Giant barrel sponges in diverse habitats: a story about the metabolome

Bayona Maldonado, L.M.

Giant barrel sponges in diverse habitats: a story about the metabolome

Lina María Bayona Maldonado

Lina María Bayona Maldonado. 2021. Giant barrel sponges in diverse habitats: a story about the metabolome. Ph.D. Thesis at Leiden University, The Netherlands 2021.

ISBN: 978-94-92597-70-0

Design of covert: Sergio Bayona

Printing and biding: Printsupport4u

This Ph.D. research was made possible with the financial support of:

Science, Technology and Innovation Ministry, Colombian government.
The Dutch Research Council (NWO).
Natural products laboratory, Leiden University.

Giant barrel sponges in diverse habitats: a story about the metabolome

Proefschrift

Ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 22 april 2021

klokke 16:15 uur

door

Lina María Bayona Maldonado

geboren te Bogotá, Colombia

in 1990

Promotoren

Prof. Dr. Peter G. L. Klinkhamer
Leiden University

Prof. Dr. Nicole de Voogd
Naturalis Biodiversity Center & Leiden University

Co-promotor

Dr. Young Hae Choi
Leiden University

Promotiecommissie:

Prof. Dr. Gilles van Wezel
Leiden University

Prof. Dr. Remko Offringa
Leiden University

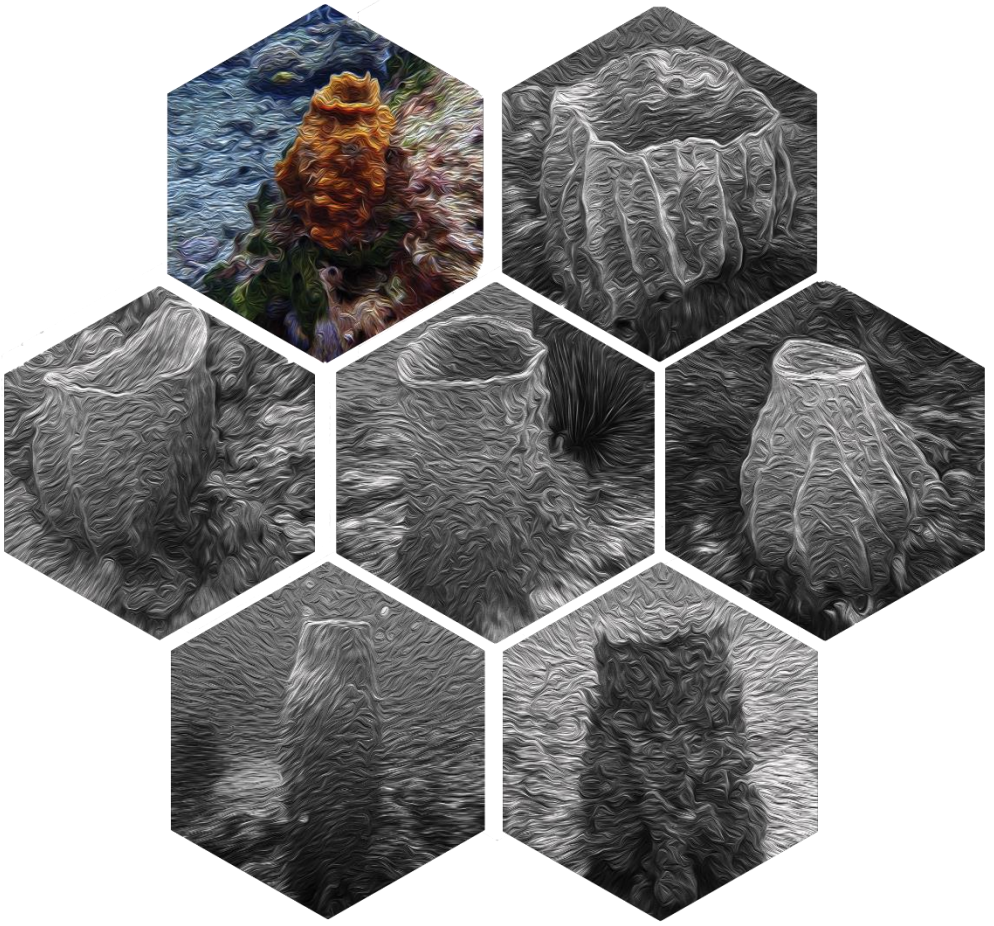
Prof. Dr. Nathaniel Martin
Leiden University

Prof. Dr. Shirley A. Pomponi
Florida Atlantic University

Dr. David Overy
Ottawa Research and Development Centre

To Laura

Content



Chapter 1

General Introduction

1. Marine sponges

Oceans possess very biodiverse and characteristic ecosystems that have remained unexplored in their vast majority. However, the rapid progress of technology in recent years has allowed more and more light to be shone on these mysterious ecosystems, revealing their exciting potential. The uniqueness of marine environments, compared with terrestrial ones, is their diversity, as 32 out of 34 of the known phyla on the planet inhabit the oceans, many more than the 17 phyla found in terrestrial environments (Snelgrove 2016). This biodiversity is not limited to the number of species, which is itself impressive at more than an estimated 200,000,000 species, but more importantly, is greatly increased by the interactions among the organisms in the ecosystems (Horton et al. 2020; Mora et al. 2011; Palumbi et al. 2009). The stocks of nutrients in the oceans are very limited and highly localized. Therefore, marine organisms have necessarily needed to develop an efficient way to access and use these resources. Ecosystems, such as coral reefs, mangroves, estuaries, and seagrass beds are examples of the tight relationships between many organisms evolved over time as successful survival strategies (de Goeij et al. 2013). The fact that many of these relations are essentially mediated through chemical compounds, has led to the highly developed chemical diversity observed in marine organisms (Naman et al. 2017). This has boosted their use as a prolific source of chemicals with unique features (Blunt et al. 2018; Carroll et al. 2019, 2020; Gerwick and Moore 2012).

Among the many marine organisms, sponges are one of the most abundant in many ecosystems, both in number and biomass.(Bell et al. 2017; Bell and Carballo 2008; Bell and Smith 2004). Despite this, their ecological roles have not yet been as extensively studied as that of other marine animals like corals. For example, it has been reported that sponges influence the quality and quantity of substrate in coral reefs and other ecosystems (Diaz and Rützler 2001). Some sponge species, known as boring sponges, actively participate in the bioerosion process which transforms the calcareous substrate into smaller particles and while this is a natural process needed in the reef, in some cases it could affect the condition and health of corals as sponges erode their skeletons (Wulff 2006, 2016). On the other hand, the presence of some sponges in the substrate has proved to increase coral survival, especially after storms, by stabilizing the substrate and binding coral fragments (Bell 2008; Swierts, et al. 2018). In addition, sponges are believed to play a vital role in bento-pelagic coupling, as they actively participate in carbon, silicon, and nitrogen cycling, the latter with help of nitrifying bacteria symbionts (Maldonado et al. 2012; Zhang et al. 2019). Due to the strong interaction of sponges with the water column resulting from their massive water pumping, they act as a

link between organisms in the water column and higher trophic levels. The relationship of sponges with other organisms can vary in very diverse ways according to the circumstances, from associations with microorganisms as an additional source of energy or chemical defense, to predation and spatial competition with other animals in the reef (Bell 2008; Webster and Taylor 2012; Wulff 2006).

Sponges are considered to be among the most ancient species. Their fossils date back to 535 million years ago that locates their first generation in the early Cambrian (Antcliffe et al. 2014; Li et al. 1998). Moreover, the species diversification of sponges has been traced to the Cambrian period. At an anatomical level, sponges are the most primitive animals, they are filter-feeding animals that use specialized flagellated cells to actively pump water through their bodies in order to absorb the nutrients from the water column (Bergquist 1978). An external layer of flattened cells (pinacocytes) called pinacoderm separates the sponge from the surrounding water. The pinacoderm has pores (ostia) that lead to chambers where flagellated cells (choanocytes) form the choanoderm. These cells are responsible for pumping the water through the sponge. The skeleton of sponges is located in the mesohyl, a layer of connective tissue between the pinacoderm and the choanoderm. This skeleton can be organic (spongin fibers), inorganic (silica or calcium carbonate) or a mix of both (Bergquist 1978). Furthermore, in the mesohyl, a specific type of cells (archeaocytes) can digest dissolved nutrients and microorganism filtered from the water column (Bergquist 1978; Van Soest et al. 2012). However, not all the microorganisms are necessarily digested, as some of them can be incorporated into the mesohyl as symbionts (Hentschel et al. 2006; Webster and Taylor 2012). The amount of symbionts such as bacteria, archaea, and fungi found in sponges can be used to classify sponges in two functionally different groups (Gloeckner et al. 2014). Sponges with bacterial counts of 10^8- 10^{10}, that is 2 to 4 times the order of magnitude than that of the microbial count of the surrounding water, are called high microbial abundance (HMA) sponges. On the other hand, those with bacterial counts of around 10^4-10^6, i.e., similar to those found in the surrounding water, are classified as low microbial abundance (LMA) sponges. These two groups differ also in their filtering capacity.

Aside from these filter -feeding mechanisms, further generalizations on sponges are remarkably complicated due to their great diversity (Bergquist 1978). In the oceans, they are found in many habitats including mangroves, coral reefs, polar regions and deep oceans (Van Soest et al. 2012). A classification based on morphology is also difficult because of the great plasticity they exhibit when adapting to local or environmental conditions (De Vos et al. 1991). So far sponges have been grouped into four classes. The Demospongiae class accounts for

above 80% of the living sponge species, and are characterized by a skeleton formed of spongin fibers sometimes complemented with siliceous structures; Hexactinellida, have a siliceous skeleton and are also known as glass sponges; Calcarea, known as calcareous sponges present calcium carbonate spicules in their skeleton; Homoscleromorpha, which was taken out of the Demospongiae class, and is the most recently accepted class (Bergquist 1978; Gazave et al. 2012).

2. Chemical components of sponges: a promising source of bioactive natural products

Sponges have resulted of great interest as a source of bioactive natural compounds from the very beginning of marine organism research (Ebada and Proksch 2012). To date, over 10,000 new metabolites have been reported to be isolated from sponges and in the past decade (2000 – 2010), compounds isolated from marine sponges in particular account for approximately 30% of all compounds isolated from marine organisms (Figure 1.1a). Although this percentage has been decreasing as the number of marine microorganism studies increased, sponges still maintain their status as an abundant source of new chemicals, contributing an average of 200 new compounds every year (Figure 1.1b) (Blunt et al. 2017; Carroll et al. 2019, 2020). Among the sponge phyla, Porifera is considered to be one of the most prolific in terms of the chemical diversity of their metabolites (Paul et al. 2019). From a chemical perspective, sponge metabolites display a wide diversity, revealed not simply by the existence of structural analogues but by the immense variety of chemical structures, which includes terpenoids, alkaloids, peptides, polyketides, steroids, quinones, and fatty acids (Figure 1.1c). Among the chemicals isolated from sponges, terpenoids and alkaloids are undoubtedly the predominant secondary metabolites groups. These two families corresponded to over 60% of all the metabolites isolated from sponges between 2014 and 2018 (Blunt et al. 2016, 2017, 2018; Carroll et al. 2019, 2020).

Sponge metabolites have been reported to display interesting responses to a broad range of bioassays for cytotoxicity, antiviral, antifungal, anti-inflammatory, and antifouling activities among others (Belarbi et al. 2003; Keyzers and Davies-Coleman 2005; Proksch et al. 2010; Sipkema et al. 2005). The presence of these metabolites with strong and diverse activities might be the consequence of the numerous ecological interactions that sponges experience in their natural environment (Perdicaris et al. 2013). Cytotoxic compounds, for example could work to keep predators at bay or as a way to increase their competitiveness in the struggle for space with other sessile organisms (Ye et al. 2015). Interestingly, more than 800 antibiotics together with several antiviral compounds have been reported from sponges, probably having a role in the defense system against pathogens present in the water (Laport et al. 2009; Sagar

et al. 2010). In addition, the production of compounds that inhibit the formation of biofilms to prevent fouling can be attributed to an important defense mechanism of sponges to maintain their capacity to filter water (Stowe et al. 2011).

