BIOSCIENCE DOCTORAL PROGRAM

UNIVERSIDAD DE LA SABANA

USE OF METABOLOMIC AND DEREPLICATION TOOLS IN THE STUDY OF THE CHEMICAL COMPOSITION OF SOME OCTOCORALS OF THE COLOMBIAN CARIBBEAN AND THEIR CYTOTOXIC ACTIVITY

Author:

MSc. Liliana Andrea Santacruz Cifuentes

Supervisors:

Dr. Edisson T. Camacho

Dr. Olivier P. Thomas

Bogotá, Colombia, 2020

UNIVERSIDAD DE LA SABANA DOCTORADO EN BIOCIENCIAS

USO DE HERRAMIENTAS METABOLÓMICAS Y DE DEREPLICACIÓN EN EL ESTUDIO DE LA RELACIÓN DE LA COMPOSICIÓN QUÍMICA DE ALGUNOS OCTOCORALES DEL CARIBE COLOMBIANO Y SU ACTIVIDAD CITOTÓXICA

Autor:

MSc. Liliana Andrea Santacruz Cifuentes

Memoria presentada para optar al título de

Doctor en Biociencias de la Universidad de La Sabana

Directores:

Dr. Edisson T. Camacho

Dr. Olivier P. Thomas

Bogotá, Colombia, 2020

A mi esposo Omar, a mis hijos Juan Felipe y Maria Jose, a mis Padres y Hermanos, los amores de mi vida.

AGRADECIMIENTOS

Agradezco a Dios por guiar mis pasos y fortalecer siempre mi espíritu; a mis directores Edisson Tello y Olivier Thomas, por creer en este proyecto, apoyarme con su acertada dirección y disposición constante durante todos los procesos de la tesis. Gracias por que, desde la codirección, viví la experiencia del conocimiento impartido por las dos Universidades que desde la excelencia apoyaron mi proceso.

Agradezco a la Dra. Carmenza Duque Beltran por su contribución en la discusión de aspectos fundamentales del proyecto. Gracias por su apoyo y conocimiento en especial en la etapa inicial del proyecto.

A la Dra. Monica Puyana por toda su colaboración en la colecta de las muestras y por su apoyo en el proceso de esta investigación.

Al Dr. Luis Eduardo Días, por su colaboración en las dudas que surgieron desde el campo de actividad citotóxica.

A la Dra. Luisa Villamil, por sus consejos y apoyo permanente.

A todas las personas que colaboraron en la fase experimental. A la ingeniera Ximena, por su apoyo en algunos análisis citotóxicos. Al personal de los laboratorios de la Universidad de La Sabana Jorge y Pilar por su disposición y ayuda. Al grupo de estudiantes de la Universidad de Galway en Irlanda donde realice mi pasantía doctoral por toda su colaboración.

A la Dra. Coralia Osorio por siempre estar presente con sus consejos.

A mis compañeros del Doctorado en Biociencias, por su amistad, paciencia, consejos y tiempo.

A mi Padre por impulsarme en la decisión de realizar el doctorado, a mi madre por el ejemplo, a mis hermanos por sus consejos.

A mi esposo Omar, a mi hijo Juan Felipe y a mi hija Maria Jose por comprenderme e inspirarme para seguir siempre adelante.

ABSTRACT

The oceans cover over two-thirds of the Earth´s surface and the organisms that live in them constitute about 2% of the organic material present in the seas. The warm waters of the Caribbean Sea contribute to the diversity of marine fauna, where soft corals are the most visible macro reef fauna. These organisms are a productive source of a great diversity of chemical compounds which serve as a chemical defense against predators, but also many of them have presented important biological activities. Therefore, it is important to investigate new tools that allow us to explore this great potential.

Hence, in this project the relationship of the cytotoxic activity of some octocorals of the Colombian Caribbean with respect to its metabolic composition was studied, using different tools such as dereplication, useful in the early identification of new or known compounds with biological activity. Additionally, the use of Natural Products Molecular Networks (GNPS) that with the Cytoscape software allowed to observe "cluster" or groups of molecules with similar fragmentation patterns and finally; the metabolomic tool, which allowed the detection of a large number of metabolites present in the studied organisms.

For the development of the this project, high efficiency chromatographic techniques were used: ultra efficient liquid chromatography coupled with high resolution mass spectrometry (UPLC-HRMS/Agilent 6540) and nuclear magnetic resonance (NMR/Agilent 600 MHz); which helped in the determination of the global profile of the metabolites present in the octocoral extracts. Also, in the models generated from the multivariate analysis of principal components (PCA) and discriminant analysis (PLSDA and OPLS) for tumor lines A549, SiHa and PC3 three diterpenes from *plexauridae* family, yielded a high score in the Variable Importance in Projection (VIP), therefore; it was possible to establish a statistically valid correlation between the chemical composition of the soft corals analyzed here and their cytotoxic activity.

This project contributed to the studies of the chemical diversity and cytotoxic activity of some octocorals of the Colombian Caribbean, using the cytotoxic activity of their extracts as a criterion.

RESUMEN

Los océanos cubren un poco mas de dos terceras partes de la superficie de la tierra y los organismos que en ellos viven constituyen cerca del 2% de la materia orgánica presente en los mares; en particular, las aguas cálidas del Mar Caribe contribuyen a la diversidad de la fauna marina, en donde, los octocorales son la macrofauna arrecifal más visible. Estos organismos, son fuente productiva de una gran diversidad de compuestos químicos los cuales les sirven como defensa química contra depredadores, pero también mucho de ellos han presentado actividades biológicas importantes para combatir diferentes enfermedades; por lo cual, es importante investigar sobre nuevas herramientas de estudio que permitan explorar este gran potencial.

Por lo anterior, en este proyecto se estudió la relación de la actividad citotóxica de algunos octocorales del Caribe Colombiano con respecto a su composición metabólica, usando diferentes herramientas como la dereplicación, útil en la identificación temprana de compuestos nuevos o conocidos con actividad biológica, el uso de Redes Moleculares de Productos Naturales (GNPS) que con el uso del software Cytoscape permitió observar "cluster" o agrupaciones de moléculas con patrones de fragmentación similares y finalmente; la herramienta metabolómica, que permitió detectar un gran número de metabolitos presentes en los organismos estudiados.

Para el desarrollo de este proyecto, se utilizaron técnicas cromatográficas de alta eficiencia: cromatografía líquida ultra eficiente junto con espectrometría de masas de alta resolución (UPLC-HRMS / Agilent 6540) y resonancia magnética nuclear (RMN / Agilent 600 MHz); que ayudó en la determinación del perfil global de los metabolitos presentes en los extractos octocorales; además, en los modelos generados a partir del análisis multivariado de componentes principales (PCA) y análisis discriminantes (PLSDA y OPLS) para las líneas tumorales A549, SiHa y PC3; tres diterpenos de la familia *plexauridae*, arrojaron una puntuación alta en la Importancia Variable en Proyección (VIP) por lo tanto; fue posible establecer una correlación estadísticamente válida entre la composición química de los corales blandos analizados aquí y su actividad citotóxica.Mediante este proyecto se contribuyó a los estudios de la diversidad química y de actividad citotóxica de algunos octocorales del Caribe Colombiano, a partir de sus extractos utilizando como criterio la actividad citotóxica de sus extractos.

CONTENT

LIST OF ABBREVIATIONS

MNP	Marine Natural Products
UV/Vis	Ultraviolet–Visible spectrophotometry
APCI	Atmospheric Pressure Chemical Ionization
CC	Chromatography Column
MS	Mass Spectrometry
TLC	Thin-Layer Chromatography
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
IARC	International Agency for Research on Cancer
ESI	Electrospray Ionization
UPLC	Ultra Performance Liquid Chromatography
UPLC-DAD	Ultra-Performance Liquid Chromatography–Diode Array Detector
UPLC-MS	Ultra performance liquid chromatography - Tandem Mass Spectrometer
NMR	Nuclear Magnetic Resonance
TOF	Time-Of-Flight
Da	Daltons
WHO	World Health Organization
FDA	Food and Drug Administration
PLS-DA	Partial Least Squares Discriminant Analysis
OPLS-DA	Orthogonal Projections to Latent Structures Discriminant Analysis
PCA	Principal Component Analysis
VIP	Variable Importance of Projection

1 INTRODUCTION

The introductory chapter of this thesis is based on the metabolomic study of some octocorals of the Colombian Caribbean and its potential application in the medicinal chemistry, as a tool for the search of new cytotoxic compounds. The approach to be presented here will be made considering these topics:

1. Marine natural products as a source of compounds that exhibit cytotoxic activity.
2. Cancer among the leading causes of mortality.
3. Metabolomic studies as an approach to find cytotoxic compounds in marine organisms.

In summary, this chapter will allow to establish in an orderly and coherent way the steps to be carried out in this research, showing the motivation to carry out this research, also the research question, the objectives of the study and, finally the structure of the document.

Considering the first point; from the marine natural products (Rodríguez, 1995; Avilés et al., 2015), a large variety of terpenes have been isolated and identified, among which are the cembranes, dolabellanes, fuscosides, cubitanes, sesquiterpenes, among others (Rodríguez et al., 1999; Berrue y Kerr, 2009; Newman y Cragg, 2012; Blunt et al., 2013, 2014, 2015), those that have presented important biological activities such as antivirals, antibacterials, acaricides, anti-inflammatories, antioxidants, and more recent their cytotoxic properties have been evaluated (Heckrodt & Mulzer, 2005) since cancer is one of the main causes of mortality worldwide (Momenimovahed & Salehiniya, 2017). In 2008, 8 million deaths were recorded as a result of malignant diseases, and this figure is estimated to reach 11 million by 2030 (Benson & Jatoi, 2012).

The importance of secondary metabolites (typically with molecular weights under 1000 Da) is their characterization help to better understand the functioning of organisms and thus be able to establish a correlation with the biological activity presented by many from them (Bennett, 2005). Metabolomic aims at the comprehensive study of small molecule (qualitative and quantitative analysis) of a biological system including cell, organ, tissue, biological fluid, organism at specific time (Fiehn, 2002). The recent advances in analytical methods (LC, MS, LC-MS, NMR, among others) have made possible the measurement of hundreds of metabolomic parameters with great sensitivity and precision. Unlike other omics, metabolomic determinations can be quantitative, opening the possibility of carrying out

statistical studies and collecting information from databases (Robosky et al., 2002). Metabolomics have high potential in the drug novel development and in medicine (Lindon et al., 2004). The metabolomics can be applied in obtaining new candidate compounds to be developed as drugs due to their powerful bioactivity shown in previous steps of experimentation. Additionally, there is another tool complementary to metabolomics, useful in the early identification of new compounds present in extracts of different organisms, called dereplicación, which is done by comparing chemical characteristics of metabolites with specialized databases (Blunt et al., 2012). His tool has been widely used in the systematic search of novel compounds from sources of marine origin as evidenced by the reviews made by Blunt and collaborators (Blunt et al., 2012, 2013, 2014, 2015, 2016, 2017, 2018).

In accordance with the above mentioned, this research was applied to discriminate the different extracts based on the metabolic fingerprints and their cytotoxic activity in a prospective framework and as a result it was found that the generated models using a metabolomic workflow that includes the use of UPLC-MS technique, allowed to establish that a diterpenoid with dolabellane type skeleton was the responsible of the cytotoxic activity showed by the extracts, against SiHa and A549 cancer cell lines. While that in the models generated using a metabolomic workflow with the NMR technique showed that the features B2_5118, B4_4965 and B4_7686 correlated with compounds asperdiol and plexaurolone was the responsible of the cytotoxic activity of the extracts, against the cancer cell line PC3. The above studies demonstrated that the soft coral extracts of the *plexauridae* family showed high cytotoxic activity against the tumor cell lines of SiHa, A549 and PC3, while the diterpene dolabellatrienone, isolated from the species *Pseudoplexaura flagellosa* presented low and moderate cytotoxic activity against different tumor cell lines, exhibited values of IC_{50} = 0.02 µg / mL against A549 and IC_{50} = 0.03 µg / mL against SiHa.

1.1 STRUCTURE OF THE DOCUMENT

This present thesis document contains five chapters.

The first chapter presents the introduction, the state of the art, the research question and the research objectives.

The second chapter shows a comparative analysis of the metabolomic footprint and cytotoxic activity of octocorals of the Colombian Caribbean.

The third chapter focuses on the metabolomic study of octocorals from the Colombian Caribbean: comparative analysis PSYCHE and ^{1}H NMR.

The fourth chapter discusses the integration of the tools: Molecular networking using data from MS/MS and metabolomics, for the identification of compounds with biological activity of the *Plexauridae* family.

Finally, chapter five presents the conclusions and suggestions for future work. The following figure shows the relationship of the published scientific articles with the objectives of the proposed theses.

1.2 STATE OF THE ART

1.2.1 Marine Natural Products as a source of bioactive compounds applicable in the production of new drugs

The medicines are of vital importance for the human beings since these contribute to the improvement of the health and therefore to the increase of the quality of life of the same ones. Currently, technological advances have allowed the development of a greater number of new medicines, which are generally chemical compounds extracted from natural sources. This is how today, one of the most successful sources in terms of new compounds used in the field of medicine, are natural products, 60% of medicines marketed today are of natural origin (Newman and Cragg, 2012, 2017), which have proven to be increasingly efficient to combat diseases such as cancer, which is the leading cause of mortality worldwide.

The process time from when scientists begin to investigate an innovative molecule until its commercialization is between 15 and 20 years (Chow & Shao, 2002). Marine natural products have attracted the attention of scientists in search of compounds with biological activity (Bhakuni & Rawat, 2005) that can become compounds of pharmaceutical and industrial use.

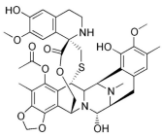
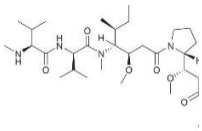
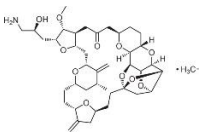
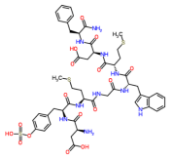
As a short historical account it can be mentioned that the first investigations carried out in order to find new bioactive compounds in marine organisms began in the 50s, then gradually expanded during the 1970s and 1980s, precisely thanks to the new techniques of diving that allowed to collect a greater quantity and greater diversity of specimens and to the development of powerful analytical techniques for the determination of the structure of their metabolites. In the period between 1990 and

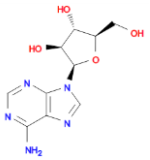
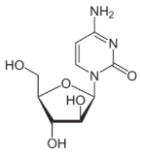
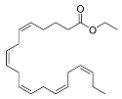
2005, an average of 800 new marine natural products were discovered each year and approximately 66% of them were isolated from marine invertebrates.

In the period between the years 2005 - 2012 the percentage dropped from 66% to 55%; However, in recent times interest has returned studies related to these compounds, mainly due to the advance in the development of more analytical tools, chemical synthesis, microbiology and molecular biology that have allowed to overcome partially the main obstacles in obtaining new compounds with biological activity, low extraction yields from natural marine sources (Karuppiah, Zhang & Li, 2015).

The advances mentioned so far have allowed the discovery of new compounds, which have been transformed into medicines by pharmaceutical industry. The marine drugs origin approved by FDA for human use are as follows:

Table 1. Approved drugs by FDA of marine origin (adapted from Marine-Derived Pharmaceuticals – Challenges and Opportunities

Compound name	Marine organism	Molecular target	Indication / Mode of Action	Approval date	Structure
Trabectedin (Ecteinascidin, ET-743, Yondelis®, PharmaMar, Colmenar Viejo, Spain)	Tunicate (*Ecteinascidia turbinate)*	anti-neoplastic compounds DNA (minor groove)	Soft tissue sarcoma and ovarian cancer	October 3, 2015	
Brentuximab Vedotin (Adcetris®, Takeda Pharma A/S, Taastrup, Denmark)	Mollusk/cyanobacteria	antibody directed against CD30, microtubules	Hodgkin lymphoma disease or systemic anaplastic large cell lymphoma T-cell	August 19, 2010	
Eribulin mesylate (Halaven®, Eisai Europe Ltd., Hatfield, Hertfordshire, UK)	Sponge (*Halichondria okadai* and *Lyssodendoryx* sp.)	Microyubules	Metastic breast cancer	November 15, 2010	
Ziconotide (Prialt®, Jazz Pharmaceuticals plc, Dublin, Ireland)	Cone Snail (*Conus magus)*	DNA polymerase	Chronic pain	December 28, 2004	

Vidarabine (arabinofuranosyladenine, adenin arabinoside, Ara-A®)	Sponge	Viral DNA polymerase	Herpes simplex virus infection	1976, current status: discontinued/ replaced by acyclovir (Zovirax®) 2001	
Cytarabine (e.g. Cytosar®-U, Pfizer, Kirkland, Quebec, Canada)	Sponge *(Cryptotethya crypta)*	DNA polymerase	Leukemia	1969	
Omega-3-acid ethyl esters	Fish	Triglyceride-producing enzymes	Hypertriglyceridemia	November 10, 2004	

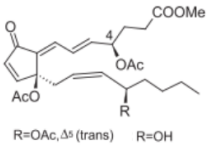
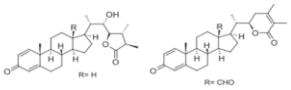
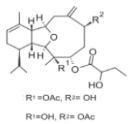
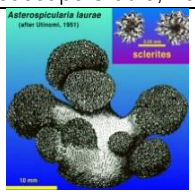
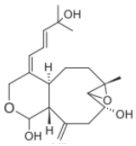
1.2.2 Bioactive compounds isolated from octocorals

After the sponges, the octocorals constitute the most studied (Chemical) marine group around the world (Heckrodt & Mulzer, 2005). Chemical compounds occurring in his organisms are an important factor in its evolution, because they use them as a chemical defense mainly, as anti predatory strategy to achieve the survival of its species (Sammarco & Coll, 1992). This reason among others, attribute the name of especialized compounds instead of secondary metabolites to these important groups of metabolites.

The chemical diversity of the octocorals is reflected in an extensive review carried out by Blunt and his collaborators in 2015, which shows that from the Cnidaria phylum until that date, 642 species had been studied (Blunt et al., 2015). In this sense, you can also consult reviews on octocoral chemistry published by (Coll, 1992; Rodríguez, 1995; Berrue & Kerr, 2009a; Blunt et al., 2012, 2013, 2014, 2015, 2016, 2017, 2018) and the reviews of diterpenes published by (Hanson, 2007).

The diterpene compounds found in these invertebrates are structurally interesting due to their novelty, which suggests new and more effective mechanisms of bioactivity. Among the isolated compounds of this order we can highlight those of the cembrane, pseudopterane, amphilectane, dolabellane, serrulatane and briarane nuclei (Berrue & Kerr, 2009b). Nowadays, cembranes, amphilectanes and briaranes continue to be the most abundant groups of compounds among the diterpenes isolated from octocorals. Several studies, have shown the potential of cytotoxic activity of these invertebrates against different cancer cell lines, as shown in table 2.

Table 2. Compounds with cytotoxic activity extracted from octocorals

Compound	Organism, photo, reference	Structure	Cytotoxic activity IC_{50}
Claviridin A-D	 *Clavularia viridis* (Quoy & Gaimard, 2015)		0.4–0.22 µM HepG2 and HeLa cell lines (Duh et al., 2007)
Paraminabeolide A-D	 *Paraminabea acronocephala* (Chao et al., 2013)		8.0 µM HepG2 (hepatocytes) (Chao et al., 2013)
Cladielloide A-B	 *Cladiella (Alcyoniidae)* (Seascape Studio, 2015)		0.009 µM leukemia tumor cells (CCRF-CEM) (Sung & et al., 2011)
Asterolaurin L	 *Asterospicularia laurae* (California Academy of Sciences, 2015)		11 µM laryngeal carcinoma (HEp-2), breast carcinoma (MCF-7) (Benayahu et al., 2004).

In Colombia there are several scientific studies which show potential bioactive of the Caribbean octocorals (Wei et al., 2004; Marrero, Rodríguez & Barnes, 2005; Tello et al., 2009; Zea, Henkel & Pawlik, 2014), nevertheless, there are few studies on octocorals as a source of compounds with cytotoxicity against cancer, considering that Colombia is a country with extend biodiversity with coasts on two oceans, and with the second largest coral reef in extension in the Caribbean Sea which allows that the octocorals grow profusely and with a marked chemo-typic variation (Duque et al., 2006; Berrue & Kerr, 2009b). Therefore, it is important to continue making studies that allow us to foster the knowledge about these invertebrates, particularly in the field of their potential as a source of anticancer agents (Vollmer et al., 2013).

1.2.3 Cancer among the leading causes of mortality / Types of treatment and methods most used to determine the anti-cancer activity of natural products

The cases of deaths by different cancer types in the World estimated by the International Agency for Research on Cancer (IARC), report the following: 9.6 million cancer deaths in 2018. In both women and men lung cancer is the most commonly diagnosed (11.6% of the total cases) and also is the leading cause of death in both sexes (18.4% of the total cancer deaths), followed by the female breast cancer (11.6%), the prostate cancer (7.1%), the colorectal cancer (6.1%), the stomach cancer (8.2%), and the liver cancer (8.2%) (International Agency for Research on, 2019)

Bearing in mind that the trend of cancer mortality in Colombia is increasing in both children and adults, it is necessary to highlight the importance of new research that allows the development of products that help to combat this problem, especially in cancer types of higher incidence and mortality for the country. It is estimated that by 2035 in Colombia deceases due to cancer will increase by 114% (Pardo & Cendales, 2010).

Although, there are many types of cancer treatment, these depend on the type of cancer and how advanced it is; most of the people who suffer from these diseases need a combination of treatments, such as surgery with chemotherapy and/or radiation therapy also immunotherapy, targeted therapy, or hormone therapy (Arruebo et al., 2011).

In recent times the precision medicine has been studied extensively due that this approach for patient care allows doctors to select treatments that are most likely to help patients based on a genetic understanding of their disease (Tran et al., 2015). Research studies are ongoing on now to test whether treating patients with treatments that target the cancer-causing genetic changes in their tumors, no matter where the cancer develops in the body, will help them. Many of these treatments are drugs known as targeted therapies (Schmidt et al., 2016).

The chemical and biological diversity of the marine environment is an extraordinary resource for the discovery of new anti-cancer drugs. As it has been exposed in the course where octocorals are an important part of biodiversity and that some of these organisms have shown to produce compounds that could be used in the treatment of cancer. It is evident the need to advance in studies related to its chemical composition using a rapid NMR and UPLC / MS based metabolomic which allow us to reduce the Time-Cost analysis in the searching molecules with cytotoxic potential.

1.2.4 Metabolomic Analysis

1.2.4.1 Definition

The word metabolomics, is derived from the Greek word (μταβολ'η) which means change and which was defined for the first time by Oliver *et al.* as "the quantitative and qualitative analysis of all molecules of low molecular weight (molecular weight less than 2000 uma) present in cells in a physiological state or particular development" (Oliver et al., 1998; Goodacre et al., 2004). Metabolisms are a record of the chemical processes that have already occurred in a biological system; therefore, they are good biomarkers of disorders or alterations in a previous phase in the system's metabolism. On the other hand, its concentration can control and/or modify the activity of proteins and genes, for these reasons the metabolomic analysis has increasing interests at present and it is becoming a popular tool in many fields of research. In the years prior to the arrival of metabolomics, most of the studies of metabolites were focused on the way in which these changed according to a given stimulus or alteration (for example, a genetic mutation or a variation in environmental temperature etc.). Currently, studies related to metabolites establish more precise descriptions of the phenotypes and bioactivity relationships of organisms with the metabolites that constitute them. (Fiehn, 2001).

The metabolomic analysis covers several steps from the experimental design, sample preparation, data acquisition through different analytical techniques generally LC-HRMS or NMR, the statistical analysis that allows understanding the results obtained in an assertive way and finally the dereplication that through the use of databases, helps the rapid identification of known compounds present in a mixture. Therefore, the metabolomic analysis should aim at all the classes of compounds, and at the same time ensure a high recovery, reproducibility and experimental robustness (Zani & Carroll, 2017).

1.2.4.2 Analytical techniques used in metabolomics

Currently, chromatography plays an important role in investigations involving metabolomic analysis. Until 2008, the most used technique was NMR; however, for 2010 the studies that used analytical technique liquid chromatography coupled to mass spectrometry (LC-MS), were equalized with respect to NMR, because this technique covers a wide range of mass molecular analysis and different characteristic polarities of the compounds that can be analyzed. HRMS and NMR are highly precise and sensitive methods, combined with liquid chromatography (LC) or gas chromatography

(GC) to increase the reliability of the results, since a new variable of time is added to the development of the metabolic process (MacNair, Lewis & Jorgenson, 1997; Shulaev, 2006).

The LC-HRMS coupling facilitates the identification and quantification of metabolites by reducing the complexity of the sample allowing the separation of these before detection. The coupling of this technique with an electrospray ionization source (ESI) is increasingly being used to detect metabolites in complex biological samples, since it is useful for the analysis of a range of non-polar and semi-polar compounds that are thermally stable. On the other hand, the most commonly used mass analyzers are those based on ion trap, quadrupoles and time of flight; Currently, the versatility of these analyzers has allowed us to make hybrid configurations (tandem) by identifying more precisely the metabolites present in a biological system by fragmenting ions generated in equipment such as the triple quadrupole (QqQ), quadrupole time flight (QTOF) or ion time trap flight (IT-TOF) among others (Patti, Yanes & Siuzdak, 2012). In LC-HRMS with ESI interface, normally reversed phase columns are used as C18 with internal particle diameters smaller than 2 µm, which allow separating compounds of medium polarity and polar columns such as aminopropyl columns, useful in separating compounds from high polarity. For its part, mass spectrometry is a highly sensitive technique (limit of detection = 1 pmol / L) that allows the identification of compounds with a high degree of confidence, based on high resolution spectra through the determination of molecular weight, fragmentation patterns, isotopic ratios and / or retention times (in the case of being coupled to a chromatograph). The coupling, allows to improve the detection of metabolites present in a biological sample (Wishart, 2008). For this reason, the LC-HRMS has allowed to solve a wide range of compounds especially of medium polarity (Webb, Bristow & Sargent, 2004). Regarding the NMR technique, the equipment contains electromagnets that work using magnetic field gradients, ranging from 90 MHz to 1000 MHz, for studies of metabolomics it is advisable to use fields higher than 600 MHz. It allows the realization of rapid and reproducible analyzes, with the possibility of identifying compounds without the need to submit the extract to cleaning processes, thus they present complex mixtures, they are non-destructive and quantitative analysis. The amount of sample needed for the analysis varies from 1 to 15 mg depending on the type of analysis and the equipment to be used; Additionally, it is a reliable tool for metabolomic studies.

1.2.4.3 Data processing

To make a comparison of the biological activity of extracts from different samples involved in a metabolomic study using LC-HRMS, specialized software is required at each stage of the process,

once the data are obtained, they must be submitted to a pretreatment which can be done using raw data on free software such as MS-DIAL that use LC-MS/MS or GC-MS, XCMS that is the most popular tool for LC-MS and also MZmine that provides an alternative solution for LC-MS raw data processing, and alignment of mz/ret. time features (Ernst et al., 2014).

Before data pretreatment, it is necessary the conversion of the data generated by the LC-HRMS equipment from the ".RAW" format to the netCDF or mzXML format, to obtain them in an extension compatible with the software that needs to be used in the pre-treatment; this can be done using the ProteoWizard software and the Command-Line Tool msconvert (Holman, Tabb & Mallick, 2014).

The data pretreatment steps for LC / MS data are:

a) The filtering stages, where the effects caused by the noise of the measurement or the baseline are eliminated.
b) The detection of peaks that allows to identify all the signals generated by real ions and avoids the detection of false positives.
c) The alignment that corrects the differences by changes in the retention times between the different chromatographic runs and allows to combine the data of the different samples.
d) Finally, the normalization which eliminates unwanted systematic variations in the samples (Katajamaa, Miettinen & Oresic, 2006). Once the data have been pretreated, the treatment process is performed, making an alignment of the chromatographic profiles obtained, given that the variations in retention times are a limitation in the use of separation techniques.

For the treatment of the data generated by NMR, it must be considered that the obtained signal is not directly usable by the bioinformatic algorithms, for which it is necessary to perform a series of basic operations that include the adjustment of the baseline level, suppression of the water signal, alignment of the peaks, normalizations, selection of variables, etc. This can be done using the NMRProcflow software (Jacob et al., 2017). Free software like MetaboAnalyst and Galaxy can be used for the result data visualization, networks, pathway mapping and integration of data (Galaxy).

Due to the high number of samples used in metabolomics, both for LC-HRMS and for NMR, it is necessary to use statistical tools, which include a dimensionality reduction process; this is precisely

what the main components analysis [(Principal Components Analisys) (PCA)] (Lindon & Nicholson, 2008).

1.2.4.4 Statistical analysis

Due to the fact that most of the samples studied in metabolomics are very complex in composition and its interpretation is difficult to interpret, a multivariate analysis of the results must be done to reduce the dimensionality of the data without losing information (Listgarten & Emili, 2005; Van den Berg et al., 2006). In metabolomics, the most used statistical analyzes are:

a) Principal component analysis (PCA), is a chemometric technique that groups large amounts of data based on the similarities and differences that exist between observations. It also allows to decrease the dimensionality while preserving the maximum amount of information possible (Metz et al., 2011).

b) Discriminant analysis of partial least squares (PLS-DA), is a classification method based on a regression technique called partial least squares (PLS), which models the association between a data table X and a response matrix Y; allows to build a model that separates the different groups of samples based on their X variables (Metz *et al.*, 2011).

c) Orthogonal discriminant analysis of partial least squares (OPLS-DA), which allows to separate the variation in the data matrix X into two parts: a part correlated with Y and a part orthogonal to Y; The OPLS model therefore comprises two modeled variations, the Y-predictive and the Y-orthogonal components.

This stage of the process is important, bearing in mind that the metabolomics approach integrates the results (called variables) obtained from the LC-HRMS technique, such as retention time (Tr), abundance of the ions and mass ratio charge (m/z) and for the case of NMR the chemical shift (ppm), with the results that are generated from the cytotoxic activity assay. This information is organized in a data matrix, to establish groupings, differences or similarities using the different statistical tools mentioned above, this matrix allows to visualize and interpret the correlations and/or the differences between observations and variables in an easier way, considering that the observations should go in the rows and the variables in the columns, in order to create a predictive model that serves as a tool that correlates cytotoxic activity vs metabolomic profile of different octocorals. Finally, for a rapid identification of natural products present in a mixture the dereplicación

is necessary by the comparison of spectral characteristics obtained from the chemical studies of the mixture (mass spectra and NMR spectra) (Buckingham, Blunt & Munro, 2012) with those shown by reference compounds in academic or commercial databases; to communicate the confidence of annotation and identification for known features, in untargeted metabolomics studies the classification system from the metabolomics community it's used. where there are five levels of confidence (Schrimpe-Rutledge et al., 2016).

a) LEVEL 5 - UNIQUE FEATURE has been established, unique features are identified via exact mass measurement accuracy. (Unique mass-to-charge ratio and retention time).

b) LEVEL 4 - MOLECULAR FORMULA Molecular formula identification of features is completed via isotope abundance distribution, charge state and adduct ion determination.

c) LEVEL 3 - TENTATIVE STRUCTURE Structural identification includes a unique match of the parent ion (MS1) data searched through literature and / or libraries and databases (in the case of marine natural products the most used database is the MarinLit database)

d) LEVEL 2 - PUTATIVE IDENTIFICATION Putative identification reveals probable structure using fragmentation data from literature and / or libraries and databases.

e) LEVEL 1 - VALIDATED IDENTIFICATION Identification is fully validated by confirming the structure and metabolite assignment using a reference standard (May & McLean, 2016).

