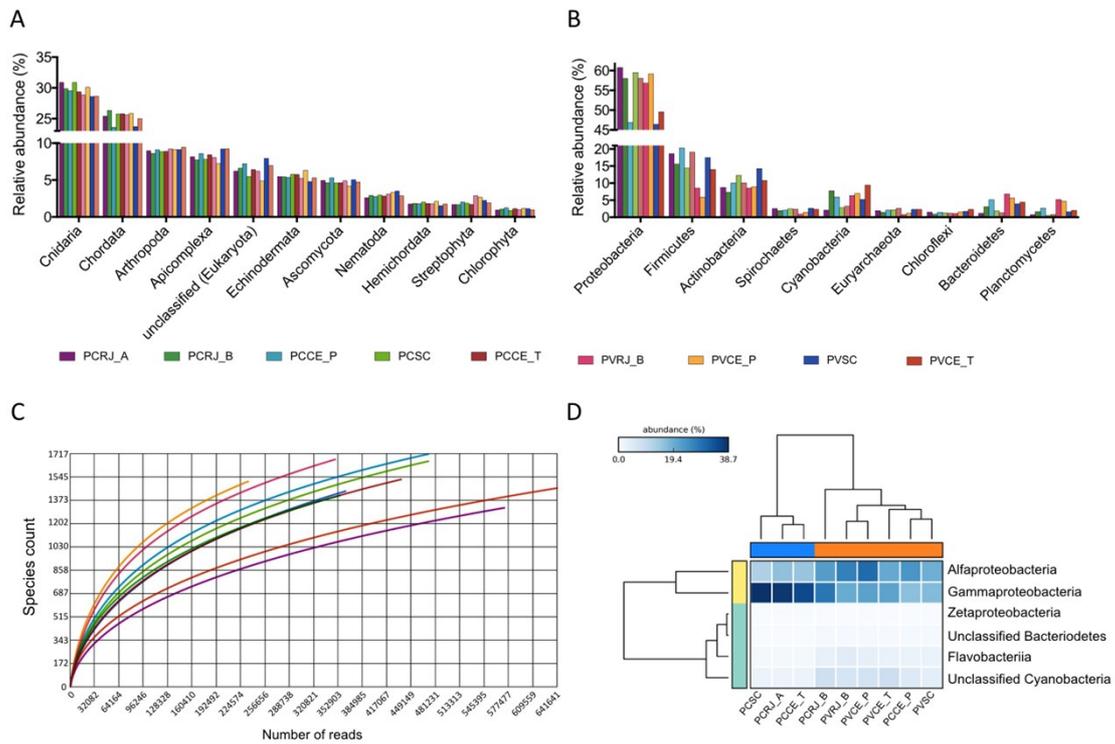


Fig. S19. Organismal classifications of zoanthid metagenomes. Bar graphs showing the eukaryotic (A) and prokaryotic (B) phyla with relative abundances $\geq 1\%$ in all metagenomic samples. C) Derived rarefaction curves presenting the number of prokaryotic species signatures per reads. The curves indicate that our sequence efforts covered a great fraction of the metagenomes microbial species richness, with up to 1717 prokaryotic species and 27 bacterial phyla being identified per sample. Taxonomic assignments were performed using M5NR database at MG-RAST. D) Heatmap plot showing bacterial phyla significantly enriched ($p < 0.05$) at *P. variabilis*-dominating (orange) and *P. caribeorum*-exclusive groups (blue). Relative abundance of each taxon to the entire sample is shown in color-coded logarithmic scale.