Figure 1.1: A: Percentage of compounds isolated from marine sponges over the total reported from marine organisms between 2001 and 2018. B: Number of new compounds reported from marine sponges between 2001 and 2018. C: Occurrence of secondary metabolites groups isolated from marine sponges between 2014 and 2018.

Marine natural products (MNP) research locations have been concentrated mainly along the continental coastlines due to the technical challenges represented by deep sea exploration. The discovery of MNP's from sponges in particular, is spread all over the world in a remarkable contrast with other marine phyla, due to their widespread distribution (Carroll et al. 2020). For example, the phylum Porifera is the most prolific source of MNP in all climate zones (tropical, subtropical, temperate, and polar), accounting for between 25 and 35% of all compounds reported in each climate zone (Carroll et al. 2020). Sponges have adapted to a large range of environmental conditions all over the world having developed a large arsenal of secondary metabolites as a result of this adaptation (Becerro et al. 2003). The combination of these circumstances, that is, their wide distribution in such distinct environmental conditions could provide a plausible explanation for the enormous variety of sponge MNPs.

Despite the potential of sponges as a great source of bioactive compounds, the number of drugs sourced from them that have been approved for commercialization is very low. So far,

only three sponge-derived drugs have been approved by the Food and Drug Administration (FDA) (Gerwick and Moore 2012; Newman and Cragg 2016). The low success rate in the output of drugs is mainly due to the lack of a stable supply of sources required for the development of a new drug. Several solutions have been proposed to overcome this problem, such as aquaculture, mariculture, cultivation of the microsymbionts, and chemical synthesis (Belarbi et al. 2003; Gomes et al. 2016; Sipkema, Osinga, et al. 2005). Moreover, understanding the driving factors involved in the production of active compounds remains too unclear to contribute to the application of some of these solutions.

3. Giant barrel sponges: conspicuous members of the reef

Giant barrel sponges are sponges belonging to the genus *Xestospongia* (order Haplosclerida: class Demospongiae) (Horton et al. 2020; Van Soest et al. 2012) and three extant species, *Xestospongia bergquistia, Xestospongia muta*, and *Xestospongia testudinaria*. Sponges from any of these species have a characteristic barrel-like shape and are large in size, exceeding 1 m in height (Van Soest et al. 2012). *Xestospongia bergquistia* is endemic to Australian reefs, a region in which *X. testudinaria* is also present. These two species can be distinguished by structural differences in their skeleton (Fromont 1991). On the other hand, morphological differences between *X. muta* and *X. testudinaria* are very subtle but are very subtle and it seems that the main difference between these two species comes from their geographical location (Setiawan et al. 2016). While *X. muta* is present in the Caribbean Sea from Florida down to South American coasts, *X. testudinaria* is present in the Indo-Pacific region from the Red Sea, the east coast of Africa up to China and New Caledonia (Van Soest et al. 2020a,b). Moreover, recent studies have shown that the previous classification in two species (*X. muta* and *X. testudinaria*) is an oversimplification of a much more complex situation species-wise, as they include several genetic groups that can be differentiated through certain mitochondrial and nuclear DNA markers (Swierts et al. 2013, 2017). Interestingly, when using this new genetic classification, some groups were found to be genetically more closely related to groups on the other side of the world than to those co-existing in the same location (Swierts et al. 2017).

Giant barrel sponges are present in a wide range of environments. As mentioned, their geographical location is widespread, being found in the Caribbean Sea and the Indo-Pacific region. This is very unusual for sponges because due to the limited traveling capacity of their larvae, many sponge species are usually endemic to specific regions (Van Soest et al. 2012). Furthermore, giant barrel sponges inhabit a very wide depth gradient, from 10 to beyond 100 m depth (Morrow et al. 2016; Olson and Gao 2013). This broad distribution suggests that these

sponges have the ability to adapt to the very diverse environmental conditions existing throughout such depth gradients, including temperature, light, pressure, nutrients, and predatory stress. Unsurprisingly, according to the previous reports, both *X. muta* and *X. testudinaria*, present large variations in their morphology, and it is possible to find individuals with a very smooth surface, with digitate or lamellate structures and with different shapes and sizes (Figure 1.2) (Kerr and Kelly-Borges 1993; Swierts et al. 2013).

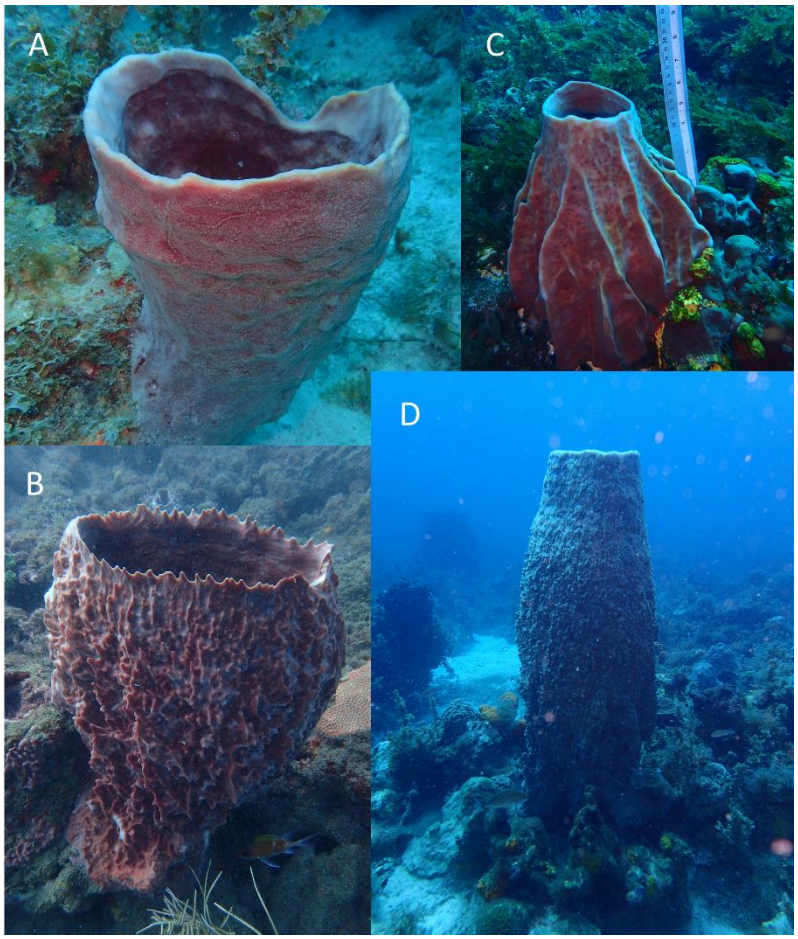


Figure 1.2: Pictures showing different morphologies (A: smooth and B: Digitate) and sizes (C and D) of giant barrel sponges (Photos by Nicole J. de Voogd and Esther van der Ent).

Besides environmental conditions, variations among sponges could also occur with age. *Xestospongia muta* are known as the redwood of the reef, due to their longevity and size (McMurray et al. 2008, 2010). Using measurements such as height, base diameter, and osculum diameter, McMurry and co-workers (McMurray et al. 2008) were able to create a

model to calculate the age of *X. muta* in the Caribbean. Using this model, a specimen found in Curaçao was determined to be about 2300 years old, indicating that *X. muta* is one of the oldest animals alive (Van Soest et al. 2012). Similar studies conducted in the Indo-Pacific on *X. testudinaria* revealed that they can grow to similar sizes as their Caribbean counterparts in a shorter time-lapse, which means that sponges in the Indo-Pacific of the same size as those in the Caribbean can be estimated to be much younger (McGrath et al. 2018).

Several studies in the Caribbean have shown *X. muta* to be one of the most abundant animals in the coral reef, covering up to 13 % of the substrate (McMurray et al. 2010; Zea 1993). Although studies on *X. testudinaria* are not as extensive as in the Caribbean, its presence has been reported in reefs throughout the Indo-Pacific (McGrath et al. 2018; Swierts et al. 2013, 2017). Besides, due to their bulky size, giant barrel sponges are very important organisms in the bento-pelagic coupling as they can overturn a 30 m deep water column in only a few days (McMurray et al. 2014). Their large size, particularly their exceptional height, results in a different interaction with the water column compared to that of other benthic organisms as the interference with water currents around them can create microenvironments for nearby organisms (Bell 2008). Furthermore, giant barrel sponges can provide shelter for other organisms in the reef, e.g. corals (Hammerman and García-Hernández 2017; Swiert et al. 2018). Among the large variety of organisms that establish a symbiotic relationship with giant barrel sponges, microsymbionts are considered to be the most influential. These sponges, classified as HMA, host millions of microorganisms in their mesohyl (Gloeckner et al. 2014). These microorganisms have been thought to play an important role in the adaptation and survival of the sponges, though their detailed functions are yet little known (Fan et al. 2012; Moitinho-Silva et al. 2017; Papale et al. 2020). Several bacteria phyla, such as proteobacteria, poribacteria, actinobacteria, and cyanobacteria have been reported to be regularly present in *Xestospongia* spp. microbiome (Fiore et al. 2013; Polónia et al. 2017). However, their relative abundance and the presence of other minor phyla are greatly influenced by environmental and biological factors (Montalvo and Hill 2011; Morrow et al. 2016; Swierts et al. 2018; Villegas-Plazas et al. 2019).

Xestospongia sponges have shown a great potential to produce a variety of chemicals. More than 300 new compounds have been reported so far from the genus including alkaloids, terpenoids, steroids, quinones, and fatty acids (Zhou et al. 2010). Along with the chemical diversity, many of the compounds have shown significant bioactivities such as cytotoxicity as well as antiviral, antibacterial, and anti-inflammatory properties. Particularly, giant barrel sponges have been characterized for the production of sterols and brominated fatty acids.

Among the sterols, some include unusual functional groups such as cyclopropane and branched side chains. In the past, the sterol content in *X. muta* was used as a chemotaxonomy marker to establish the presence of three chemotypes that could be related to the presence of cryptic species among individuals of this sponge (Kerr and Kelly-Borges 1993). Additionally, the presence of acetylenic groups and very long fatty acids have been identified as characteristic of these sponges (Zhou et al. 2015).

4. Metabolomics: a way to gain an insight into the overall metabolome of organisms

Metabolomics is the most recently developed 'omics' technology. Its aim it to obtain a comprehensive overview of metabolome present in biological samples (tissues, organs or biofluids) under a given set of conditions (Kim et al. 2010). One of the biggest challenges in metabolomics studies is how to deal with the great chemical diversity exhibited by small molecules in the samples. In practice, all the metabolites have very different physical and chemical characteristics which makes impossible for one single analytical technique to cover all of them (Dunn and Ellis 2005). Therefore, various analytical platforms must be employed in metabolomics studies in order to widen the coverage of the metabolome. Undoubtedly, liquid chromatography hyphenated to a mass spectrometry detector (LC-MS) and nuclear magnetic resonance (NMR) are recognized as the most inclusive analytical platforms in metabolomics studies (Emwas et al. 2019; Goulitquer et al. 2012; Markley et al. 2017). Naturally, each of them has several advantages and some limitations. Therefore, to achieve the goal of metabolomics, i.e., to obtain a comprehensive understanding of the chemical diversity present in an organism, the best option is a combination of both techniques.

The (sponge) metabolome is composed not only of regularly synthesized compounds but also of those that are induced as a response to a wide range of internal or external factors. Thus, the metabolome is not static, but highly susceptible to the change of environmental, biological, or genetic factors (Bundy et al. 2008; Shulaev et al. 2008). These qualitative and/or quantitative alterations, could affect metabolic levels which in turn could induce changes in genomic and proteomic expression or silencing (Reverter et al. 2016; 2018; Saito and Matsuda 2010). In the case of marine organisms, temperature, salinity, light intensity, pressure, and predatory stress are among the many environmental factors that can induce changes in the metabolome, However, the factor which could be by far the most influential, is the interaction with co-existing organisms. For example, as an HMA sponge, *Xestospongia*'s metabolome is composed not only of the metabolites produced by the sponge *per se*, but also by those that result from the interaction of their metabolism with that of associated microorganisms. Thus,

modifications in the microbiome of the sponges might be reflected in alterations in the metabolome of the holobiont.