1.3 RESEARCH QUESTION AND ITS JUSTIFICATION

The studies cited in the present theoretical framework show that the majority of drugs developed from natural products in commerce have been isolated from terrestrial organisms, followed by marine organisms and recently by microorganisms (Newman & Cragg, 2012; Blunt et al., 2013, 2014, 2015). A 2012 report from the Federal Drug Administration (FDA) (Marine Pharmacology) shows that so far there are a total of 28,500 marine natural products (MNPs) had been identified by the end of 2016 (Jiménez, 2018) of which, 262 compounds are in preclinical trials, 11 in clinical trials and have been approved and marketed for the treatment of various diseases (Malve, 2016). Table 2 of this manuscript shows examples of notable advances in the development of drugs from marine

organisms such as the Yondelis® (Trabectedina or ET-743), isolated from *Ecteinascidia turbinata*, a compound of marine natural origin approved for the treatment of two types of advanced human soft tissue sarcoma (Dias, Urban & Roessner, 2012), and Caliculon A isolated from octocoral *Eunicea* sp. which presented a strong cytotoxic activity against SR leukemia cell lines, UACC-62 melanoma and RXF 393 renal cancer among others (Wei, Nieves & Rodríguez, 2012).

The Colombian Caribbean Sea is one of the places in the world that contains very high biodiversity, which makes it an attractive source for the search for bioactive compounds. Among the organisms that inhabit the Caribbean Sea, the octocorals have been shown as one of the most abundant species (INVEMAR, 2010) and with great potential to supply compounds with important biologicals activities, which can be studied and identified using metabolomic and replication tools due that metabolomic methods can be utilized to screen diverse biological sources of potentially novel and pharmacologically-active drugs(Macintyre et al., 2014) Additionally, dereplication studies by ultraperformance liquid chromatography-tandem mass spectrometry (UPLC / MS) and nuclear magnetic resonance (NMR) spectroscopy can establish the chemical profile from soft coral extracts to identifying the compounds of interest at an early stage will aid in the isolation of the bioactive components (Floros et al., 2016).

1.3.1 Research question

Based on the premises discussed earlier in this manuscript, it is proposed to study the relationships of the chemical composition of some octocorals of the Colombian Caribbean with their cytotoxic activity using metabolomic and dereplication tools.

Consequently, the following research question is established in a bioprospective framework:

Is it possible to create a model based on the use of metabolomic and dereplication tools that allows the correlation of the cytotoxic activity with the chemical composition of the octocorals studied?

1.4 OBJECTIVES

1.4.1 General objective

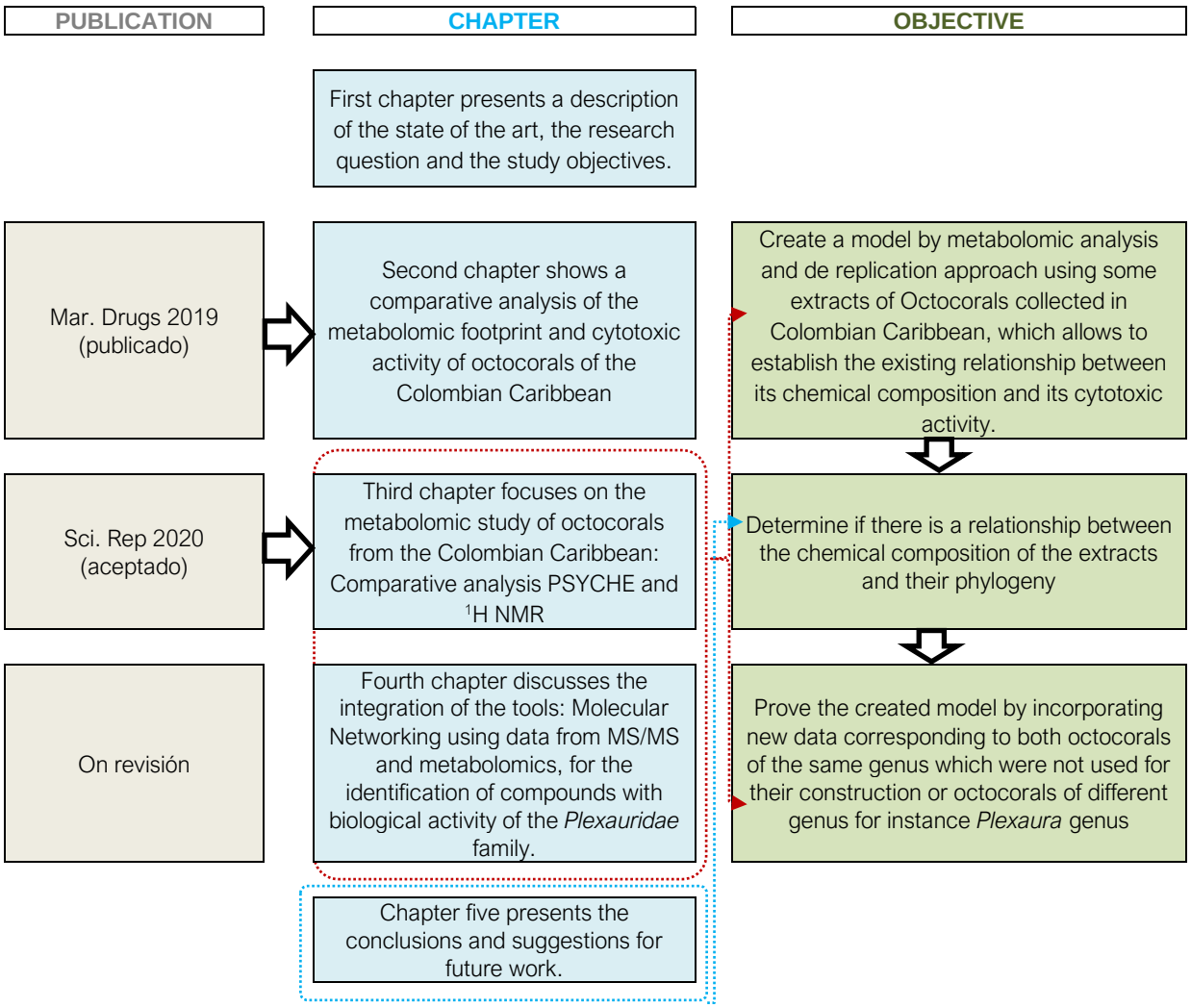
For bioprospective purposes, to establish a model that correlates the cytotoxic activity and chemical composition of octocorals of the Colombian Caribbean, within the framework of a

metabolomic and dereplication approach that also allows the identification of compounds with promising cytotoxic activity.

1.4.2 Specific Objectives

1. To create a model by metabolomic analysis and dereplication approach using some extracts of Octocorals collected in Colombian Caribbean, which allows to establish the existing relationship between its chemical composition and its cytotoxic activity.
2. Determine if is there a relationship between the chemical composition of the extracts and their phylogeny.
3. Prove the created model by incorporating new data corresponding to both octocorals of the same genus which were not used for their construction or octocorals of different genus for instance *Plexaura* genus.

The following figure represents the estructure of this thesis:



1.5 PUBLISHED PAPERS

The Research papers published as a result of the present work are:

1. Santacruz, L.; Thomas, O.P.; Duque, C.; Puyana, M.; Tello, E. Comparative analyses of metabolomic fingerprints and cytotoxic activities of soft corals from the colombian caribbean. Mar. Drugs 2019, 17, 37.
2. Santacruz, L; Hurtado, D; Doohan, R; Thomas, O; Puyana, M; Tello, E. Metabolomic study of soft corals from the Colombian Caribbean: PSYCHE and ^{1}H-NMR comparative analysis. Sci. Rep. 2020.

On Revisión:

3. Santacruz, L.; Thomas, O.P.; Duque, C.; Puyana, M.; Tello, E. Integration of molecular networking using ms/ms data and metabolomics tools for identification of compounds with biological activity from *plexauridae* family. 2020

1.6 REFERENCES

Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M, González-Fernández Á. 2011. Assessment of the evolution of cancer treatment therapies. *Cancers* 3:3279–3330. DOI: 10.3390/cancers3033279.

Avilés E, Prudhomme J, Le KG, Franzblau SG, Chandrasena K, Mayer AMS, Rodríguez AD. 2015. Bioorganic & Medicinal Chemistry Letters Synthesis and preliminary biological evaluation of a small library of hybrid compounds based on Ugi isocyanide multicomponent reactions with a marine natural product scaffold. :6–10. DOI: 10.1016/j.bmcl.2015.09.033.

Benayahu Y, Jeng MS, Perkol-Finkel S, Dai CF. 2004. Soft corals (Octocorallia: Alcyonacea) from Southern Taiwan. II. Species diversity and distributional patterns. *Zoological Studies* 43:548–560.

Bennett D. 2005. Growing Pains for Metabolomics. *Scientist* 19:25–28. DOI: 10.1016/S1081-1206(10)60393-6.

Benson JR, Jatoi I. 2012. The global breast cancer burden. *Future Oncology* 8:697–702.

Van den Berg R, Hoefsloot H, Westerhuis J, Smilde A, Van der Werf M. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics* 7:1–15. DOI: 10.1186/1471-2164-7-142.

Berrue F, Kerr RG. 2009a. Diterpenes from gorgonian corals. *Natural Product Reports* 26:681–710. DOI: 10.1039/b821918b.

Berrue F, Kerr RG. 2009b. Diterpenes from gorgonian corals. *Natural product reports* 26:681–710. DOI: 10.1039/b821918b.

Bhakuni D, Rawat D. 2005. Bioactive metabolites of marine algae, fungi and bacteria. *Bioactive marine natural products*.

Blunt JW, Carroll AR, Copp BR, Keyzers RA, Davis RA. 2018. Marine natural products. *Natural product report*. DOI: 10.1039/C7NP00052A.

Blunt J, Copp B, Hu W. 2008. Marine natural products. *Natural product reports* 25:35–94. DOI: 10.1039/b701534h.

Blunt J, Copp B, Keyzers R, Munro M, Prinsep M. 2012. Marine natural products. *Natural product reports* 29:144–222. DOI: 10.1039/c2np00090c.

Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2013. Marine natural products. *Natural product reports* 30:237–323. DOI: 10.1039/c2np20112g.

Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2014. Marine natural products. *Natural product reports* 31:160–258.

Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2015. Marine natural products. *Natural product reports* 32:116–211. DOI: 10.1039/c4np00144c.

Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2017. Natural Product Reports. *Natural product reports* 34:235–294. DOI: 10.1039/C6NP00124F.

Blunt JW, Copp BR, Keyzers RA, Munto MHG, Prinsep MR. 2016. Marine natural products. *Natural Product Reports* 33:382–431. DOI: 10.1039/c8np00092a.

Buckingham J, Blunt J, Munro M. 2012. The role of databases in Marine Natural Products research. In: *Handbook of Marine Natural Products*. Christchurch, New Zealand, 1–4. DOI: 10.1007/978-90-481-3834-0.

California Academy of Sciences. 2015.*asterospicularia laurae*. *Available at http://researcharchive.calacademy.org/research/izg/asterospicularia.htm* (accessed November 7, 2015).

Chao C-H, Wu Y-C, Wen Z-H, Sheu J-H. 2013. Steroidal carboxylic acids from soft coral Paraminabea acronocephala. *Marine drugs* 11:136–145. DOI: 10.3390/md11010136.

Chow S-C, Shao J. 2002. *Statistics in Drug Research: Methodologies and Recent Developments*.

Coll JC. 1992. The chemistry and chemical ecology of octocorals (Coelenterata, Anthozoa, Octocorallia). *Chemical Reviews* 92:613–631. DOI: 10.1021/cr00012a006.

Dias DA, Urban S, Roessner U. 2012. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* 2:303–336. DOI: 10.3390/metabo2020303.

Duh C, Lo Iw, Wang S, Dai C. 2007. New cytotoxic steroids from the soft coral Clavularia viridis. *Steroids* 72:573–9. DOI: 10.1016/j.steroids.2007.03.010.

Duque C, Puyana M, Castellanos L, Arias A, Correa H, Osorno O, Asai T, Hara N, Fujimoto Y. 2006. Further studies on the constituents of the gorgonian octocoral Pseudopterogorgia elisabethae collected in San Andrés and Providencia islands, Colombian Caribbean: isolation of a putative biosynthetic intermediate leading to erogorgiaene. *Tetrahedron* 62:4205–4213. DOI: 10.1016/j.tet.2006.02.032.

Ernst M, Silva DB, Silva RR, Vêncio RZN, Lopes NP. 2014. Mass spectrometry in plant metabolomics strategies: From analytical platforms to data acquisition and processing.

Natural Product Reports 31:784–806. DOI: 10.1039/c3np70086k.

Fiehn O. 2001. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and functional genomics*.

Fiehn O. 2002. Metabolomics – the link between genotypes and phenotypes. :155–171.

Floros DJ, Jensen PR, Dorrestein PC, Koyama N. 2016. A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* 12. DOI: 10.1007/s11306-016-1087-5.

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. 2004. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in biotechnology* 22:245–52. DOI: 10.1016/j.tibtech.2004.03.007.

Hanson JR. 2007. Diterpenoids. *Natural product reports* 24:1332–41. DOI: 10.1039/b705951p.

Heckrodt TJ, Mulzer J. 2005. Marine Natural Products from Pseudopterogorgia elisabethae: Structures, Biosynthesis, Pharmacology, and Total Synthesis. 244:1–41. DOI: 10.1007/b96886.

Holman JD, Tabb DL, Mallick P. 2014. Employing ProteoWizard to convert raw mass spectrometry data. *Current Protocols in Bioinformatics*:1–9. DOI: 10.1002/0471250953.bi1324s46.

International Agency for Research on C. 2019. Globocan 2018 World Report. 876:2018–2019.

INVEMAR. 2010. *Biodiversidad del Margen Continental del Caribe Colombiano Biodiversidad del Margen Continental del Caribe Colombiano*.

Jacob D, Deborde C, Lefebvre M, Maucourt M, Moing A. 2017. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. *Metabolomics* 13:36. DOI: 10.1007/s11306-017-1178-y.

Jiménez C. 2018. Marine Natural Products in Medicinal Chemistry. *ACS Medicinal Chemistry Letters* 9:959–961. DOI: 10.1021/acsmedchemlett.8b00368.

Karuppiah V, Zhang F, Li Z. 2015. Natural Products with Anticancer Activity from Marine Fungi. :253–267. DOI: 10.1007/978-3-319-07145-9.

Katajamaa M, Miettinen J, Oresic M. 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics (Oxford, England)* 22:634–6. DOI: 10.1093/bioinformatics/btk039.

Lindon JC, Holmes E, Bollard ME, Stanley EG, Nicholson JK. 2004. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* 9:1–31. DOI: 10.1080/13547500410001668379.

Lindon JC, Nicholson JK. 2008. Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics. :45–69. DOI: 10.1146/annurev.anchem.1.031207.113026.

Listgarten J, Emili A. 2005. Statistical and Computational Methods For Comparative Proteomic Profiling Using Liquid Chromatography-Tandem Mass Spectrometry. *Bioinformatics*:1–48. DOI: 10.1074/mcp.R500005-MCP200.

Macintyre L, Zhang T, Viegelmann C, Martinez IJ, Cheng C, Dowdells C, Abdelmohsen UR,

Gernert C, Hentschel U, Edrada-Ebel RA. 2014. *Metabolomic tools for secondary metabolite discovery from marine microbial symbionts*. DOI: 10.3390/md12063416.

MacNair JE, Lewis KC, Jorgenson JW. 1997. Ultrahigh-Pressure Reversed-Phase Liquid Chromatography in Packed Capillary Columns. *Analytical Chemistry* 69:983–989. DOI: 10.1021/ac961094r.

Malve H. 2016. Exploring the ocean for new drug developments: Marine pharmacology. *Journal of Pharmacy and Bioallied Sciences* 8:83–91. DOI: 10.4103/0975-7406.171700.

Marrero J, Rodríguez AD, Barnes CL. 2005. Intricarene, an unprecedented trispiropentacyclic diterpene from the Caribbean Sea plume Pseudopterogorgia kallos. *Organic letters* 7:1877–80. DOI: 10.1021/ol0505961.

May JC, McLean JA. 2016. Advanced Multidimensional Separations in Mass Spectrometry: Navigating the Big Data Deluge. *Annual Review of Analytical Chemistry* 9:387–409. DOI: 10.1146/annurev-anchem-071015-041734.

Metz TO, Metabolomics L, Bajad S, Shulaev V. 2011. Metabolic Profiling. *Methods and Protocols* 708:345. DOI: 10.1007/978-1-61737-985-7.

Momenimovahed Z, Salehiniya H. 2017. Incidence, mortality and risk factors of cervical cancer in the world. *Biomedical Research and Therapy* 4:1795. DOI: 10.15419/bmrat.v4i12.386.

Newman DJ, Cragg GM. 2012. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of natural products* 75:311–35. DOI: 10.1021/np200906s.

Oliver S, Winson M, Kell D, Baganz F. 1998. Systematic functional analysis of the yeast genome. *Trends in biotechnology* 16:373–378. DOI: 10.1016/s0167-7799(98)01214-1.

Pardo C, Cendales R. 2010. *Incidencia estimada y mortalidad por cáncer en Colombia 2002-2006*. Bogotá: Instituto Nacional de Cancerología.

Patti GJ, Yanes O, Siuzdak G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews. Molecular cell biology* 13:263–9. DOI: 10.1038/nrm3314.

Quoy, Gaimard. 2015. Clavularia viridis. *biolib*.

Robosky LC, Robertson DG, Baker JD, Rane, S. & Reily M. 2002. In vivo toxicity screening programs using metabonomics. *Combinatorial Chemistry & High Throughput Screening* 5:651–662. DOI: 10.2174/1386207023329932.

Rodríguez AD. 1995. The natural products chemistry of West Indian gorgonian octocorals. *Tetrahedron* 51:4571–4618. DOI: 10.1016/0040-4020(95)00216-U.

Sammarco PW, Coll JC. 1992. Chemical adaptations in the Octocorallia: evolutionary considerations. *Marine Ecology Progress Series* 88:93–104. DOI: 10.3354/meps088093.

Schmidt KT, Chau CH, Price DK, Figg WD. 2016. Precision Oncology Medicine: The Clinical Relevance of Patient-Specific Biomarkers Used to Optimize Cancer Treatment. *Journal of Clinical Pharmacology* 56:1484–1499. DOI: 10.1002/jcph.765.

Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. 2016. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry* 27:1897–1905. DOI: 10.1007/s13361-016-1469-y.

Seascape Studio. 2015.Cladiella Alcyoniidae. *Available at*

http://www.seascapestudio.net/reference/cnidarian.php?id=156 (accessed November 7, 2015).

Shulaev V. 2006. Metabolomics technology and bioinformatics. *Briefings in bioinformatics* 7:128–39. DOI: 10.1093/bib/bbl012.

Sung P-J, et al. et al. 2011. ChemInform Abstract: Cladieunicellins A-E, New Eunicellins from an Indonesian Soft Coral *Cladiella* sp. *ChemInform* 42:no-no. DOI: 10.1002/chin.201134191.

Tello E, Castellanos L, Arevalo-Ferro C, Duque C. 2009. Cembranoid diterpenes from the Caribbean sea whip *Eunicea knighti*. *Journal of Natural Products* 72:1595–1602. DOI: 10.1021/np9002492.

Tran NH, Cavalcante LL, Lubner SJ, Mulkerin DL, Loconte NK, Clipson L, Matkowskyj KA, Deming DA. 2015. Precision medicine in colorectal cancer: The molecular profile alters treatment strategies. *Therapeutic Advances in Medical Oncology* 7:252–262. DOI: 10.1177/1758834015591952.

Vollmer S V, Baker AC, Coffroth M-A, Harvell CD, Medina M. 2013. Understanding the coral holobiont through science and scuba. *Smithsonian Contributions to the Marine Sciences* 39:173–186.

Webb K, Bristow T, Sargent M. 2004. *Methodology for Accurate Mass Measurement of Small Molecules Best Practice Guide*.

Wei X, Nieves K, Rodríguez AD. 2012. Bioactive cubitane diterpenoids from a Colombian gorgonian species of the genus *Eunicea*. *Pure and Applied Chemistry* 84:1847–1855. DOI: 10.1351/PAC-CON-11-10-20.

Wei X, Rodríguez AD, Baran P, Raptis RG, Sánchez JA, Ortega-Barria E, González J. 2004. Antiplasmodial cembradiene diterpenoids from a Southwestern Caribbean gorgonian octocoral of the genus *Eunicea*. *Tetrahedron* 60:11813–11819. DOI: 10.1016/j.tet.2004.09.108.

Wishart DS. 2008. Quantitative metabolomics using NMR. *TrAC Trends in Analytical Chemistry* 27:228–237. DOI: 10.1016/j.trac.2007.12.001.

Zani CL, Carroll AR. 2017. Database for Rapid Dereplication of Known Natural Products Using Data from MS and Fast NMR Experiments. *Journal of Natural Products* 80:1758–1766. DOI: 10.1021/acs.jnatprod.6b01093.

Zea S, Henkel TP, Pawlik JR. 2014.The Sponge Guide: a picture guide to Caribbean sponges. *Available at http://www.spongeguide.org* (accessed December 15, 2015).

marine drugs

MDPI

2 COMPARATIVE ANALYSES OF METABOLOMIC FINGERPRINTS AND CYTOTOXIC ACTIVITIES OF SOFT CORALS FROM THE COLOMBIAN CARIBBEAN

Liliana Santacruz [1], Olivier P. Thomas [2,*], Carmenza Duque [3], Mónica Puyana [4] and Edisson Tello [1,*]

[1] Bioprospecting Research Group and Bioscience Doctoral Program, Faculty of Engineering, Campus Puente del Común, Universidad de La Sabana, 250001 Chía, Colombia; lilianasaci@unisabana.edu.co

[2] Marine Biodiscovery, School of Chemistry and Ryan Institute, National University of Ireland Galway (NUI Galway), University Road, H91 TK33 Galway, Ireland

[3] Departamento de Química, Universidad Nacional de Colombia, Carrera 30 # 45-03, 111321 Bogotá, Colombia; cduqueb@unal.edu.co

[4] Departamento de Ciencias Biológicas y Ambientales, Universidad Jorge Tadeo Lozano, Carrera 4 # 22-61, 110311 Bogotá, Colombia; monica.puyana@utadeo.edu.co

* Correspondence: olivier.thomas@nuigalway.ie (L.S.); edisson.tello@unisabana.edu.co (E.T.); Tel.: +353-(0)9-149-3563 (L.S.); +57-32-0699-5696 (E.T.)

Abstract: Soft corals (Cnidaria, Anthozoa, Octocorallia) are a diverse group of marine invertebrates that inhabit various marine environments in tropical and subtropical areas. Several species are recognized as prolific sources of compounds with a wide array of biological activities. Recent advances in analytical techniques, supported by robust statistical analyses, have allowed the analysis and characterization of the metabolome present in a single living organism. In this study, a liquid chromatography-high resolution mass spectrometry metabolomic approach was applied to analyze the metabolite composition of 28 soft corals present in the Caribbean coast of Colombia. Multivariate data analysis was used to correlate the chemical fingerprints of soft corals with their cytotoxic activity against tumor cell lines for anticancer purpose. Some diterpenoids were identified as specific markers to discriminate between cytotoxic and non-cytotoxic crude extracts of soft corals against tumor cell lines. In the models generated from the comparative analysis of PLS-DA for tumor lines, A549 and SiHa, the diterpene 13-keto-1,11-dolabell-3(E),7(E),12(18)-triene yielded a high score in the variable importance in projection. These results highlight the potential of metabolomic approaches towards the identification of cytotoxic agents against cancer of marine origin. This workflow can be useful in several studies, mainly those that are time consuming, such as traditional bioprospecting of marine natural products.

Keywords: metabolomics; soft-corals; *Pseudoplexaura flagellosa*; diterpenes; cytotoxic activity; LC-HRMSs

2.1 INTRODUCTION

Oceans cover around 71% of the Earth surface and host a great diversity of species, to which the marine environments have exerted a driving force, leading to new adaptive strategies and the synthesis of new metabolites [1]. Sessile and soft-bodied invertebrates, such as soft orals (Cnidaria, Anthozoa), have evolved particular metabolic pathways, leading to the production of chemical compounds as mechanisms of defense [2].

To date, seven compounds based on natural products of marine origin have been approved by the Food and Drug Administration (FDA) as pharmaceuticals. Among these, four have been specifically developed for the treatment of cancer: Cytarabine (Cytosar-U®, 1969 for the treatment of leukemia), eribulin mesylate (Halaven®, 2010 for the treatment of metastatic breast cancer), brentuximab vedotin (Adcetris®, 2011 for the treatment of anaplastic large T-cell lymphoma, and Hodgkin's lymphoma), and trabectedin (Yondelis®, 2015 for the treatment of soft tissue sarcoma and ovarian cancer) [3]. Considering that cancer is still a major health concern worldwide [4], it is necessary to find new and effective cytotoxic agents against cancer. In this context, marine organisms are still largely unexplored and are promising sources of bioactive compounds with potential cytotoxic activity.

Chemical studies over the last 60 years have revealed that soft corals have developed an extraordinary ability to produce a large variety of compounds with unique chemical structures, usually associated with a broad range of biological activities [5]. Among some of the most interesting bioactive metabolites from soft corals are eleutherobin, originally isolated from the Australian soft

coral, *Eleutherobia* sp., that exhibit anticancer properties [6] and the pseudopterosins from Caribbean *Antillogorgia elisabethae* that exhibit potent cytotoxic against five human cells lines (HeLa, PC-3, HCT116, MCF-7, and BJ), and anti-inflammatory and antimicrobial activity [7–11]. Due to the outstanding chemical diversity of soft corals, new tools that encompass the broad metabolome of a particular species belonging to this group will definitely open new opportunities to quickly target bioactive metabolites. Recent developments in analytical chemistry techniques, especially HPLC, MS, and NMR, have allowed the detection of thousands of metabolites with great sensitivity and specificity in a short time period [12]. Still, there is a prevailing question of whether metabolomic approaches could be used to quickly identify known or unknown metabolites as potential candidates for pharmaceutical applications [13].

Statistical analyses, such as multivariate data analyses, and particularly PCA (principal component analysis), group different samples through clustering or by determining outliers, whereas multivariate analysis methods allow the classification of the samples based on their cytotoxic activity against tumor cell lines. Thus, multivariate analyses can be used to correlate a set of metabolomic data to the results of a specific assay, making it possible to ascribe the metabolites that are likely involved in the detected cytotoxic activity against tumor cell lines [14].

The aim of this study was to establish a correlation between cytotoxic activity using three tumor lines, SiHa: Human cervical cancer, A549: Human lung adenocarcinoma, and PC3: Human prostatic carcinoma, and the chemical composition of 28 crude extracts from soft corals from the Colombian Caribbean. UPLC-HRMS (QToF) data were used to yield a metabolomic fingerprint of each crude extract based on terpenoids, [15,16], which are characteristic and well represented in these organisms and can be detected by the selected analytical method [17].

Orthogonal projections to latent structures discriminant analysis (OPLS-DA) was applied to the metabolomic data, considering that this analysis uses information in the Y matrix to decompose the X matrix into blocks of structured variation correlated and orthogonal to Y, respectively. In this investigation, OPLS-DA was used to discriminate the different extracts based on the metabolic fingerprints of all extracts and their cytotoxic activity against tumor cell lines, showing that extracts of *Eunicea clavigera* and *Pseudoplexaura flagellosa* were the extracts that mainly contributed to the separation. Additionally, partial least squares discriminant analysis (PLS-DA) was used to calculate the variable importance in projection (VIP) [18]. However, OPLS-DA can be used analogously to PLS-DA for discrimination, where the main benefit in interpretation using OPLS-DA compared to PLS-DA lies in the ability of OPLS-DA to separate predictive from non-predictive (orthogonal) variation [19]. Finally, the multivariate statistical analyzes established that a diterpenoid with a dolabellane type skeleton was responsible for the cytotoxic activity shown by the extracts against the SiHa and A549 cancer cell lines.

2.2 RESULTS

2.2.1 Processing and Untargeted Data Acquisition

A metabolomics workflow pathway was established for this research as is shown in Figure 1. In that way, a fingerprint metabolomics approach was applied to 28 soft coral extracts belonging to five different genera (*Plexaura*, *Antillogorgia*, *Eunicea*, *Plexaurella*, and *Pseudoplexaura*) by UPLC-HRMS in the positive ion mode, leading to 18,290 features analyzed with Galaxy 4.0 software [20]. The Script used for data processing is described in the Supplementary Table S3. The LC–MS data sets can be downloaded from the metabolomics repository, MetaboLights [21], with the reference study code, MTBLS777. Following this analysis, we established some quality criteria by which any variation of Quality Control (QC) samples around their mean (CVQC) > 30% was removed from the dataset; this reduced the matrix data from 18,290 to 12,060 features, representing a reduction of 34% of the total matrix.

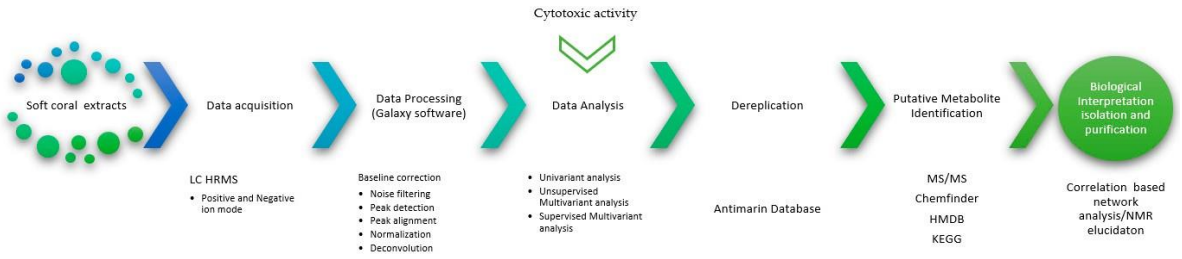


Figure 1. Metabolomic workflow pathway for 28 soft coral extracts. Data were processed with the Galaxy 4.0 software [20]. Metabolomic workflow included baseline correction, noise filtering, peak detection, peak alignment, normalization, deconvolution, and deisotoping.

Following our metabolomic workflow, three data matrices were generated. These data sets had the same number of features, but differed in the classification groups, following the in vitro cytotoxic activity of the soft coral extracts against the three cancer cell lines evaluated, human cervical cancer (SiHa), human lung adenocarcinoma (A549), and human prostatic carcinoma (PC3). L929 fibroblasts (L929; ATCC®CCL-1™) were used as the non-tumor cell line for toxicity control, however, the data obtained from L929 were not considered for the elaboration of the metabolomic matrix, due to the interest of this research.

For this research, it was established that an extract is considered active if it exhibited an inhibition of the tumor cell lines ≥50% at 20 µg/mL, following the guidelines of the National Cancer Institute (NCI) [22]. Supplementary Table S2 shows the results of the cytotoxicity presented by each extract against the tumor cell line, where the extract of the species, *Pseudoplexaura flagellosa* (code G17Ef), presented a value of 66.8% against SiHa and 71.5% against A549. The extract of the species, *Eunicea clavigera* (code C17Ec), presented a value of 61.3% against SiHa and 57.5% against A549. In agreement with Figures 2 and 3 and using the multivariate statistical analysis, PCA and OPLS-DA, it was visualized that these two extracts presented greater separation of the group

of extracts that presented cytotoxic activity for the tumor cell lines of A549 and SiHa.

2.2.2 Statistical Analysis

Using the data matrix (12,060 features), statistical analyses were run in the MetaboAnalyst platform [23]. PCA, orthogonal projection to latent structures discriminant analysis (OPLS-DA), and correlation analyses allowed the generation of variable importance in projection (VIP) scores. Figure 2 shows the results of the multivariate PCA considering their exposure to three different tumor lines, SiHa, A549, and PC3.

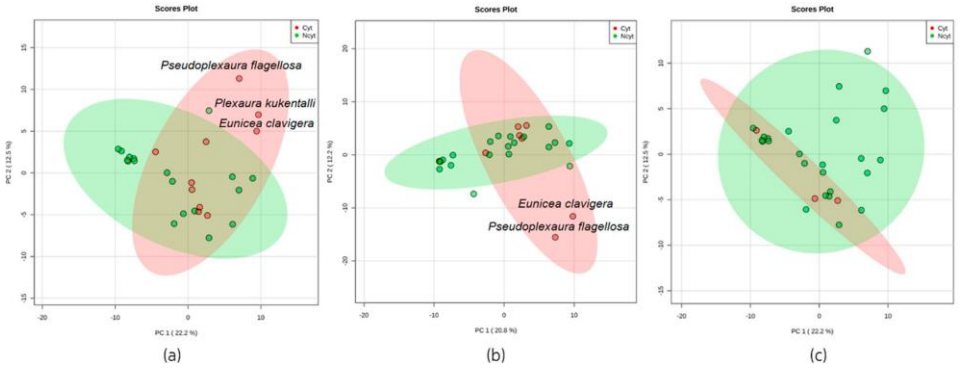


Figure 2. Principal component analysis score plot of metabolomics data from 28 extracts based on their cytotoxicity against three different cancer cell lines: (**a**) Human cervical cancer, SiHa, (**b**) human lung adenocarcinoma, A549, and (**c**) human prostatic carcinoma, PC3. Red dots indicate active extracts, and green dots represent extracts that were not active. An extract was considered active if it exhibited an inhibition of the tumor cell lines ≥50% at 20 µg/mL) [22]. The ellipses indicate confidence intervals of 95%.