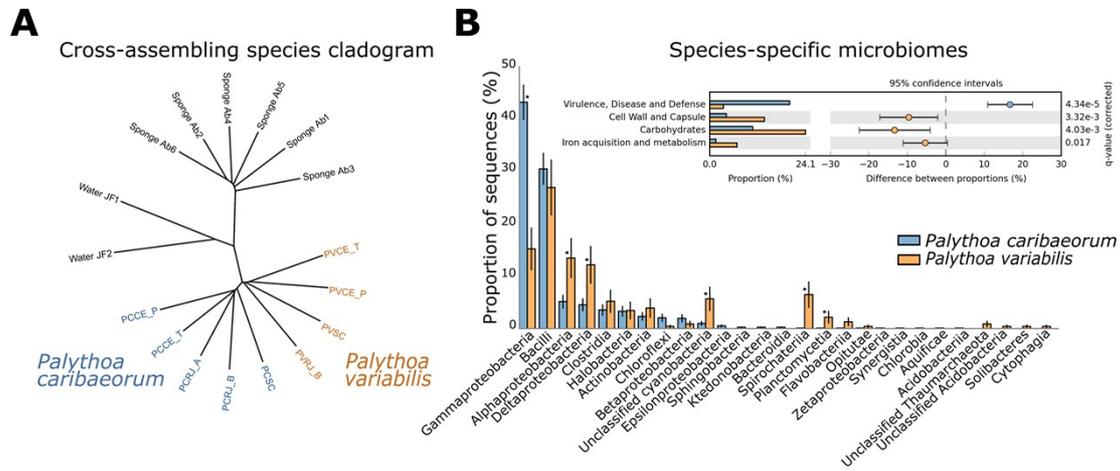


Fig. S20. A) A cladogram representing the shared contigs after cross-assembly of metagenomes from *P. variabilis* (orange) and *P. caribaeorum* (blue) and previously published samples.³⁶ B) Relative abundances of prokaryotic phyla signatures retrieved from *P. caribaeorum* and *P. variabilis* clusters (*, p-value < 0.05). Extended error bar plot (*inset*) shows functional annotations with Subsystems (hierarchical level 1) differing significantly.

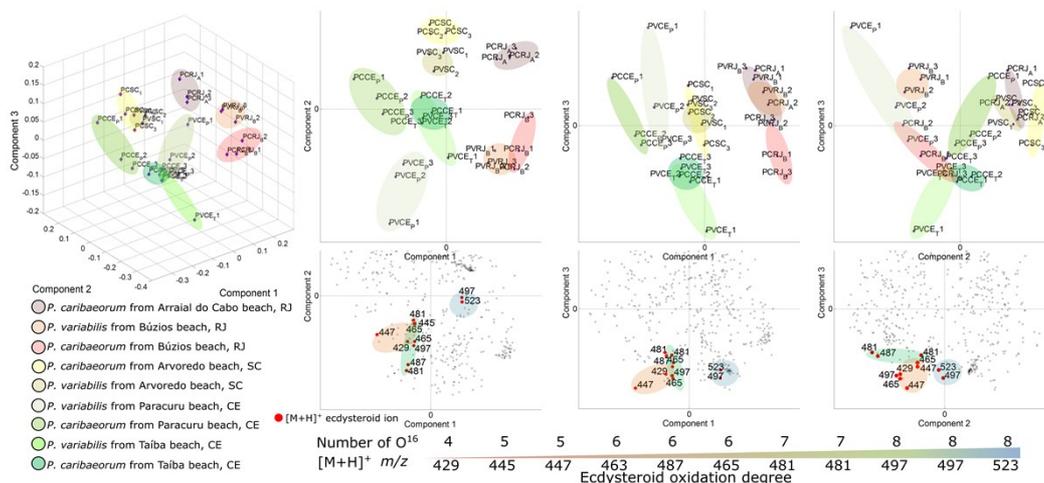


Fig. S21. **Ecdysteroid parent ion distribution.** The multivariate analysis reveals the distribution of polyhydroxylated ecdysteroids in terms of oxidation pattern and geographic location. Score plots of the multisource augmented data matrix from *Palythoa* species were split into graphs to support significant differences among all 3 components. Comparison with loading plots showed that ecdysteroids had higher oxidation states in both species from Taíba beach, Ceará State, whereas a higher content of ecdysteroids with lower oxidation states was observed in *P. variabilis* from Paracuru beach, Ceará State.

Supplementary Tables

Table S1. LC-MS Settings

Source MS conditions	Ionization ESI positive and negative
	Enhanced resolution mode (8100 $m/z/s$)
	200-1500 m/z scan range
	100000 ICC target
	5 Spectra Averages (2 Rolling Averaging)
<hr/>	
MS/MS conditions	Target Mass 1,200 m/z (CID) and 900 m/z (ETD)
	UltraScan mode (32500 $m/z/s$)
	200-1500 m/z scan range
	100000 ICC target
	Abs. threshold 25000
CID	Rel. threshold 5.0%
	2 precursor
	5 Spectra Averages
<hr/>	
CID	Fragmentation amplitude: 70% (SmartFrag active)