The study of marine natural products, whether undertaken using a traditional or metabolomics approach, has focused mostly on two fields: the discovery of new compounds with a potential biological activity, or chemical ecology investigations into the role of a certain compound. However, synergistic efforts to study sponge secondary metabolites in general between these two fields have been scarce, and even more so in the case of giant barrel sponges. The holistic overview of the chemical space provided by the metabolomics approach can be used to detect new potential active compounds and at the same time provide indications of the ecological and/or biological drivers that lead to the production of these compounds, seamlessly combining both fields in one collaborative effort.

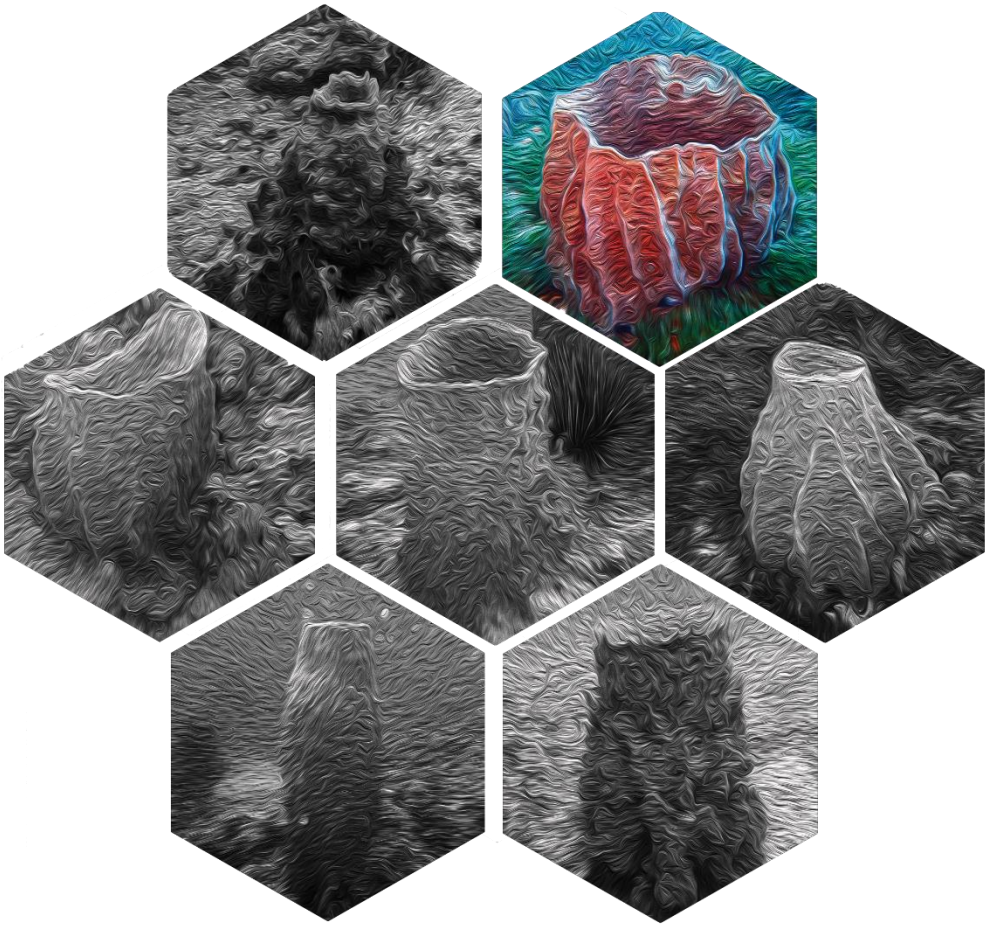
5. Outline

The chemistry of marine sponges has been widely studied over the past 60 years leading to the discovery of a large number of bioactive compounds. However, the driving factors that control the production of these compounds remain mostly unknown. A deeper understanding of the factors that potentially modify the metabolome of sponges would improve their reliability as sources of bioactive compounds. Furthermore, the rapid change experimented by marine ecosystems in the past decades has made the need for this understanding more imperative as the prediction of the extent of the alterations these changes in the environmental conditions have on biological processes has become critical. Metabolomics, with its capacity to provide a comprehensive overview of the metabolome, can contribute to an increased insight into the biological process while assisting in the discovery of new active compounds.

Using multiplatform-based metabolomics analyses, this thesis will focus on the following research questions: 1) How do extraction parameters influence the chemical diversity of extracts for metabolomics studies? 2) Which environmental factors cause variation in the metabolome of giant barrel sponges? 3) Do aging processes cause changes in the metabolome of giant barrel sponges? 4) Can genetic groups of giant barrels sponges also be distinguished by their metabolome?

This thesis aims to unravel the influence of environmental conditions and genetic background on the chemical production of giant barrel sponges using a metabolomics approach. Chapter 2 provides an extensive literature review of metabolomics studies of marine organisms. It includes an overview of every step involved in a metabolomics study from the collection of the

sample until the identification of compounds. Several examples of applications of metabolomics to the study of marine organisms are shown, and future perspectives of the field are discussed. Chapter 3 illustrates the effects of extraction conditions (temperature, pressure, number of cycles, and solvent polarity) on the metabolic diversity of *Xestospongia* extracts using a pressurized extraction system. The objective is to optimize the extraction conditions in order to obtain extracts with a composition that accurately reflects the chemical content of the sample, a most important requirement for metabolomics studies. Chapter 4 shows the differences in the chemical composition of giant barrel sponges collected in four geographical locations, two in the Caribbean Sea (Martinique and Curaçao) and two in the Indo-Pacific region (Tanzania and Taiwan) using LC-MS and NMR based metabolomics. It also explores the relationship between the variation in chemical composition of sponges collected in different locations and their biological activity against *Staphylococcus aureus* and *Escherichia coli*. Chapter 5 presents the changes in the metabolome of giant barrel sponges belonging to two genetic groups, associated with the age of sponges. It also reveals a differential change in the metabolome of these two genetic groups along a depth gradient (from 7 to 42 m), showing that even very genetically close sponges can react differently to changes in environmental conditions. The changes in the metabolome are established using LC-MS based metabolomics together with molecular networking. Chapter 6 describes how different genetic groups of giant barrel sponges can be distinguished not only by their genetic markers, but also by their metabolome. This is exemplified by the results of a NMR- and LCMS-based metabolomics study conducted on sponges belonging to three genetic groups collected in the Spermonde archipelago, SW Sulawesi, Indonesia. In addition, changes in the metabolome due to environmental conditions such as pH and temperature, are reported. Chapter 7 summarizes the main findings of this thesis regarding the chemical production of giant barrel sponges and presents general conclusions and future perspectives about the research in this field.



Chapter 2

Recent progress in marine organism metabolomics

Lina M. Bayona [1], Nicole J. de Voogd [2,3], Young Hae Choi [1,*]

[1] Natural Products Laboratory, Institute of Biology, Leiden University, 2333 BE Leiden, The Netherlands.

[2] Naturalis Biodiversity Center, Marine Biodiversity, 2333 CR Leiden, The Netherlands.

[3] Institute of Environmental Sciences, Leiden University, 2333 CC Leiden, The Netherlands.

[*] Corresponding author e-mail: y.choi@chem.leidenuniv.nl

Manuscript submitted to Metabolomics

Abstract

Marine ecosystems are hosts to a vast array of organisms, being among the most richly biodiverse locations on the planet. The study of these ecosystems is very important, as they are not only a significant source of food for the world but have also, in recent years, become a prolific source of compounds with therapeutic potential. Studies of aspects of marine life have involved diverse fields of marine science, and the use of metabolomics as an experimental approach has increased in recent years. As part of the "omics" technologies, metabolomics has been used to deepen the understanding of interactions between marine organisms and their environment at a metabolic level and to discover new metabolites produced by these organisms.

This review provides an overview of the developments in the field of marine organism metabolomics, and applications of metabolomics common to other fields that could potentially contribute to its use in marine organism studies. It deals with the entire process of a metabolomic study, from sample collection considerations, metabolite extraction, analytical techniques and data analysis. It also includes an overview of recent applications of metabolomics in fields such as marine ecology and drug discovery and future perspectives of its use in the study of marine organisms.

Keywords: Marine organisms, metabolomics, drug discovery, marine chemical ecology.

1. Introduction

Marine ecosystems show remarkable differences with their terrestrial counterparts. The first and most evident is that they develop in an aqueous media, with significant implications on the dynamics of the ecosystem. One of the most influential factors is the distribution of species and chemical nutrients in three dimensions, as depth is an additional factor to be considered in the study of marine organisms. Another key difference is the distribution of chemical nutrients and other compounds in a medium which is in continual movement due to waves and currents, resulting in very dynamic systems (Carr et al. 2003). Possibly as a result of this, marine organisms are more sensitive to environmental changes, and react relatively fast. Thus, while changes in terrestrial ecosystems caused by large-scale disruptive events (e.g., global warming) can take centuries, changes in marine ecosystems can be observed within a few decades.

Metabolism is the first to response to environmental changes experienced by a marine organism (Viant 2007). In this context, a comprehensive analysis of the metabolites produced by these organisms is crucial to understand the dynamics of marine ecosystems. Such a comprehensive analysis can be achieved through metabolomics, an "omics" technology defined as the study of all metabolites or small molecules present in an organism, cell or tissue under certain conditions (Bundy et al. 2008). Metabolomics has an incomparable edge over conventional analysis, offering several tools for chemical profiling-related studies such as metabolic fingerprinting and footprinting (Dunn and Ellis 2005). Using a systems biology approach, metabolomics studies have played an essential role in the reconstruction of molecular networks, providing renewed insights into the relationship between genes and the final products of metabolism (Kell 2004; Weckwerth 2003).

In addition to offering a full picture of molecular networks, one of the most important fields in which metabolomics could be useful is in marine chemical ecology, the discipline that investigates the individual biological/physiological roles of marine natural products. This field focuses on the interactions between marine organisms mediated by chemical compounds (Hay 2009). Despite its rapid growth over the past years, the comprehension of the possible roles of metabolites as well as the environmental, genetic and phenotypic factors affecting the biosynthesis of metabolites has remained marginal considering the vast number of metabolites that have been isolated (Paul et al. 2006; Paul and Ritson-Williams 2008). It is expected that metabolomics will provide useful information by identifying the metabolites that vary during the interaction between organisms and assigning a putative function to these metabolites.

Apart from in-depth information in advances in physiological and ecological issues, metabolomics could potentially give some key indications on the way to go for an efficient exploitation of marine resources. Unbiased metabolomic analysis allows the identification of compounds that can be used as biomarkers for toxicity, diseases or the detection of novel molecules with biological activity in big sample sets (Harvey et al. 2015; Viant 2007). The sea offers a vast array of secondary metabolites produced by marine organisms, an unexplored source of potential new drug candidates. These molecules introduce a new set of molecular features that greatly differ from compounds produced by terrestrial organisms and have thus been used as leads to develop new anticancer or antiretroviral drugs and painkillers (Gerwick and Moore 2012). For this reason, an unbiased metabolomics approach is essential to provide an extensive overview of the metabolism of marine organisms facilitating the process of discovery of new compounds that can be used for the development of novel drugs.

Despite the success of the application of metabolomics to marine organisms, its implementation has suffered a number of practical setbacks that have thwarted the development of its full potential. The collection of significant amounts of sample is often complicated by an extremely low availability, not only due to scarce natural abundance but to the difficulties encountered when accessing the collection sites and/or in their transportation in appropriate conditions from these remote collection locations to laboratories. Another challenging aspect for the experimental design is marine organism chemistry itself. Marine organism metabolomes are characterized by the massive predominance of salt and lipids, requiring specific sample pre-treatment to remove these and obtain extracts that comply with analytical requirements and can provide data sets suitable for metabolomic analysis.