From the unsupervised statistical analysis method using PCA, it was possible to determine that both *Eunicea clavigera* and *Pseudoplexaura flagellosa* (Figure 2a,b) were the extract of the species that mainly contributed to the separation from others that did not exhibit cytotoxicity against the cell lines, SiHa and A549. Therefore, this analysis suggests that both extracts might share some metabolites responsible for the detected cytotoxicity. Additionally, it was also found that the extract of *Plexaura kuekenthali* (Figure 2a) was also included in the group of separate species because it exhibited cytotoxicity against the cell line, SiHa. The PCA analysis for the extract of this species that exhibited cytotoxicity against the PC3 cell line did not show a clear separation that allowed identification of the clusters or outliers responsible for the observed cytotoxicity as seen in Figure 2c.

For better separation between classes in the hyperspace plot and visualization of the features responsible for the discrimination between bioactive and non-bioactive extracts as cytotoxic against tumor cell lines, the statistical method, OPLS-DA, was performed (Figure 3). Regarding the OPLS-DA models observed in Figure 3b,c, a clear separation is observed between the extract of the species that presented cytotoxicity from those that did not present. Additionally, in Figure 3a,b, it can be observed that the extract of the species, *Pseudoplexaura flagellosa* and *Eunicea clavigera*, are closely related to each other again for the models using the SiHa and A549 tumor cell lines.

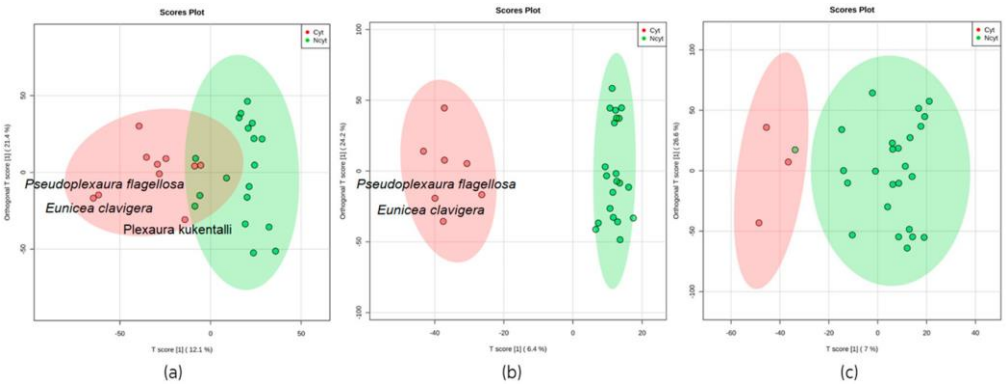


Figure 3. Supervised statistical analysis OPLS-DA score plots of metabolomic data of 28 soft coral extracts based on their cytotoxicity against three different cancer cell lines: (**a**) Human cervical cancer, SiHa, (**b**) human lung adenocarcinoma, A549, and (**c**) human prostatic carcinoma, PC3. Red dots indicate active extracts, and green dots represent extracts that were not active. An extract was considered active if it exhibited an inhibition of the tumor cell lines ≥50% at 20 µg/mL mL) [22]. The ellipses indicate confidence intervals of 95%.

To infer statistically significant discrimination (p-value ≤ 0.05) between the two classes (cytotoxic and non-cytotoxic against tumor cell lines), a cross validation test was performed for the classification model with the three tumor cell lines, where the models were evaluated using both R2 and Q2 metrics. R2 values report the total amount of variance explained by the model in both the data (R2X) and independent variables (R2Y); the Q2 reports model accuracy and the ratio, Q2/R2, is a measure of cross-validation reproducibility, when the value is above 0.5 is considered with relevant associations [24]. The results showed values of Q2/R2 greater than 0.5 for the lines, A549 and SiHa, but not when the model was made with the tumor line of PC3 (see Table 1). These results show that for the first two cell lines, a good classification model was obtained, indicating the reliability of the models. Considering that the statistical analysis produced unreliable results for the model established with the PC3 tumor line, the subsequent analyzes will contemplate the data derived from the models when the lines of SiHa and A549 were used.

Table 1. Partial least squares-discriminant analysis (PLS-DA) parameters and permutation test for distinguishing between cytotoxic and non-cytototoxic groups from 28 extracts of soft corals tested against three cancer cell lines: SiHa: Human cervical cancer, A549: Human lung adenocarcinoma, PC3: Human prostatic carcinoma.

	PLS-DA Parameters		
Cell Lines	Q2	R2	Q2/R2
SiHa	0.43	0.67	0.64
A549	0.34	0.65	0.52
PC3	−0.23	0.78	−0.29

Parameters based on Q2 indicate the best classifier of PLS-DA analyses using a 10-fold cross-validation method. PLS-DA: partial least squares-discriminant analysis, Q2: predictive capability, R2: correlation coefficients.

To complement the PLS-DA analyses, the molecular formula of the compounds was generated using the Agilent formula generator platform (MFG). Features that significantly contributed to clustering and discrimination were selected according to a threshold value of VIP ≥ 2.0 and a p value

< 0.05 [25].

According to the VIP analysis, 110 VIP were selected; Figure 4 shows the first 15 VIP for each cell line that mainly contributed to the separation of the extracts according to their cytotoxic potential.

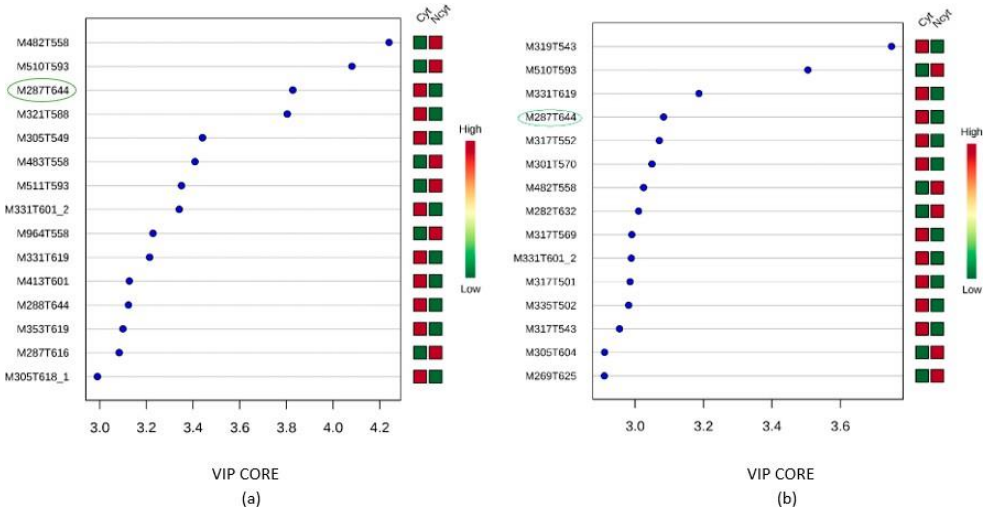


Figure 4. Results of the variable influence on projection (VIP) analyses to determine molecular features that contributed to extract clustering and discrimination in PLS-DA models against two cancer cell lines. Features were selected according to a threshold value of VIP ≥2.0 and a *p* value <0.05. The figures present the first 15 and most important features [26]. (**a**) SiHa; (**b**) A549.

As shown in Figure 4, the main feature that exhibited the greatest scores (SiHa score 3.1, A549 score 4.3) was M287T644, which was found in both models and seemed to be partly responsible for the separation. Other important features were M331T601 (SiHa score 3.0, A549 score 3.5) and M331T619 (SiHa score 3.2, A549 score 3.5).

2.2.2.1 Annotation, Dereplication, and Identification of the Feature M287T644

One of the great opportunities of metabolomics studies is the potential to identify new compounds by untargeted methods. A range of "dereplication" procedures are currently emerging to meet this challenge as key strategies for the identification of already known bioactive compounds and to improve the performance of natural product screening programs [27]. Here, we use a combinatorial approach for features selection with cytotoxic potential against tumor cell lines, using metabolomics analysis to establish the chemical profiles of soft coral extracts and dereplication using the AntiMarin® data base to obtain a putative identification of the structures. By this way, we established that the characteristic, M287T644, corresponds to *m/z* of 287.2374 $[M + H]^+$ consistent with the molecular formula, $C_{20}H_{30}O$, which was found in the extract of *Pseudoplexaura flagellosa*. Comparison with the database allowed us to propose two possible compounds for this molecular formula, eduenone [28] and dolabellatrienone [29].

To more precisely identify this feature, the MS/MS data of *m/z* 287.2374 $[M + H]^+$ were recorded

and analyzed using the MetFrag software [30]. The possibility that this feature corresponds to any of these two compounds, eduenone with a Diff of 0.59 ppm and dolabellatrienone with a Diff of 0.57 ppm, was confirmed. Considering that this feature was detected in the models made using the two tumor cell lines as one of the main features responsible for the separation of the group that presented cytotoxic activity, we decided to isolate the feature using repetitive HPLC-UV purifications. This compound was isolated as a yellowish oil with a molecular formula of $C_{20}H_{30}O$ assigned based on HRESI-MS. The NMR spectroscopic features of this compound were indicative of a dolabellane compound similar to eduenone and dolabellatrienone. Some key ^{1}H NMR signals allowed a quick identification of this

compound: The chemical signals of the olefinic methine group in H-3 δ_H 5.24, dd, *J* = 11.4, 5.3 Hz, (eduenone: δ_H 6.30, br s; dolabellatriene: δ_H 5.24, dd, *J* = 11.2, 5.0 Hz), one cyclic-bearing methine in H-11 δ_H 2.83, d, *J* = 11.8 Hz, (eduenone: δ_H 2.99, br d, *J* = 12.0 Hz; dolabellatriene: δ_H 2.83, br d, *J* = 12.2 Hz), and one methyl group located in H-15 δ_H 1.23, s, (eduenone: δ_H 1.14, s; dolabellatriene: δ_H 1.23, s) were all consistent with the dolabellatriene. Therefore, based on the above results and the use of ^{1}H-^{1}H COSY and HMBC spectra, the structure of this compound was established as shown

in Figure 5, a known dolabellane diterpenoid named 13-keto-1,11-dolabell-3(*E*),7(*E*),12(18)-triene previously isolated from *Eunicea calyculata* [31].

To evaluate the cytotoxicity of the compound, 13-keto-1,11-dolabell-3(E),7(E),12(18)-triene, two tumor cell lines, A549 and SiHa, were used in the MTT assay under five different concentrations, the dolabellatrienone exhibited values of IC_{50} = 0.02 µg/mL against A549 and IC_{50} = 0.03 µg/mL against SiHa.

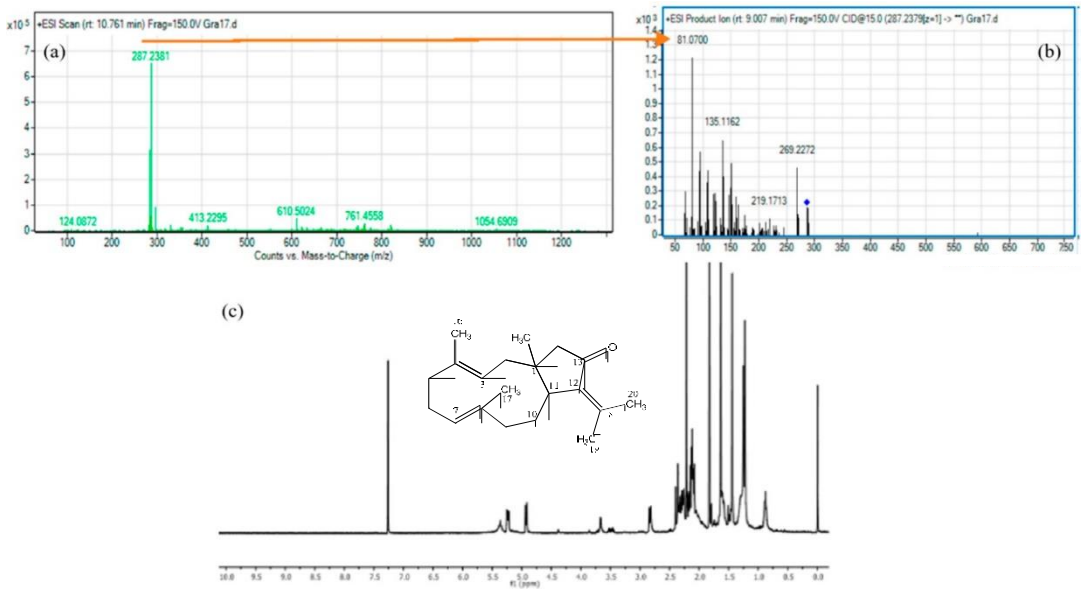


Figure 5. (**a**) MS spectrum of 13-keto-1,11- dolabell-3(E),7(E),12(18)-triene, isolated from the extract of *Pseudoplexaura flagellosa*

(**b**) MS/MS spectrum of compound *m/z*: 287.2374 $[M + H]^+$. (**c**). ^1H NMR spectrum of 13-keto-1,11-dolabell-3(E),7(E),12(18)-triene and chemical structure of $C_{20}H_{30}O$.

2.3 DISCUSSION

In the search for bioactive substances, marine organisms, such as soft corals, have led to metabolites with significant cytotoxic activities against different cancer cell lines [32]. In that way, this work established a metabolomic workflow pathway (Figure 2) that allowed correlation between the cytotoxic activities of 28 crude extracts from soft corals with their chemical composition. As seen in Figure 3, the statistical analysis, OPLS-DA, which seeks for maximal variance between the latent components, showed discrimination between the extracts that presented potential cytotoxicity against the extracts with lower cytotoxic potential. In addition, the data of the selected features were found to contribute to the separation of the extracts that presented cytotoxic potential and was found to be in common for the two models that presented a significant statistical difference (SiHa and A549), as shown in Figure 4. Considering this, it was decided to isolate the compound that corresponds to this feature from the extract identified as *Pseudoplexaura flagellosa*. This compound was identified by NMR and HRESI-MS analysis as 13-keto-1,11-dolabell-3(*E*),7(*E*),12(18)-triene $C_{20}H_{30}O$, *m/z* 287.2374 $[M + H]^+$, a dolabellane compound previously isolated from *Eunicea calyculata* by Look and Fenical 1982 [31].

According with these metabolomics results, it is important to show some studies on compounds with dolabellane skeletons obtained from different marine organisms, which have shown cytotoxic potential against different tumor cell lines, scilicet: The compound, clavirolide G, isolated from the soft coral, *Clavularia viridis*, collected from the Xisha Islands in the South China Sea showed moderate cytotoxic activity against KB and HL-60 cells with IC_{50} values of 5.12 μg/mL and 5.92 μg/mL, respectively [33]. The compounds, clavinflols A and B, from the *Clavularia inflata* collected in Green

Island, exhibited cytotoxicity against human oral epidermoid carcinoma (KB) cells (ED_{50} = 0.35 μg/mL) and showed selective activity towards human Hepa cells (ED_{50} = 1.2 μg/mL), respectively [34]. The compound, casearimene A, isolated from the species, *Casearia membranacea*, showed marginal activity against the tumor cell line, A549 (ED_{50} > 50 μg/mL) [35]. Also, five cytotoxic dolabellane diterpenes isolated in 2001 by Chang et al. from the Formosan soft coral, *Clavularia inflata*, showed moderate cytotoxicity against the A549 tumor cell line (ED_{50} = 7.74 to 50 μg/mL). Additionally, the compound, 7-hydroperoxydolabella-4(16),8(17),11(12)-triene-3,13-dione, was the most promising (ED_{50} = 0.57 μg/mL) [36]. The above studies demonstrated that diterpenes with a dolabellane skeleton

have presented low and moderate cytotoxic activities against different tumor cell lines, as was presented against the tumor cell line, A549, which agrees with the results obtained in this research in which the compound, dolabellatrienone, exhibited values of IC_{50} = 0.02 µg/mL against A549, IC_{50} = 0.03 µg/mL against SiHa, and IC_{50} = 64.95 µg/mL against L929.

Finally, the main prospective application of this fingerprint metabolomic analysis in soft corals was aimed at the identification of potential metabolites with cytotoxic activity against cancer cell lines. This kind of metabolomic process can be useful in a several studies, mainly those that are time consuming, such as traditional bioprospecting of marine natural products; this takes around 1 to 3 years (depending on the sample and the metabolites), while the workflow established in this investigation took around four to six months to predict the presence and the identification of biologically active compounds (VIP) from soft coral extracts.

2.4 MATERIALS AND METHODS

2.4.1 Materials

Solvents used for extraction, methanol, and dichloromethane were purchased from Merck (Darmstadt, Germany). For the chromatographic and spectrometric analysis, Acetonitrile, methanol, and formic acid of LC-MS were purchased from Sigma Aldrich (Dublin, Ireland). For use in cell culture, D-MEM (Dulbecco's Modified Eagle Medium) (1X) and RPMI 1640 (Roswell Park Memorial Institute, Darmstadt, Germany were made by Gibco / Invitrogen, Paisley, UK. Fetal bovine serum (FBS), brand Eurobio (Les Ulis, France). Trypticase soy broth (TSB) and trypticase soy agar (TSA) brand Scharlau Co (Barcelona, Spain), PC3 cell line (prostate cancer), extracted from prostatic adenocarcinoma of a Caucasian man (ATCC® CRL1435™), SiHa cancer cervical cell line (ATCC® HTB-35™), and A549 cancer lung cell line (ATCC® CCL-185™), and L929 fibroblast (ATCC® CCL-1™).

2.4.2 Methods

2.4.2.1 Soft Coral Material Collection and Identification

Twenty-eight samples of soft corals (Supplementary Table S1) were collected at Santa Marta Bay, Colombia, in Punta Venado (N = 11° 16.26′ 87″; W = 74° 12.24′ 58″) at depths between 10 to 20 m. Small terminal fragments (of approximately 30 cm) were cut off the main soft coral colony with sharp scissors. Samples were air dried and then kept frozen until the moment of extraction at −80 °C.

Soft corals were identified by morphological and sclerite analyses [37–39]. For sclerite preparations, a small distal fragment of each sample was treated with 5% sodium hypochlorite. Once the organic matter was removed, sclerites were washed with distilled water and centrifuged

at least four times. A final wash with ethanol followed by oven drying yielded sclerite preparations. Analysis of the sclerites was performed by microscopy [39,40]. Vouchers of all samples are stored at the Invertebrate collection of Instituto de Ciencias Naturales at Universidad Nacional de Colombia (Bogotá, Colombia) (Supplementary Table S1).

2.4.2.2 General Experimental Procedures

UV measurements were obtained by the extraction of the diode array detector (DAD) signal in a PerkinElmer HPLC-DAD-ELSD FLEXAR LC(r) SYSTEM. High-resolution mass spectra (HRESIMS) were obtained with an Agilent 6540 mass spectrometer. Compound purification was carried out in a JASCO HPLC equipment, supplied with a PU4087 pump and a UV4070 UV/Vis detector using a preparative Phenyl-Hexyl OBD Column.

2.4.2.3 Soft Coral Extraction, Sample Preparation, and UPLC/MS Analyses

The data acquisition was performed in an Agilent 6540 that generated a mass spectra zip file of 596 megabytes. Then, all data was preprocessed using the platform, Galaxy, which allows the automation of pipelines, ensuring reproducibility [20]. Afterwards, 1.0 g of dried powder from each soft coral was extracted at room temperature with a mixture of 1:1 DCM/MeOH, three times (30 mL) using an ultrasonic bath for 20 min. Debris were removed by centrifugation two times at 12,000× *g* for 5 min. Solvents were evaporated and dried extracts were passed through a C18 cartridge, eluting with MeOH to remove salts. Subsequently, the extracts were concentrated. The samples corresponding to the specie, *Pseudoplexaura flagellosa* (Gra 17) (one responsible for the separation according to OPLS-DA analysis), was purified by preparative RP-HPLC with a Phenyl-Hexyl OBD Column (XSelect CSH, 19 mm × 250 mm, 5 µm) and the optimization of gradient profiles was performed by selecting the mobile phase used for HPLC UV detection at λ 254 nm during 40 min. of acquisition time. Elution was done using water (A) and acetonitrile (B), both containing 0.1% formic acid with a gradient elution of 20%–50% B over 20 min, 50%–80% B over 10 min, 80%–100% B in 1 min, and holding for 5 min.

2.4.2.4 Metabolomic Procedure

All samples used for metabolomics studies were analyzed using an Agilent 6540 mass spectrometer, which performs MS acquisition (*m*/*z* 300–3000) at 10 spectra/sec in high resolution mode at a resolution of 35,000 and mass accuracy of 1 ppm.

Electrospray Ionization (ESI) in positive and negative mode was used to ionize and detect compounds after chromatographic separations. General parameters of the MS1 mode source were gas flow of 12 L min^{-1}, gas temperature of 300 °C, voltage charge of 2000 V, fragmentor of 150 V, capillary voltage of 3500 V, nebulizer pressure of 30 psi, and octopole RF Peak of 700 V.

The QC samples were analyzed intermittently during the analytical studies to assess the variance observed in the data. Chromatographic separation was achieved using a Phenyl-Hexyl HPLC column (150 mm × 3.0 mm, 1.9 µm Poroshell, Agilent Technologies, Santa Clara, CA, USA). The mobile phase consisted of (A) water with 0.1% formic acid to improve ionization and (B) methanol (MeOH) with 0.1% formic acid.

The UPLC injection volume on each run was 1.0 µL. The UPLC run contained blanks. QC samples and pooled samples [41] were intercalated throughout the UPLC run to control for any acquisition-dependent variation. Samples were filtered using a 0.2 µm Whatman® membrane filter with a pore size of 0.2 µm (Merck, Germany) prior to injection.

Data were analyzed using the Agilent MassHunter Qualitative software (Version B.07.00). For formula generation, the Molecular Formula Generator algorithm (MFG) was used, which can automatically eliminate unlikely candidate compounds and rank the putative molecular formula according to their mass deviation, isotopic pattern accuracy, and elemental composition.

2.4.2.5 Statistical Analyses

When analyzing large data sets from different samples, data variability can become a relevant complication: Hence, it might be to judge the quality of the data and assess their analytical variance. Hence, the use of a quality control (QC) allowed the minimization of this variability [42]. The QC samples contain the average of all metabolites within all the samples that were analyzed (0.1 µL of each octocoral extract was placed in a 1 mL vial), which were homogenized before the injection. QC samples were analyzed intermittently for the duration of the analytical study to assess the variance observed in the data throughout the sample preparation, data acquisition, and data pre-processing. Replicate injections should provide identical data for each injection, however, analytical variance is observed. Replicate QC injections can be used to measure this variance throughout the analytical study. We used a PCA to quickly assess the reproducibility of QC samples in an analytical run, and to determine the variance of the metabolite feature [43], pools were used and QC samples greater than 30% standard deviation were removed from the dataset [44]. An exploratory data analysis was first performed using PCA (Supplementary Figure S1). Additionally, the obtained signals from the blank were considered as interference and were subtracted (sn threshold = 3) [45].

For an untargeted approach, UPLC-MS chromatograms in negative and positive ion mode were pre-treated using an open source, platform independent software called Galaxy [20]. With this software, it was possible to exclude noise from LC-MS profiles (Noise level 5.0 E3, all data points below this intensity level were ignored). Parameters for data processing included the centwave method, 15 ppm, mzwid 0.015, and minfrac 0.3. After exporting the processed data in tabular

format (.cvs file), further analysis of the data matrix was performed by MetaboAnalyst version 3.0 [23].

After PCA analyses, a PLS-DA (partial least square-discriminant analysis) and OPLS-DA (orthogonal partial least-squares-discriminant analysis) were performed with MetaboAnalyst software. OPLS-DA is a powerful tool for the analysis of qualitative data structures, while the prediction results are equivalent to classification using standard PLS-DA [19]. The PLS approach is a robust regression technique used for investigations of the relationship between two data sets. In this experiment, a discriminant classification was carried out using a PLS-DA based on the PLS algorithm, in which the discriminating variable was the cytotoxic activity. The combined application of PCA and OPLS-DA to spectral datasets yields valuable insights on both general spectral trends (PCA) and group-predictive spectral features (PLS) [46].

2.4.2.6 Cytotoxicity Assays

Cell lines, human cervical cancer SiHa, human lung adenocarcinoma A549, human prostatic carcinoma PC3 and L929 fibroblasts, which was used as non-tumor cell line for toxicity control, were cultured in DMEM and RPMI 1640 media supplemented with 10% heat-inactivated fetal bovine serum (FBS) and penicillin-streptomycin (1%), at 37 °C in a 5% CO_2 humidified atmosphere, until 100% confluence was achieved [47].

The in vitro cytotoxicity of the soft corals extracts was evaluated using the MTT method, which is a colorimetric assay based on the capacity of mitochondrial succinate dehydrogenase enzymes in live cells to reduce the yellow, water-soluble substrate, 3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT), into an insoluble, colored formazan product, which is measured spectrophotometrically following Mosmann [48] with modifications proposed by Denizot and Lang [49]. Cells were allowed to attach in 96-well plates (4.0×10^4 cells/well) during 24 h. After 24 h, the medium was replaced with the extracts to be tested resuspended in DMSO (at a concentration of 2 mg/mL).

Soft coral extracts were evaluated at a concentration of 20 µg/mL. The greatest DMSO concentration was 0.1%, which was not cytotoxic to any of the cell lines. Prior to the assay, the supernatant was removed and 100 µL of 12 mM MTT solution in sterile PBS was added to each well and incubated at 37 °C for 4 h. The solution was removed, and extracts in dimethyl sulfoxide (DMSO) were added to each well, followed by incubation at 37 °C for 15 min. Optical density at 595 nm was read in an iMarkTM Microplate Reader. Cells cultured without extracts and doxorubicin were used as controls [50,51]. All tests were performed in triplicate. The viability percentage was calculated with Equation (1) and the cell inhibition was calculated with Equation (2) [32].

$$\%viability = \frac{Abs_{sample}}{Abs_{control}} \quad 100 \qquad (1)$$

Where Abs_{sample} is the absorbance of the cells treated with the test extract and $Abs_{control}$ is the absorbance of cells not treated with the test extract.

$$\%cell\ inhibition = 100 - Cell\ Survival \qquad (2)$$

Extract toxicity is demonstrated by the inhibition of cell growth and division.

2.4.2.7 Structure Elucidation

Fraction SP-9 obtained from the extract of *Pseudoplexaura flagellosa* (Gra17) was purified and then analyzed by LC–MS to establish the *m*/*z* of this purified fraction, yielding a (*m*/*z* of 287.2373). NMR spectra were acquired in an Agilent 600 MHz spectrometer equipped with a cryoprobe with pulse field gradient, and signals were referenced in ppm to the residual solvent signals (CDCl3 at 7.26 ppm, TMS at 0.00 ppm) (Table S3, NMR data of feature M287T644).

2.5 CONCLUSIONS

In this work, metabolomics tools were successfully applied to compare the metabolomic profile of 28 extracts of soft corals. Data analyses using PCA, PLS-DA, and OPLS-DA were valuable to determine and highlight the potential cytotoxic metabolites from soft corals extracts. The Feature M287T644 was found as VIP in some extracts tested against SiHa and A549 cancer cell lines. Additionally, a dereplication process was important to putatively identify compounds suggested by statistical analyses as the main features explaining potential cytotoxicity in the tested extracts. The workflow established in this work led to the identification of the compounds 13-keto-1,11-dolabell-3(*E*),7(*E*),12(18)-triene as the main feature responsible for the separation of extracts with major cytotoxic potential, IC_{50} = 0.02 µg/mL against A549 and IC_{50} = 0.03 µg/mL against SiHa cancer cell lines, being in agreement with that reported in the literature for dolabellane type compounds.

Additionally, as shown in Table S2, the G17 extract (*Pseudoplexaura flagellosa*) showed cytotoxic activity against the SiHa and A549 cell lines and was also cytotoxic against L929; however, the metabolite (VIP), $C_{20}H_{30}O$ (13-keto-1,11-dolabell-3(E),7(E),12(18)-triene), isolated from this extract, which was the most responsible for the separation of the extracts of the soft corals that were classified as cytotoxic, did not reveal cytotoxicity against the normal cell line, L929, but presented moderate cytotoxicity against the cell line, A549.

The application of this fingerprint metabolomic analysis in soft corals was aimed at identifying potential metabolites with cytotoxic activity against tumor cell lines. This approach can be useful in several studies, like traditional bioprospecting of marine natural products, which are time consuming.

In that way, the workflow established in this investigation allows the prediction of the presence of biologically active compounds from soft coral extracts in short periods of times.

Supplementary Materials: The following are available online at http://www.mdpi.com/1660-3397/17/1/37/s1, Figure S1: Principal component analysis (PCA); in the triangle the pools (QC) are shown. Table S1: Name of species of soft corals used in this research a Collection Code (ICN), assigned by Collection of the Institute of Natural Sciences of the National University of Colombia, (Bogotá, Colombia). Table S2: Percentage of cytotoxic activity against three cancerous cell lines. (1 if it is above 50% 0 if it is below). Table S3: Scrip using Galaxy software to UPLC-MS data extraction.

Author Contributions: L.S., O.P.T., C.D. and E.T. conceived and designed the experiments; L.S., M.P. and E.T. collected the biological material; M.P. identified the soft corals, L.S. and E.T. performed the chemical experiments;

L.S. analyzed the data; L.S. and E.T. wrote the article; L.S., M.P., O.P.T. and E.T. edited and reviewed the article. All the authors have contributed to and approved the final manuscript.

Funding: Doctoral studies of L.S. were supported by grants by Colciencias Beca de Doctorado Nacional 647/2014 and Universidad de la Sabana. Universidad de la Sabana supported the Project INGPHD-5-2015 “Bioprospección de octocorales del Caribe Colombiano, como una fuente potencial de compuestos con actividad citotóxica”. Part of this project (Grant-Aid Agreement No.PBA/MB/16/01) was carried out with the support of the Marine Institute and was funded under the Marine Research Programme supported the Irish Government. Sergio Zapata Lopera from UTADEO, provided great help in obtaining, processing and measuring soft coral sclerites for soft coral identification purposes.

Acknowledgments: The authors acknowledge the Universidad de La Sabana for support towards the cytotoxic bioassays and Galway University towards NMR and MS equipment. We also thank Invertebrate collection of Instituto de Ciencias Naturales at Universidad Nacional de Colombia (Bogotá, Colombia) for curation of the octocoral samples and coded assigned. E.T. and L.S. would like to acknowledge L.D., D.X.H. and C.R. for the support in this project.

The collection of the softcorals used in this research were covered under the permission granted by the Ministerio de Ambiente y Desarrollo Sostenible de Colombia to Universidad de La Sabana, through the Contrato Marco de Acceso a Recursos Genéticos y sus Productos Derivados No. 117 de 2015 (Otrosí No. 3).

Conflicts of Interest: The authors declare no conflict of interest.