Table S2. Cumulative explained variance by PCA components

	PC1	PC2	PC3
GC/MS_R2	0.61004	0.13016	0.0971
GC/MS_R2cum	0.61004	0.7402	0.8373
MALDI_R2	0.34834	0.17598	0.09373
MALDI_R2cum	0.34834	0.52432	0.61805
LC/MS (ESI+)_R2	0.11214	0.09056	0.08547
LC/MS (ESI+)_R2cum	0.11214	0.2027	0.28817
LC/MS (ESI-)_R2	0.15044	0.09391	0.08501
LC/MS (ESI-)_R2cum	0.15044	0.24435	0.32936

R2 = explained variance, R2cum = cumulative explained variance

Table S3. Cumulative explained variance by PLS components

	1 comps	2 comps	3 comps	4 comps	5 comps
GC/MS_R2cum	0.409753	0.694295	0.830537	0.893897	0.925713
GC/MS_Q2cum	0.262497	0.537346	0.633733	0.642903	0.706004
MALDI_R2cum	0.727316	0.846513	0.929341	0.959667	0.97909
MALDI_Q2cum	0.607938	0.753874	0.782272	0.79845	0.795179
LC/MS (ESI+)_R2cum	0.914749	0.988556	0.99808	0.999831	0.999979
LC/MS (ESI+)_Q2cum	0.692865	0.766311	0.777551	0.781621	0.781547
LC/MS (ESI-)_R2cum	0.842669	0.968553	0.997385	0.999879	0.999989
LC/MS (ESI-)_Q2cum	0.635964	0.758516	0.785028	0.783826	0.786939

R2cum = cumulative explained variance, Q2cum = cumulative explained variance in prediction

Table S4. Summary of non-polar metabolites detected by GC-MS.

Retention Time (min.)	compounds	Species	location
8.6	2-pentadecyn-1-ol	PC, PV	B, P
13.5	myristic acid	PC, PV	A, B, T, P
14.0	2-nonadecanone	PC, PV	A, B, P, Ar
14.2	tetradecanal	PC, PV	A, B, P, Ar
14.8	1-heptadecanol	PC	T
14.8	1-hexadecanol	PC	A, B, T
15.4	palmitoleic acid	PC, PV	B, P, Ar
15.2	palmitic aldehyde	PC	P
15.3	methyl palmitate	PC, PV	T, P, Ar
15.6	9-octadecenal	PC, PV	T, P, Ar
15.7	palmitic acid	PC, PV	A, B, T, P, Ar
15.8	ethyl palmitoleate	PC	Ar
16.0	ethyl palmitate	PC, PV	A, B, T, P, Ar
16.3	oleic acid	PC	B, T
16.4	stearaldehyde	PC	Ar
18.6	stearic acid	PC, PV	A, B, Ar
18.7	ethyl oleate	PC	B
21.7	arachidonic acid	PC, PV	A, B, T, P, Ar
27.2	gamolenic acid	PC, PV	A, B, T, P, Ar
30.2	oleamide	PC, PV	A, B, P, T, Ar
31.9	cholesta-4,6-dien-3-ol	PC	A, B, P, T, Ar
32.8	cholesta-5,22-dien-3-ol	PC, PV	P, Ar
33.2	lanol	PC, PV	A, B, P, T, Ar
33.5	stearyl palmitate	PV	Ar
33.7	myristyl palmitate	PC	P, Ar
34.4	fucosterol	PV	T
34.5	campesterol	PC, PV	A, P, T, Ar
35.5	6-methyl-cholestan-3-ol	PV	P
35.7	8,11,14-eicosatrienoic acid	PC, PV	P, T, Ar
35.8	methyl stearate	PC, PV	A, P, Ar
36.1	stearyl palmitate	PC, PV	P, Ar
36.3	palmityl palmitate	PC, PV	P, Ar
37.8	gorgost-5-en-3-ol	PC, PV	A, B, T, P, Ar
39.6	oleyl oleate	PV	P

PC = *Palythoa caribaeorum*, PV = *Palythoa variabilis*, A = Arraial do Cabo beach (RJ), B = Buzios beach (RJ), T = Taiba beach (CE), P = Paracuru beach (CE), Ar = Arvoredo beach (SC)