Marine organism metabolomics can lead to the discovery of many bioactive compounds and an increased understanding of the interaction between certain organisms and their environment. In the last few years, a number of technological advances in existing analytical tools or the development of new related ones have contributed to pave the way for exciting applications for metabolomics in this field. This paper reviews all the required steps in the workflow of marine organism metabolomics based on recent studies: sample preparation including extraction, data analysis, and identification of metabolites. It also provides examples of the application of metabolomics to a large diversity of organisms, and the discussion of scientific and technological advances.

2. Sample preparation and extraction: how to maximize the number of intact metabolites in a simple manner

The complexity of marine metabolomics is intrinsic to both the organisms and the environment that they inhabit. Marine organisms establish tight relationships in order to survive the harsh conditions of the marine environment. Thus, when designing an experiment, it is important to bear in mind that this complexity implies that access to marine environments is much harder, requiring more complicated logistics to collect reliable sample sets. Marine ecosystems also differ from terrestrial ones in key features and salinity, water currents, depths, etc. must all be considered apart from common ones such as temperature, light, oxygen concentration, for example, among the significant environmental conditions.

Once samples are collected, i.e., separated from their natural state, it is generally highly recommended to quench the metabolism immediately, as the collection or harvesting process can *per se* rapidly produce changes in their metabolism. Many protocols, designed in general for plant or animal tissues, indicate flash freezing the samples using liquid nitrogen (Kim et al. 2010). This, however, is often impossible in the case of marine organisms due to practical constraints related to the installations and available technical equipment at collection locations which are often very distant from a laboratory or similar facilities. Existing alternatives include freezing at -20° using dry ice or adding organic solvents that can stop or reduce metabolic reactions. Representative quenching techniques that have been used in marine organism metabolomic studies are listed in Table 2.1.

Sample treatment of the collected samples involves the extraction of metabolites and eventually the removal of impurities that may be incompatible with the analytical method to be used. The selection of the extraction method, includes the choice of a solvent and conditions, is driven by two basic aims: the extraction of as many metabolites as possible and ensuring their integrity by the elimination or at least reduction of time-consuming steps and manipulation that can lead to the degradation or modifications in the chemical structure of the metabolites. Similarly to common protocols used in plant natural products studies, organic solvents such as methanol and ethanol or their aqueous mixtures are used, but the selection of solvents varies according to the predominant or targeted metabolites in the studied organism (Ebada et al. 2008). The efficiency of some extraction systems or related technology and conditions (e.g. temperature and pressure), have been evaluated using algae (Heavisides et al. 2018) and sponges (Bayona et al. 2018) as examples. In these cases, extraction conditions for automatic pressure-assisted extractions systems were evaluated to optimize yields of certain groups of metabolites or the chemical diversity of the extracts. Factors such as

temperature, solvent polarity and the number of cycles were found to greatly influence the diversity of chemical composition of the extracts. However, it is still too early to propose general guidelines for optimum conditions for the extraction of a wider range of marine organism metabolites.

Table 2.1: Examples of quenching methods applied in marine organism metabolomics.

Quenching method	Organisms
Centrifugation of the liquid culture media and filtration	Bacteria (Forner et al. 2013; Hou et al. 2012; Viegelmann et al. 2014).
Addition of solvent and freezing at -20 °C	Cyanobacteria (Luzzatto-Knaan et al. 2017; Winnikoff et al. 2014).
Freeze-drying after harvesting	Cyanobacteria (Kleigrewe et al. 2015)
Addition of solvent and centrifugation	Bacteria (Bose et al. 2015)
Frozen and stored at -20 °C	Sponge (Ali et al. 2013; Olsen et al. 2016; Reverter et al. 2018) Corals (He et al. 2014)
Snap frozen using liquid nitrogen	Coral (Farag et al. 2016; Sogin et al. 2014, 2016a) Bacteria (Boroujerdi et al. 2009)

As mentioned, marine organism samples usually contain high amounts of salts and lipids. The removal of salt from their extracts is essential since it interferes with most analytical procedures used for data acquisition. For example, samples with high salt concentration are incompatible with common analytical methods such as LC-MS, HPLC or HPTLC. One of the most common sample pre-treatment methods for desalting marine organism extracts is solid phase purification with Diaion HP-20 resins preequilibrated with methanol (Houssen and Jaspars 2012), C18 and PS-DVB SPE cartridges (Cutignano et al. 2015; Ivanišević et al. 2011). Unless specifically targeted, lipids must also be removed since especially in chromatographic analysis, they build up in lipophilic stationary phases used most commonly, causing serious reproducibility problems. This can be avoided by defatting the extract on Sephadex LH-20 with a mobile phase of methanol and dichloromethane (1:1) or C18 SPE cartridges that retain these highly lipophilic compounds (Houssen and Jaspars 2012).

3. Analytical Methods: to detect a variety of intact metabolites in a simple and robust manner

Metabolomics studies aim to detect all the metabolites present in an organism under a given set of conditions. Nevertheless, due to the limitations of each analytical platform, it is now accepted that no single method can provide a complete picture of the metabolome. In the case of marine organisms, the limitation is more critical given the vast diversity of the metabolites. Furthermore, there are relatively fewer identified metabolites to date. This difficulty has been quite acceptably solved with the increase in available single or combined spectroscopic or spectrometric and chromatographic techniques. In this section, the potential and limitations of each analytical method are discussed with various examples.

3.1 Nuclear Magnetic Resonance spectroscopy: a way of visualizing the overall metabolome

NMR spectroscopy is a widely used analytical technique, mainly for structure elucidation. Since the emergence of metabolomics, NMR, mainly ^{1}H-NMR, has been one of the most common analytical platforms used for data acquisition together with MS-based methods. Several advantages are hard-wired into NMR-based metabolomics: its high signal robustness, ease of quantification of detected metabolites and most importantly, the structural information provided directly by the spectra themselves. In the past, the evident edge of ^{1}H-NMR over other analytical platforms for metabolomics resulted in its application to the study of a wide range of organisms such as mussels (Wu and Wang 2010), shrimps (Schock et al. 2013), sponges (Ali et al. 2013), corals (Sogin et al. 2014) and various fish (Cappello et al. 2018; Melis et al. 2017). In these applications, NMR-based methods showed strong advantages over other methods, particularly in monitoring the levels of primary metabolites submitted to changes in environmental, biological or ecological conditions. Overall, NMR allowed the identification of metabolites present in high concentrations such as amino acids, ketone bodies, organic acids and the energy-related compounds ATP and carbohydrates, which could piece together a metabolic picture of primary metabolism. Notwithstanding its strength in profiling highly concentrated metabolites, ^{1}H-NMR is severely limited when it comes to the detection and identification of the less abundant secondary metabolites. This occurs due to its extremely low sensitivity, the congestion of signals in the spectra (Kim et al. 2010; Viant 2007), the difficulty in the identification of minor compounds in databases or the concurrence of one or more of these situations.

There have been many attempts to compensate for the weak points of ^{1}H-NMR. The information lost due to densely overlapping peaks in ^{1}H-NMR spectra has been partially compensated using the ^{13}C- and 2D NMR spectra. Up to now, J-resolved (Ludwig and Viant

2010), HSQC (Bingol et al. 2014) and ^{13}C-NMR (Clendinen et al. 2014) have been used in metabolomics studies. Conventionally, 2D NMR has been used to obtain additional information on the structural connectivity of molecules which is crucial for their structural elucidation. When applied to mixture analysis for profiling, 2D NMR experiments such as heteronuclear single quantum coherence spectroscopy (HSQC) and J-resolved have proved to be valuable tools for metabolomics applications. Additionally, the projection of some 2D NMR spectra such as J-resolved can be useful to reduce the signal congestion observed in ^{1}H-NMR as the multiplicity of the proton signals is removed (Verpoorte et al. 2007). However, in these experiments one of the main advantages of ^{1}H-NMR, which is the proportionality of the signals to the molar concentration of compounds, is lost, thus resulting in an information vs quantification trade-off.

Despite the limitations of the multidimensional NMR methods, such as longer acquisition times, larger file size and problems with accurate quantification, there have been some interesting advances in the use of these platforms in metabolomics analysis. Recently, improvements in J-resolved pulse experiments have allowed a sensitivity similar to ^{1}H-NMR in close to 20 minutes, making it possible to use 2D J-resolved experiments for metabolomics analysis (Ludwig and Viant 2010). A J-resolved experiment follows a workflow similar to that of ^{1}H-NMR experiments. It includes the removal of regions of the spectra that are not useful (e.g., a solvent signal), as well as normalization, scaling and bucketing to obtain the matrix used for multivariate data analysis. For the scaling step a generalized logarithm (glog) transformation has been specifically developed for the processing of J-resolved spectra. The efficacy of this method for metabolomic studies was confirmed by its use for the classification of mussel samples according to the environmental conditions to which they were exposed (Parsons et al. 2007).

As the backbone of all organic compounds is composed of carbon atoms, ^{13}C-NMR spectra also contain detailed structural information of molecules. However, due to the low natural abundance of ^{13}C and long T2 relaxation times as well as non-quantitative features in the common broadband decoupling method, it has been less used in metabolomics. In the past years, a workflow using ^{13}C-NMR spectra at natural abundance was developed by Clendinen and co-workers (Clendinen et al. 2014). With the use of an optimized ^{13}C probe, the acquisition times for mixtures were reduced to just 2 hours. With this methodology, it was possible to improve the differentiation of groups and increase the number of metabolites identified from the mixture. To date however, this method has not been applied to the metabolomics of

marine organisms studies but could nevertheless be considered as a complementary tool for metabolic studies.

3.2 Liquid chromatography-based methods with mass spectrometer detectors: A closer look at minor metabolites

Liquid chromatography hyphenated to mass spectrometry is the most commonly used technique in marine organism metabolomics. Among its several strong points is the separation provided by the chromatographical step, which also provides additional chemical information as metabolites can be distinguished by their retention time as well as by their m/z value. Moreover, the separation step avoids saturation of the ionization system, allowing more metabolites to be ionized and detected. In LC-MS analysis, hundreds of thousands of metabolites can be detected including some that are present at trace levels (Goulitquer et al. 2012). This higher sensitivity is ideal for the detection of bioactive and ecologically relevant compounds, since in many cases these are present in very small quantities (Belarbi et al. 2003).

The term LC-MS refers generically to all hyphenated systems consisting of a liquid chromatograph connected to a mass spectrometer detector. However, there are diverse types of liquid chromatographs and mass detectors. Currently, the chromatographic instrument in most LC-MS setups is an ultrahigh performance liquid chromatography system (UHPLC). This system reduces analysis time without losing resolution as compared to conventional high-performance liquid chromatography (HPLC) systems. The most common stationary phase used in LC is C18, but in recent times new normal phases such as HILIC have been developed, extending the range of metabolites that can be efficiently analyzed to include very hydrophilic and ionic compound mixtures which cannot be resolved using C18 (Belghit et al. 2017; Borras et al. 2017). Different types of mass detectors have been coupled to chromatography systems for marine metabolomics studies. Among the most commonly used are triple quadrupole (QQQ), quadrupole/ time of flight (Q-TOF), ion trap (IT) and Orbitrap. The potential and drawbacks of each system have been widely discussed in the literature (Allwood and Goodacre 2010; Forcisi et al. 2013; Lei et al. 2011). In general, Orbitrap has a higher resolution, but when it comes to quantification, QQQ and Q-TOF systems are expected to provide higher accuracy. All these MS instruments can be used for MS^n experiments, but ion trap is considered to be the most powerful, generating robust fragmentation patterns which are essential for structure elucidation.

Despite its popularity in the field of metabolomics, one of the main drawbacks of LC-MS-based studies is the complicated data mining process given the size of the generated data sets. In

practice, molecules cannot be identified simply on the basis of their molecular mass. Even with high-resolution molecular mass data, reliable fragmentation patterns are needed to confirm the molecular structure. Thus, target compounds must be submitted to tandem MS, i.e., MS^n experiments, to generate fragmentation patterns which can then be matched with a reference spectrum, much like a fingerprint. Furthermore, recent development of the GNPS (Global Natural Products Social network) platform which uses MS/MS data has enabled the deduction of structural similarities between metabolites, allowing them to be grouped accordingly (Wang et al. 2016). The detailed procedure and potential of this platform are discussed in the data analysis section.