2.6 REFERENCES

1. Blunt, J.; Copp, B.; Keyzers, R.; Munro, M.; Prinsep, M. Marine natural products. *Nat. Prod. Rep.* **2012**, *29*, 144–222. [CrossRef] [PubMed]
2. Jiménez, J.; Marfil, A.; Francesch, C.; Cuevas, M.; Alvarez, A.; Albericio, F. Productos naturales de origen marino. *Investig. Cienc.* **2007**, *365*, 75–83.
3. Dyshlovoy, S.; Honecker, F. Marine Compounds and Cancer: 2017 Updates. *Mar. Drugs* **2018**, *16*, 41. [CrossRef] [PubMed]
4. Ruiz Torres, V.; Encinar, J.A.; Herranz López, M.; Pérez Sánchez, A.; Galiano, V.; Barrajón Catalán, E.; Micol, V. An updated review on marine anticancer compounds: The use of virtual screening for the discovery of small-molecule cancer drugs. *Molecules* **2017**, *22*, 1037. [CrossRef] [PubMed]
5. Bennett, D. Growing Pains for Metabolomics. *Scientist* **2005**, *19*, 25–28. [CrossRef]
6. Ojima, I.; Chakravarty, S.; Inoue, T.; Lin, S.; He, L.; Horwitz, S.B.; Kuduk, S.D.; Danishefsky, S.J. A common pharmacophore for cytotoxic natural products that stabilize microtubules. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4256–4261. [CrossRef] [PubMed]
7. Correa, H.; Valenzuela, A.L.; Ospina, L.F.; Duque, C. Anti-inflammatory effects of the gorgonian Pseudopterogorgia elisabethae collected at the Islands of Providencia and San Andrés (SW Caribbean). *J. Inflamm.* **2009**, *6*, 5. [CrossRef] [PubMed]
8. Correa, H.; Aristizabal, F.; Duque, C.; Kerr, R. Cytotoxic and antimicrobial activity of pseudopterosins and seco-pseudopterosins isolated from the octocoral Pseudopterogorgia elisabethae of San Andrés and Providencia Islands (Southwest Caribbean Sea). *Mar. Drugs* **2011**, *9*, 334–343. [CrossRef][PubMed]
9. Correa, H. Estudios de Bioprospección del coral blando Pseudopterogorgia Elisabethae como Fuente de Sustancias con Actividad Biológica Fase IV. Ph.D. Thesis, National University of Colombia, Bogotá, Colombia, 2012.
10. Marrero, J.; Rodríguez, A.D.; Baran, P.; Raptis, R.G. Ciereszkolide: Isolation and structure characterization of a novel rearranged cembrane from the caribbean sea plume *Pseudopterogorgia kallos*. *Eur. J. Org. Chem.* **2004**, 3909–3912. [CrossRef]
11. Look, S.A.; Fenical, W.; Jacobst, R.S.; Clardyt, J.O.N. The pseudopterosins: Anti-inflammatory and analgesic natural products from the sea whip *Pseudopterogorgia elisabethae*. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 6238–6240. [CrossRef] [PubMed]
12. Deng, L.; Gu, H.; Zhu, J.; Nagana Gowda, G.A.; Djukovic, D.; Chiorean, E.G.; Raftery, D. Combining NMR and LC/MS using backward variable elimination: Metabolomics analysis of colorectal cancer, polyps, and healthy controls. *Anal. Chem.* **2016**, *88*, 7975–7983. [CrossRef] [PubMed]
13. Lindon, J.C.; Holmes, E.; Bollard, M.E.; Stanley, E.G.; Nicholson, J.K. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **2004**, *9*, 1–31. [CrossRef] [PubMed]
14. Tistaert, C.; Chataigné, G.; Dejaegher, B.; Rivière, C.; Hoai, N.N.; Van, M.C.; Quetin-leclercq, J.; Heyden, Y. Vander Multivariate data analysis to evaluate the fingerprint peaks responsible for the cytotoxic activity of *Mallotus* species. *J. Chromatogr. B* **2012**, *910*, 103–113. [CrossRef] [PubMed]
15. Leal, M.C.; Madeira, C.; Brandão, C.A.; Puga, J.; Calado, R. Bioprospecting of marine invertebrates for new natural products—A chemical and zoogeographical perspective. *Molecules* **2012**, *17*, 9842–9854. [CrossRef] [PubMed]
16. Wei, X.; Rodríguez, A.D.; Baran, P.; Raptis, R.G. Dolabellane-type diterpenoids with antiprotozoan activity from a southwestern Caribbean gorgonian octocoral of the genus *Eunicea*. *J. Nat. Prod.* **2010**, *73*, 925–934. [CrossRef] [PubMed]
17. Maille, G.; Qin, C.; Siuzdak, G. Nonlinear Data Alignment for UPLC—MS and HPLC—MS Based Metabolomics: Quantitative Analysis of Endogenous and Exogenous Metabolites in Human Serum. *Anal. Chem.* **2006**, *78*, 3289–3295. [CrossRef]
18. Szymanska, E.; Saccenti, E.; Smilde, A.K.; Westerhuis, J.A. Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **2012**, *14*. [CrossRef] [PubMed]
19. Rantalainen, M.; Cloarec, O.; Nicholson, J.K.; Holmes, E.; Trygg, J. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *J. Chemiometr.* **2006**, *20*, 341–351. [CrossRef]

20. Goecks, J.; Nekrutenko, A.; Taylor, J.; Afgan, E.; Ananda, G.; Baker, D.; Blankenberg, D.; Chakrabarty, R.; Coraor, N.; Goecks, J.; et al. Galaxy: A comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biol.* **2010**, *11*, 13. [CrossRef]
21. Haug, K.; Salek, R.M.; Conesa, P.; Hastings, J.; De Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; et al. MetaboLights—An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **2013**, *41*, 781–786. [CrossRef]
22. Hostettman, K. *Methods in Plant Biochemistry. Assays for Bioactivity*; Academic Press: London, UK, 1991.
23. Xia, J.; Psychogios, N.; Young, N.; Wishart, D.S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, *37*, 652–660. [CrossRef] [PubMed]
24. Moltu, S.J.; Sachse, D.; Blakstad, E.W.; Strømmen, K.; Nakstad, B.; Almaas, A.N.; Westerberg, A.C.; Rønnestad, A.; Brække, K.; Veierød, M.B.; et al. Urinary metabolite profiles in premature infants show early postnatal metabolic adaptation and maturation. *Nutrients* **2014**, *6*, 1913–1930. [CrossRef] [PubMed]
25. Chiu, C.Y.; Yeh, K.W.; Lin, G.; Chiang, M.H.; Yang, S.C.; Chao, W.J.; Yao, T.C.; Tsai, M.H.; Hua, M.C.; Liao, S.L.; et al. Metabolomics reveals dynamic metabolic changes associated with age in early childhood. *PLoS ONE* **2016**, *11*, e0149823. [CrossRef] [PubMed]
26. Xia, J.; Sinelnikov, I.V.; Han, B.; Wishart, D.S. MetaboAnalyst 3.0—Making metabolomics more meaningful. *Nucleic Acids Res.* **2015**, *43*, W251–W257. [CrossRef] [PubMed]
27. Hubert, J.; Nuzillard, J.M.; Renault, J.H. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev.* **2017**, *16*, 55–95. [CrossRef]
28. National Center for Biotechnology Information Eduenone. Available online: https://pubchem.ncbi.nlm.nih.gov/compound/10424127 (accessed on 25 September 2018).
29. National Center for Biotechnology Information Dolabellatrienone. Available online: https://pubchem.ncbi.nlm.nih.gov/compound/10469260 (accessed on 25 September 2018).
30. Wolf, S.; Schmidt, S.; Müller Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform.* **2010**, *1*, 43. [CrossRef]
31. Look, S.A.; Fenical, W. New Bicyclic Diterpenoids from the Caribbean Gorgonian Octocoral *Eunicea* calyculata. *J. Org. Chem.* **1982**, *47*, 4129–4134. [CrossRef]
32. Patel, S.; Gheewala, N.; Suthar, A.; Shah, A. In-Vitro cytotoxicity activity of Solanum Nigrum extract against Hela cell line and Vero cell line. *Int. J. Pharm. Pharm. Sci.* **2009**, *1*, 38–47.
33. Gao, Y.; Xiao, W.; Liu, H.C.; Wang, J.R.; Yao, L.G.; Ouyang, P.K.; Wang, D.C.; Guo, Y.W. Clavirolide G, a new rare dolabellane-type diterpenoid from the Xisha soft coral *Clavularia viridis*. *Chin. Chem. Lett.* **2017**, *28*, 905–908. [CrossRef]
34. Shen, Y.C.; Pan, Y.L.; Ko, C.L.; Kuo, Y.H.; Chen, C.Y. New dolabellanes from the Taiwanese soft coral clavularia inflata. *J. Chin. Chem. Soc.* **2003**, *50*, 471–476. [CrossRef]
35. Chang, K.C.; Duh, C.Y.; Chen, I.S.; Tsai, I.L. A cytotoxic butenolide, two new dolabellane diterpenoids, a chroman and a benzoquinol derivative formosan *Casearia membranacea*. *Planta Med.* **2003**, *69*, 667–672. [CrossRef] [PubMed]
36. Duh, C.Y.; Chia, M.C.; Wang, S.K.; Chen, H.J.; El-Gamal, A.A.H.; Dai, C.F. Cytotoxic dolabellane diterpenes from the Formosan soft coral *Clavularia inflata*. *J. Nat. Prod.* **2001**, *64*, 1028–1031. [CrossRef] [PubMed]
37. Frederickm, M. *Bayer the Shallow-Water Octocorallia of the West Indian Region: A Manual for Marine Biologists*; Smithsonian Institution: Washington, DC, USA, 1988.
38. Sánchez, J.A.; Wirshing, H.H. A field key to the identification of tropical western Atlantic zooxanthellate octocorals (Octocorallia: Cnidaria). *Caribb. J. Sci.* **2005**, *41*, 508–522.
39. Sánchez, J.A. Sistemática Filogenética del Género *Eunicea Lamouroux*, 1816 (Octocorallia: Gorgonacea: Plexauridae) con Aspectos Sobre la Historia Natural de Algunas Especies en el Caribe Colombiano. Master's Thesis, Universidad Nacional de Colombia, Bogotá, Colombia, 1998.
40. Sánchez, J.A.; Lasker, H.R. Patterns of morphological integration in marine modular organisms: Supra-module organization in branching octocoral colonies. *R. Soc.* **2003**, *270*, 2039–2044. [CrossRef] [PubMed]
41. Sangster, T.; Major, H.; Plumb, R.; Wilson, A.; Wilson, I. A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst* **2006**, *131*, 1075. [CrossRef]
42. Godzien, J.; Alonso-Herranz, V.; Barbas, C.; Armitage, E.G. Controlling the quality of metabolomics data: New strategies to get the best out of the QC sample. *Metabolomics* **2015**, *11*, 518–528. [CrossRef]

43. Dunn, W.; Broadhurst, D.; Edison, A.; Guillou, C.; Viant, M.; Bearden, D.; Beger, R. Quality assurance and quality control processes: Summary of a metabolomics community questionnaire. *Metabolomics* **2017**, *13*, 6. [CrossRef]
44. Gorrochategui, E.; Jaumot, J.; Lacorte, S.; Tauler, R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. *TrAC Trends Anal. Chem.* **2016**, *82*, 425–442. [CrossRef]
45. 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. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics. *Analyst* **2009**, *134*, 1322. [CrossRef]
46. Gromski, P.S.; Muhamadali, H.; Ellis, D.I.; Xu, Y.; Correa, E.; Turner, M.L.; Goodacre, R. A tutorial review: Metabolomics and partial least squares-discriminant análisis—A marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23. [CrossRef]
47. Iwagawa, T.; Hashimoto, K.; Yokogawa, Y.; Okamura, H.; Nakatani, M.; Doe, M.; Morimoto, Y.; Takemura, K. Cytotoxic biscembranes from the soft coral *Sarcophyton glaucum*. *J. Nat. Prod.* **2009**, *72*, 946–949. [CrossRef] [PubMed]
48. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983**, *65*, 55–63. [CrossRef]
49. Denizot, F.; Lang, R. Rapid colorimetric assay for cell growth and survival. *J. Immunol. Methods* **1986**, *89*, 271–277. [CrossRef]
50. Al-Ghamdi, S.S. Time and dose dependent study of doxorubicin induced DU-145 cytotoxicity. *Drug Metab. Lett.* **2008**, *2*, 47–50. [CrossRef] [PubMed]
51. Shaikh, K.S.; Pawar, A.; Aphale, S.R.; Moghe, A.S. Effect of vesicular encapsulation on in-vitro cytotoxicity of ciclopirox olamine. *Int. J. Drug Deliv.* **2012**, *4*, 139–146. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

2.7 SUPPLMENTARY INFORMATION

Name assigned by collection area	Name of the species	Accession	Código ICN-UN
Cali 2	*Plexaura kukenthali*	C2P	ICN-MHN (Po)-CO-271
Cali 4	*Pseudopterogorgia albatrossae*	C4P	ICN-MHN (Po)-CO-272
Cali 6	*Eunicea succinea*	C6E	ICN-MHN (Po)-CO-273
Cali 8	*Eunicea succinea plantaginea*	C8E	ICN-MHN (Po)-CO-274
Cali 9	*Eunicea clavigera*	C9E	ICN-MHN (Po)-CO-275
Cali 11	*Eunicea flexuosa*	C11E	ICN-MHN (Po)-CO-276
Cali 12	*Plexaura cf. nina*	C12P	ICN-MHN (Po)-CO-279
Cali 13	*Eunicea fusca*	C13E	ICN-MHN (Po)-CO-277
Cali 15	*Pseudopterogorgia albatrossae*	C15Pst	ICN-MHN (Po)-CO-278
Cali 16	*Eunicea fusca*	C16E	ICN-MHN (Po)-CO-277
Cali 17	*Eunicea clavigera* thin form	C17E	ICN-MHN (Po)-CO-280
Cali 19	*Eunicea flexuosa*	C19E	ICN-MHN (Po)-CO-276
Cali 24	*Plexaurella nutans*	C24Plr	ICN-MHN (Po)-CO-281
Gra 1	*Plexaurella fusifera*	G1P Plr	ICN-MHN (Po)-CO-282
Gra 2	*Plexaura homomalla*	G2P	ICN-MHN (Po)-CO-283
Gra 3	*Eunicea clavigera* thin form	G3E	ICN-MHN (Po)-CO-284
Gra 4	*Eunicea asperula*	G4E	ICN-MHN (Po)-CO-285
Gra 5	*Eunicea clavigera*	G5E	ICN-MHN (Po)-CO-275
Gra 6	*Plexaura kukenthali*	G6P	ICN-MHN (Po)-CO-292
Gra 8	*Eunicea clavigera*	G8E	ICN-MHN (Po)-CO-275
Gra 9	*Plexaura* sp.	G9P	ICN-MHN (Po)-CO-286
Gra 10	*Eunicea knightii*	G10E	ICN-MHN (Po)-CO-287
Gra 13	*Eunicea clavigera*	G13E	ICN-MHN (Po)-CO-275
Gra 14	*Eunicea cf. calyculata*	G14E	ICN-MHN (Po)-CO-288
Gra 17	*Pseudoplexaura flagellosa*	G17Ps	ICN-MHN (Po)-CO-289
Gra 18	*Plexaura kukenthali*	G18P	ICN-MHN (Po)-CO-292

Gra 19	*Eunicea tayrona*	G19E	ICN-MHN (Po)-CO-290
Gra 22	*Plexaurella* sp.	G22P	ICN-MHN (Po)-CO-291

Table S1. Name of species of soft corals used in this research a Collection Code (ICN), assigned by Collection of the Institute of Natural Sciences of the National University of Colombia. (Bogotá, Colombia)

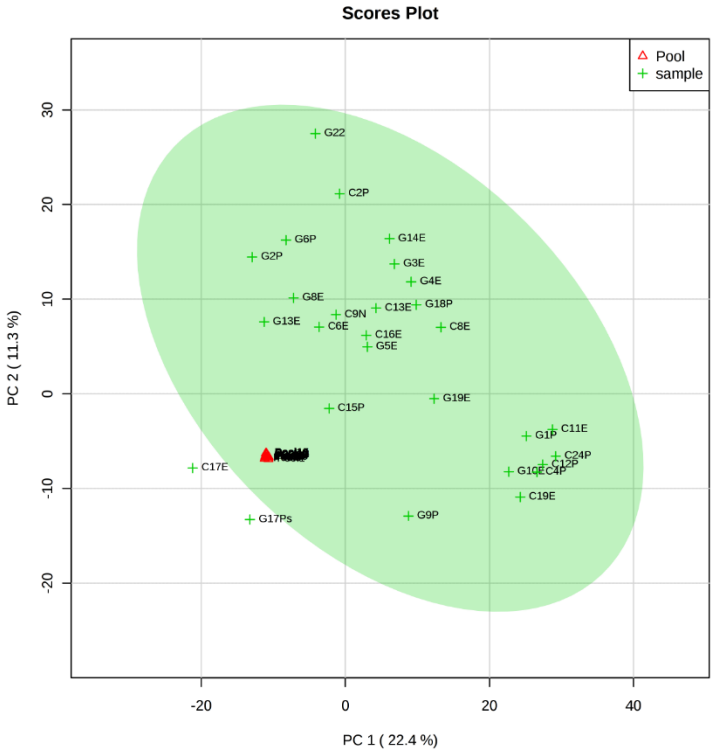


Figure S1. Principal component analysis (PCA); in the triangle the pools (QC) are shown.

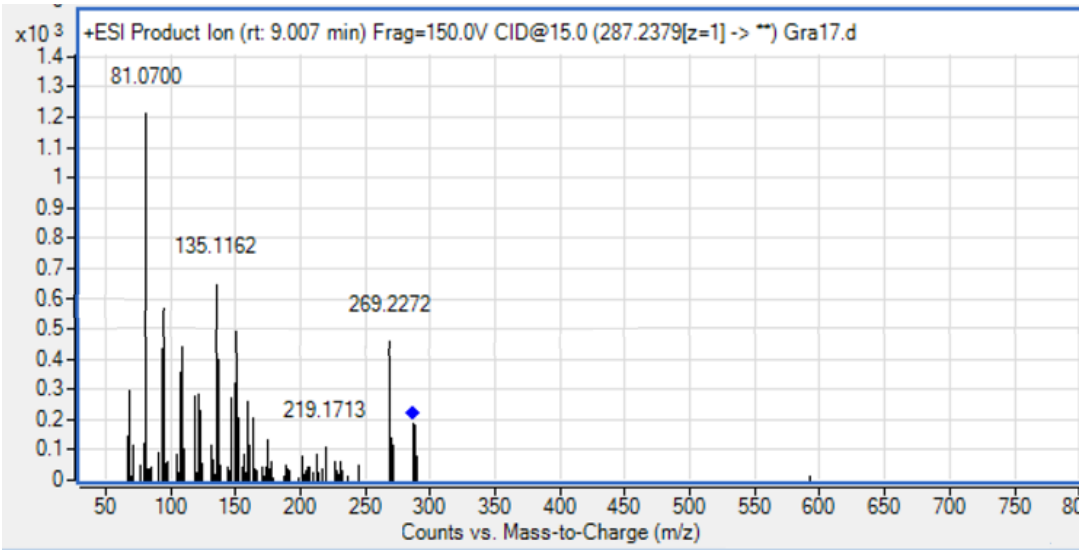


Figure S2. Tandem mass spectra of feature M287T644

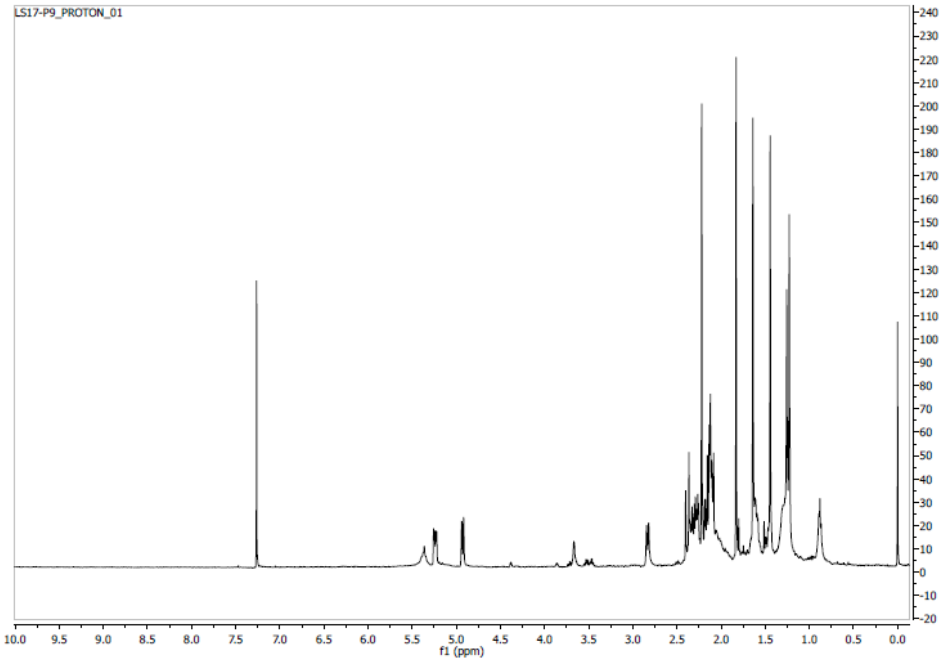


Figure S3. (a) The representative ^{1}H-NMR spectra of fraction F-9 in $CDCl_3$ from *Pseudoplexaura flagellosa* (G17Ps).

Obs	name	Siha (%)	Pc3 (%)	A549 (%)	CSIHA	CPC3	CA549
1	G9Px	84.5	20.4	31.0	1	0	0
2	C6Es	69.0	61.7	35.1	1	1	0
3	G17Ef	66.8	40.3	71.5	1	0	1
4	G14Ec	63.5	22.0	65.7	1	0	1
5	C17Ec	61.3	17.0	27.5	1	0	0
6	C16Ee	60.5	30.0	39.1	1	0	0
7	G3Ec	59.9	33.1	45.3	1	0	0
8	G6Pk	57.0	40.3	48.5	1	0	0
9	G4Ea	53.9	35.5	67.6	1	0	1
10	C9NE	53.3	32.1	14.7	1	0	0
11	C8Ec	48.4	35.9	43.6	0	0	0
12	C2Pk	46.2	7.8	62.4	0	0	1
13	G10Ek	46.1	1.3	45.4	0	0	0
14	G19Et	45.9	39.7	67.8	0	0	1
15	G8Ec	45.9	38.5	49.9	0	0	0
16	Cali3	45.9	40.9	44.0	0	0	0
17	G5Ec	39.8	50.6	46.2	0	1	0
18	G1Pf	39.4	23.8	33.0	0	0	0
19	G22P	37.6	33.4	41.9	0	0	0
20	G13Ec	36.9	30.9	49.7	0	0	0
21	C11Ef	36.8	12.1	49.3	0	0	0
22	C13EA	36.4	27.2	43.8	0	0	0
23	G2Po	34.8	37.8	21.1	0	0	0
24	G18Pk	34.0	44.0	52.5	0	0	1
25	C4Pa	27.1	21.7	33.1	0	0	0
26	C12Pn	23.7	16.0	43.3	0	0	0
27	C19Ef	15.1	16.1	36.1	0	0	0
28	C24Pa	13.1	65.1	39.0	0	1	0

Table S2. Percentage of cytotoxic activity against three cancerous cell lines.

3 METABOLOMIC STUDY OF SOFT CORALS FROM THE COLOMBIAN CARIBBEAN: PSYCHE AND ^{1}H NMR COMPARATIVE ANALYSIS

Liliana Santacruz[1], Diana X. Hurtado[1], Roisin Doohan[2], Olivier P. Thomas[2], Mónica Puyana[3] and Edisson Tello[*1]

[1] Bioprospecting Research Group and Biosciences Doctoral Program, Faculty of Engineering, Campus Puente del Común, Universidad de la Sabana, 250001 Chía, Colombia
[2] Marine Biodiscovery, School of Chemistry and Ryan Institute, National University of Ireland Galway (NUI Galway), University Road, H91 TK33 Galway, Ireland
[3] Departamento de Ciencias Biológicas y Ambientales, Universidad Jorge Tadeo Lozano, Carrera 4 # 22-61, 110311 Bogotá, Colombia; monica.puyana@utadeo.edu.co
*Correspondence: edisson.tello@unisabana.edu.co (E.T.)
Tel.: +57-32-0699-5696 (E.T.)

Abstract

Marine organisms have evolved to survive against predators in complex marine ecosystems via the production of chemical compounds. Soft corals (Cnidaria, Anthozoa, Octocorallia) are an important source of chemically diverse metabolites with a broad spectrum of biological activities. Herein, we perform a comparative study between high-resolution proton nuclear magnetic resonance (^{1}H-NMR) and pure shift yielded by chirp excitation (PSYCHE) experiments to analyze the metabolic profile of 24 soft corals from the Colombian Caribbean to correlate chemical fingerprints with their cytotoxic activity against three cancer cell lines (human cervical carcinoma (SiHa), human prostatic carcinoma (PC3) and human lung adenocarcinoma (A549)). All data obtained were explored using multivariate analysis using principal components analysis (PCA) and orthogonal partial least squares (OPLS) analysis. The results did not show a significant correlation between clusters using ^{1}H-NMR data in the PCA and OPLS-DA models and therefore did not provide conclusive evidence; on the other hand, a metabolomic analysis of PSYCHE data obtained under the same parameters revealed that when a decoupled experiment is performed, it was possible to establish a statistically valid correlation between the chemical composition of soft corals and their cytotoxic activity against the PC3 cancer cell line, where the asperdiol and plexaurolone markers were putatively identified and related to the cytotoxic activity presented by extracts of *Plexaurella* sp. and *Plexaura kukenthali,* respectively. These results increase the speed, effectiveness and reliability of analyses for the study of this type of complex matrices.

3.1 INTRODUCTION

Metabolomics studies allow a complete analysis of a set of metabolites that are the substrates and products of metabolism driving essential cellular functions in a given biological system[1]. This research has applications in different fields, such as pharmacology, environmental sciences, chemotaxonomy, nutrition and medicine[2]. Recently, metabolomic approaches have allowed the understanding of complex biological systems and the biochemical composition of organisms that live in diverse environments, such as marine areas[3]. Goulitquer *et al*. demonstrated that metabolites are

important links between genotype and phenotype and are important for studying several biological processes and for analyzing interactions between organisms within communities via mass spectrometry (MS)-based metabolomics[4]. In addition, Mohamed A. Farag *et al*. comparative metabolomics results obtained with liquid chromatography coupled to mass spectrometry (LC-MS) with those obtained with nuclear magnetic resonance (NMR) to investigate the metabolism of 16 Sarcophyton species in the context of their genetic diversity and growth habitats[3].

The importance of studying marine invertebrates lies in their extraordinary ability to produce a broad variety of chemical compounds with unique chemical structures that in most cases have been correlated with significant biological activities, which has led to the successful development of commercial drugs. One of the best-known compounds is ziconotide (Prialt®), a synthetic derivative of a peptide originally isolated from the venom of the marine snail *Conus magus;* this compound is a neuronal calcium antagonist useful for treating severe chronic pain[5]. Trabectedin ET-743 (Yondelis™), a complex peptide originally isolated from *Ecteinascidia turbinata,* is effective in the treatment of many types of cancer, including melanoma, sarcoma, and lung, breast, ovarian, endometrial and prostate cancer[6]. Brentuximab vedotin®, originally isolated from the marine opisthobranch *Dolabella auricularia,* is used against breast and Hodgkin's lymphomas[7]. Eribulin mesylate (Halaven®), originally isolated from the marine sponge *Halichondria okadai,* is a potent microtubule inhibitor used in the treatment of breast cancer[8], and vidarabine (Vira-A®), a modified nucleoside originally isolated from the Caribbean sponge *Cryptotethya crypta,* is an inhibitor of viral DNA polymerases and other enzymes and is used against varicella zoster and herpes viruses[9]. This information allows us to infer that marine organisms are a prolific source of bioactive and novel molecules that can be used as potential agents against different human illnesses.

Bioprospecting marine research in the Colombian Caribbean has established organisms with various biological activities, mainly anti-inflammatory, cytotoxic and antiviral activities, e.g., gorgonian *Antillogorgia (Pseudopterogorgia) elisabethae* collected at the Islands of Providencia and San Andrés showed anti-inflammatory properties[10]; additionally, the octocorals *Eunicea laciniata* and *Eunicea asperula* have been evaluated for their cytotoxic and antiviral activities, where dolabellane diterpenes isolated from the soft corals *E. laciniata* and *E. asperula* showed anti HSV-1 activity[11]. In addition, the prostaglandins isolated from the soft coral *Plexaura homomalla* presented anti-inflammatory activity[12].

There is no single analytical technology or protocol to analyze the overall metabolome of an organism and obtain a complete metabolic profile[13]. However, metabolomics approaches use hyphenated and high-throughput techniques to perform chromatographic separation of metabolites using either liquid chromatography (LC) or gas chromatography (GC) coupled with mass spectrometry or nuclear magnetic resonance (NMR) to analyze complex mixtures of metabolites from different organisms[14]. Furthermore, each technique has advantages and drawbacks, and although the analytical technique most often used for metabolite profiling is liquid chromatography-mass spectrometry, due to its high sensitivity and the wide range of molecules that can be analyzed, the use of NMR experiments has expanded rapidly over the past ten years[15]. Regardless of the low sensitivity of NMR, this technique presents advantages, such as being nondestructive and requiring no sample preparation and relatively short acquisition times. NMR is a highly reproducible method for metabolomics studies. At present, it is possible to record NMR spectra from crude extracts[16] to perform preliminary studies of the metabolic composition of marine organisms. NMR analyses give a global overview of all metabolites present in complex biological samples such as soft coral extracts that produce compounds such as prostaglandins, sterols and a wide range of terpenes, and these compounds represent the main chemical defense of these organisms against predators[17,18]. On the other hand, metabolomic studies using ^{1}H-NMR on mixtures may experience signal overlap, especially in samples that contain significant amounts of fatty acids and terpenoids, hampering comparative metabolomic studies of extracts from soft corals[19,20]. The table S1 in supplementary information shows the advantages and disadvantages of the most commonly used methods for metabolomics analysis.

And the gathering of metabolomics data, new NMR techniques such as Pure Shift Yielded by Chirp Excitation (PSYCHE) are proposed. This method proposed by Morris and coworkers[21] contains two low flip angle (β) swept-frequency pulses in the presence of a weak pulsed field gradient. The advantage of this technique over ^{1}HNMR is that it resolves overlapping ^{1}H-^{1}H scalar coupling multiplets, which improves chemical shift analysis of complex natural products[22]. PSYCHE suppresses the effects of homonuclear coupling and allows observation of decoupled ^{1}H-NMR spectra with chemical shifts only, helping in the identification the overlapping signals in ^{1}H-NMR experiments are a complex issue, and to greatly improve spectral resolution of potential biomarkers in metabolomic studies[16].

Here, we used ^{1}H and PSYCHE NMR experiments to perform a metabolomic comparison of 24 soft coral extracts (complex biological samples) to examine whether there is a correlation between the chemical composition of the extracts and their cytotoxicity against SiHa, PC3 and A549 tumor cancer cell lines[23]. When large data sets are analyzed, a multivariate analysis (MVA) is a valuable approach for the identification of potentially bioactive metabolites in complex mixtures. Due to this, MVA was developed using the principal component analysis (PCA) algorithm to compress a dataset onto a lower-dimensional subspace with the goal of maintaining most of the relevant information and identifying chemical differences from high-dimensional spectra in the NMR experiments[24]. This approach was followed by the orthogonal projection to latent structures discriminant analysis (OPLS-DA), which is a supervised model that filters out orthogonal metabolite variables that are not related to categorical variables to discriminate and separate predictive from nonpredictive (orthogonal) variation. A response matrix Y (containing toxicity data) was correlated to a descriptor matrix X (containing spectral data) that is orthogonal (noncorrelated) to Y to identify the markers that contributed to the discrimination of cytotoxic activity and to determine outliers[25]. The reliability of the model was verified by the cross validation method, and the parameters for the OPLS model, R2 and Q2, were calculated (varying from 0 to 1), where R2 corresponded to the fraction of the variance explained by the model, Q2 suggests the predictive capability of the model[26], and the ratio Q2/R2, which is a measure of cross-validation reproducibility, was also considered indicative of relevant associations[27].

Furthermore, a partial least squares discriminant analysis (PLS-DA) model was used to calculate the variable importance in projection (VIP)[28], which describes a quantitative estimation of the discriminatory power of each individual feature and summarizes the contribution a variable makes to the model.[29] A combination of univariate and multivariate statistical approaches ($VIP > 3$, $p < 0.05$) was used to determine the markers that can discriminate between the fold changes in cytotoxicity level[30].