Table S5. Summary of polar metabolites detected by LC-DAD-IT and LC-DAD-TOF

number	observed <i>m/z</i>	ion form	error (ppm)	observed <i>m/z</i>	error (ppm)	ion form	UV (nm)	Final Annotation	Compound group	Group
1	243.0966	[M-H] ⁻	3.7	245.1128	1.6	[M+H] ⁺	320	palythine	Amino acids and derivatives	C,R,S
2	-	-	-	246.0967	4.5	[M+H] ⁺	320	mycosporine-glc	Amino acids and derivatives	C,R,S
3	-	-	-	285.1443	2.5	[M+H] ⁺	360	palythene	Amino acids and derivatives	C,R,S
4	-	-	-	260.1167	78.8	[M+H] ⁺	-	methyl-palythine	Amino acids and derivatives	R
5	251.1047	[M-H] ⁻	-6.0	253.1176	2.8	[M+H] ⁺	320	palythazine	Amino acids and derivatives	C,R,S
6	-	-	-	218.1384	3.7	[M+H] ⁺	-	<i>O</i> -propanoylcarnitine	Amino acids and derivatives	C,R,S
7	-	-	-	347.1478	-6.9	[M+H] ⁺	320	porphyrin-334	Amino acids and derivatives	C,R,S
8	480.2707	[M-H] ⁻	9.0	482.2885	4.4	[M+H] ⁺	270	norzoanthamine	zoanthid alkaloid	C,R,S
9	522.2473	[M-H] ⁻	3.6	524.2625	4.4	[M+H] ⁺	270	zoanthamide	zoanthid alkaloid	C,R,S
10	-	-	-	510.2864	1.6	[M+H] ⁺	270	zoanthamine or zoanthamine	zoanthid alkaloid	C,R,S
11	-	-	-	496.2671	-5.6	[M+H] ⁺	270	norzoanthamine	zoanthid alkaloid	C,R,S
12	-	-	-	465.3201	2.1	[M+H] ⁺	-	ponasterone A or ecdysone	Ecdysteroid	C,R,S
13	-	-	-	447.3106	-0.2	[M+H] ⁺	-	dehydro-deoxyecdysone	Ecdysteroid	C,R,S
14	-	-	-	513.3050	1.6	[M+H] ⁺	-	trihydroxyecdysone	Ecdysteroid	C,R,S
15	-	-	-	481.3138	4.5	[M+H] ⁺	-	20-hydroxyecdysone*	Ecdysteroid	C,R,S
16	-	-	-	481.3151	1.9	[M+H] ⁺	-	zoanthosterone*	Ecdysteroid	C,R,S
17	-	-	-	497.3105	1.8	[M+H] ⁺	-	dihydroxyecdysterone	Ecdysteroid	C,R,S
18	-	-	-	495.3345	5.9	[M+H] ⁺	-	makisterone isomer	Ecdysteroid	C,R,S
19	-	-	-	523.3275	-1.9	[M+H] ⁺	-	hydroxyecdysone 2-acetate*	Ecdysteroid	C,R,S
20	-	-	-	523.3250	3.0	[M+H] ⁺	-	hydroxyecdysone 3-acetate*	Ecdysteroid	C,R,S
21	-	-	-	467.3147	3.8	[M+H] ⁺	-	ecdysterone	Ecdysteroid	C,R,S
22	-	-	-	429.2994	1.4	[M+H] ⁺	-	C ₂₇ H ₄₁ O ₄	Ecdysteroid	C,R,S
23	-	-	-	445.2937	2.1	[M+H] ⁺	-	C ₂₇ H ₄₁ O ₅	Ecdysteroid	C,R,S
24	-	-	-	463.3046	1.8	[M+H] ⁺	-	C ₂₇ H ₄₃ O ₆	Ecdysteroid	C,R,S
25	-	-	-	497.3093	4.3	[M+H] ⁺	-	palythalone B	Ecdysteroid	R
26	-	-	-	481.3147	3.8	[M+H] ⁺	-	C ₂₇ H ₄₄ O ₇	Ecdysteroid	R
27	-	-	-	496.6683	2.2	[M+H] ⁺	-	PC(16:0/18:1)	phosphatidylcholine derivative	C,R,S
28	-	-	-	440.3121	3.0	[M+H] ⁺	-	PC(O-12:0/O-1:0)	phosphatidylcholine derivative	C,R,S
29	-	-	-	274.2744	1.4	[M+H] ⁺	-	C16 sphinganine	phosphatidylcholine derivative	C,R,S
30	-	-	-	482.3602	0.6	[M+H] ⁺	-	PC(O-8:0/O-8:0)	phosphatidylcholine derivative	C,R,S
31	-	-	-	452.2768	0.4	[M+H] ⁺	-	PE(16:1/0:0)	phosphatidylcholine derivative	C,R,S
32	-	-	-	524.2641	1.9	[M+H] ⁺	-	PC(18:0/0:0)	phosphatidylcholine derivative	C,R,S
33	-	-	-	454.3279	3.0	[M+H] ⁺	-	PC(O-14:0/0:0)	phosphatidylcholine derivative	C,R,S
34	-	-	-	544.3393	0.8	[M+H] ⁺	-	PC(20:4/0:0)	phosphatidylcholine derivative	C,R,S
35	-	-	-	482.2879	0.0	[M+H] ⁺	-	PC(6:0/8:0)	phosphatidylcholine derivative	C,R,S
36	-	-	-	480.3436	2.7	[M+H] ⁺	-	PC(O-16:1/0:0)	phosphatidylcholine derivative	C,R,S
37	-	-	-	464.2764	1.6	[M+H] ⁺	-	PA(19:3/0:0)	phosphatidylcholine derivative	C,R,S