3.3 Gas chromatography-based-mass detector hyphenated methods: Detailed metabolic analysis for primary metabolites.

Gas chromatography hyphenated to mass spectrometry is one of the earliest analytical platforms used in metabolomics studies. Even today, among all the MS-based techniques, GC-MS is still considered to offer the highest resolution, reproducibility and robustness. These two last features of the fragmentation patterns of compounds generated with electronic impact ionization (EI), the most common ionizer in GC-MS, makes their identification by matching their spectra to data libraries relatively easy (Kopka 2006). In marine metabolomics however, GC-MS has been used less frequently, mainly because GC analysis requires the metabolites to be thermostable and volatile, limiting the number and kind of compounds that can be analyzed directly using this method. This is overcome by the derivatization of samples (Sumner et al. 2003), but even with the improvements made in derivatization methods, this is still a limitation to the application of GC-MS in metabolomics and it has thus been mainly used for the detailed analysis of primary metabolism, especially for amino acids and organic acids.

Although the number of reports of metabolomics studies using GC-MS is notably lower than NMR or LC-MS, some examples show the potential of this technique for marine organism metabolomics. The GC-MS analysis of scallops (*Chlamys farreri*), an important organism in the food industry, was used to evaluate modifications in metabolic pathways related to their survival rates in different preservation conditions. The presence of metabolites indicative of a trade-off between aerobic and anaerobic metabolism detected in live scallops preserved in semi-anhydrous conditions provided essential information for the marketing of scallops, due to its direct incidence on the final value of the product (Chen et al. 2015). The metabolomes of other organisms, such as starfish and marine polychaeta have also been investigated. These studies focused on establishing the best extraction and analytical conditions for the optimized detection of metabolites. In both examples, GC-MS analysis provided detailed information of

the primary metabolites such as amino acids, lipids, and carbohydrates in the samples (Fernández-Varela et al. 2015; Pereira et al. 2012).

Along with primary metabolites, volatiles have historically been the most frequent targets of GC-MS analysis. In the past, volatiles were suspected to play various key physiological roles. In recent decades, they have also come to be considered interesting vectors in the interactions between organisms given their unique physical characteristics and high diffusion rate. Recently, *in situ* extraction methods such as headspace extraction and solid phase microextraction (SPME) have been applied to the analysis of volatile compounds present in marine organisms, such as algae (de Alencar et al. 2017; Jerković et al. 2018) as well as in some marine microorganisms (Barra et al. 2017; Salvatore et al. 2018). The use of *in situ* methodologies for the extraction and further analysis of this kind of compounds can provide additional information in metabolomics studies.

3.4 Direct infusion mass spectrometry: Getting a picture of the metabolome

Although less popular, MS-based metabolomics can be performed with data generated directly by the MS-spectrometer. A wide range of direct infusion methods are used in this case, all of which have their own application fields in metabolomics.

In studies involving marine organisms, the amount of available sample tends to be extremely low and this is aggravated by the fact that they require some preparation, starting with extraction. Ultimately, no extraction would be the ideal method for metabolomics studies. Nowadays, this is possible if using direct infusion MS techniques such as matrix-assisted laser desorption/ionization (MALDI) imaging, direct analysis in real time (DART) and desorption electrospray ionization (DESI) all of which offer the possibility of analyzing samples without an extraction step (Esquenazi et al. 2009; Parrot et al. 2018). In marine organisms, MALDI-TOF imaging has been used to detect metabolites such as viridamides, jamaicamides and curacin in filaments of marine cyanobacteria and to determine their location within sponge cells (Esquenazi et al. 2008). This study is of particular interest because while MALDI-TOF has been generally applied to analyze macromolecules, this application proves that MALDI-TOF imaging is capable of detecting low molecular mass molecules despite the interference of the matrix, eventually providing valuable information on the spatial distribution of secondary metabolites in the tissues of marine organisms.

Notwithstanding their evident value, the widespread implementation of mass imaging techniques has had its setbacks: the lack of reproducibility between samples, the difficulty in sample preparation, the low resolution of the mass measurements and complexity of data

processing. These limitations, however, have not diminished its potential as a powerful approach for studying ecological interactions involving for example surface-related phenomena and spatial location of metabolites. Recently, the combination of LC-MS based metabolomics with mass imaging has been used to study the chemical defense mechanism of eelgrass against microbial foulers (Papazian et al. 2019). In this study, untargeted metabolomics of the marine plant *Zostera marina* revealed different concentrations of a group of fatty acids and phenolic compounds in a leaf surface extract and the whole leaf. These compounds were targeted in a DESI-MS imaging analysis, showing that some fatty acids were more abundant on the surface of the leaf while phenolic compounds were more abundant within the leaf. With this information a defense mechanism of *Zostera marina* against microbial foulers was proposed. This is an example of the potential of direct infusion-mass imaging methods as an approach for marine organism metabolomics.

3.5 Other analytical platforms: To complete the metabolome puzzle

There is no single method that can cover the full metabolome of an organism, despite the increasing analytical range achieved over the past years. It has therefore become increasingly important to develop novel analytical tools to complement the information obtained by current techniques. In this context, SFC coupled to mass spectrometry has been used in metabolomic studies, aiming to increase the metabolic coverage and decrease the time of analysis (Shulaev and Isaac 2018). The particular physicochemical characteristics of the mobile phase, usually supercritical CO_2, which features the advantages of a gas and a liquid, gives SFC a niche in the field of metabolomics. Using a simple gaseous-like mobile phase, SFC fills the gap between LC- and GC-based methods allowing the analysis of compounds of a wider range of polarity, with shorter run times (mostly within 10 min), and with very high resolution and reproducibility between analyses without the need to volatilize the samples. Despite its potential, one of the main snags in the use of SFC has been the availability of a stationary phase that could cover a wide range of metabolites in a single analysis. Recent studies using plant extracts have shown that stationary phases such a DIOL and C18 are suitable for the analysis of crude extracts, resolving 88% of the compounds present in the mixture (Grand-Guillaume Perrenoud et al. 2016). Another recently used chromatographic technique in the field of metabolomics is capillary electrophoresis (CE) coupled to mass spectrometry. The developments in CE have allowed the detection of up to 900 molecular features in a single run, making this technique ideal for metabolomics studies (Zhang W. et al. 2017). However, despite the remarkable sensitivity of CE for specific metabolic groups, variation in the migration time has been one of the main obstacles to its development as a robust

metabolomics analysis tool. There have been successful efforts to overcome this issue (Hirayama et al. 2014) and though there are no reports to our knowledge of marine metabolomics studies using CE-MS, it is likely that in the foreseeable future, both SFC-MS and CE-MS may be used as supplementary metabolomics tools, increasing the chemical space.

Infrared spectroscopy has been considered a very versatile, inexpensive and robust tool for the analysis of samples. However, the use of FT-IR in metabolomics has been confined to fingerprinting studies due to its limitations in the identification or dereplication of metabolites. This can be appreciated in the study of the effect of the exposure of the green algae *Ulva lactuca* to different concentrations of oil fuels (Pilatti et al. 2017). This showed FT-IR to be a reliable tool for fingerprinting or footprinting applied to the detection of changes induced by marine organism exposure to fossil fuels, even at low concentrations. Another example of the use of FT-IR for fingerprinting is a study performed with the mussels *Mytilus edulis* (Gidman et al. 2007). In this case, FT-IR was used to observe chemical changes in the hemolymph caused by environmental stress.

Thin layer chromatography (TLC) was among the first separation techniques used in analytical chemistry and it is still being applied for the rapid evaluation of chemical profiles. Recent improvements of this technique aimed at increasing its relatively low resolution, have resulted in the development of high-performance thin-layer chromatography (HPTLC). This renewed technique has recently been applied to metabolic fingerprinting, mainly in plant metabolomic studies (Bayona et al. 2019). Several other studies had revealed the potential of HPTLC for data acquisition in marine organism metabolomics, e.g. sponges and algae. In these cases, the possibility of performing *in situ* bioactivity tests provided by HPTLC separation allowed the rapid recognition of bioactive metabolites. Moreover, the integration of HPTLC with a MS-detector can facilitate the identification of compounds, providing, eventually, guidance for preparative separation work.

4. Data analysis and multivariate data analysis: to extract hidden information from a raw data matrix

The data acquired using any of the described analytical platforms is generally massive and must be profiled in an untargeted manner. This is a key step in any metabolomics study. Data processing involves the conversion of raw spectroscopic and chromatographic data to a numerical matrix which can be used for statistical or multivariate data analysis. It is an essential step for the extraction of hidden valuable information from the bulky data sets and normally requires several steps depending on the source (analytical method) of the data (Figure 2.1 and

2.2). The preprocessing of NMR data is somewhat simpler than that required for other analytical tools and consists of phase correction, baseline adjustment, shift adjustment, binning (also

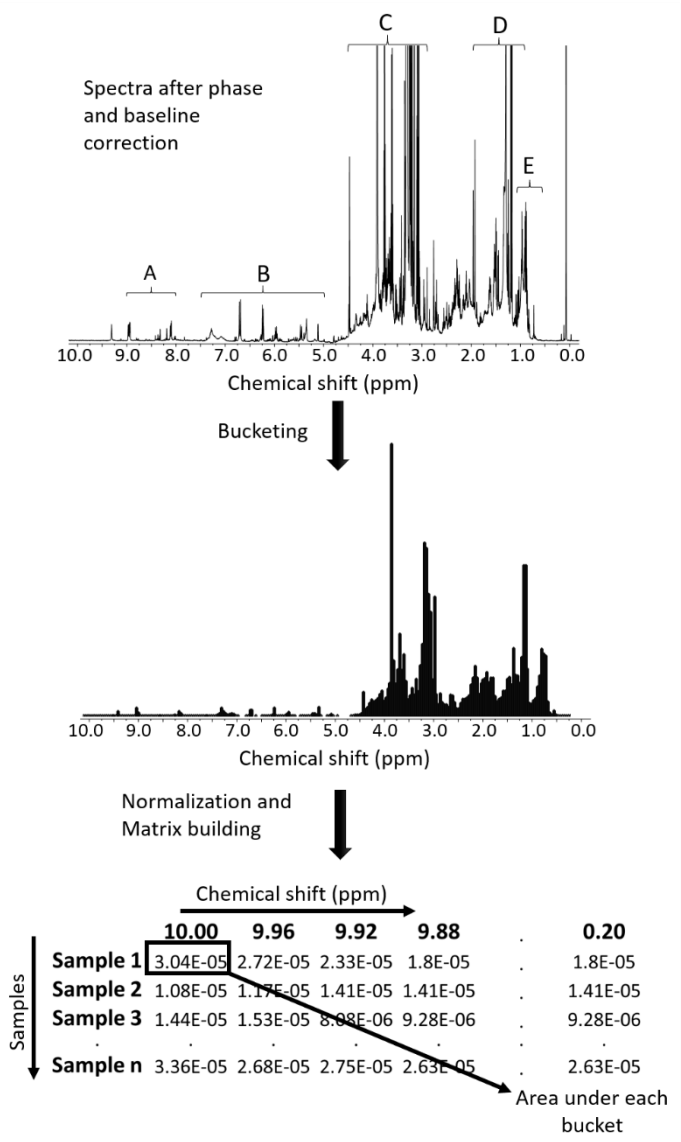


Figure 2.1: Schematic diagram of the workflow to build a data matrix of ^{1}H-NMR spectra using an example of the marine sponge *Xestospongia* sp. in CH_3OH-d_4. The regions pointed in the NMR spectra correspond to A) aromatic quinones, B) unsaturated brominated fatty acids, C) hydroxylated fatty acids and glycerol derivates D) methylene groups E) Sterols

and the flexibility of the binning range have a substantial impact on the final results. Ideally, a smaller binning size should result in a more precise representation of the original spectrum, but this can cause the shifting of signals which eventually may result in a loss of important information. In the past, the smallest binning size was suggested to be between 0.01-0.04 ppm for 400 - 600 MHz NMR as long as no further alignment was applied to. However, the signal robustness gained from binning is a trade-off for resolution that is decreased compared to the raw data (Figure 2.1). To overcome this dilemma, many alternative algorithms have been developed to improve the binning procedure or the alignment of the spectra, e.g. dynamic adaptative binning (Anderson et al. 2011), fuzzy warping (Wu et al. 2006) and peak alignment using a genetic algorithm (Forshed et al. 2003).