OPLS-DA may be considered equivalent to PLS-DA for sample discrimination, where the main benefit in data interpretation using OPLS-DA over PLS-DA lies in the ability of OPLS-DA to separate predictive from nonpredictive (orthogonal) variation[31]. Finally, MVA and its loading and score plots, which are closely linked such that features (chemical shifts) that are highly loaded in a specific direction in the loading plot contribute to an increased degree to the observations (soft corals) that are located in that direction in the score plot. This allowed us to establish that the markers B2_5118,

B4_4965 and B4_7686 were responsible for the cytotoxic activity shown by the extracts of the soft corals *Plexaurella* sp. and *P. kukenthali* against the tumor cancer cell line PC3. This article describes a workflow that helped to estimate the confidence levels of compound annotations in reported metabolites that were used as established by the Compound Identification workgroup of the Metabolomics Society at 2017[32]. In summary, this work achieved a confidence level of 2 for features (VIPs) most responsible for cytotoxic activity considering that the probable structure were matched to literature data or databases by diagnostic evidence[33].

3.2 RESULTS

3.2.1 ^{1}H-NMR metabolomic fingerprints

Correlations between cytotoxic activity against SiHa, PC3 and A549 tumor cancer cell lines and the chemical composition using ^{1}H-NMR spectra of the 24 crude extracts of soft corals (in the genera *Plexaura, Pseudopterogorgia*, *Eunicea*, *Plexaurella* and *Pseudoplexaura* Table S2 in Supplementary Information) were explored over 168 spectral bins (variables) describing metabolic profiles using the open access NMRProcFlow v1.2 software[34] developed by INRA Science & Impact in France. The complexity of the spectra can be visualized in some regions due to overlapping of signals (Fig. 1). The script used for data processing is described in the Supplementary Information (Table S4).

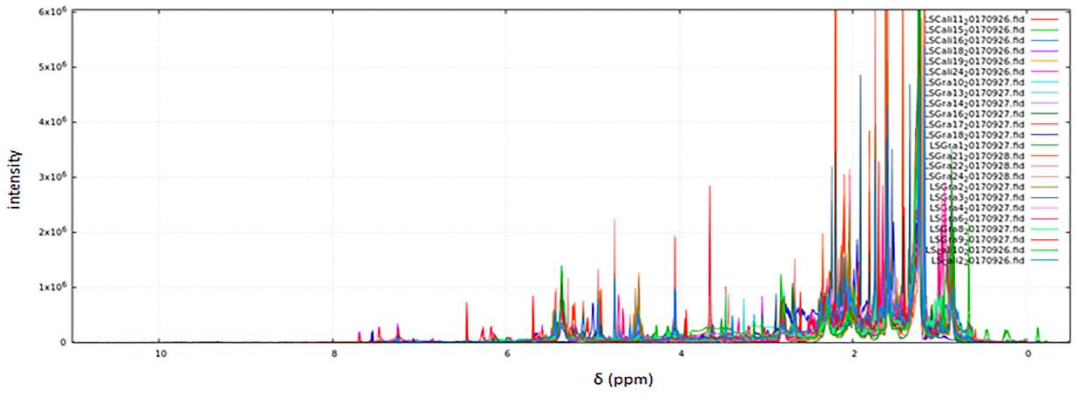


Figure 1 ^{1}H-NMR data of 24 soft coral extract samples using the open access software NMRProcFlow v1.2 25. Spectra of soft coral extracts are presented in different colors.

Exploratory data analysis was performed by PCA using the MetaboAnalyst version 3.0 web application[35] (Supplementary Fig. S1). For the tumor cell lines A549, PC3 and SiHa, extracts of *P. kukenthali, E. asperula* and *Plexaurella* sp. were most distant from those groups that exhibited the

highest cytotoxic activity against the three tumor cell lines. Additionally, we used partial least squares-discriminate analyses (OPLS-DA) (Fig. 2) to visualize increased separation of the groups that presented cytotoxic activity from those that did not. Finally, to demonstrate that the observed PCA was valid, the results from the plots (loading and observations) were contrasted, with the results obtained in the PLS-DA model. Additionally, the statistical parameters for the PLS-DA models for the classification of the ^{1}H-NMR data were determined, as shown in Table 1.

In this study, an extract was considered active if it showed inhibition of tumor cell lines ≥ 40% at 20 µg/mL. This designation included extracts that showed moderate to strong cytotoxic activity[36] (Table S3).

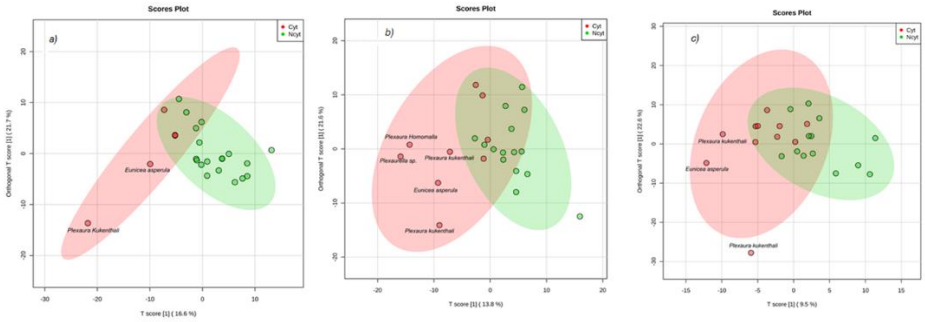


Figure 2. Supervised OPLS-DA score plots of ^{1}H-NMR metabolomic data of 24 soft coral extracts based on their cytotoxicity against three different cancer cell lines. a) human lung adenocarcinoma, A549 b) human prostatic carcinoma, PC3 and c) human cervical carcinoma, SiHa. Red dots indicate active extracts, and green dots represent extracts that were not active. The ellipse represents the 95% confidence interval[37].

With the purpose of discriminating between metabolite profiles (analysis of a large group of metabolites that are a related class of compounds[38]) and cytotoxicity of all the extracts analyzed, a validation of the model was carried out using a PLS-DA based on the PLS algorithm. The discriminating variable was cytotoxic activity. A cross validation test was performed for the classification model with the three tumor cell lines, where the models were evaluated using both R2 and Q2 metrics. R2 values report the total amount of variance explained by the model, Q2 values report model accuracy, and the Q2/R2 ratio is a measure of cross-validation reproducibility. Predictive relevance is considered when values are greater than 0.5[39]. The resulting models for each tumor cell line revealed an overfitting in the separation of the metabolic profiles. Discrimination of the samples according to their cytotoxic activity is shown in Table 1.

Table 1. Statistical parameters of PLS-DA models for classification of ^{1}H-NMR experiments according to cytotoxic and noncytotoxic groups from 24 extracts of soft corals tested against three tumor cell lines; A549: human lung adenocarcinoma, PC3: human prostatic carcinoma and SiHa: human cervical carcinoma.

PLS-DA parameters			
Cell lines	Q2	R2	Q2 /R2
A549	0.14	0.49	0.28
PC3	-0.02	0.49	-0.04
SiHa	0.03	0.59	0.05

Parameters based on Q2 indicate the best classifier of PLS-DA analyses using a 10-fold cross-validation method. PLS-DA: partial least squares-discriminant analysis, Q2: predictive capability, R2: correlation coefficient.

3.2.2 Pure shift experiments (PSYCHE) in metabolomic fingerprints

To compare the results (cytotoxic activity vs. metabolic profile) between the ^{1}H and PSYCHE NMR experiments, the same protocol and the same number of samples were used for both experiments. The PSYCHE data was analyzed over 113 spectral bins (variables) describing metabolic profiles using the NMRProcFlow v1.2 software[34]. The PSYCHE spectra of the 24 soft coral extracts analyzed are shown in Fig. 3a. Fig. 3b shows the same 24 spectra but separated into two groups by color: cytotoxic extracts appear in red, and noncytotoxic extracts appear in green. To establish bins responsible for group separation, it was necessary to perform a statistical analysis to show the relevant chemical shift (VIP) that allowed the discrimination.

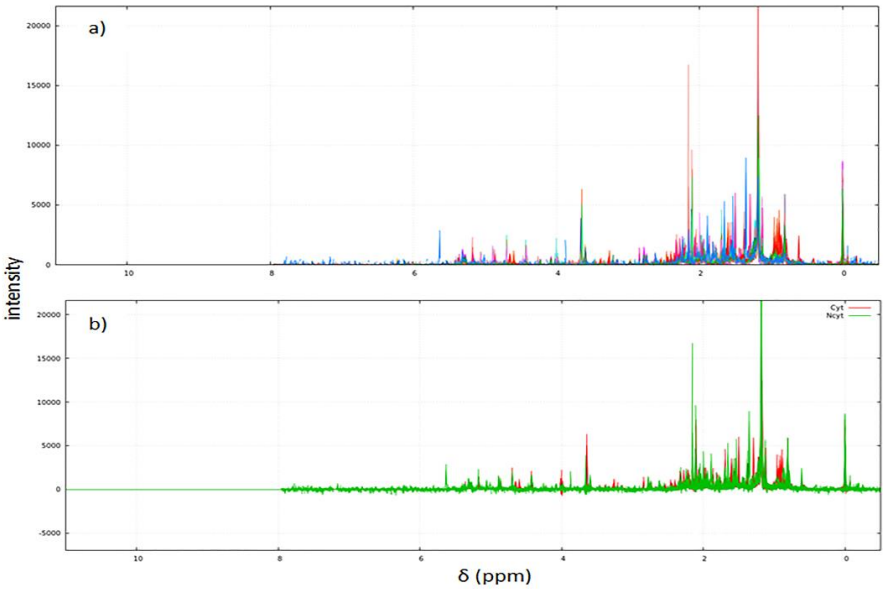


Figure 3. a) PSYCHE spectra of the 24 soft coral extracts analyzed. The spectrum of each soft coral extract is represented in a different color. b) Spectra separated into two groups, by color. Red (cytotoxic extracts), green (not cytotoxic extracts).

PCA[35] showed that the cytotoxic extracts of *P. kukenthali* and *E. asperula* are separated in PC2 for the tumor line A549, as shown in Supplementary Fig. S2a. On the other hand, the PC2 shown in the model for the PC3 tumor line showed that cytotoxic extracts of *P. kukenthali*, *E. asperula* and *Plexaurella* sp., were the most separated, which agrees with the results obtained for this same tumor cell line using the ^{1}H-NMR experiment (See Supplementary Fig. S2b). Finally, PC1 and PC2 from PCA generated for the SiHa tumor cell line did not show a clear separation between the extracts and their cytotoxic activity, as shown in Supplementary Fig. S2c.

It is appropriate to apply PCA as a first step for exploratory studies where differences between experimental groups may be unknown or unpredictable; however, the spectral noise and high within-group variation do not show a separation between groups in many cases[40]; therefore, to overcome this problem, a supervised model OPLS-DA was constructed for classification of the samples and to improve discrimination between metabolic profiles and their cytotoxic activity (Fig. 4) because the OPLS-DA algorithm is normally applied when there are only two classes, improving the class discrimination and robustness of important feature identification[41].

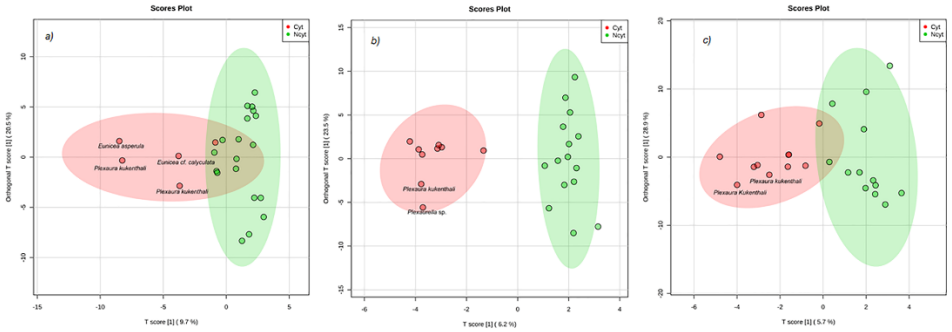


Figure 4. OPLS-DA Scores Plot[35] of PSYCHE experiment of metabolomics data from 24 soft coral extracts based on their cytotoxicity against three different cancer cell lines. a) human lung adenocarcinoma A549, b) human prostatic carcinoma PC3 and c) human cervical carcinoma SiHa. Red dots indicate cytotoxic extracts, green dots represent extracts that were not cytotoxic. An extract was considered active if it exhibited inhibition of the tumor cell line ≥ 40% at 20 µg/mL)[42]. Ellipses indicate confidence intervals of 95%[37]

As performed for the ^{1}H-NMR experiments, the validation of the PSYCHE models (Table 2) was carried out using PLS-DA, where the models were evaluated using R2, Q2 and Q2/R2 metrics. A model is considered predictive if the Q2/R2 ratio is greater than 0.5[43]. When the results of the models obtained for the ^{1}H-NMR experiments were compared with those from the PSYCHE NMR experiment (Tables 1 and 2), an improvement in the total amount of variance (R2) and in the accuracy (Q2) was observed for the latter experiment. This may be because the soft coral extracts contain many

overlapping peaks from multiple compounds and that such overlaps obscure peak assignments and compromise the quantitative analysis, which explains the overfitting observed in the OPLS-DA model using ^{1}H-NMR data compared to that in the decoupled PSYCHE model; all this is reflected in the total reproducibility explained by the ratio Q2/R2[44]. However, a reasonable separation between the groups (cytotoxic and noncytotoxic) was only evident in the model generated for the PC3 tumor cell line, with a Q2/R2 value of 0.59 (Table 2). These results confirm that cluster separations in the OPLS-DA score plot shown in Fig. 4b for the PC3 tumor cell line were statistically significant.

The cytotoxic activity of each extract against the tested tumor cell lines showed that the extracts of *P. kukenthali* (code G18P) and *Plexaurella* sp. (code G22P) exhibited inhibition percentages of 64.0% and 63.5% against the PC3 tumor cell line, respectively. In Fig. 4b, the OPLS-DA of PSYCHE experiments show that those extracts were part of the group that presented cytotoxic activity against the tumor line PC3.

Table 2. Statistical parameters of PLS-DA models for classification of PSYCHE experiment according to cytotoxic and non-cytotoxic groups from 24 extracts of soft corals tested against three tumor cell lines; A549: human lung adenocarcinoma, PC3: human prostatic carcinoma and SiHa: human cervical carcinoma.

PLS-DA parameters			
Cell lines	Q2	R2	Q2 /R2
A549	0.14	0.76	0.19
PC3	0.39	0.66	0.59
SiHa	0.07	0.72	0.09

Parameters based on Q2 indicate the best classifier of PLS-DA analyses using a 10-fold cross-validation method. PLS-DA: partial least squares-discriminant analysis, Q2: predictive capability, R2: correlation coefficient.

Information on the metabolites responsible for the separation between cytotoxic and noncytotoxic groups (VIP) for the tumor cell line PC3 (statistically significant model) was extracted from the PLS-DA. This process allowed us to identify key discriminatory metabolites through a VIP analysis, which revealed several distinguishing patterns. High values indicate the increased discriminatory power of particular metabolites. Variables with VIP > 1.0 are considered potential biomarker candidates for group discrimination[45]. However, for this study, only variables with VIP > 3.0 were considered because these variables play important roles in the discrimination of cytotoxic activity, as three markers were the major variables responsible for group separation and the chemical

shifts corresponding to these VIP markers in the NMR experiments were well resolved, which allowed an improved interpretation of the results.

The three markers that showed the highest scores in the VIP analysis (Fig. 5a) were B2_5118 (score = 5.1), marker B4_4965 (score = 4.8) and marker B4_7686 (score = 3.5). Additionally, in the OPLS S-plot (Fig. 5b), each of the three markers selected as VIPs can be clearly distinguished, with each coordinate representing a single NMR signal (contributing variables to the classification). This result shows that features (VIP) correlated with extracts of *P. kukenthali* (Gra 18) and *Plexaurella* sp. (Gra 22) have cytotoxic activity against the PC3 tumor cell line, in accordance with the analysis obtained from the OPLS-DA (Fig. 4b).

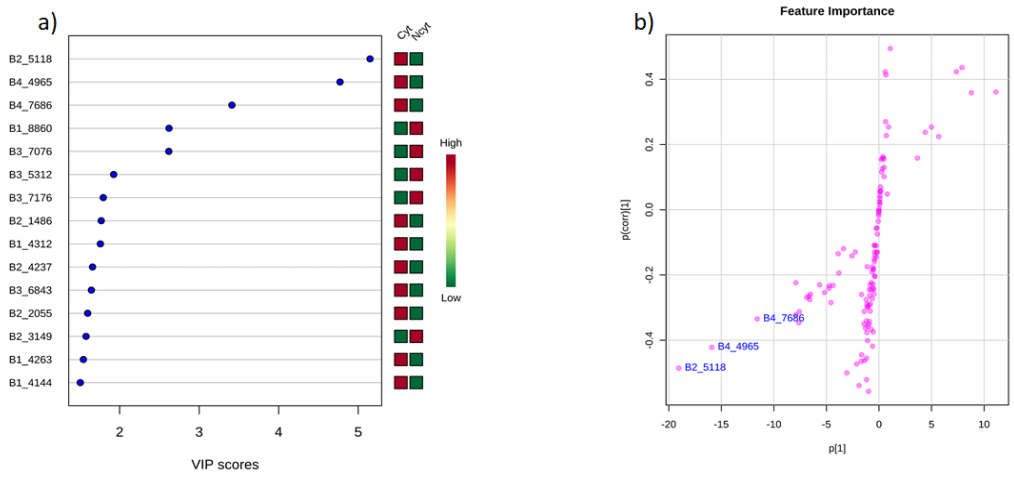


Figure 5. a) VIP (variable importance in the projection) scores / $p < 0.05$[35] obtained from the PSYCHE PLS-DA model b). Feature importance from the OPLS S-plot[37].

The previously described markers B2_5118, B4_4965 and B4_7686 showed a high correlation with 25 additional features according to "Pattern Hunter" a tool from MetaboAnalyst software[35] (See Supplementary Fig. S3). This analysis allowed a putative identification of the possible compounds that corresponded to those features. In addition, PeakForest 2.0.1® software functions were used with a correlation coefficient of ± 0.5 to help in the identification of metabolites.

3.2.3 Confidence annotation of compound identification

Confidence annotation consisted of three steps. First, a literature review of the compounds that have been isolated from the species was performed using the SciFinder® database. Cembrane-type diterpenoids have been isolated from *Plexaurella* sp. and *P. kukenthali*[46]. From a biomedical

perspective, some of those cembranes have shown cytotoxic activity against several tumor cell lines and have also been reported as anti-inflammatory, neuroprotective and antimicrobial compounds[47].

In a second step, a verification of chemical formulas and exact masses was developed using both LC-MS and LC-MS/MS with the purpose of yielding a putative identity of the compounds. The presence of the compound asperdiol in the extract of *Plexaurella* sp., with a molecular formula $C_{20}H_{32}O_3$ corresponding to the ion *m/z* 321.2426 [M+H], was confirmed (Supplementary Figs. S5a and S5b). In addition, analysis of MS/MS data fragmentation using MetFrag software [48] matched the data for the compound asperdiol (score 7.47). On the other hand, from a high-resolution electrospray ionization mass spectrometry (HRESIMS) analysis of the extract from *P. kukenthali*, the formula $C_{20}H_{34}O_3$ was calculated, which confirmed the possible presence of the compound plexaurolone with an ion *m/z* of 323.25860 [M + H] and an *m/z* of 345.24038 [M + Na] (Supplementary Figs. S6a and S6b)[49].

In the third step, chemical shifts were analyzed using the AntiMarin® database to check signal patterns that corresponded to the kind of compounds (diterpenoid-type metabolites) expected to be present in the mixture (diterpenoid-type compounds)[50]. As a consequence, in the ^{1}H-NMR spectra of the extracts from *Plexaura kukenthali* and *Plexaurella* sp. (Fig. S4 in Supplementary Information), singlets were found at a range of chemical shifts from approximately 0.5 to 2.0 ppm, indicating the presence of several methyl groups, which is characteristic of terpenoid-type compounds and in agreement with the literature[51]. Additionally, chemical signals of VIP markers B4_4965 and B4_7686 were found in the ^{1}H-NMR spectrum of the extract from *Plexaurella* sp. When comparing these data with those reported in the literature, we detected similarities with the NMR data published for the antineoplastic compound asperdiol[52]. Additionally, an analysis of the marker B2_5118 in the ^{1}H-NMR spectrum of the extract from *P. kukenthali* and further comparisons with NMR data reported in the literature allowed us to make a putative identification of this compound as plexaurolone. Table S5 shows the summarizing the key results of this study.

Finally, in accordance with recent discussions in the metabolomics community[53], the confidence grade in the identification of these metabolites (asperdiol and plexaurolone shown in the Supplementary Information) is level 2 because this level reveals probable structure using fragmentation data from literature and/or libraries and databases.

3.3 DISCUSSION

There are two analytical techniques commonly used in metabolomics, each with advantages and disadvantages: mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. There is an increasing growth in NMR-based metabolomics over the last few years due to the advantages that this technique offers: high reproducibility and quantification ability over a wide dynamic range of compounds and support for the identification of unknown structures; moreover, NMR is nondestructive and can be utilized *in vivo*[50].

NMR is a very suitable method for analysis that allows the simultaneous detection of diverse groups of secondary metabolites in complex matrices, as evidenced by a study conducted by Mohamed *et al*. where some cembrane diterpenoids were isolated from the *Sarcophyton ehrenbergi* soft coral and identified using NMR techniques[54]. In another study, Zhang *et al*., also using NMR, found cytotoxic diterpenoids from the soft coral *Sinularia microclavata*[55].

One-dimensional (1D) ^{1}H-NMR is the most widely used NMR approach in metabolomics. However, this approach has several limitations, such as low sensitivity and difficulties in key signal identification and accurate peak integration resulting from the ^{1}H-NMR spectrum containing hundreds of overlapping signals[56]. The improvement in sensitivity and overlapping of peaks can be solved to a great extent if the density of signals in a spectrum is decreased by almost an order of magnitude. A pure shift experiment (PSYCHE) is an alternative that offers superior sensitivity and spectral purity. It has the potential to find wide application in the NMR spectroscopy of small molecules and NMR-based metabolomics[21,57], as we demonstrate in this research by using the PSYCHE experiment in soft coral metabolomics for the first time.

Here, the application of OPLS-DA to 1D ^{1}H-NMR data (Table 1) revealed an excessive data adjustment in the model that could be attributed to signal overlap in the ^{1}H-NMR spectra of soft-coral extracts, which was principally due to olefinic protons of diterpenes and the high-field region between 3.2 and 0.8 ppm having a high density of signals due to terpene methyl, methylene, and methine resonances, revealing that soft coral extracts contain large amounts of terpenoids[58]. The noise present in each spectrum of this type of biological sample hinders the alignment and creation of the bins used to reduce the data dimensionality in the NMR spectrum from several thousand points[34]. Therefore, the results obtained from this experiment were not conclusive, which agrees with the reports by Farag *et al*. 2017.

Validations using OPLS-DA of PSYCHE NMR experiment data (Table 2), on the other hand, showed an improvement in the total amount of variance, accuracy and reproducibility of the models, particularly for the PC3 tumor cell line, which was predictive according to the validation results. This result showed that metabolomic studies using PSYCHE NMR experiments allowed us to obtain a more reliable correlation between the chemical composition and the cytotoxic activity of soft coral extracts against the PC3 tumor cell line than metabolomic studies using solely ^{1}H-NMR data. The results obtained throughout the metabolomic process always showed a correlation in the data that was identified by the PCA, OPLS-DA, the markers displayed in the VIP table, and an S-plot obtained from PLS-DA of the extracts from *P. kukenthali* and *Plexaurella* sp. These organisms were highlighted in all the analyses as those responsible for the separation of the groups that presented cytotoxic activity and those indicating that the variables B2_5118, B4_4965 and B4_7686 (chemical shifts) that were present in these species were the main markers related to the activity shown; also, we can derive from this analysis that these extracts, in addition to exhibiting the greatest cytotoxicity, were the closest in the orthogonal t score, indicating they may have a common chemical composition. This was evidenced from MS and MS/MS spectra (Supplementary Figs. S5 and S6), which showed that the chemical formulas $C_{20}H_{32}O_3$ (with an *m/z* 321.2426 [M+H]) and $C_{20}H_{34}O_3$ (with an *m/z* of 323.25860 [M + H]) corresponded to the compounds asperdiol and plexaurolone, respectively. They also showed that both compounds were present in the two soft coral species. However, the ^{1}H-NMR signals corresponding to asperdiol were not clearly observed in the extract of *P. kukenthali*, and in the same way, the chemical shifts of plexaurolone could not be observed in *Plexaurella* sp. extracts due to plexaurolone abundance.

Here, the fact that the extracts from *P. kukenthali* and *Plexaurella* sp. were mainly responsible for the separation of the group that presented cytotoxic activity was statistically validated using the cross-validation method (Table 2), which agreed with the cytotoxicity presented by these species as reported by Honda *et al*. (1987)[59] and Rueda *et al*. (2001)[60], who showed that some terpenoids obtained from *Plexaurella grisea* and *P. kukenthali* exhibited cytotoxic activity against myeloid leukemia and P-388 cancer cell lines.

From analyses of the VIP and the OPLS score plot (Fig. 5), it was established that the most important characteristics in the separation of extracts that showed cytotoxic activity were B2_5118, B4_4965 and B4_7686. Based on their ^{1}H-NMR chemical signals, it was possible to make a putative identification of asperdiol in the extract of *Plexaurella* sp. due to the signals evident in the extract;

these signals agree with the chemical shifts for this compound reported by Weinheimer and Matson (1977)[52]. Chemical shifts at 4.76 and 4.95 ppm correspond to an exomethylene group characteristic of the cembrane skeleton. Signals located at 4.50 and 4.05 ppm are characteristic of carbinolic methine and methylene groups. Additionally, three signals, 1.20, 1.62 and 1.77 ppm, were assigned to three methyl groups, and the signal located at 2.70 ppm corresponds to an epoxy proton characteristic of asperdiol. The remaining chemical shifts were consistent with those reported in the literature[61]. In addition, chemical shifts assigned to the marker B2_5118 confirmed that the compound plexaurolone was present in the extract of *P. kukenthali*, in agreement with NMR data reported for this compound[49]. The signal located at 2.53 ppm was assigned to a typical methine that supports the isopropenyl group in the cembrane skeleton. Additionally, the signals at 0.99, 1.03 and 1.04 ppm were in agreement with three methyl groups, and the signal at 4.72 ppm together with the signal 1.68 ppm were assigned to an isopropenyl group typical in this kind of nucleus. All the remaining signals were consistent with data reported in the literature[49].

Finally, the main prospective application of metabolomic analyses using PSYCHE NMR experiments in soft corals was identifying metabolites with potential cytotoxic activity. This metabolomic approach may be useful in various scenarios but mostly in untargeted studies in which the identification of compounds is a challenge that involves a long time isolating and identifying biologically active metabolites. In addition, the advantage of using the decoupled PSYCHE experiments over the ^{1}H-NMR experiments due to the overlapping of signals present in the latter is corroborated.

3.4 MATERIALS AND METHODS

3.4.1 Materials

Methanol and dichloromethane used for extraction were bought from Merck (Darmstadt, Germany). Cell culture reagents, D-MEM (Dulbecco's Modified Eagle Medium (1X), RPMI 1640 Roswell Park Memorial Institute, Darmstadt, Germany were made by Gibco/Invitrogen, Paisley, UK. Other reagents were Fetal bovine serum (FBS), Eurobio brand (Les Ulis, France), trypticase soy broth (TSB) and trypticase soy agar (TSA) Scharlau Co. brand (Barcelona, Spain). All cancer cell lines were acquired from ATCC, PC3 human prostatic carcinoma (ATCC® CRL1435™), SiHa human cervical carcinoma (ATCC® HTB-35™) and A549 human lung adenocarcinoma (ATCC® CCL-185™).

3.4.2 Methods

3.4.2.1 Soft coral collection and identification

Small portions (30 cm) of soft corals (N=24) (Table S2, Supplementary Information) were collected by SCUBA diving at Punta Venado (11°16.26′ 87″N, 74°12.24′58″W), Santa Marta, Colombian Caribbean. Samples were collected at a depth range between 10–20 m. Collected samples were stored in dry ice and transported to the laboratory. These were kept frozen at -80 ° C until extraction.

Samples were identified by colony morphology and sclerite shape, dimensions and distribution. Sclerites were obtained from a distal fragment of each soft coral portion that was treated with 5% sodium hypochlorite. Once the organic material was dissolved, sclerites were observed under the microscope. A voucher of each sample is stored at the collection of the Instituto de Ciencias Naturales (ICN) of Universidad Nacional de Colombia (Bogotá, Colombia) (Table S2 in Supplementary Information).

3.4.2.2 General experimental procedures

All NMR data were acquired using nonspinning samples on a 600 MHz Agilent DD2 NMR at 25°C equipped with a 5 mm C13 enhanced HCN cold probe. Signals were referenced in ppm in reference to the residual solvent signals (CDCl3, at δH 7.26). The ^{1}H spectra were acquired with 32 transients, a 1 s relaxation delay (d1), a 90 degree pulse of 6.70 µs and a 9.6 kHz (16 ppm) spectral window (sw). The transmitter offset was set to 6 ppm.

The PSYCHE spectra were acquired with 6 transients, a 1 s relaxation delay (d1), steady state scans of 2 (ss) and a pure shift tau delay of 6.3 ms. The spectral window was set to 6 kHz (10 ppm) with the transmitter offset at 3 ppm. A WURST180 pulse with a 6.0 deg flip angle, a 531(1.0 G/cm) gradient and a pulse width of 30.0 ms was used. A Grad-90-Grad option for the PSYCHE steady state with a G-strength of 6372 and a G-time of 2 ms was used along with Echo gradients (Encode Grad (5.0 G/cm)) using a 1.0 ms time, a 1.5 ms recovery time and a strength of 2659.

3.4.2.3 Sample extraction

Prior to extraction, each soft coral fragment was ground. Subsequently, 10 g of dried powder of each soft coral sample was extracted at room temperature with a mixture of solvents (DCM/MeOH, 1:1) three times (300 mL) using an ultrasonic bath for 20 minutes. Remaining debris was removed

by centrifugation twice at 12,000×g for 5 minutes, the solvent was then evaporated, and the dried samples were passed through a C18 cartridge with MeOH to remove salts. Extracts were concentrated under vacuum using rotary evaporation.

3.4.2.4 Cytotoxicity assay

Tumor cell lines A549, SiHa and PC3 were cultured in DMEM supplemented with 10% heat-inactivated fetal bovine serum (FBS) and penicillin-streptomycin (1%) at 37°C in a 5% CO_2 humidified atmosphere until 100% cell confluence was achieved [48]

The *in vitro* cytotoxicity of soft coral extracts was evaluated using the MTT method following Mosmann (1983) with modifications by Denizot and Lang (1986). This is a colorimetric assay based on the capacity of mitochondrial succinate dehydrogenase enzymes, found in living cells, to reduce the yellow, water-soluble substrate 3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) into an insoluble, colored, formazan product, which is measured spectrophotometrically. Cells were grown and left to attach in 96-well plates (4.0E+04 cells/well) for 48 hours.

A preliminary screening was performed using the 24 soft coral extracts evaluated at 20 µg/mL and DMSO at 0.1%, which was not cytotoxic to any of the cell lines. Then, the supernatant was removed, and 100 µL of 12 mM MTT solution in sterile PBS was added to each well and incubated at 37°C for 4 hours. The solution was removed, and dimethyl sulfoxide (DMSO) was added to each well, followed by incubation at 37°C for 15 minutes. Cell density was read in an iMarkTM Microplate Reader at a wavelength of 595 nm. Controls consisted of cells cultured without extract, and cells exposed to doxorubicin (25 ppm) were a positive control [62,63]. All tests were performed in triplicate. The cell viability percentage was calculated with equation (1), and the cell inhibition percentage was calculated with equation (2)[64]:

$$\%viability = \frac{(Abs_{sample})}{Abs_{control}} * 100 \qquad (1)$$

where Abs_{sample} is the absorbance of cells treated with the test extract, and $Abs_{control}$ is the absorbance of untreated cells.

$$\%cell\ inhibition = 100 - Cell\ Survival \quad (2)$$

3.4.2.5 Metabolomic analyses

^{1}H and PSYCHE NMR experiments of the 24 soft coral extracts were run in the NMRProcFlow v1.2 program[34]. This software performs all the spectral processing steps, including baseline correction, chemical shift calibration and alignment, and allows metabolic fingerprinting and targeted metabolomics[65].