38	436.2835	[M-H] ⁻	3.9	438.2982	5.9	[M+H] ⁺	-	terpendole E	indole-diterpene	C,R,S
39	450.2608	[M-H] ⁻	8.0	452.2792	2.0	[M+H] ⁺	-	terpendole G	indole-diterpene	C,R,S
40	655.5150	[M-H] ⁻	-9.3	657.5239	-0.6	[M+H] ⁺	-	palyosulfonoceramide A*	ceramide and derivative	C,R,S
41	657.5278	[M-H] ⁻	-4.9	659.5394	-0.4	[M+H] ⁺	-	palyosulfonoceramide B*	ceramide and derivative	C,R,S

C = Ceara State; R = Rio de Janeiro State; S = Santa Catarina State.

* Metabolites previously isolated and elucidated by NMR and HRMS.

Table S6. Summary of metabolites detected by MALDI-TOF.

observed m/z	ion form	MS/MS	Final Annotation	Compound group	Group
1008.6	[M+N a] ⁺	741, 612, 544, 184, 132, 104, 74	PC (34:0/16:0)	phosphatidylcholine	C,R ,S
1008.6	[M+N a] ⁺	950, 907, 846, 760, 718, 666, 482, 233, 184, 104, 74	PC (34:0/16:0)	phosphatidylcholine	C,R ,S
1024.6	[M+K] ⁺	-	PC (34:0/16:0)	phosphatidylcholine	C,R ,S
1026.6	[M+N a] ⁺	-	PC (20:5/32:0)	phosphatidylcholine	C,R ,S
1042.6	[M+K] ⁺	984, 880, 794, 754, 482, 184, 104, 74	PC (20:5/32:0)	phosphatidylcholine	C,R ,S
1058.6	-	1006, 955, 869, 482, 184, 104, 74	PC derivative	phosphatidylcholine	C,R ,S

C = Ceara State; R = Rio de Janeiro State; S = Santa Catarina State.

Table S7. Metagenomes overall features.