Compared to the simplicity of the NMR dataset, LC-MS datasets are inherently more complex due to the multi-dimensional features provided by the response of analytes to both the chromatographic separation and mass spectrometry. Moreover, the higher sensitivity of this method results in large data sets that can be more challenging to work with. Several software programs are available for data processing, some of which are open access, such as MZMine 2 (Pluskal et al. 2010), OpenMS (Röst et al. 2016), XCMS (Smith et al. 2006), and MS-DIAL (Tsugawa et al. 2015). Although each software uses its own algorithms with different steps, they share the goal of creating a matrix that contains the intensity of all molecular features in each sample and removes possible technical variations (Figure 2.2). Alignment of the molecular features in both retention time and mass values is seen as the most essential part of LC-MS data processing, and variability in retention times, mass values and isotopic patterns must all be considered.

Regardless of the analytical methods used for metabolomics analyses, a huge number of variables will always be generated. This can range from several hundred up to tens of thousands of variables. Therefore, the first important step in data mining is the reduction of data dimensionality to extract hidden information from the raw data matrix. Multivariate data analysis (MVDA) is usually divided into two categories: unsupervised and supervised methods. For unsupervised MVDA, many approaches have been developed to reduce the dimensionality of the data, out of which principal component analysis (PCA) is currently the most commonly applied in the field of metabolomics. The aim of PCA is to collect most of the variance present in a data set using new orthogonal variables, known as principal components (Worley and Powers 2013). By analyzing the first few components of the PCA, it is possible to observe

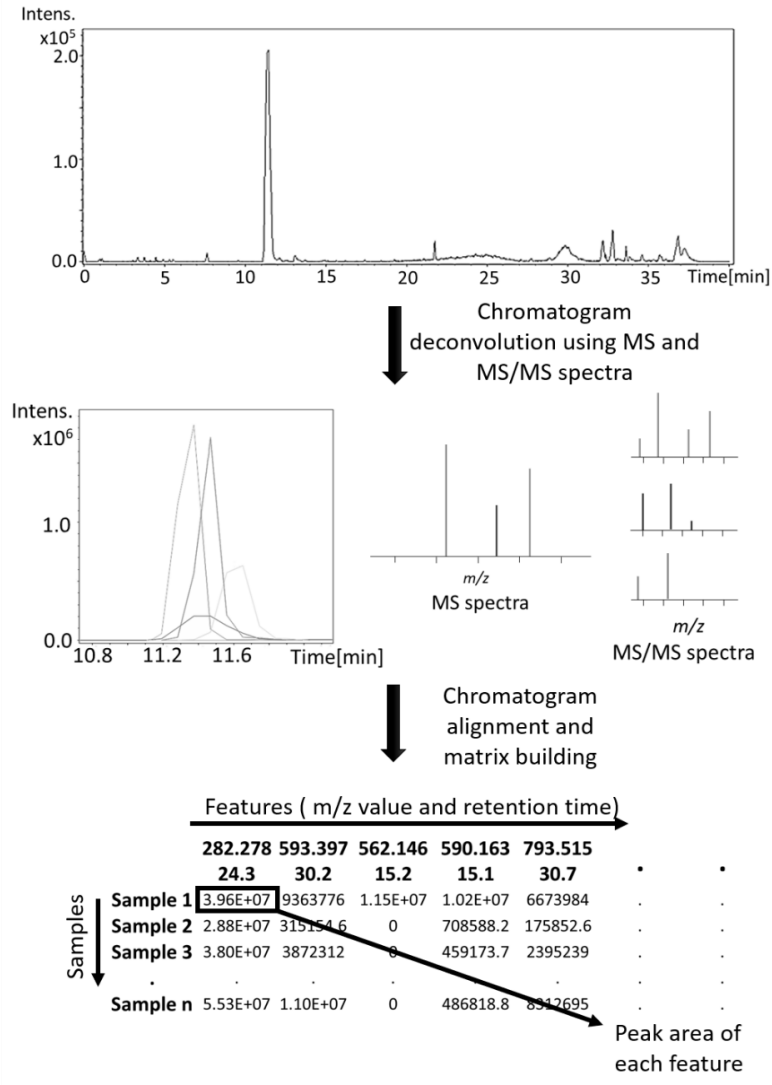


Figure 2.2: Schematic diagram of the workflow to build a matrix of LC-MS data, as an example the chromatogram of a sample from the marine sponge *Xestospongia* sp. The chromatograms deconvolution is done using both MS and MS/MS data. Followed by the alignment of the features present in each sample using m/z values and retention time criteria.

whether samples that belong to a specific class are grouped together. The metabolites associated with the classes or factors of interest can then be singled out using the loading plots. There are a number of examples of marine organism studies in which PCA was successfully employed: the metabolic discrimination of three species of crabs (*Callinectes sapidus*, *Eriphia verrucosa* and *Cancer pagurus*) (Zotti et al. 2016), the differentiation of

bacterial strains that produce new compounds (Macintyre et al. 2014), and the detection of metabolic changes caused by exposure to steroids in mussels (Cappello et al. 2017).

Similarly to unsupervised methods, there are numerous methods that can be applied for supervised approaches in many biological systems in marine organisms. Typical examples of the applications of various unsupervised methods are shown in Table 2.2 and 2.3. Although there is no universal way to perform an supervised MVDA, currently, partial least square (PLS) modeling, or its variations such as PLS-discriminant analysis (PLS-DA), orthogonal partial least square (OPLS) modeling and OPLS-discriminant analysis (OPLS-DA) are the most common methods applied after completing a PCA. In PLS (for continuous Y-variables) or PLS-DA (for discrete Y variables) the new variable will be calculated to maximize the covariance between the X- data matrix (e.g., NMR, LC-MS, GC-Ms data) and Y-data matrix (e.g., age, distance, specie, and biological activity) (Liland 2011). However, even after the use of PLS modeling, it may be difficult to find patterns in the data set or in the validation of the model due to prevailing variation caused by factors that are not included in the study. To solve this problem, OPLS and OPLS-DA have been developed. These analyses are a rotation of the original solution of the PLS and PLS-DA analyses. Although these models do not represent an improvement in the predictive capacity of the model, the interpretation of the data is more straightforward (Liland 2011). Especially OPLS-DA is being applied to discriminate groups such as active and non-active samples in a data set, after which the metabolites likely to be responsible for the activity are selected using the samples as a guideline.

5. Databases: both for the dereplication of known metabolites and for a repository for future identification

The most efficient way to annotate the metabolites presumably associated with certain treatments is to dereplicate the compounds using available data bases (Blunt et al. 2012). There are two types of databases used in metabolomics studies; general chemical entities and spectral data libraries. The general chemical databases such as PubChem (PubChem, NIH), ChemSpider (ChemSpider, RCS), the CAS registry through SciFinder (SciFinder, CAS), Dictionary of Natural Products (CRC Press), and AntiBase (AntiBase, Wiley) are available online. These databases provide information on nomenclature, molecular formula and mass, source and physicochemical properties of the compounds. Databases such as PubChem and ChemSpider, are non-specialized and cover all kinds of chemical entities regardless of their origin. On the other hand, databases like Dictionary of Natural Products and Antibase focus on compounds isolated from specific natural sources such as plants, animals or microorganisms. In the field of marine natural products, the unique chemical features that characterize compounds

isolated from marine organisms, particularly marine invertebrates, together with the wide chemical space that is covered by marine natural products, demands the use of specialized databases. MarinLit is perhaps the most important database in the marine natural products field containing more than 27500 compounds reported from marine sources (MarinLit, RSC). Another of its advantages is the possibility of including a wide range of features in its search criteria. In particular, compounds can be searched using spectroscopic data that include UV maximum, NMR chemical shifts and the exact mass.

The second type of databases focus on spectral information and are very useful for dereplication in the field of metabolomics. These databases are usually specialized in one analytical technique. NMR databases such as SpecInfo-Consortium Member-NMR, IR and MS (SpecInfo-Consortium Member-NMR, IR and MS, Wiley) and Chenomx contain information on the ^{1}H-NMR and/or ^{13}C-NMR spectra of several thousand compounds. Particularly for ^{1}H-NMR spectra, complete databases are required for full identification, as signals in a specific region of the spectra can be assigned to different compounds or kinds of compounds and detailed information on specific compounds is very helpful. In the case of GC-MS data, the widespread use of electronic impact (EI) as an ionization method generates highly reproducible fragmentation patterns that allow reliable identification using databases. In contrast, LC-MS analysis has produced less reproducible fragmentation data. In the past, analyses focused on mass of adduct ions because of their soft-ionization modes. However, recently the necessity of tandem mass information (MS/MS and MS^{n}) for identification has become evident. Databases such as NIST 20 (NIST 20), MassBank (Horai et al. 2010), and METLIN (Guijas et al. 2018) have available MS/MS spectra that can be used to perform dereplication of compounds, which include different ionization methods such as EI, ESI and MALDI coupled to various mass analyzers. These databases enable the matching of putative compounds using both their molecular mass and the MS/MS spectra. Additional search features, such as fragment losses in the MS/MS spectra can help to find analogs of unknown compounds.

Marine metabolomic studies, as any other field in metabolomics, greatly rely on the information available in databases for the annotation of compounds. The availability of specialized databases such as MarinLit and Antibase is particularly useful for the identification of secondary metabolites. However, the use of generic databases should not be dismissed, as several metabolomics studies of marine organisms have revealed changes in primary metabolism that may not be available yet in specialized libraries. Mass spectrometry databases are continually updated with additions of new information / molecules, making them a powerful tool for the dereplication of compounds, even if the standardization of

experimental conditions is still a distant concept. In the case of NMR metabolomics, it is important to note that there is a need for the inclusion of more NMR spectroscopic data of marine natural products into databases to improve the annotation rate of compounds. Moreover, the *in silico* predictions of NMR spectra, already available in software such as ACD/Labs, could benefit from the addition of more experimental data into the databases. Lastly, with the increase in public spectroscopic data that is now required in metabolomics studies and the inclusion of this information into databases, an increase in the annotation of compounds in marine metabolomics studies can be expected.

6. Statistical total correlation (STOCSY), small molecule accurate recognition technology (SMART) and molecular networking (MN): *in silico* identification of metabolites from complex mixtures

The final aim of any metabolomics study is the identification of metabolites from the signals selected from the MVDA. However, the process between the selection of a signal and the identification of the corresponding metabolites is arduous and time-consuming and usually considered to be the bottleneck in metabolomics studies. A reasonable approach for identification is to use currently available databases as shown in the previous section. However, the limited number of entities in databases (e.g. NMR and MS spectra) was an incentive for the development of *in silico* identification strategies using both NMR and MS based methods. The different nature of the features obtained from NMR and MS data have led to the development of diverse approaches.

One of the NMR-based computational tools used in some marine organism metabolomic studies is statistical total correlation spectroscopy (STOCSY). This method uses the correlation between the intensity of the signals in the ^{1}H-NMR spectra to distinguish which signals belong to the same compound or family of compounds (Figure 2.3a), facilitating the identification of some molecules (Lindon and Nicholson 2008). STOCSY was used to identify compounds such as alanine, trigonelline, threonine and lactate, present in coral extracts, in a study that aimed to distinguish the chemical profiles of reef building corals (Sogin et al. 2014). It was also used to identify the changes in the metabolic profile of breast cancer cells treated with candidate anticancer marine natural products (Bayet-Robert et al. 2010). It was possible to determine that when treated with ascididemin, a compound isolated from the marine tunicate *didemnum* sp., the cells accumulated unusual amounts of gluconic acid.