All spectral data were compressed with a basic Zip format and processed in the NMRProcFlow v1.2 software, which allows the input of raw data in FID format. Data processing comprises calibration of the PPM scale, baseline correction, alignment, binning and scaling.

1. Calibration of the ppm scale was performed to adjust chemical shifts according to a known reference compound. The reference compound used for the calibration of the ^{1}H-NMR spectra was trimethylsilane (TMS), with a chemical shift of δ = 0.00 ppm, and for the PSYCHE experiment, the reference compound was the solvent $CDCl_3$, with a chemical shift of δ = 7.26 ppm.

2. Baseline correction was performed using the global baseline correction algorithm[66], where the correction level was chosen as the 'soft' level. For increased efficiency of the method, spectral level noise ranges between 10.2 and 10.5 ppm for the ^{1}H-NMR experiment and between 7.7 and 8.0 ppm for the PSYCHE experiment were considered.

3. Alignment steps were very tedious to solve. Misalignments are a result of changes in the chemical shifts of NMR peaks largely due to differences in pH, ionic strength or other physicochemical interactions [67]. To perform this step, we use the algorithm based on least squares.

4. Binning. An NMR spectrum may contain several thousands of variables. Binning is used to reduce data dimensionality. When binning, spectra are divided into bins (so-called buckets), and the total area (sum of each resonance intensity) within each bin is calculated to represent the original spectrum. The approach we chose was the adaptive 'Intelligent Binning' method[68]. This allowed us to split the spectra so that each area, common to all spectra, contained the same single resonance, i.e., belonged to the same metabolite.

5. Normalization. Before bucket data export, to make all spectra comparable with each other, variations in the total concentrations of samples must be considered. We used the constant sum normalization, which consists of normalizing the total intensity of each individual spectrum to the same

value[69]. After the data matrices of each of the selected regions were exported, they were used for statistical analysis using MetaboAnalyst version 3.0[35].

3.4.2.6 Multivariate data analysis

All variables of the data matrix were scaled using a self-scaling algorithm prior to multivariate data analysis. Principal component analysis (PCA), an unsupervised pattern recognition tool that explains the maximum amount of variation inherent in a multidimensional dataset[70], was carried out to detect patterns in the variables matrix and to detect outliers.

Orthogonal partial least squares discriminate analysis (OPLS-DA) is a supervised pattern recognition technique that aims at finding the maximum separation between a priori established groups[71]. Therefore, OPLS-DA was applied to discriminate spectral data obtained from the ^{1}H and PSYCHE NMR experiments and to discriminate between groups that exhibited or did not exhibit cytotoxicity.

The resulting models were evaluated using both R2 and Q2 metrics. R2 values reported the total amount of variance explained by the model in both the data (R2X) and independent variables (R2Y). Q2 reported the model accuracy and was calculated by cross-validation[72].

3.5 CONCLUSION

In this study, it was possible to correlate the chemical compositions of extracts from soft coral with their cytotoxic activity against the tumor cell line PC3 using a metabolomics workflow and PSYCHE NMR experiments. By using this approach, it was possible to resolve the overlapping of ^{1}H-^{1}H scalar coupling multiplets, yielding an adequate matrix for reliable statistical and chemical shift analyses of complex natural products. Additionally, a preliminary identification of features responsible for the separation of the groups of extracts was possible due to the chemical shifts observed in the VIP analysis, which were deemed the most important projection variables. The PSYCHE NMR experiment, combined with metabolomics studies, allowed the development of a procedure/methodology to establish which extracts from complex biological samples were most active and allowed the identification of compounds responsible for the activity. Extracts from *P. kukenthali* and *Plexaurella* sp. were the most cytotoxic and were responsible for separation between the groups. Asperdiol and plexaurolone were the compounds responsible for the cytotoxicity exhibited by the most-active extracts.

Author Contributions: L.S., O.P.T., and E.T. conceived and designed the experiments; L.S., M.P. and E.T. collected the biological material; M.P. identified the soft corals, L.S., R.D., D.X and E.T. performed chemical experiments; L.S. and E.T. analyzed the data; L.S. and E.T. wrote the article; L.S., M.P., O.P.T. and E.T. edited and reviewed the article. All the authors have contributed to and approved the final manuscript.

Funding: Doctoral studies of L.S. were supported by grants by Colciencias Beca de Doctorado Nacional 647/2014 and Universidad de la Sabana. Universidad de la Sabana supported Project ING-177-2016 "Búsqueda de Compuestos Bioactivos. Fase II: Análogos sintéticos de diterpenos con actividad citotóxica". Part of this project (Grant-Aid Agreement No.PBA/MB/16/01) was carried out with the support of the Marine Institute and was funded under the Marine Research Programme supported the Irish Government.

Acknowledgments: The authors acknowledge Universidad de La Sabana for support for running cytotoxic bioassays and Galway University for performing NMR analyses. Sergio Zapata Lopera from UTADEO, provided great help in obtaining, processing and measuring soft coral sclerites for soft coral identification purposes. We also thank the Invertebrate collection at Instituto de Ciencias Naturales, Universidad Nacional de Colombia (Bogotá, Colombia) for curation, coding and storage of soft-coral samples.

Competing interests: The authors declare no conflict of interest.

3.6 REFERENCES

1. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).

2. Pontes, J. G. M., Brasil, A. J. M., Cruz, G. C. F., de Souza, R. N. & Tasic, L. NMR-based metabolomics strategies: plants, animals and humans. *Anal. Methods* **9**, 1078–1096 (2017).

3. Farag, M. a *et al.* Soft Corals Biodiversity in the Egyptian Red Sea: a Comparative MS and NMR Metabolomics Approach of Wild and Aquarium Grown Species. *J. Proteome Res.* **15**, 1274–1287 (2016).

4. Goulitquer, S., Potin, P. & Tonon, T. Mass spectrometry-based metabolomics to elucidate functions in marine organisms and ecosystems. *Mar. Drugs* **10**, 849–80 (2012).

5. Miljanich, G. P. Ziconotide: neuronal calcium channel blocker for treating severe chronic pain. *Curr. Med. Chem.* **11**, 3029–3040 (2004).

6. Ebada, S. S., Edrada, R. A., Lin, W. & Proksch, P. Methods for isolation, purification and structural elucidation of bioactive secondary metabolites from marine invertebrates. *Nat. Protoc.* **3**, 1820–1831 (2008).

7. Younes, A. *et al.* Brentuximab vedotin (SGN-35) for relapsed CD30-positive lymphomas. *N. Engl. J. Med.* **363**, 1812–1821 (2010).

8. Cortes, J., Montero, A. & Glück, S. Eribulin mesylate, a novel microtubule inhibitor in the treatment of breast cancer. *Cancer Treat. Rev.* **38**, 143–151 (2012).

9. Sagar, S., Kaur, M. & Minneman, K. P. Antiviral lead compounds from marine sponges. *Mar. Drugs* **8**, 2619–2638 (2010).

10. Correa, H., Valenzuela, A. L., Ospina, L. F. & Duque, C. Anti-inflammatory effects of the gorgonian Pseudopterogorgia elisabethae collected at the Islands of Providencia and San Andrés (SW Caribbean). *J. Inflamm. (Lond).* **6**, 5 (2009).

11. Amaya García, F. *et al.* Dolabellane diterpenes from the Caribbean soft corals *Eunicea laciniata* and *Eunicea asperula* and determination of their anti HSV-1 activity. *Rev. Colomb. Química* **46**, 5 (2017).

12. Reina, E., Ramos, F. A., Castellanos, L., Aragõn, M. & Ospina, L. F. Anti-inflammatory R-prostaglandins from Caribbean Colombian soft coral *Plexaura homomalla*. *J. Pharm. Pharmacol.* **65**, 1643–1652 (2013).

13. Van Der Greef, J., Stroobant, P. & Van Der Heijden, R. The role of analytical sciences in medical systems biology. *Curr. Opin. Chem. Biol.* **8**, 559–565 (2004).

14. Patel, K., Patel, J., Patel, M., Rajput, G. & Patel, H. Introduction to hyphenated techniques and their applications in pharmacy. *Pharm. Methods* **2**, 1–13 (2010).

15. Pettersson, L. & Elvingson, K. Studies of Vanadate-Organic Ligand Systems Using Potentiometry and NMR Spectroscopy. *ACS Symp. Ser.* 30–50 (1998) doi:10.1021/bk-1998-0711.ch002.

16. Panda, A., Parida, A. K. & Rangani, J. Advancement of Metabolomics Techniques and Their Applications in Plant Science: Current Scenario and Future Prospective. in *Plant Metabolites and Regulation Under Environmental Stress* 1–36 (Elsevier Inc., 2018). doi:10.1016/B978-0-12-812689-9.00001-7.

17. Lages, B. G., Fleury, B. G., Ferreira, C. E. L. & Pereira, R. C. Chemical defense of an exotic coral as invasion strategy. *J. Exp. Mar. Bio. Ecol.* **328**, 127–135 (2006).

18. Han, A. R. *et al.* Cytotoxic constituents of the octocoral *Dendronephthya gigantea*. *Arch. Pharm. Res.* **28**, 290–293 (2005).

19. Mahrous, E. A. & Farag, M. A. Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review. *J. Adv. Res.* **6**, 3–15 (2015).

20. Farag, M. A., Fekry, M. I., Al-hammady, M. A., Khalil, M. N. & Wessjohann, L. A. Cytotoxic Effects of *Sarcophyton* sp. Soft Corals — Is There a Correlation to Their NMR Fingerprints ? *Mar. Drugs* **15**, 1–13 (2017).

21. Foroozandeh, M. *et al.* Ultrahigh-resolution NMR spectroscopy. *Angew. Chemie - Int. Ed.* **53**, 6990–6992 (2014).

22. Straßburger, D. *et al.* Hadamard homonuclear broadband decoupled TOCSY NMR: Improved efficacy in detecting long-range chemical shift correlations. *ChemPhysChem* **17**, 4037–4042 (2016).

23. Santacruz, L., Thomas, O. P., Duque, C., Puyana, M. & Tello, E. Comparative analyses of metabolomic fingerprints and cytotoxic activities of soft corals from the Colombian Caribbean. *Mar. Drugs* **17**, 1–14 (2019).

24. De Ketelaere, B., Hubert, M. & Schmitt, E. Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data. *J. Qual. Technol.* **47**, 318–335 (2015).

25. Tistaert, C. *et al.* Multivariate data analysis to evaluate the fingerprint peaks responsible for the cytotoxic activity of Mallotus species. *J. Chromatogr. B* **910**, 103–113 (2012).

26. Triba, M. N. *et al.* Molecular BioSystems PLS / OPLS models in metabolomics : the impact of permutation of dataset rows on the K-fold. *Mol. Biosyst.* **11**, 13–19 (2014).

27. Chiu, C. Y. *et al.* Metabolomics reveals dynamic metabolic changes associated with age in early childhood. *PLoS One* **11**, 14 (2016).

28. Szymanska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 14 (2012).

29. Cho, H. W. *et al.* Discovery of metabolite features for the modelling and analysis of high-resolution NMR spectra. *Int. J. Data Min. Bioinform.* **2**, 176–192 (2008).

30. Abdullah, M. *et al.* Non-targeted metabolomics analysis of Golden Retriever Muscular Dystrophy-affected muscles reveals alterations in arginine and proline metabolism, and elevations in glutamic and oleic acid in vivo. *Metabolites* **7**, 1–19 (2017).

31. Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E. & Trygg, J. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *J. Chemiometrics* **20**, 341–351 (2006).

32. Blaženović, I., Kind, T., Ji, J. & Fiehn, O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* **8**, 31 (2018).

33. Rochat, B. Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data. *J. Am. Soc. Mass Spectrom.* **28**, 709–723 (2017).

34. Jacob, D., Deborde, C., Lefebvre, M., Maucourt, M. & Moing, A. NMRProcFlow: a graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. *Metabolomics* **13**, 36 (2017).

35. Xia, J., Sinelnikov, I., Han, B. & Wishart, D. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.* **43**, 251–257 (2015).

36. Olsen, E. K. *et al.* Cytotoxic activity of marine sponge extracts from the sub-Antarctic Islands and the Southern Ocean. *S. Afr. J. Sci.* **112**, 1–5 (2016).

37. Chong, J. *et al.* MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).

38. Rasmussen, S., Lane, G. A. & Villas-bo, S. G. Metabolomics or metabolite profiles ? *Sci. Direct* **23**, 385 (2005).

39. Moltu, S. J. *et al.* Urinary metabolite profiles in premature infants show early postnatal metabolic adaptation and maturation. *Nutrients* **6**, 1913–1930 (2014).

40. Powers, R. The current state of drug discovery and a potential role for NMR metabolomics. *J. Med. Chem.* **57**, 5860–70 (2014).

41. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **1**, 92–107 (2012).

42. Hostettman, K. Methods in Plant Biochemistry. Assays for Bioactivity. *New Phytol.* **119**, 474 (1991).

43. Fay, M. P. & Shih, J. H. Permutation Tests Using Estimated Distribution Functions. *J. Am. Stat. Assoc.* **93**, 387–396 (2014).

44. Gromski, P. S. *et al.* A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Anal. Chim. Acta* **879**, 10–23 (2015).

45. Yang, Z. *et al.* Exploring potential biomarkers of early stage esophageal squamous cell carcinoma in pre-and post-operative serum metabolomic fingerprint spectrum using ^{1}H-NMR method. *Am. J. Transl. Res.* **11**, 819–831 (2019).

46. Sponholtz, W. R., Bianco, M. A. & Gribble, G. W. Isolation and structure determination of the

cembranoid eunicin from a new genus of octocoral, *Pseudoplexaura*. *Nat. Prod. Res.* **22**, 441–448 (2008).

47. Sawant, S. S. *et al.* Bioactive rearranged and halogenated semisynthetic derivatives of the marine natural product sarcophine. *J. Nat. Prod.* **67**, 2017–2023 (2004).

48. Yang, S.-F., Weng, C.-J., Sethi, G. & Hu, D.-N. Natural Bioactives and Phytochemicals Serve in Cancer Treatment and Prevention. *Evidence-Based Complementary and Alternative Medicine* vol. 2013 112 (2013).

49. Tello, E. *et al.* Absolute stereochemistry of antifouling cembranoid epimers at C-8 from the Caribbean octocoral *Pseudoplexaura flagellosa*. Revised structures of plexaurolones. *Tetrahedron* **67**, 9112–9121 (2011).

50. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **43**, 34–40 (2017).

51. Rodríguez, A. D. The natural products chemistry of West Indian gorgonian octocorals. *Tetrahedron* **51**, 4571–4618 (1995).

52. Alfred J. Weinheimer & James A. Matson. Marine Anticancer Agents: Asperdiol, Cembranoid Prom Gorgonians, *Eunicea Asperula* And *E. Tourneforti*. *Tetrahedron* 2926 (1977).

53. Creek, D. *et al.* Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* **10**, 350–353 (2014).

54. Hegazy, M. E. F. *et al.* Cembrene diterpenoids with ether linkages from sarcophyton ehrenbergi: An anti-proliferation and molecular-docking assessment. *Mar. Drugs* **15**, 2–15 (2017).

55. Zhang, C. *et al.* Cytotoxic diterpenoids from the soft coral Sarcophyton crassocaule. *J. Nat. Prod.* **69**, 1476–1480 (2006).

56. Kim, H. K., Choi, Y. H. & Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **5**, 536–549 (2010).

57. Huang, Y. *et al.* A Pure Shift-Based NMR Method for Transverse Relaxation Measurements on Complex Samples. *IEEE Trans. Instrum. Meas.* **PP**, 1–11 (2019).

58. He, Q. *et al.* NMR-based metabolomic analysis of spatial variation in soft corals. *Mar. Drugs* **12**, 1876–1890 (2014).

59. Honda, A., Mori, Y., Iguchi, K. & Yamada, Y. Antiproliferative and cytotoxic effects of newly discovered halogenated coral prostanoids from the Japanese stolonifer *Clavularia viridis* on human myeloid leukemia cells in culture. *Mol. Pharmacol.* **32**, 530–535 (1987).

60. Rueda, A., Zubía, E., Ortega, M. J. & Salvá, J. Structure and cytotoxicity of new polyhydroxylated sterols from the Caribbean gorgonian *Plexaurella grisea*. *Steroids* **66**, 897–904 (2001).

61. Weinheimer, A. & Matson, J. Marine Anticancer Agents: Asperdiql, Cembranoid Prom Gorgonians, Eunicea Asperula And E. Tourneforti. *Tetrahedron Lett.* **18**, 1295–1298 (1977).

62. Al-Ghamdi, S. S. Time and dose dependent study of doxorubicin induced DU-145 cytotoxicity. *Drug Metab. Lett.* **2**, 47–50 (2008).

63. Shaikh, K. S., Pawar, A., Aphale, S. R. & Moghe, A. S. Effect of vesicular encapsulation on in-vitro

cytotoxicity of ciclopirox olamine. *Int. J. Drug Deliv.* **4**, 139–146 (2012).

64. Patel, S., Gheewala, N., Suthar, A. & Shah, A. In-Vitro cytotoxicity activity of *Solanum Nigrum* extract against Hela cell line and Vero cell line. *Int. J. Pharm. Pharm. Sci.* **1**, 38–47 (2009).

65. Sogin, E. M., Anderson, P., Williams, P., Chen, C. S. & Gates, R. D. Application of ^{1}H-NMR metabolomic profiling for reef-building corals. *PLoS One* **9**, 4–8 (2014).

66. Bao, Q. *et al.* A new automatic baseline correction method based on iterative method. *J. Magn. Reson.* **218**, 35–43 (2012).

67. Vu, T. N. *et al.* An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics* **12**, 14 (2011).

68. Anderson, P. E. *et al.* Dynamic adaptive binning: An improved quantification technique for NMR spectroscopic data. *Metabolomics* **7**, 179–190 (2011).

69. Kohl, S. M. *et al.* State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* **8**, 146–160 (2012).

70. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *R. Soc.* 16 (2016).

71. Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput. Struct. Biotechnol. J.* **14**, 135–153 (2016).

72. Lussu, M. *et al.* ^{1}H NMR spectroscopy-based metabolomics analysis for the diagnosis of symptomatic E. coli-associated urinary tract infection (UTI). *BMC Microbiol.* **17**, 1–8 (2017).

3.7 SUPPLMENTARY INFORMATION

Most commonly used methods for metabolomics analysis - advantages and disadvantages		
Technology	**Advantages**	**Disadvantages**
MS (LC/MS and GC/MS)	• High analytical sensitivity • Robust technique • give more information (TR, MS)	• Volume Large sample is required (500 L) • sample Destructive • Extensive sample preparation, (GC-MS) requires derivatization
NMR Spectroscopy (^{1}H/PSYCHE)	• Less sample is required (1–10 L) • Robust and reproducible • Non-destructive sample • Minimal sample preparation • Time acquisition (≈7min)	• Low analytical sensitivity • Signal overlap • give less information (δppm) • Time acquisition (≈20 - 30min)

Table S1. Advantages and disadvantages of the most commonly used methods for metabolomics analysis

Name assigned by collection area	Name of the species	Accession	Código ICN-UN
Cali 10	*Muriceopsis flavida*	C10m	ICN-MHN (Po)-CO-292
Cali 11	*Eunicea flexuosa*	C11E	ICN-MHN (Po)-CO-276
Cali 15	*Pseudopterogorgia albatrossae*	C15Pst	ICN-MHN (Po)-CO-278
Cali 16	*Eunicea fusca*	C16E	ICN-MHN (Po)-CO-277
Cali 18	*Muriceopsis flavida*	C18m	ICN-MHN (Po)-CO-292
Cali 19	*Eunicea flexuosa*	C11E	ICN-MHN (Po)-CO-276
Cali 2	*Plexaura kukenthali*	C2P	ICN-MHN (Po)-CO-271
Cali 24	*Plexaurella nutans*	C24p	ICN-MHN (Po)-CO-281
Gra 10	*Eunicea knightii*	G10Ek	ICN-MHN (Po)-CO-287
Gra 1	*Plexaurella fusifera*	G1P Plr	ICN-MHN (Po)-CO-282
Gra 13	*Eunicea clavigera*	G13E	ICN-MHN (Po)-CO-275
Gra 14	*Eunicea cf. calyculata*	G14E	ICN-MHN (Po)-CO-288
Gra 16	*Eunicea knightii*	G16Ek	ICN-MHN (Po)-CO-287
Gra 17	*Pseudoplexaura flagellosa*	G17Ps	ICN-MHN (Po)-CO-289
Gra 18	*Plexaura kukenthali*	G18P	ICN-MHN (Po)-CO-292
Gra 21	*Plexaura homomalla*	G21P	ICN-MHN (Po)-CO-283
Gra 2	*Plexaura homomalla*	G2P	ICN-MHN (Po)-CO-283
Gra 22	*Plexaurella* sp.	G22P	ICN-MHN (Po)-CO-291
Gra 24	*Plexaura homomalla*	G2P	ICN-MHN (Po)-CO-283

Gra 3	*Eunicea clavigera*	G13E	ICN-MHN (Po)-CO-284
Gra 4	*Eunicea asperula*	G4E	ICN-MHN (Po)-CO-285
Gra 6	*Plexaura kukenthali*	G6Pk	ICN-MHN (Po)-CO-271
Gra 8	*Eunicea clavigera*	G8E	ICN-MHN (Po)-CO-275
Gra 9	*Plexaura sp.*	G9P	ICN-MHN (Po)-CO-286

Table S2. Name of species of soft corals used in this research a Collection Code (ICN), assigned by Collection of the Institute of Natural Sciences of the National University of Colombia. (Bogotá, Colombia)

Obs	name	Siha (%)	Pc3 (%)	A549 (%)	L929 (%)
1	C10m	23.1	20.8	31.0	49.8
2	C11E	36.8	12.1	49.3	21.7
3	C15Pst	26.8	20.3	71.5	25.9
4	C16E	60.5	30.0	39.1	21.3
5	C18m	21.2	15.0	26.5	13.8
6	C19E	15.1	16.1	36.1	20.7
7	C2P	46.2	7.8	62.4	16.2
8	C24p	13.1	65.1	39.0	13.3
9	G10Ek	46.1	1.3	35.4	14.5
10	G1P Plr	39.4	23.8	33.0	11.5
11	G13E	46.9	30.9	39.7	6.8
12	G14E	63.5	22.0	65.7	19.2
13	G16Ek	26.1	15.3	30.4	14.5
14	G17Ps	66.8	40.3	71.5	65.9
15	G18P	44.0	64.0	52.5	31.2
16	G21P	25.5	20.9	34.0	18.4
17	G2P	34.8	47.6	21.1	14.3
18	G22P	37.6	63.5	31.9	20.7

19	G24P	37.6	33.5	31.9	20.7
20	G3E	59.9	43.7	35.3	5.8
21	G4E	53.9	41.5	67.6	14.6
22	G6Pk	57.0	40.3	38.5	23
23	G8E	54.8	47.6	22.1	14.3
	G9P	34.5	20.4	31.0	49.8
29	Doxorubicin (25ppm)	60.1	46.9	58.2	16.0

Table S3. Percentage of cytotoxic activity against three cancerous cell lines and fibroblasts L929 (ATCC®CCL-1 ™) which was used as non-tumor cell line for toxicity control.

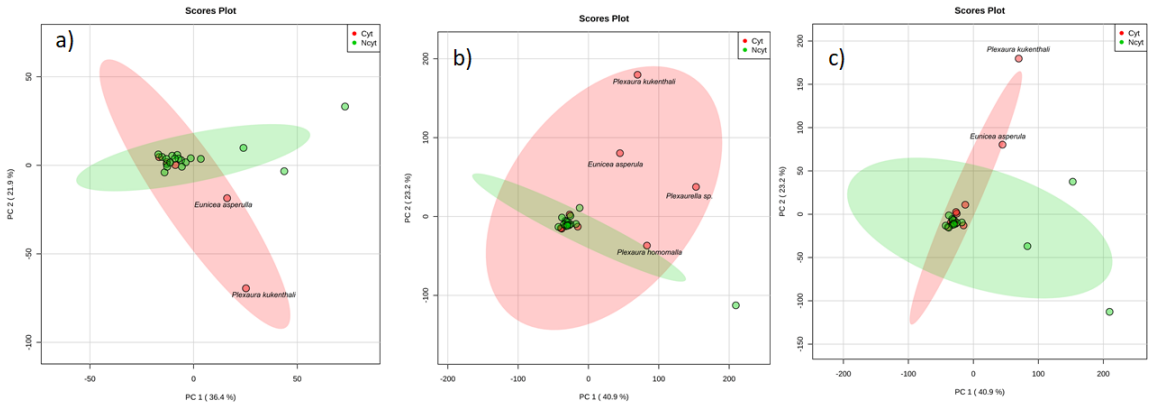


Figure S1. Principal component analysis score plot of ^{1}HNMR/metabolomics data from 24 extracts based on their cytotoxicity against different cancer cell lines (a) human lung adenocarcinoma, A549 (b) human prostatic carcinoma, PC3 and (c) human cervical cancer, SiHa. Red dots indicate active extracts, green dots represent extracts that were not active. An extract was considered active if it exhibited an inhibition of the tumor cell lines ≥ 40% at 20 µg/mL) (Hostettman, 1991). The ellipses indicate confidence intervals of 95%.

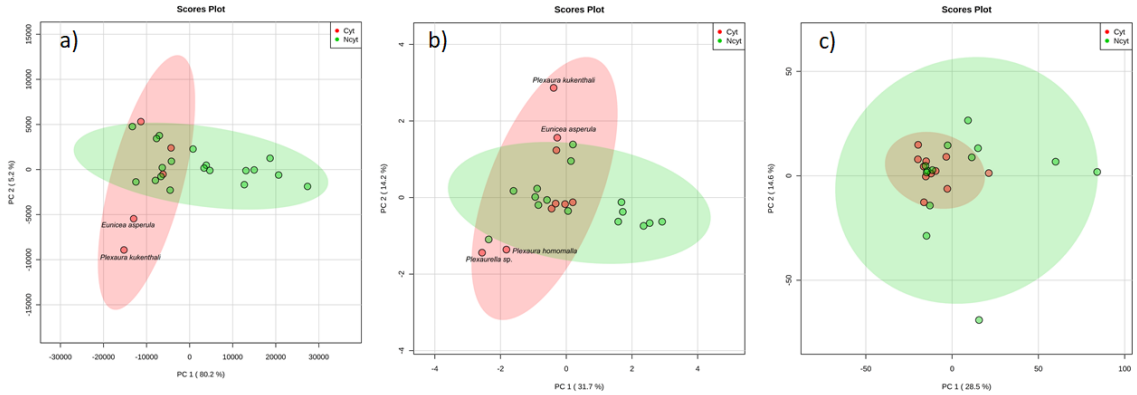


Figure S2. Principal component analysis score plot of PSYCHE/metabolomics data from 24 extracts based on their cytotoxicity against three different tumor cancer cell lines (a) human lung adenocarcinoma A549 (b) human prostatic carcinoma PC3 and (c) human cervical cancer SiHa. Red dots indicate active extracts, green dots represent extracts that were not active. An extract was considered active if it exhibited an inhibition of the tumor cell lines ≥ 40% at 20 µg/mL(Hostettman, 1991). The ellipses indicate confidence intervals of 95% (Chong et al., 2018).

Supplementary text for ^{1}H-NMR and PSYCHE experiments data extraction

Instrument/Vendor/Format = Varian
Spectra type = fid
The original name of the Zip file = testPscomple31hasta8ppm sincali8nigra20.zip
The number of Spectra = 24
Calibration: PPM REF = 0, Zone Ref = (0.045,0.071) to PSYCHE experiment
calibration 0.045 0.071 0 10.2 10.5
Normalisation (CSN) of the Intensities based on the selected PPM ranges 0.348 7.847
Global Baseline Correction: PPM Range = (-0.499764773120937, 10.999765269097)
gbaseline 10.2 10.5 -0.499764773120937 10.9997652690972 100 70
Alignment of the selected zones (0.18,7.972)
Rbuc1D: --- BUCKETING MODULE ---
Rbuc1D: Read the SpecProcpar.ini file ...
Rbuc1D: Read the spectra processing parameters file (list_pars.csv) ...
Rbuc1D: Read the specs.pack.
Rbuc1D: Bucketing method = AIBIN
Rbuc1D: Noise Zone= (10.5 ,10.2), (818 - 1310)
Rbuc1D: Factor Resolution = 0.3
Rbuc1D: Zone 1 = (0.133, 7.905), Nb Buckets = 1230
Rbuc1D: Total Buckets = 113 to PSYCHE experiment
The parameters for ^{1}H-NMR are the same, only the calibration varies using CDCl3:
Calibration: PPM REF = 7.26, Zone Ref = (7.25, 7.26)

Table S4. Scrip using NMRProcFlow software to ^{1}H-NMR and PSYCHE experiments to data extraction

Main results of the study					
Experiment	Tumoral cell line	Cytotoxic activity (statistically significant)	Extracts that exhibit activity	VIP Features	Compounds / Putative identification
^{1}H NMR	**A549**	No	-	-	-
	PC3	No			
	SiHa	No	-	-	-
Psyche	**A549**	No			
	PC3	Si	• *Plexaura kukenthali* • *Plexaurella* sp.	• *B2_5118* • *B4_4965* • *B4_7686*	• Asperdiol • Plexaurolone
	SiHa	No	-	-	-

Table S5. Summarizing the key results of the study

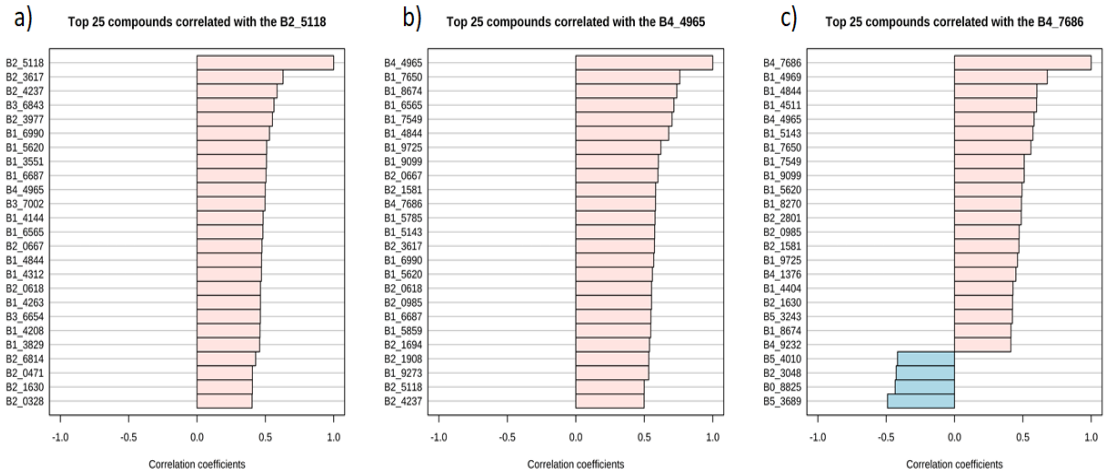


Figure S3. Twenty five features that most correlate with the features B2_5118, B_4965 and B4_7686 , established using "Pattern Hunter" a tool of MetaboAnalyst software (Chong et al., 2018).