Features / Metagenome	PV cross-contigs	PC cross-contigs	PVCE_P	PVCE_T	PVRJ_B	PVSC	PCCE_P	PCCE_T	PCRJ_B	PCSC	PCRJ_A
Uploaded Seqs.	67067	212796	234911	436149	349696	363219	472196	641934	355591	472141	572519
Post QC¹	64446	203907	219864	393285	291667	328988	438784	580112	318476	449017	509466
bp Count²	17528810	66464386	36794576	67169356	38943974	54422081	84473129	98755204	53724456	106069574	80909182
Mean seqs. Length (bp)	271 ± 103	325 ± 124	167 ± 85	170 ± 92	133 ± 76	165 ± 89	192 ± 95	170 ± 90	168 ± 89	236 ± 94	158 ± 87
Mean GC %	38 ± 6	38 ± 5	39 ± 8	38 ± 7	39 ± 8	39 ± 7	38 ± 7	38 ± 7	39 ± 7	39 ± 6	38 ± 7
Predicted Protein Features (%)[*]	53946 (83,7)	182233 (89,4)	152538 (69,3)	260699 (66,3)	176038 (60,3)	220292 (66,9)	303704 (69,2)	378008 (65,2)	212998 (66,8)	341246 (76,0)	323235 (63,4)
Predicted rRNA Features (%)[*]	588 (0,9)	1338 (0,7)	3642 (1,6)	7310 (1,8)	6828 (2,3)	6691 (2,0)	6960 (1,6)	10247 (1,8)	5658 (1,7)	4802 (1,1)	10085 (2,0)
Identified Protein Features (%)[§]	3167 (4,9)	12037 (5,9)	8717 (5,7)	14205 (5,4)	9790 (5,6)	10713 (4,9)	18799 (6,2)	19833 (5,2)	12130 (5,7)	20680 (611)	15554 (4,8)
Classified by MG-Rast (M5NR) (%)	10416 (16,2)	40839 (20)	33832 (15,4)	55671 (14,2)	8601 (13,2)	40440 (12,3)	73430 (16,7)	84259 (14,5)	46810 (14,7)	79916 (17,8)	62866 (12,3)
Classified under Domain											
Eukaryota (%)	8318 (79,9)	33433 (81,9)	22149 (65,5)	44134 (79,3)	21286 (55,1)	32127 (79,4)	56977 (77,6)	69004 (81,9)	35201 (75,2)	64590 (80,8)	52242 (83,1)
Bacteria (%)	1734 (16,6)	6063 (14,8)	10030 (29,6)	9425 (16,9)	15156 (39,3)	6678 (16,5)	13517 (18,4)	12644 (15)	9952 (21,3)	12435 (15,6)	8516 (13,5)
Archaea (%)	200 (0,5)	56 (0,5)	126 (0,4)	248 (0,4)	243 (0,6)	247 (0,6)	348 (0,5)	364 (0,4)	162 (0,3)	341 (0,4)	193 (0,3)
Others (%)	1133 (2,8)	308 (3,0)	1527 (4,5)	1864 (3,3)	1916 (5,0)	1388 (3,4)	2588 (3,5)	2247 (2,7)	1495 (3,2)	2550 (3,2)	1915 (3,0)

Supplementary References

24. S. E. Stein, *J Am Soc Mass Spectrom*, 1999, **10**, 770–781.
25. V. I. Babushok, P. J. Linstrom, J. J. Reed, I. G. Zenkevich, R. L. Brown, W. G. Mallard and S. E. Stein, *J. Chromatogr. A*, 2007, **1157**, 414–421.
26. F. Carnevale Neto, A. Pilon, D. Selegato, R. Freire, H. Gu, D. Raftery, N. Lopes and I. Castro-Gamboa, *Front Mol Biosci*, 2016, **3**, 1–13.
27. C. A. Smith, G. O’Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak, *Ther. Drug Monit.*, 2005, **27**, 747–751.
28. J. W. Blunt and M. H. G. Munro, *Dictionary of marine natural products with CD-ROM*, Chapman & Hall/CRC Press, Boca Raton, FL., 2007.
29. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res*, 2003, **13**, 2498–2504.
30. M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt and J. Egertson, *Nat. Biotechnol.*, 2012, **30**, 918–920.
31. C. A. Smith, E. J. Want, G. O’Maille, R. Abagyan and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 779–787.
32. S. Gibb and K. Strimmer, *Bioinformatics*, 2012, **28**, 2270–2271.
33. J. Xia, I. V Sinelnikov, B. Han and D. S. Wishart, *Nucleic Acids Res.*, 2015, gkv380.
34. J. G. L. Almeida, A. I. V Maia, D. V Wilke, E. R. Silveira, R. Braz-Filho, J. J. La Clair, L. V Costa-Lotufo and O. D. L. Pessoa, *Mar drugs*, 2012, **10**, 2846–2860.
35. A. Suksamrarn, A. Jankam, B. Tarnchompoo and S. Putchakarn, *J Nat Prod*, 2002, **65**, 1194–1197.
36. G. D. Garcia, G. B. Gregoracci, E. de O. Santos, P. M. Meirelles, G. G. Z. Silva, R. Edwards, T. Sawabe, K. Gotoh, S. Nakamura, T. Iida, R. L. de Moura and F. L. Thompson, *Microb. Ecol.*, 2013, **65**, 1076–1086.
37. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass and M. Kubal, *BMC Genomics*, 2008, **9**, 1.
38. R. A. Edwards, R. Olson, T. Disz, G. D. Pusch, V. Vonstein, R. Stevens and R. Overbeek, *Bioinformatics*, 2012, **28**, 3316–3317.
39. G. G. Z. Silva, K. T. Green, B. E. Dutilh and R. A. Edwards, *Bioinformatics*, 2015, btv584.
40. R. Overbeek, T. Begley, R. M. Butler, J. V Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz and R. Edwards, *Nucleic Acids Res*, 2005, **33**, 5691–5702.