Other computational tools based on NMR data, combine HSQC spectra with deep convolutional neural networks in a new platform called Small Molecule Accurate Recognition

Technology (SMART) for the dereplication of natural products (Zhang et al. 2017). This platform enables rapid dereplication of molecules into a family of compounds. In combination with metabolomics platforms based on MS data, such as molecular networking, it can lead to the fast and efficient discovery of new molecules. This was the case for viqueamide C, a new cyclopeptide isolated from marine cyanobacteria *Rivularia* sp. and *Moorea producens* (Gerwick 2017). This molecule was highlighted by both molecular networking and SMART approaches as a new compound from the family of viqueamides, leading to its isolation and confirmation of its structure.

Molecular networking (MN) is a metabolomics workflow based on MS-data sets, which has emerged in the last few years (Yang et al. 2013). When it was first introduced, it was used as a dereplication methodology, but this was later extended to the visualization of complex connections between all metabolites, which could reveal metabolic correlations in biosynthesis. The construction of molecular networks is based on the analysis of MS/MS spectra of compounds, assuming that molecules with similar chemical structures should display similar fragmentation patterns. The basic idea of MN data calculation is to give cosine scores to MS/MS spectral similarity: The closer the score is to 1 the higher the similarity between spectra with a threshold usually set at 0.7 (Figure 2.3b) (Quinn et al. 2017). Recently, a new workflow called feature-based molecular networking was developed (Nothias et al. 2020). This network has some advantages compared to traditional Molecular Networking as isomers can be distinguished based on differences in retention times and the quantification of the features in the samples is more accurate, as it uses peak area instead of ion intensity.

The power of MN lies in grouping the metabolites with structural similarities. This has proved to be very useful for metabolomics and particularly in marine organism metabolomics, such as the study of three marine cyanobacteria strains *Trichodesmium erythraeum*, *Okeania* sp, and *Oscillatoria* sp. that allowed the identification of the aplysiatoxins family from the metabolomic pools (Ding et al. 2018). The *T. erythraeum* strains collected in different locations displayed highly characteristic clusters depending on the location, while those from *Okeania* sp. and *Oscillatoria* sp. appeared to be similar even at a distance of 100 m. Ecologically, this investigation is an example of the strong influence of environmental factors on the metabolic production of these cyanobacteria. From a chemical perspective, it provided an example of the possibility of identifying compounds within clusters using the structural similarity to known compounds present in this same cluster (Yang et al. 2013).

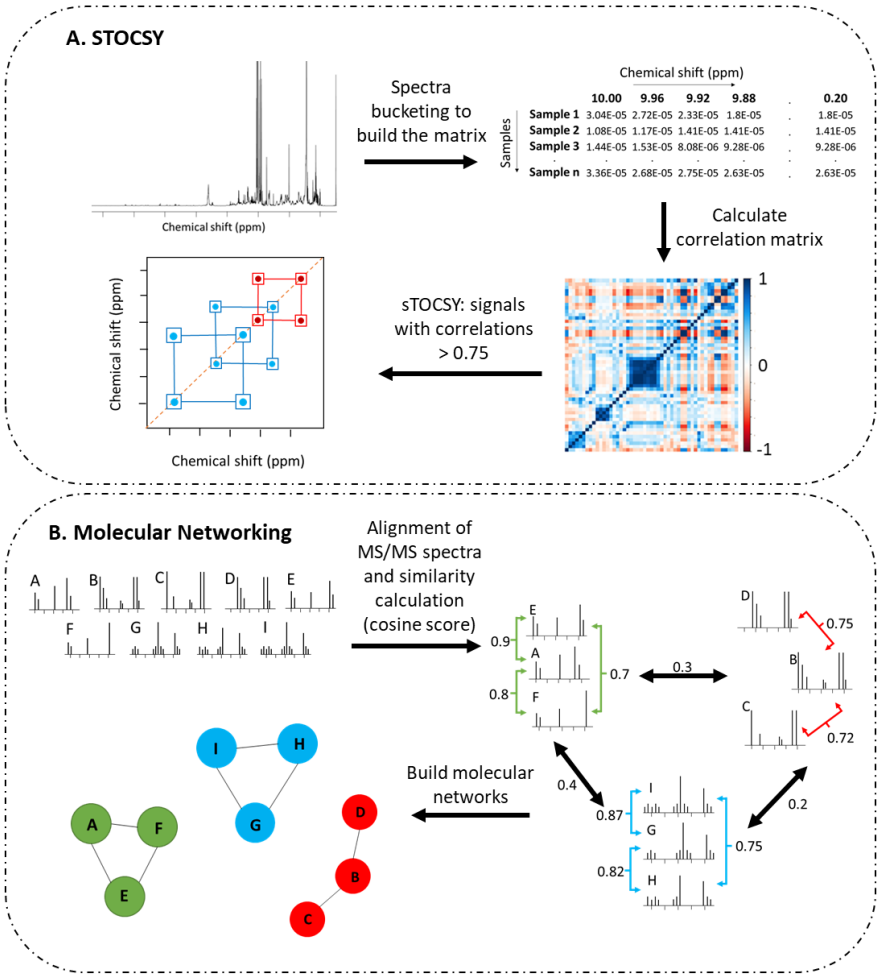


Figure 2.3: Comparison of the workflows of A) STOCSY (statistical total correlation spectroscopy) and B) Molecular networking. A: the correlation matrix shows the correlation values. Those with correlation values above 0.75 are plotted in the spectra. In red are the correlations that are also present in COSY spectra as a result of chemical bonds, and in blue are the correlations between signals that are not chemically bonded B: shows the alignment of MS/MS spectra and calculation of the similarity cosine score, features with scores above 0.7 are connected adapted from (Watrous et al. 2012).

The use of computational tools has increased the amount of information that can be extracted from NMR spectra and consequently the identification of compounds in a mixture. One of the

Even with existing databases, the chance of success in the identification of metabolites would still be low, mainly due to the limitation of data entities, which require time-consuming measurements. To circumvent this limitation, *in silico* dereplication tools, especially *in silico*

MS fragmentation, appeared a few years ago as a solution to the lack of experimental information. Following the proposal of the Wolfender group to use *in silico* dereplication together with molecular networks (Allard et al. 2016), many researchers have followed suit and developed platforms with this purpose. Platforms such as network annotation propagator NAP (da Silva et al. 2018), MS2LDA (van der Hooft et al. 2016) and MolNEtEnhancer (Ernst et al. 2019) have all been implemented in the GNPS webpage, to be used together with molecular networks for the identification of molecules or families of molecules.

7. Metabolomics applications to marine organisms: biological, ecological, pharmaceutical and physiological studies

7.1 Marine natural products and discovery of new bioactive compounds

The structural novelty of marine natural products shows great potential, as they can provide a multitude of leads for the development of novel drugs. There are, however, a few critical limitations for the isolation of natural products from marine organisms: the structural complexity of their metabolites, the extremely low concentration of active molecules and the limited amount of source materials. Considering that these organisms cannot be easily obtained simply by cultivating or farming, the only point that can be optimized is the deconvolution of chemical structures from small amounts of complex mixtures. Therefore, any successful project for drug discovery from marine sources must adopt an approach that can speed up identification of bioactive molecules in these relatively adverse circumstances.

Metabolomics has been increasingly used in recent years for the study of marine natural products as a new tool for metabolic mining. Metabolomics platforms have improved the ability to prioritize which samples within a sample set have either novel chemical structures or possible bioactive compounds. Moreover, it is even possible to target putative active compounds before starting the isolation process. This pre-profiling procedure based on metabolomics saves time and effort, highlighting valuable sources of new structures and activities. In general, the procedure for the discovery of bioactive molecules using metabolomics in marine organisms consists of metabolic profiling and bioactivity testing, correlation of features by MVDA, selection of signals responsible for positive correlations and identification of the metabolites corresponding to the signal. Table 2.2 lists applications of metabolomics platforms for the discovery of new molecules from various marine organisms.

Table 2.2: Examples of application of metabolomics for the discovery of new bioactive compounds

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Bacteria *Vibrio* sp. QWI-06	UHPLC-ESI-Orbitrap-MS	Molecular networking	no	Discovery of vitroprocines A-J active against *Acinetobacter baumannii.*	(Liaw et al. 2015)
Cyanobacteria *Moorea bouillonii* and *Moorea producens*	HPLC-ESI-Q-TOF-MS	Molecular networking	no	Discovery of three new metabolites columbamides A-C.	(Kleigrewe et al. 2015)
Cyanobacteria *Moorea producens* JHB	LC-ESI-LTQ-FTICR-MS	Molecular networking	no	Discovery of new compounds of the Jamaicamide and Hectochlorins family.	(Boudreau et al. 2015)
Bacteria *Salinispora arenícola*	UHPLC-ESI-Q-TOF-MS	PCA	OPLS-DA	The incubation time and salinity of the culture media of *Salinispora arenicola* change the quantity and type of Rifamycins produced.	(Bose et al. 2015)
Fungi *Aspergillus terreus*	LC-ESI-TOF-MS	PCA	no	Using diverse cultivation media lead to the discovery of the new compound 7-desmethylcitreoviridin.	(Adpressa and Loesgen 2016)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Bacteria 1000 marine microorganisms	UHPLC-ESI-Q-TOF-MS	Molecular networking and PCoA[a]	no	Increment in the chemical space by using diverse extraction methods and the discovery of two compounds: maridric acids A and B.	(Floros et al. 2016)
Sponge *Geodia macandrewii* and *Geodia baretti*	UHPLC-ESI-TOF-MS	PCA	OPLS-DA	The differentiation of two sponges of the genera *Geodia* and the isolation of the new compound geodiataurine.	(Olsen et al. 2016)
Bacteria *Micromonospora* sp and *Rhodococcus* sp	UHPLC-ESI-Q-TOF-MS	PCA	no	Discovery of keycin, a new antibiotic from the co-culture of the bacteria.	(Adnani et al. 2017)
317 marine cyanobacteria and benthic algae	UHPLC-ESI-Q-TOF-MS	Molecular networking and PCoA[a]	no	Differences in the chemical composition of cyanobacteria collected in a different location giving and the discovery of yuvalamine A, a new compound.	(Luzzatto-Knaan et al. 2017)
Bacteria *Streptomyces* sp. WU20	HPLC-ESI-QQQ-MS	no	PLS	The cultivation of *Streptomyces* sp WU 20 under nickel stress lead to the discovery of a new compound with antibacterial activity.	(Shi et al. 2017)
Bacteria 24 *actinobacteria* like strains	HPLC-ESI-IT-MS	HCA	OPLS-DA	24 strains were group depending on their metabolic production. Putative active compounds in quorum sensing assays were selected and dereplicated.	(Betancur et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Sponge *Spongia officinalis*	HPLC-ESI-Q-TOF-MS	PCA and molecular networking	PLS-DA	Composition of furanoterpenes in *S. officinalis* is changed depending on the geographical location and season. Three new compounds: Furofficin and Spongialactam A and B were isolated.	(Bauvais et al. 2017)
Fungi 21 isolated fungi	UHPLC-ESI-Q-TOF-MS	Molecular networking	PLS-DA	The co-culture of isolated fungi with phytopathogenic bacteria and fungi trigger the production putative active secondary metabolites. One new putative peptide from the emerimicin family was annotated.	(Oppong-Danquah et al. 2018)
Alga *Fucus vesiculosus*	UHPLC-ESI-Q-TOF-MS	Molecular networking	no	Seasonal changes in the metabolome the algae were observed and the variation in the concentration of some metabolites was related to changes in the bioactivity of the alga extracts.	(Heavisides et al. 2018)

[a] Principal coordinate analysis

7.2 Biological and Ecological Applications

The discovery of new molecules from marine organisms focusses mainly on sedentary organisms such as sponges, corals, algae, and microorganisms. In contrast, biological or ecological studies pose a greater variety of research questions and include wide variety of animals such as fish, mussels, and turtles. The diversity of organisms in these studies results in a great heterogenicity in the type of samples. The preanalytical and analytical protocols confined to sponges, corals, and other sessile organisms might be similar to the ones used in general drug discovery studies. However, for other organisms such as, fish, mussels and marine mammals, samples from organs, tissues and biofluids require preanalytical treatment similar to the ones used in human metabolomics studies.