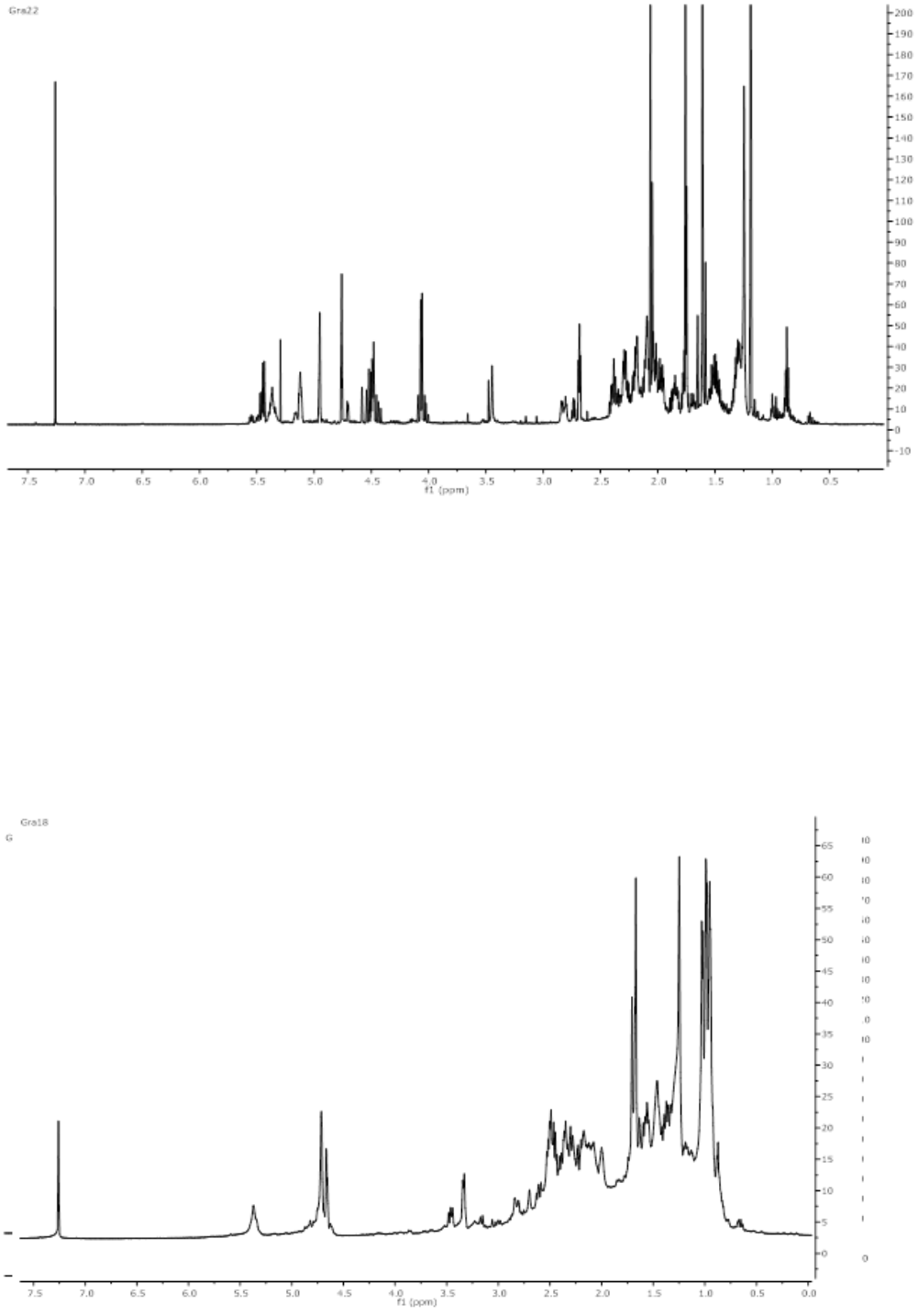


Figure S4. a) ^{1}H NMR spectrum of crude extract from specie *Plexaurella* sp, (Gra22) in $CDCl_3$ at 600 MHz. b) ^{1}H NMR spectrum of crude extract from specie *Plexaura kukenthali* (Gra18) in $CDCl_3$ at 600 MHz.

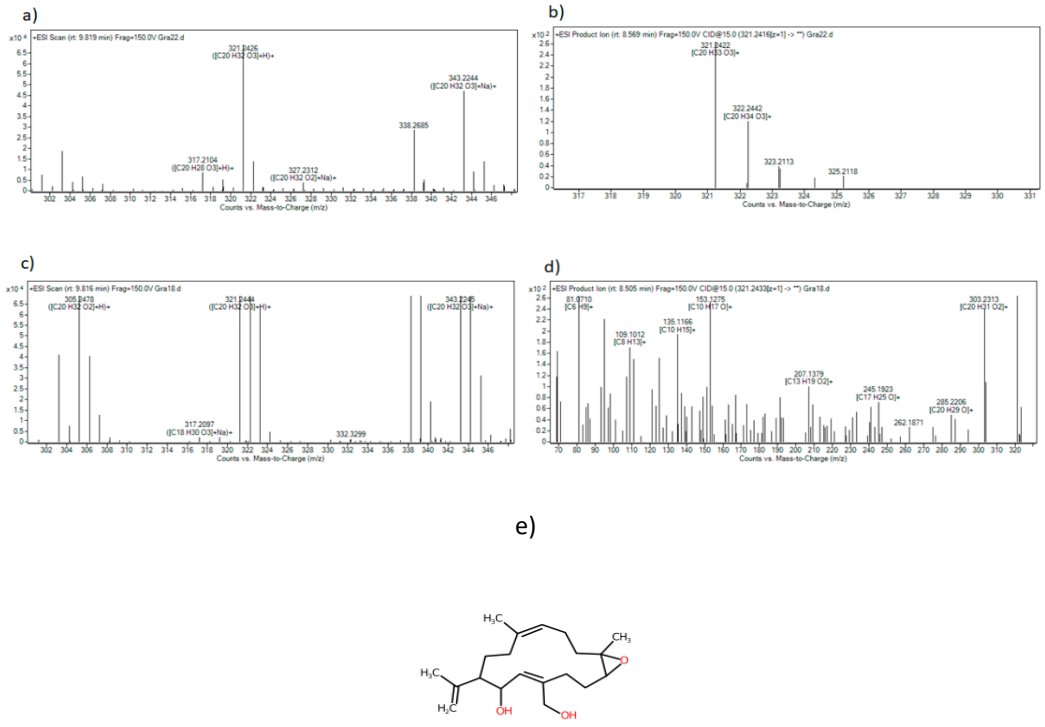
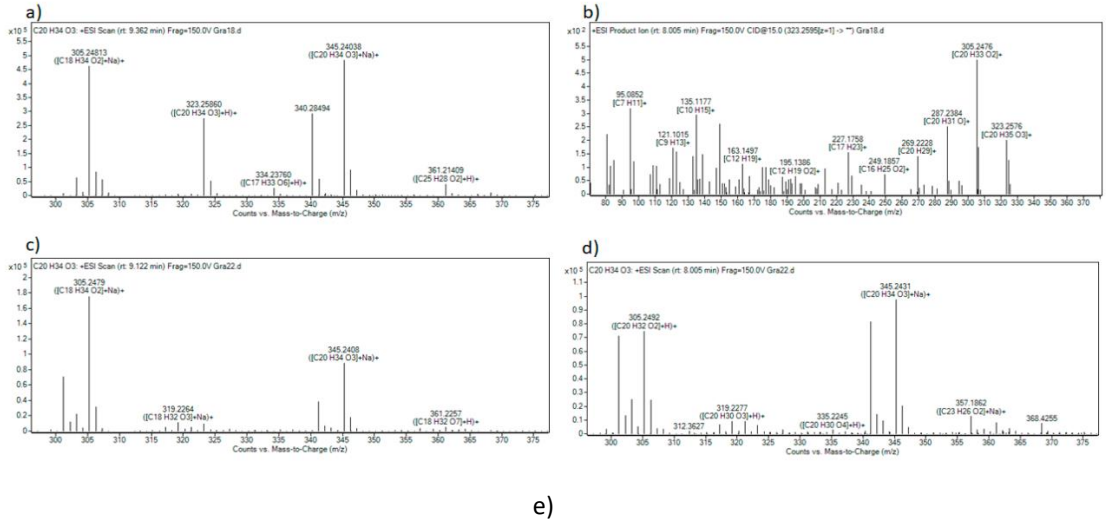


Figure S5. a) MS from *Plexaurella* sp (Gra 22) extract, b) MS/MS spectrum of 321.2416 ion from *Plexaurella* sp, c) MS from *Plexaura kukenthali* (Gra 18) extract, d) MS/MS spectrum of 321.2433 ion from *Plexaura kukenthali*. These spectrums confirming the possible presence of the compound asperdiol showing a molecular formula of $C_{20}H_{32}O_3$, e) asperdiol compound structure.



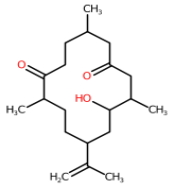


Figure S6. a) MS from *Plexaura kukenthali* (Gra 18) extract, b) MS/MS spectrum of 323.2595 ion from *Plexaurella* sp, c) MS from *Plexaurella* sp, (Gra 22) extract, d) MS/MS spectrum of 345.2431 ion from *Plexaurella* sp. These spectrums confirming the possible presence of the compound plexaurolone showing a molecular formula of $C_{20}H_{34}O_3$, e) plexaurolone compound structure.

4 Integration of Molecular Networking and metabolomics data for the identification of bioactive metabolites from octocorals of the Plexauridae family

Liliana Santacruz[1], Olivier P. Thomas[2], Mónica Puyana[3] and Edisson Tello[1]

[1] Bioprospecting Research Group and Doctorado en Biociencias Program, Faculty of Engineering, Campus Puente del Común, Universidad de La Sabana, 250001 Chía, Colombia
[2] Marine Biodiscovery, School of Chemistry and Ryan Institute, National University of Ireland Galway (NUI Galway), University Road, H91 TK33 Galway, Ireland
[3] Departamento de Ciencias Biológicas y Ambientales, Universidad Jorge Tadeo Lozano, Carrera 4 # 22-61, 110311 Bogotá, Colombia

Corresponding Author:
Edisson Tello[1]
Bioprospecting Research Group and Doctorado en Biociencias Program, Faculty of Engineering, Campus Puente del Común, Universidad de La Sabana, 250001 Chía, Colombia.
Email address: edisson.tello@unisabana.edu.co (E.T.)
Tel.: +57-32-0699-5696 (E.T.)

Abstract

This study focuses on cytotoxic activity exhibited by the 11 samples of soft corals extract from *Plexaura* and *Plexarella* genus against the tumor cell lines A549, SiHa and PC3. The discriminant analysis PLS-DA, found that the VIPs features (responsible for the separation of the group of extracts that showed the highest cytotoxic activity) were M321T579 (asperdiol), M323T549 (plexaurolone) and M287T644 (13 keto -1) (S), 11 (R) -dolabell -3 (E), 7 (E), 12 (18) -triene). The molecular networking (MN), was carried out, taking into account that the formation of the clusters was done only by the similarities in fragmentation of the chemical compounds that would be present in this genus. The molecular network showed the formation of a cluster that has the cytotoxic VIPs features found in the metabolomic analysis. This workflow allowed to determine that the *Plexaura* and *Plexaurella* genus owe their cytotoxic activity not only to the presence of the compounds identified as VIP cytotoxic but also to the correlation of chemical structure that exists between them. From the compounds identified as cytotoxic, was possible assigned a putative chemical structure for other compounds that were in the same cluster.

4.1 INTRODUCTION

The successful search for bioactive compounds from marine organisms has turned the marine environment into an immeasurable source for the discovery of new biomedical drugs [1]. Soft corals have been recognized as a rich source of metabolites with potential cytotoxic application [2]. In particular, the family Plexauridae (Cnidaria: Anthozoa: Gorgonacea) provides a wealth of original chemical structures where terpenoids and especially diterpenes are the most representative [3,4].

Some diterpenes of the family Plexauridae such as edunone and edudione [5] have been reported as cytotoxic against Hela tumor cell line (cervical cancer) with values of IC5o 25 pg /ml and

E51 IC5o> l00 pg / ml respectively. On the other hand, Nieto et al, isolated three cembranolides which showed moderate cytotoxic activity against (Lung cancer) A549, (colon cancer) H116 and (pancreatic cancer) PSN1 [6].

The metabolomics studies which is defined as the comprehensive analysis of metabolites within cells, biofluids, tissues, or organisms to identify and quantify low-molecular-weight (<1 kDa) small metabolites in a biological system [7,8]. Metabolomics studies have been used in marine natural product drug discovery (art 3. Drug discovery) has also contributed to metabolome variability In marine organisms as demonstrated by Reverter et, al, in a study of two Mediterranean sponge species of the genus Haliclona [9] and in the discrimination of taxonomic diversity of some soft corals [10].Commonly, Liquid Chromatography - Mass Spectrometry (LC-MS) and Nuclear Magnetic Resonance (NMR) data are used in metabolomics experiments [11,12] and the combination of cytotoxic activity assay of crude extracts with metabolomics can accelerate the discovery of new and natural drugs [13].

Metabolomics studies provide large amount of data, therefore, multivariate projection methods for data exploration (MVDA) have been applied to study the variability of the metabolome [14]. Partial Least Squares-Discriminant Analysis (PLS-DA) is a supervised technique were PLS regression method with it is commonly used for classification of chemical markers in metabolomics studies [15]. Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) has the property to discriminate and separate predictive from non-predictive (orthogonal) variation, integrating an OSC-filter that removes systematic spectral variation that does not agree with the assigned group memberships this allows a visualization of the information of the response matrix Y and the descriptor matrix X orthogonal (non-correlated) to Y, thus helping to discriminate the different samples through clusters [16]. Which is very useful for metabolomic studies that relate biological activities to chemical composition [17].

Some metabolomic studies in soft corals have allowed the identification of the compounds responsible for the cytotoxicity of these marine organisms, for example, the compound flexibilide isolated from the soft coral Sinularia flexibilis [18] which involved in five metabolic pathways against HCT-116 cells and the dolabellatrienone compound isolated from the *Pseudoplexaura flagellosa* species that exhibited cytotoxic activity IC50 = 0.02 µg / mL against A549 and IC50 = 0.03 µg / mL against SiHa [19].

However, the main bottleneck of a metabolomics approach often lies in the identification (annotation) of the important features responsible for the discriminations. Global Natural Products Social Molecular Networking (GNPS) tool has appeared recently as a rapid comparison tool of MS/MS profiles from complex crude extracts [20]. This approach helps in the annotation and clustering of structurally related compounds, using high-resolution mass spectrometry of parent ion fragmentation data (MS/MS) [21]. The compounds included in the clusters will share similar fragmentation pattern (minimum of 6 fragments) under identical ionization conditions [22]. The parent ions (nodes) are then connected by edges and the width of the edges will represent the cosine score value ranging from 1 (identical fragmentation spectra) to 0 (completely different parent ions) [20].

In this study we applied a high-resolution mass spectrometry metabolomic approach on *Plexaura* and *Plexaurela* genus to identify the variables importance of projection (VIP) (responsible for the cytotoxic activity. In addition, Molecular network was performed using MS / MS data in a process of "pattern based on genus" with the MS-Cluster algorithm implemented on the Global Natural Product Social Molecular Networking platform (GNPS) and Cytoscape was applied for analysing molecules clustering with a similar fragmentation profile. Finally, it was combined both the metabolomics outcome and Molecular network result in order to verify if there is a positive relation between the chemical structure presented in the studied genus and their cytotoxic activity

4.2 RESULTS

4.2.1 Relationship between Metabolomics profiling of *Plexaura* and *plexaurella* extracts and their cytotoxic activity

The workflow used allowed the visualization of composition/activity correlations for the extracts of the Plexauridae soft corals using a combination of information from multivariate analyses and molecular networks (Fig. 1). In this study, the data matrix was built with the extracts only of these two genus *plexaurella* and *plexaura* (11 samples) and to the multivariate analysis of OPLS-DA, the cytotoxicity was used as a discriminative parameter. In addition, the information obtained from the MS/MS data was used to determine if the greater activity presented by the extracts of these genus was due to a correlation between the compounds VIPs identified as cytotoxic or to correlation with other compounds that could be present in the extracts studied. The MS/MS data were incorporated into the GNPS platform where the groups were separated by genus (*Plexaura* and *Plexaurella*) and the visualization of the network and clusters, was possible through the use of Cytoscape software.

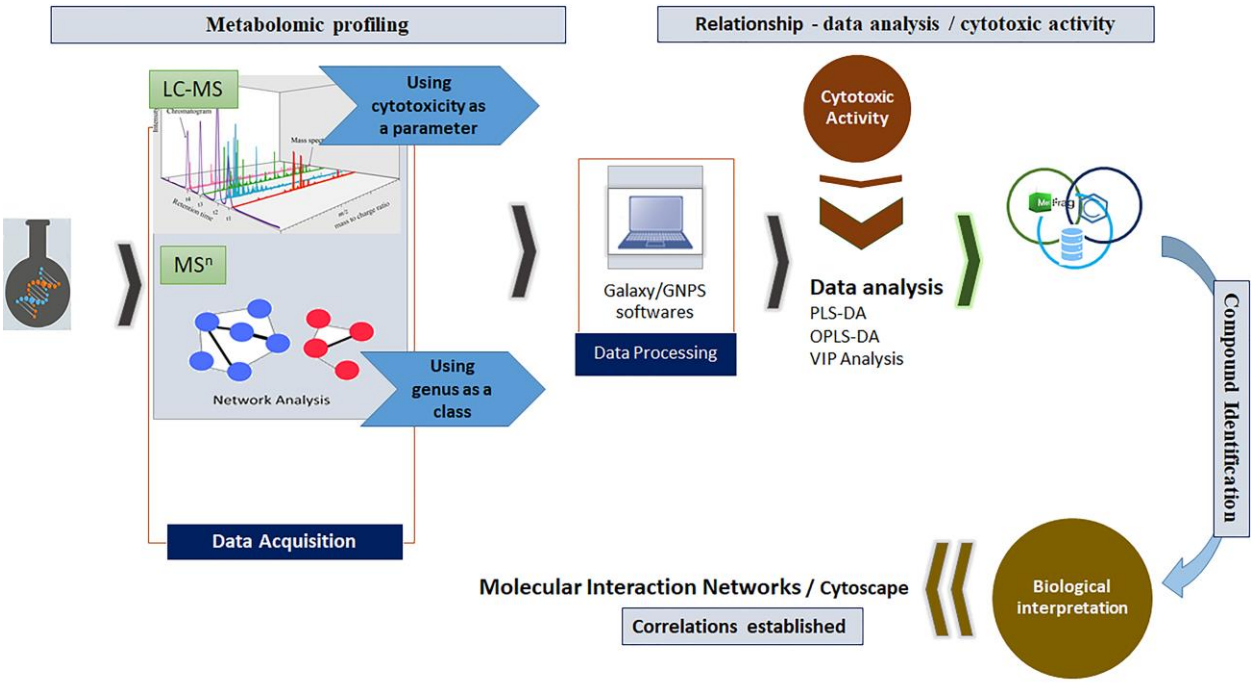


Figure 6. Workflow that allowed to establish a correlation between the cytotoxic activity-the chemical composition of the soft corals plexauridae family.

Taking into account that the extracts of the genus *Plexaura* and *Plexaurella* studied here showed cytotoxic activity some against the tumor lines of PC3, A549 and SiHa (table S2); for this study the "class cytotoxicity" (extract active or low active) was used in the data matrix when the inhibition value was greater than 50% at 20 µg/mL concentration [13], against two of tumor cell lines. With these data matrix, the multivariate analysis was elaborated, and the results were visualized by score plot and loading plot from OPLS-DA analysis shown in Fig. 2a and 2b. In addition, the VIP analysis from PLS-DA (Fig. 2C), validated what was observed in the loading plot where the features M321T579, M323T549 and M287T644 were the furthest from the group of features present in the most cytotoxic extracts. Therefore the analysis of VIP, revealed several distinguishing patterns considering a higher score value as discriminatory of main variables. [23]. From the analysis of the Molecular Network (MS / MS data) (Fig. 2d) a cytotoxic cluster was highlighted, because of the presence of cytotoxic features in the cluster (identified by their mass and retention time) linked to nodes/metabolites that presented a similar fragmentation pattern; (see figure S1) [24].

The structural grouping (cluster) observed in the molecular network is extremely useful when a node has been annotated to a particular structure. Indeed it can then help identifying connected

nodes in the network by proposing similar structures [25]. Thus, annotations can be propagated trough connected nodes. But the Molecular Network approach is highly dependent on tandem MS/MS data available in databases, and it is currently limited by the size of available fragmentation data libraries especially for marine organisms [26]. Therefore, this workflow includes, mainly, the putative identification of the metabolites, which are in the cytotoxic cluster and which agrees with the analyzes of PLSDA and VIPS previously carried out (Fig. 2d).

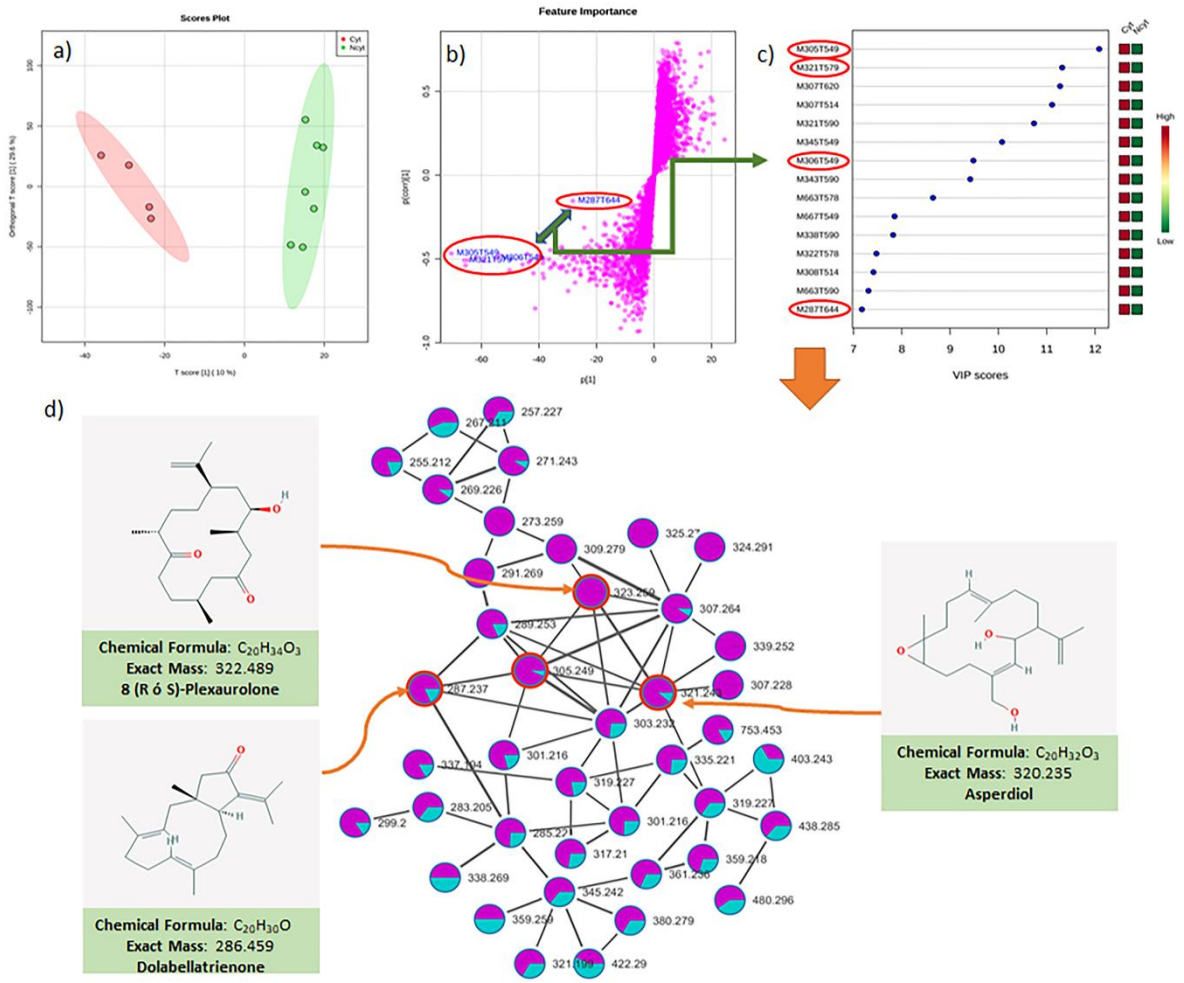


Figure 7. Structure/Activity correlation of soft coral extracts from plexauridae family a) Scores Plot from OPLS-DA b) Loading plot from OPLS-DA of 11 soft coral extracts from genus Plexaura and Plexaurella based on their cytotoxicity. In Red dot: active extracts; green dot: low-active extracts. The ellipse determines the 95% of confidence interval [27]. c) The VIP (Variable Importance in the Projection) scores [28]. From OPLSDA analysis. The VIP features visualized correspond to the main compounds responsible for the separation of the group of extracts that showed cytotoxic activity. d) Molecular Network/Cytotoxic Mapping Cluster (*plexaura* purple and *plexaurella blue*) [29]

A permutation test which involves randomly reassigning the class labels was undertaken and the value of R2 was also determined. Finally the Q2 that indicates the predictive relevance of the path or structural model OPLS-DA for distinguishing between cytotoxic and low-cytotoxic groups soft corals was measured [27].

The R2 values can be classified as weak (R2: 0.19), moderate (R2: 0.33) and substantial (R2: 0.67) [30]. The indicator the cross-validated Q2, [31,32] indicates the predictive relevance of the path or structural model [30]. In the structural model, Q2 values that are greater than zero indicate predictive relevance in the path model, while Q2 values less than 0 suggest a lack of predictive relevance in the model for that particular variable [33].

In this research Although the Q2 value for the PLS-DA model were low (0.251) and with an R2Y value of 0.972, the scores scatter plot indicated a significant variation in the chemical profile of the soft coral extracts while clearly separating the cytotoxic extracts from low-cytotoxic extracts.

4.2.2 Correlations established for the Compound Identification

Using the information obtained from the cytotoxic cluster of the molecular network it was possible to propose structures to the compounds directly correlated with the features /VIPs as shown in the following table:

Table 1. Correlated features with features/VIPs - using information from cytotoxic cluster.

VIP Feature	Putative Identification using MarinLit database and NMR Data/absolute identification [19]	Main correlations observed in network (cosine score > 7.0)/unidentified compounds*	Mass shifts for which the mass difference between network pairs of known chemical groups (VIP feature – Feature from main correlation)/putative element or group [42]
M321T579	Asperdiol ($C_{20}H_{32}O_3$)	M303T579 M319T579 M307T579 M289T514* M307T620* M305T549	18 (H_2O) -2 (H_2) 14 (CH_2) M323T549-H_2O
M323T549	8 (R ó S)-Plexaurolone ($C_{20}H_{34}O_3$)	M303T579 M309T549 M305T549 M307T549 M321T579	M321T579 – H_2O 14 (CH_2) 18 (H_2O) -16 (O) Plexaurolone

M287T644	13-keto-1,11-dolabell-3,7,12(18)-triene* ($C_{20}H_{30}O$)	M285T644 M289T514* M305T549 M303T579	-2 (H_2) M323T549-H_2O M321T579 – H_2O

N.O: not observable; N.A: Not apply. RT. Retention time.

In accordance with table 1, the features directly related to the features/VIPs (dolabellatrienone, asperdiol, plexaurolone) found in molecular network using MS/MS data correspond mainly to water losses or to the formation of unsaturation in VIPs except for the feature M289T514 and M307T620 which appears on the table as unknown identification.

The Analisys on KEEG database confirmed that the chemical characteristics (exact mass and retention time) of these two structurally related compounds (by molecular network) with others that presented cytotoxic activity (by metabolomic study), may possibly correspond to Abietinol ($C_{20}H_{32}0_2$) and 2,7,11-Cembratrien-4,6-diol ($C_{20}H_{34}O_2$) respectively. Additionally, the tandem mass spectral data of these compounds incorporated in MetFrag web server confirmed (score above 7.0) their putative identification as observed in figure 3.

Pathway analysis was performed using the KEEG database [34] to verify the structural relationship of the compounds of Abietinol and 2,7,11-Cembratrien-4,6-diol taking into account their common biogenetic origin (terpenoid type compounds related) and we were able to establish that the compounds that were investigated were formed by the diterpenoid route (Shikimate / acetate-malonate pathway derived compounds --- >Terpenoids --- >Diterpenoids (C20) --- >Grandfathers --- > C11882 Abietinol --- >Cembrenes --- > C09072 2,7,11-Cembratrien-4,6-diol).

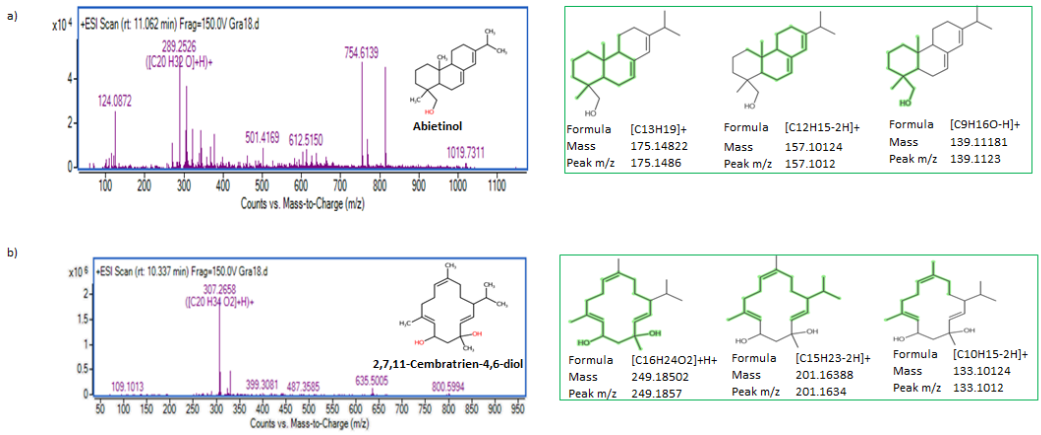


Figure 8. a) Spectrum mass of abietinol b) Spectrum mass of 2,7,11-Cembratrien-4,6-diol. In green box the structures formed by fragmentation (tamden mass by MetFrag).

4.3 DISCUSSION

The workflow established in this study, allowed to filter the information and identify that the eleven extracts from the soft corals *plexaura* and *plexaurella* genus showed cytotoxic activity either low and high against three tumor cell lines (PC3, SiHa and A549). From the multivariate analysis using as discriminatory variable the cytotoxicity of these organisms, it was possible to establish in the loading plot that the previously identified features/VIPs M321T579 (asperdiol), M323T549 (Plexaurolone) and M287T644 were the responsible for the separation of the group of extracts of the *plexaura* and *plexaurella* genus which presented more cytotoxic activity and were visualized as VIPs from PLS-DA analysis.

Furthermore, to answer the hypothesis established in this study that the activity / structure correlation between the VIPs (plexaurolone, dolabellatriene and asperdiol) present in the plexauridae family; the MS / MS data were incorporated into the GNPS platform and its visualization was done using the CytoScape software, which helped to identify a closer in the molecular network that contained the mass of the VIPs together with other masses that corresponded mainly to the VIPs with losses of water (-18), or to the formation of unsaturation (-2). The information produced by this type of study can be used if the synergy of these compounds allows obtaining a greater cytotoxic activity, as has been shown in several studies, such as that of Williamson et al [35] in the that synergy between different constituents of extracts. has been documented in pharmacological activities or by what Gilbert and collaborators [36] demonstrated in their work on Synergy in plant medicines.

From analysis for the putative identification and metabolic pathways for the two compounds discovered in the cytotoxic cluster using KEEG, was possible to identify the compounds abietinol and 2,7,11-Cembratrien-4,6-diol, formed via the diterpene biosynthetic pathway. Therefore, it can help us to infer if the biosynthetic pathways are involved in the production of potential metabolites with interesting pharmacological activities such as antitumor activity.

4.4 MATERIALS AND METHODS

4.4.1 Samples

Small portions (of approximately 30 cm) were cut off the main soft coral colony with sharp scissors (N=11) (Table S1, Supplementary Information) were collected by SCUBA diving at Punta Venado (11°16.26′ 87″N, 74°12.24′58″W), Santa Marta, Colombian Caribbean. Samples were collected at a depth range between 10–20 m. Collected samples were stored in dry ice and transported to the laboratory. These samples were air dried and then kept frozen until the moment of extraction at −80 ◦C.

Samples were identified by colony morphology and sclerite shape, dimensions and distribution. Sclerites were obtained from a distal fragment of each soft coral portion that was treated with 5% sodium hypochlorite. Once the organic material was dissolved, sclerites were observed under the microscope. A voucher of each sample is stored at the collection of the Instituto de Ciencias Naturales (ICN) of Universidad Nacional de Colombia (Bogotá, Colombia) (Table S1 in Supplementary Information).

4.4.2 General Experimental Procedures

Data produced by tandem mass spectrometry (MS/MS) and high-resolution electrospray ionization mass spectrometry detection (HRESIMS) were obtained with an Agilent 6540 mass spectrometer. UV measurements were obtained by the extraction of the diode array detector (DAD) signal in a PerkinElmer HPLC-DAD-ELSD FLEXAR LC(r) SYSTEM.

The general protocols used for the extractions of the samples and to cytotoxicity assays were established in two previous investigations [19], using methanol and dichloromethane as extraction solvents purchased from Merck (Darmstadt, Germany) and D-MEM (Dulbecco's Modified Eagle Medium (1X), RPMI 1640 Roswell Park Memorial Institute, Darmstadt, Germany were made by Gibco/Invitrogen, Paisley, UK. Other reagents were Fetal bovine serum (FBS), Eurobio brand (Les

Ulis, France), trypticase soy broth (TSB) and trypticase soy agar (TSA) Scharlau Co. brand (Barcelona, Spain). All cancer cell lines were acquired from ATCC, PC3 human prostatic carcinoma (ATCC® CRL1435™), SiHa human cervical carcinoma (ATCC® HTB-35™) and A549 human lung adenocarcinoma (ATCC® CCL-185™) For use in cell culture.

4.4.3 Metabolomic Analysis

Sample preparation for metabolomic analysis was performed using 1.0 g of dried powder from each soft coral was extracted at room temperature with a mixture of 1:1 DCM/MeOH, three times (30 mL) using an ultrasonic bath for 20 min. Debris were removed by centrifugation two times at 12,000× g for 5 min. Solvents were evaporated, and dried extracts were passed through a C18 cartridge, eluting with MeOH to remove salts. Subsequently, the extracts were concentrated and the samples were purified by preparative RP-HPLC with a Phenyl-Hexyl OBD Column (XSelect CSH, 19 mm × 250 mm, 5 µm), the optimization of gradient profiles was performed by selecting the mobile phase used for HPLC UV detection at λ 254 nm during 40 min. of acquisition time. Elution was done using water (A) and acetonitrile (B), both containing 0.1% formic acid with a gradient elution of 20%–50% B over 20 min, 50%–80% B over 10 min, 80%–100% B in 1 min, and holding for 5 min.

The UHPLC-HRESIMS and MS/MS data from 11 extracts of soft corals from plexauridae family (See table S1 in supplementary material) were acquired on an Agilent 6540 mass spectrometer using, MS acquisition (*m/z* 300–3000) at 10 spectra/sec and mass accuracy of 1 ppm. Electrospray Ionization (ESI) in positive mode was used to ionize and detect compounds after chromatographic separations. General parameters of the MS1 mode source were gas flow of 12 L min^{-1}, gas temperature of 300 ◦C, voltage charge of 2000 V, fragmentor of 150 V, capillary voltage of 3500 V, nebulizer pressure of 30 psi, and octopole RF Peak of 700 V.

Auto MS/MS fragmentation was carried out for the four most intense ions per spectrum and using fixed collision energies of 15 eV.

Chromatographic separation was achieved using a Phenyl-Hexyl HPLC column (150 mm×3.0 mm, 1.9 µm Poroshell, Agilent Technologies, Santa Clara, CA, USA). The mobile phase consisted of (A) water with 0.1% formic acid and (B) methanol (MeOH) with 0.1% formic acid. The UPLC injection volume on each run was 1.0 µL. Samples were filtered using a 0.2 µm Whatman® membrane filter with a pore size of 0.2 µm (Merck, Germany) prior to injection. Data were analysed using the Agilent

MassHunter Qualitative software (Version B.07.00). For formula generation, the Molecular Formula Generator algorithm (MFG) was used, which can automatically eliminate unlikely candidate compounds and rank the putative molecular formula according to their mass deviation, isotopic pattern accuracy, and elemental composition, with the number of hits limited by assigning the maximum number of atoms expected of each element (C, H, O, N) [37].

4.4.4 Molecular Networking

Molecular networks were created using the chemical information of the 11 extracts of soft corals of the family plexauridae in online workflow at GNPS (http://gnps.ucsd.edu). The MS/MS data were converted into mzXML files using the MSConvert from ProteoWizard software (ProteoWizard, Palo Alto, CA, USA) [38] and a molecular network was created using the online workflow at GNPS. The data were then clustered with MS-Cluster with a parent mass tolerance of 0.02 Da and a MS / MS fragment ion tolerance of 0.02 Da to create consensus spectra. Further, consensus spectra that contained less than 2 spectra were discarded. network was then created where edges were filtered to have a cosine score above 0.7 and more than four matched peaks.

The spectra in the network were then searched against GNPS spectral libraries. All matches between network spectra and library spectra were selected with a cosine score above 0.7. For visualization and more specific analysis, the network data was exported and analysed into Cytoscape (Version 3.6, Cytoscape consortium, San Diego, CA, USA) [39].

4.4.5 Multivariate data analysis

All variables of the data matrix were scaled using the mean center algorithm prior to multivariate analysis [40]. To identify variables showing a particular pattern of change the "pattern hunter" he MetabolAnalyst tool was used, which help in the identification of molecules showing a correlation of 0.5 or less than 0.5 with important features such as VIP features [41]. In this way, molecules that contribute to pattern recognition of the main features responsible for the separations of the groups are evidenced according to established classes (e.g. Cytotoxic activity) [42].

PLS-DA analysis was performed to the 11 extracts of soft coral using MetaboAnalyst software [27,28] due to classification purposes of PLS-DA[16]. as a discriminant variable, cytotoxic activity was used taking into account that for this study, it was considered cytotoxic if it complied with the

premise of showing a percentage of cytotoxic activity greater than 50% at 20 µg / mL concentration [43], for two of the studied cell tumor lines.

Pattern matching method used was based on the default distance measure "Pearson r" (the Pearson correlation coefficient), which calculates the strength of linear dependence between the two variables [42].This analyses were run in MetaboAnalyst (http://www.metaboanalyst.ca) [27,28].

4.4.6 Analysis of the metabolic pathway

The prediction of the metabolic pathway and the putative identification of the two metabolites (M289T514 and M307T620 features), was made by using KEGG Kyoto database [34]; which is included in Galaxy web server [44]. Additional the putative identification was supported by incorporating mass tamdem spectral information in MetFrag web server [45].

4.5 CONCLUSIONS

From this research, it was possible to establish a structure/activity correlation of the soft corals extract of the *plexaura* and *plexaurella* genus from the plexauridae family and to determine that the greatest activity presented by these marine organisms is due to the fact that the compounds found such as VIPs (plexaurolone, asperdiol and dolabelatriene) correlate with each other as observed in the MS/MS data analysis using GNPS platform and CytoScape software and additionally the related compounds directly to these features (VIPs), correspond to water losses and insaturations formation.

Additional Information and Declarations

Competing Interests

Author Contributions: L.S., O.P.T., and E.T. conceived and designed the experiments; L.S., M.P. and E.T. collected the biological material; M.P. identified the soft corals, L.S. and E.T. performed the chemical experiments; L.S. and E.T. analyzed the data; L.S. and E.T. wrote the article; L.S., M.P., O.P.T. and E.T. edited and reviewed the article. All the authors have contributed to and approved the final manuscript.

Funding: Doctoral studies of L.S. were supported by grants by Colciencias Beca de Doctorado Nacional 647/2014 and Universidad de la Sabana. Universidad de la Sabana supported the Project ING-177-2016 “Búsqueda de Compuestos Bioactivos. Fase II: Análogos sintéticos de diterpenos con actividad citotóxica”. Part of this project (Grant-Aid Agreement No.PBA/MB/16/01) was carried out with the support of the Marine Institute and was funded under the Marine Research Programme supported the Irish Government. Sergio Zapata Lopera from UTADEO, provided great help in obtaining, processing and measuring soft coral sclerites for soft coral identification purposes.The collection of the softcorals used in this research were covered under the permission granted by the Ministerio de Ambiente y Desarrollo Sostenible de Colombia to Universidad de La Sabana, through the Contrato Marco de Acceso a Recursos Genéticos y sus Productos Derivados No. 117 de 2015 (Otrosí No. 3).

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgements
The authors acknowledge the Universidad de La Sabana for support towards the cytotoxic bioassays and Galway University towards NMR and LC-MS equipment. We also thank Invertebrate collection of Instituto de Ciencias Naturales at Universidad Nacional de Colombia (Bogotá, Colombia) for curation of the octocoral samples and coded assigned.

4.6 REFERENCES

1. Dyshlovoy, S.; Honecker, F. Marine Compounds and Cancer: 2017 Updates. *Mar. Drugs* **2018**, *16*, 41.
2. Al-Lihaibi, S.S.; Alarif, W.M.; Abdel-Lateff, A.; Ayyad, S.E.N.; Abdel-Naim, A.B.; El-Senduny, F.F.; Badria, F.A. Three new cembranoid-type diterpenes from Red Sea soft coral Sarcophyton glaucum: Isolation and antiproliferative activity against HepG2 cells. *Eur. J. Med. Chem.* **2014**, *81*, 314–322.
3. Wang, L.H.; Sheu, J.H.; Kao, S.Y.; Su, J.H.; Chen, Y.H.; Chen, Y.H.; Su, Y. Di; Chang, Y.C.; Fang, L.S.; Wang, W.H.; et al. Natural Product Chemistry of Gorgonian Corals of the Family Plexauridae Distributed in the Indo-Pacific Ocean. *Mar. Drugs* **2012**, *10*, 2415–2434.
4. Berrue, F.; Kerr, R.G. Diterpenes from gorgonian corals. *Nat. Prod. Rep.* **2009**, *26*, 681–710.
5. Rodríguez, A.D.; González, E.; González, C. Additional dolabellane diterpenes from the caribbean gorgonian octocoral *Eunicea* laciniata. *J. Nat. Prod.* **1995**, *58*, 226–232.
6. Nieto, M.I.; González, N.; Rodríguez, J.; Kerr, R.G.; Jiménez, C. New cytotoxic cembranolides: isolation, biogenetic studies, and synthesis of analogues. *Tetrahedron* **2006**, *62*, 11747–11754.
7. Clish, C.B. Metabolomics: an emerging but powerful tool for precision medicine. *Mol. Case Stud.* **2015**, *1*, a000588.
8. Nalbantoglu, S. Metabolomics: Basic Principles and Strategies. *Mol. Med.* **2019**, 1–15.
9. Reverter, M.; Tribalat, M.A.; Pérez, T.; Thomas, O.P. Metabolome variability for two Mediterranean sponge species of the genus Haliclona: specificity, time, and space. *Metabolomics* **2018**, *14*, 0.
10. Farag, M.A.; Porzel, A.; Al-Hammady, M.A.; Hegazy, M.E.F.; Meyer, A.; Mohamed, T.A.; Westphal, H.; Wessjohann, L.A. Soft Corals Biodiversity in the Egyptian Red Sea: A Comparative MS and NMR Metabolomics Approach of Wild and Aquarium Grown Species. *J. Proteome Res.* **2016**, *15*, 1274–1287.
11. Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788.
12. Suarez, M.; Rodríguez, M.A.; Vinaixa, M.; Arola, L.; Beltran, A.; Correig, X.; Samino, S.; Yanes, O. Assessment of Compatibility between Extraction Methods for NMR- and LC/MS-Based Metabolomics. *Anal. Chem.* **2012**, *84*, 5838–5844.
13. Bingol, K.; Bruschweiler-Li, L.; Li, D.; Zhang, B.; Xie, M.; Brüschweiler, R. Emerging new strategies for successful metabolite identification in metabolomics. *Bioanalysis* **2016**, *8*, 557–573.
14. Nyamundanda, G.; Brennan, L.; Gormley, I.C. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* **2010**, *11*, 571.

15. Szymanska, E.; Saccenti, E.; Smilde, A.K.; Westerhuis, J.A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* 2012, *8*, 14.

16. Tistaert, C.; Chataigné, G.; Dejaegher, B.; Rivière, C.; Hoai, N.N.; Van, M.C.; Quetin-leclercq, J.; Heyden, Y. Vander Multivariate data analysis to evaluate the fingerprint peaks responsible for the cytotoxic activity of Mallotus species. *J. Chromatogr. B* 2012, *910*, 103–113.

17. Wang, M.; Lamers, R.-J.A.N.; Korthout, H.A.A.J.; van Nesselrooij, J.H.J.; Witkamp, R.F.; van der Heijden, R.; Voshol, P.J.; Havekes, L.M.; Verpoorte, R.; van der Greef, J. Metabolomics in the context of systems biology: bridging traditional Chinese medicine and molecular pharmacology. *Phytother. Res.* 2005, *19*, 173–82.

18. Gao, D.; Wang, Y.; Xie, W.; Yang, T.; Jiang, Y.; Guo, Y.; Guan, J.; Liu, H. Metabolomics study on the antitumor effect of marine natural compound flexibilide in HCT-116 colon cancer cell line. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 2016, *1014*, 17–23.

19. Santacruz, L.; Thomas, O.P.; Duque, C.; Puyana, M.; Tello, E. Comparative analyses of metabolomic fingerprints and cytotoxic activities of soft corals from the Colombian Caribbean. *Mar. Drugs* 2019, *17*, 1–14.

20. Wang, M..; Carver, J.J..; Phelan, V.V..; Sanchez, L.M..; Garg, N..; Peng, Y..; Nguyen, D.D..; Watrous, J..; Kapono, C.A..; Luzzatto-Knaan, T..; et al. Sharing and community curation of mass spectrometry data with GNPS. *Nat. Biotechnol.* 2017, *34*, 828–837.

21. Quinn, R.A.; Nothias, L.F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P.C. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol. Sci.* 2017, *38*, 143–154.

22. Sedio, B.E.; Boya P., C.A.; Rojas Echeverri, J.C. A protocol for high-throughput, untargeted forest community metabolomics using mass spectrometry molecular networks. *Appl. Plant Sci.* 2018, *6*, 1–13.

23. Bryan, K.; Brennan, L.; Cunningham, P. MetaFIND: A feature analysis tool for metabolomics data. *BMC Bioinformatics* 2008, *9*, 1–13.

24. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B.S.; Yang, J.Y.; Kersten, R.D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J.M.; et al. Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci.* 2012, *109*, E1743–E1752.

25. Mohimani, H.; Pevzner, P.A. Dereplication, sequencing and identification of peptidic natural products: From genome mining to peptidogenomics to spectral networks. *Nat. Prod. Rep.* 2016, *33*, 73–86.

26. Kind, T.; Liu, K.-H.; Lee, D.Y.; DeFelice, B.; Meissen, J.K.; Fiehn, O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* 2013, *10*, 755–8.

27. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D.S.; Xia, J. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 2018, *46*, W486–W494.

28. Xia, J.; Sinelnikov, I.; Han, B.; Wishart, D. MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res.* 2015, *43*, 251–257.

29. Wang, M.; Carver, J.J.; Phelan, V. V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 2016, *34*, 828–837.

30. Chin, W.W. The partial least squares approach to structural equation modeling 1998, 334.

31. Geisser, S. Sample Reuse Method The Predictive with Applications. *J. Am. Stat. Assoc.* 1975, *70*, 320–328.

32. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B* 1974, *36*, 111–147.

33. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariat Data Analysis*; 2014; ISBN 9781292021904.

34. KEGG; GenomeNet; Laboratories, K. KEGG COMPOUND Database Available online: https://www.genome.jp/kegg/compound/.

35. Williamson, E.M. Synergy and other interactions in phytomedicines. *Phytomedicine* 2001, *8*, 401–409.

36. Gilbert, B.; Alves, L. Synergy in Plant Medicines. *Curr. Med. Chem.* 2005, *10*, 13–20.

37. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 2007, *8*, 105.

38. Turcu Francisca E. Reyes, Ventii Karen H., W.K.D. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Annu Rev Biochem* 2010, 363–397.

39. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.; Wang, J.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, *13*, 2498–504.

40. Salkind, N. Principal Components Analysis. *Encycl. Res. Des.* 2012, *1*, 1–6.

41. Das, M.K.; Arya, R.; Debnath, S.; Debnath, R.; Lodh, A.; Bishwal, S.C.; Das, A.; Nanda, R.K. Global Urine Metabolomics in Patients Treated with First-Line Tuberculosis Drugs and Identification of a Novel Metabolite of Ethambutol. *Antimicrob. Agents Chemother.* 2016, *60*, 2257–2264.

42. Xia, J.; Wishart, D.S. Using metaboanalyst 3.0 for comprehensive metabolomics data analysis. *Curr. Protoc. Bioinforma.* 2016, *2016*, 14.10.1-14.10.91.

43. Chakravarti, S.K.; Klopman, G. A structural analysis of the differential cytotoxicity of chemicals in the NCI-60 cancer cell lines. *Bioorganic Med. Chem.* 2008, *16*, 4052–4063.

44. NHGRI; NSF; The Huck Institutes of the Life Sciences, T.I. for C. at P.S.; University, E. Workflow4metabolomics Available online: https://galaxy.workflow4metabolomics.org/.

45. DENBI; Solutions, E. MetFrag Available online: https://msbi.ipb-halle.de/MetFrag.

4.7 SUPPLEMENTARY INFORMATION

Name assigned by collection area	Name of the species	Accession	Código ICN-UN
Cali 12	*Plexaura cf. nina*	C12m	ICN-MHN (Po)-CO-279
Cali 2	*Plexaura kukenthali*	C2P	ICN-MHN (Po)-CO-271
Cali 24	*Plexaurella nutans*	C24p	ICN-MHN (Po)-CO-281
Gra 1	*Plexaurella fusifera*	G1P	ICN-MHN (Po)-CO-282
Gra 18	*Plexaura kukenthali*	G18P	ICN-MHN (Po)-CO-292
Gra 21	*Plexaura homomalla*	G21P	ICN-MHN (Po)-CO-283
Gra 2	*Plexaura homomalla*	G2P	ICN-MHN (Po)-CO-283
Gra 22	*Plexaurella* sp.	G22P	ICN-MHN (Po)-CO-291
Gra 24	*Plexaura homomalla*	G24P	ICN-MHN (Po)-CO-283
Gra 6	*Plexaura kukenthali*	G6Pk	ICN-MHN (Po)-CO-271
Gra 9	*Plexaura sp.*	G9P	ICN-MHN (Po)-CO-286

Table S1. Name of species of soft corals used in this research a Collection Code (ICN), assigned by Collection of the Institute of Natural Sciences of the National University of Colombia. (Bogotá, Colombia)

Obs	name	Siha (%)	Pc3 (%)	A549 (%)	L929 (%)
1	C12m	23.1	16.0	43.3	45.2
2	C2P	46.2	7.8	62.4	16.2
3	C24p	13.1	65.1	39.0	13.3
4	G1P	39.4	23.8	33.0	11.5
5	G18P	44.0	64.0	52.5	31.2
6	G21P	25.5	20.9	34.0	18.4
7	G2P	34.8	47.6	21.1	14.3
8	G22P	37.6	63.5	31.9	20.7
10	G24P	37.6	33.5	31.9	20.7

11	G6Pk	57.0	40.3	38.5	23
12	G9P	34.5	20.4	31.0	49.8
29	Doxorubicin (25ppm)	60.1	46.9	58.2	16.0

Table S2. Percentage of cytotoxic activity against three cancerous cell lines and fibroblasts L929 (ATCC®CCL-1 ™) which was used as non-tumor cell line for toxicity control.

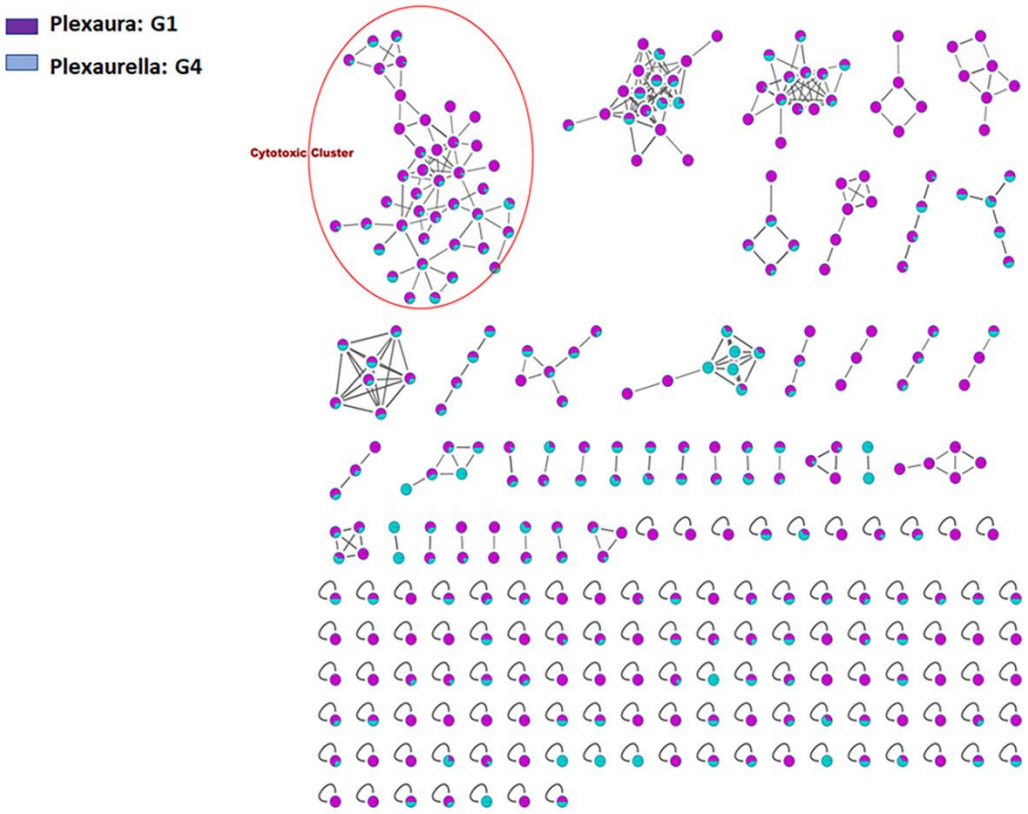


Figure S1. Massive Molecular Network from soft corals of the genera: *plexaura* and plexaurella, the colors indicate producers of the nodes*: plexaura* Red and *plexaurella* green. Nodes found in more than one producer are represented as the combination of their colors (Wang et al., 2017).

5 GENERAL DISCUSSION

Compounds from natural sources (plants, animals and microorganisms), are defined as natural products (Jha & Zi-rong, 2004) and marine organisms are sources of many unique low molecular weight compounds (metabolites), with cytotoxic properties that have been studied for their potential as medicines (Jha & Zi-rong, 2004; Hong-Fang Ji, 2009).

Due to major and recent advances in the technologies as a Nuclear Magnetic Resonance and mass spectrometry used to extract and analyze this type of data the metabolomic studies have increased (Alonso, Marsal & Julià, 2015). To date, metabolomics is envisaged as one of the major "omics" tools (Hubert, Nuzillard & Renault, 2017) and the advantages of using metabolomics with these high performance techniques, in challenging research like the determination of correlations between chemical composition and biological activities compared to the other omics, such as transcriptomics and proteomics it is that the metabolomics, provides a representation of the cellular metabolite state of an organism at the time of sampling and shows the influence of any perturbation induced by the environment, altered genes, or disease (Weckwerth, 2003).

To our knowledge, there are few metabolite studies and their correlation with biological activity on soft corals using Nuclear Magnetic Resonance (NMR) spectroscopy and hyphenated techniques like liquid chromatography coupled by mass spectrometry approach; most of the studies relate to the information on coral physiology, particularly as the chemotype that corelate with genotype and environment where the metabolomics can provide a snapshot of an organism's physiological state (Sogin et al., 2014; Farag et al., 2016c; Januar et al., 2017; Lohr et al., 2019); however, in the few studies that involve the search for bioactive substances, from marine organisms such as soft corals have been found metabolites with significant cytotoxic activities against different cancer cell lines (Patel et al., 2009) an example of this is a study conducted by Gao and coworkers who showed tha metabolomics approach provides new insights in studies of flexibilide compound isolated from the soft coral Sinularia flexibilis which has presented antitumor activity agains HCT116 (ATCC® CCL-247™) colorectal carcinoma cell line (Gao et al., 2016). In another study, the cytotoxic effects of soft Coral, sarcophyton sp. was evaluated using metabolomics to establish a correlation between the cytotoxicity and chemotype of these octocorals using ^{1}H NMR concluding that the result observed in the study is not clear due possibly to the to an overlap of the signals produced by diterpenes when analyzed by this technique (Farag et al., 2017).

In addition, in metabolomic studies to search compounds of marine origin with cytotoxic activities, The importance of statistic is evident due to the most of these studies use of multivariate models to report their main findings, taking into account that the reduction in the dimensionality of the data allows researchers to visualize patterns, biomarkers, outliers, and correlations (Liland, 2011). The use the multivariate models require intensive validation work to avoid overfitting the data (Malinsky, Simpson & Durbin, 2016), e.g. In the supervised multivariate modelling "partial least squares analysis (PLS)" which is often used to cope with complex data and assess the importance of variables, besides facilitating selection of relevant variables into biologically meaningful interpretations (Vinaixa et al., 2012). It requires double cross validation procedures or permutation tests to ensure the reliability of the results (Szymanska et al., 2012). That is how in this research, by UPLC/MS the statistical analysis OPLS-DA, showed discrimination between the extracts that presented greater potential cytotoxic against the extracts with lower cytotoxic potential and was found that the compound identified as 13-keto-1,11-dolabell-3(*E*),7(*E*),12(18)- he was the responsible for the cytotoxic activity with Q2 values of 0.43 and 0.34 against the tumor cell lines of SiHa and A549 respectively. On the other hand, by NMR technique, the validation of OPLS-DA using PSYCHE experiments allowed to obtain a reliable correlation between the chemical composition and the cytotoxic activity of the soft corals extracts against PC3 tumor cell line where the putatively identified compounds as asperdiol and dehydroxyplexaurolone, were the responsible for the activity evidenced against this tumor line, this is because the PSYCHE technique is a new method for Pure Shift NMR Spectroscopy which suppresses the effects of homonuclear coupling, allowing ^{1}H spectra to be produced that contain chemical shift only, with no multiplet structure which makes it possible to separate and visualize the features (chemical shifts) responsible for cytotoxic activity (Foroozandeh, Morris & Nilsson, 2018).Our study reported Q2 values of 0.66 against the PC3 cell tumor line indicating that with this technique the search for bioactive compounds in some soft corals of the Colombian Caribbean is reproducible and reliable.

Therefore, the statistic in metabolomics studies, contributes to the construction of parsimonious models, meaning simple models with great and robust explanatory predictive power (Saeys, Inza & Larrañaga, 2007), which can be helpful in several studies, mainly those that are time-consuming as traditional bioprospecting of marine natural products. The metabolomics process is helping in the rapid identification of compounds with important cytotoxic activities

On the other hand, the organization of MS / MS data based on chemical similarity using the approach of molecular networking, is a powerful complement to the visualization of the results of the multivarent analyzes observed (Yang et al., 2013a). Molecular networking not only allows to observe the relation of compounds structurally, it also allows to visualize the formation of clusters of families of compounds which is important to establish structure / activity correlations, representing a challenge for many other identifications and corralation strategies (Philippus et al., 2018).
As was evidenced in chapter 5 where it was determined that the activity/structure correlation was given by the relationship between the VIPs (plexaurolone, dolabellatriene and asperdiol) found in the *plexauridae* family.

6 OVERALL CONCLUSION AND PERSPECTIVES

The metabolites from marine organisms are a promising source of unexploited drugs that have shown a variety of biological activities among them the antitumor activities due their compounds have a wide structural diversity. These compounds are produced to the harsh and competitive conditions in marine environment as a defense against some predators.

Metabolomics studies have become a valuable scientific discipline, helping to identify biomarkers in multiple studies with a wide range of research fields, such as the identification of compounds from marine organisms with promising biological as the antitumor activities. Its expansion is due to a significant increase in the number of computational tools available to process and analyze metabolomics data. These tools help to create metabolomics workflows that include data processes, use of univariate and multivariate statistical analyzes, annotation of metabolites and use of molecular networks to infer different hypotheses, such as the one proposed in this research on the possible correlation between chemical composition of the soft corollas of the Colombian Caribbean and its cytotoxic activity against the cancer cells lines of lung (A549), cervix (SiHa) and prostate (PC3).

The use of the UPLC/MS analytical tool together with the statistical analysis of data using PCA, PLS-DA and OPLS-DA and the process of dereplication, allowed to identify the compound dolabellatrienone (13-keto-1,11-dolabell-3 (E),7(E),12(18)-triene) as the VIP responsible for the cytotoxic activity when the extracts of soft corals were exposed against the cell lines of A549 and SiHa.

On the other hand, the PSYCHE experiment, allowed to correlate the chemical composition of extracts from soft coral with their cytotoxic activity against the cell line PC3, which was not observed when using the data matrix obtained from the ^{1}H NMR experiment. Due to that the psyche experiments allow to solve the overlapping of ^{1}H-^{1}H scalar coupling multiple yielding an adequate matrix for reliable statistical and chemical shift analyzes. In addition, this comparative analysis, allowed to found the putative identification of features responsible for the separation of the most cytotoxic groups.

Based on the information seeped from previous investigations (relationship between chemical composition and cytotoxic activity using UPLC/MS and NMR studies) it was possible to select softcoral extracts from the plexauridae family as the most cytotoxic against the SiHa tumor lines, PC3

and A549. Therefore, based on a metabolomic workflow that incorporates molecular networking (MN), in silico MS/MS databases, and manual dereplication it was possible to correlate the features responsible for the separation of the group of the extracts belonging to eleven most bioactive from *Plexaura* and *Plexaurella* sp. Genus. In conclusion, here, we present the importance of developing workflows that allow the prospective application of the fingerprint metabolomic analysis in soft corals, aimed to identify metabolites with cytotoxic activity against cancer cell lines.

The workflows shown here, allowed us to find a dolabelane-type compound as responsible for the cytotoxic activity against the tumor lines of SiHa and A549 when UPLC/MS analytical technique was used; while in metabolomics workflow using NMR technique, the putative identification of two diterpenoids as responsible for the activity presented by the extracts against the PC3 tumor line was possible. Additionally, the concept of "cytotoxic molecular networking" to find candidate active molecules directly from bioactive soft coral extracts was introduced, by establishing the correlations that allowed the putative identification of the compounds that were part of the cytotoxic cluster.

Our research has revealed the potential and importance of metabolomic studies in extracts from marine organisms as a source of citotoxic agents agains A549, SiHa and PC3 tumor cell lines. Future investigations could evaluate these same extracts against other tumor lines, which would allow to extend the range of response of the cytotoxic activity that they present.

6.1 REFERENCES

Alonso A, Marsal S, Julià A. 2015. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Frontiers in Bioengineering and Biotechnology* 3:1–20. DOI: 10.3389/fbioe.2015.00023.

Farag MA, Fekry MI, Al-hammady MA, Khalil MN, Wessjohann LA. 2017. Cytotoxic Effects of *Sarcophyton* sp. Soft Corals — Is There a Correlation to Their NMR Fingerprints ? *Marine drugs* 15:1–13. DOI: 10.3390/md15070211.

Farag MA, Porzel A, Al-Hammady MA, Hegazy MEF, Meyer A, Mohamed TA, Westphal H, Wessjohann LA. 2016. Soft Corals Biodiversity in the Egyptian Red Sea: A Comparative MS and NMR Metabolomics Approach of Wild and Aquarium Grown Species. *Journal of Proteome Research* 15:1274–1287. DOI: 10.1021/acs.jproteome.6b00002.

Foroozandeh M, Morris GA, Nilsson M. 2018. PSYCHE Pure Shift NMR Spectroscopy. :13988–14000. DOI: 10.1002/chem.201800524.

Gao D, Wang Y, Xie W, Yang T, Jiang Y, Guo Y, Guan J, Liu H. 2016. Metabolomics study on the

7 ACKNOWLEDGMENTS

The collection of the softcorals used in this research were covered under the permission granted by the Ministerio de Ambiente y Desarrollo Sostenible de Colombia to Universidad de La Sabana, through the Contrato Marco de Acceso a Recursos Genéticos y sus Productos Derivados No. 117 de 2015 (Otrosí No. 3).