The investigation of the effect of environmental changes in the metabolome of marine organisms resulting from human activities has been growing in recent years. Currently, many industries produce residues such as pesticides, antibiotics, hormones, and heavy metals that end up polluting the oceans. In the past few years, a number of studies concerning the effect of artificial chemical residues on the metabolome of a wide range of marine organisms have been actively carried out. Studies on the connection between environmental changes and metabolomes have been well reviewed (Bundy et al. 2008). Additionally, reports of studies that apply metabolomics to explain physiological processes of marine organisms such as mating, predator defense, and growth development, all of which are likely mediated by chemical compounds, have been growing gradually, especially related to sessile animals. A summary of recent applications of metabolomics to biological and environmental sciences are shown in Table 2.3.

Lastly, one of the central and most substantial issues in marine biology is taxonomical classification. Many of the traditionally used taxonomic characteristics are not useful for the classification of marine organisms and the available knowledge on genetic markers is still scarce. Classical characterization has in many cases resulted in misclassifications of many species, generating a constant creation and reclassification of species. The use of metabolomics in this field could provide an additional taxonomical marker or at least a holistic overview of the metabolome, which is unlikely to be achieved with conventional methods. For example, metabolic fingerprinting was implemented to distinguish two morphotypes of the zoanthid, *Parazoanthus axinellae,* collected in different locations of the Mediterranean Sea. This resulted in the detection of taxonomic marker type metabolites: ecdysteroids, zoanthoxanthins, and parazoanthines (Cachet et al. 2015). Parazoanthines were found to be

present only in the "slender" morphotype during the whole year but not in the other morphotype. This result endorsed the revision of the classification of *Parazoanthus axinellae.*

8. Summary and perspectives

Although knowledge of marine organism and their ecosystems has been rapidly increasing over the past decades, with more species and thousands of new molecules reported every year, the scale of unexplored organisms is thought to be astounding. Moreover, even with the developments in technologies such as SCUBA diving, ROV's (Remotely operated vehicles) and research submersibles, which allow a study of the environment that could not even be imagined in the past, the study of marine organisms remains an arduous task. The application of metabolomics in marine organism studies provides a new approach for the discovery of compounds that can be used for the benefit of humans in different ways. Perhaps even more importantly, it has increased the understanding of the function that all these metabolites have within the producing organism as well as the interactions of these organisms with their environment.

Recent studies have revealed the association of microorganisms to be the real synthesizers of many of the secondary metabolites isolated from marine invertebrates. This has prompted the study of marine associated microorganisms, where new active novel compounds can be discovered using metabolomics as a criterion to select strains for the study. The main focus of marine chemical ecology and environmental metabolomics is the study of interactions of marine organisms with their environment and with other organisms. There is particular interest in the study of the effect of modifications in environmental conditions resulting from human activities. Within this field, metabolomics-based analyses could be expected to lead to increased understanding of the interactions between some organisms and their symbionts, or biosynthetic pathways of compounds and how these are transferred between organisms. Another interesting topic to study is the biotransformation of secondary metabolites along the trophic chain, since some metabolites are known to originate through this complex mechanism. In this case, metabolomics studies together with biosynthetic pathway studies can lead to the discovery of how compounds are transformed.

The ultimate goal of metabolomics, that is, acquiring a picture of the whole metabolome of an organism is still a dream, as none of the existing analytical platforms can detect all metabolites in a single analysis. The addition of new analytical tools, such as CE-MS, SFC-MS, HPTLC and IR- based metabolomics, is a step in this direction but even then, the use of more analytical platforms does not necessarily guarantee more useful information. Rather, it is

Table 2.3: Examples of the application of metabolomics to environmental and biological studies of marine organisms

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Coral *Pocillopora damicornis*	UHPLC-ESI-Q-TOF-MS and GC-EI-TOF-MS	no	OPLS-DA	Exposure to different conditions of temperature and pCO_2 shift metabolic pathways including carbohydrate metabolism, cell structural maintenance, defense mechanisms among others.	(Sogin et al. 2016)
Coral *Sarcophyton* spp., *Lobophytum pauciflorum*, and *Sinularia polydactyla*	UHPLC-ESI-LCQ-MS and ^{1}H-NMR	PCA	OPLS	Coral growth in the wild exhibit higher levels of cembranoids, the most common group of diterpenes reported for soft corals, while corals growing in an aquarium have a higher content of oxylipins.	(Farag et al. 2016)
Bacteria *Persicivirga* (Nonlabens) *mediterranea* TC4 and TC7 *Pseudoalteromonas lipolytica* TC8 and *Shewanella* sp. TC11	UHPLC–ESI-Q-TOF-MS	PCA and molecular networking	PLS-DA	Different culture conditions are reflected in the metabolome of the four bacteria studied. Compounds of the family of hydroxylated ornithine lipids, diamine lipids and glycine lipids are putative biomarkers.	(Favre et al. 2017)
Alga *Ulva mutabilis*	GC-EI-TOF-MS and UHPLC-ESI-TOF-MS	PCoA[a]	CAP[b]	The exo-metabolome and development of *Ulva mutabilis* is changed by the presences of symbionts.	(Alsufyani et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Clams *Ruditapes philippinarum*	^{1}H-NMR	no	PLS-DA	The analysis of *R. philippinarum* gills after been exposed to hypoxia showed changes in the concentration of some amino acids and energy related metabolites.	(Y. Zhang et al. 2017)
Sea snail *Haliotis diversicolor*	^{1}H-NMR	PCA	OPLS-DA	The exposure of *H. diversicolor* to organotin contaminant cause changes in the energy metabolism, osmotic balance oxidative stress. Moreover, the metabolic response is different depending on the sex and the tissue analyzed.	(Lu et al. 2017)
Mussels *Mytilus galloprovincialis*	^{1}H-NMR	PCA	no	The exposure of *M. galloprovincialis* to drospirenone has no effect in their sexual development but disrupt energy, amino acids, and glycerophospholipid metabolism.	(Cappello et al. 2017)
Macroalga *Asparagopsis taxiformis* Coral *Astroides calycularis*	UHPLC–ESI-QTOF-MS	PCA	no	The interaction between a coral and an invasive alga revealed no changes in the metabolome of the coral while the metabolome of the alga was changed when in contact with the coral.	(Greff et al. 2017)
Fish *Salvelinus alpinus*	^{1}H-NMR	PCA	PLS-DA	The metabolic changes caused by a test diet were reflected in the metabolome of the plasma, liver and muscle. Based on this, improvements in the diet were proposed.	(Cheng et al. 2017)

Organisms	Analytical method	Pattern recognition		Finding	Reference
		Unsupervised	Supervised		
Sponge *Haliclona mucosa* and *Haliclona fulva*	UHPLC-ESI-Q-TOF	PCA	no	Differences in the metabolome between the two species were observed. Additionally, a decrease in the diversity of the metabolome during the period between April and May was observed and variation due to the location was detected over a 200 km ratio in the Mediterranean Sea.	(Reverter et al. 2018)
Fish *Thunnus thynnus*	^{1}H-NMR	PCA	no	Study of the liver tissue of bluefin tuna showed differences in the metabolic changes according to the gender, caused by the accumulation of environmental contaminant. Energy- related metabolites, amino acids and lipids were identified as the most affected metabolites.	(Cappello et al. 2018)

[a] Principal coordinate analysis

[b] Canonical analysis of principle coordinates

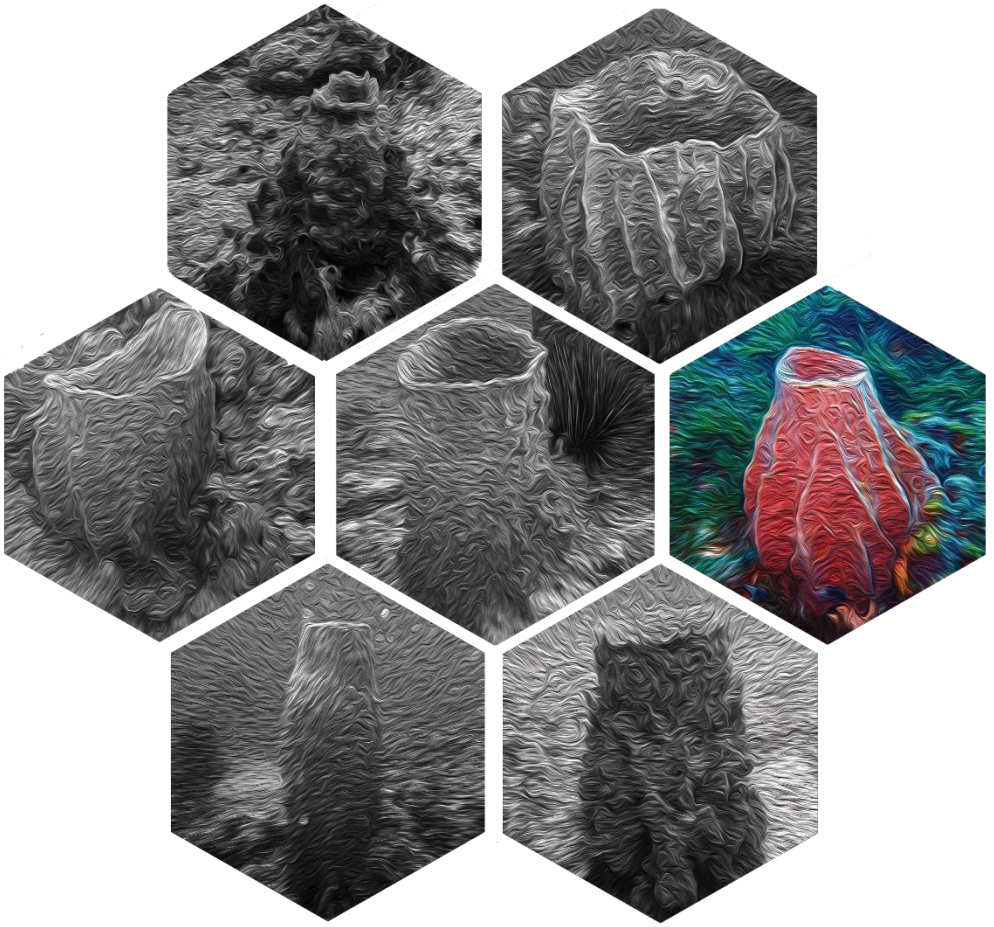
increasingly clear that the solution to this problem may come via the combination of available metabolomics tools in a way that can provide the most useful information about the particular organism under study using bioinformatic tools.

Identification of the metabolites of interest continues to be the major restriction in metabolomics studies. Although dereplication can be performed in the initial steps, on many occasions the compounds cannot be detected within the mixture and have to be isolated for their identification. In the isolation process, the limited amount of sample is always an obstacle and this has been circumvented in some cases by the implementation of microscale isolation protocols for low amounts of crude extracts. Additionally, other dereplication tools such as *in silico* MS/MS comparison can be useful for the identification of compounds, as reference compounds needed for identity confirmation are generally very difficult, if not impossible to acquire. The use of these strategies can facilitate metabolomics studies and speed up the process of discovering new compounds and thereby reveal the function of these compounds in the organisms.

Until now studies have focused on the metabolites present inside the organisms, ignoring those that are released into the ocean. To achieve a deeper understanding of the interactions occurring in marine organisms, it is crucial to undertake the study of metabolites that are released into the water, as it is highly likely that these metabolites are possible mediators in these interactions. However, the study of exuded metabolites represents a major experimental challenge, as they are rapidly diluted in the ocean making their analysis very difficult. As a first approach, studies could be performed in the laboratory, in a controlled environment that reproduces the conditions is in the field. In addition, water filtering systems could allow to concentrate the metabolites using methods such as solid-phase extraction. The use of this kind of experiment enables the collection of metabolites in a sufficient quantity to perform further metabolomics analyses. These kinds of studies will become an indispensable tool to understand the transfer of metabolites between organisms, providing valuable information for fields such as aquaculture, marine chemical ecology, and environmental metabolomics.

Acknowledgements

The authors greatly appreciate the contribution of Dr. Erika G. Wilson in the scientific discussion of this paper. This work was supported by the COLCIENCIAS (#728, Science Technology and Innovation Ministry, Colombian Government)



Chapter 3

Increasing Metabolic Diversity in Marine Sponges Extracts by Controlling Extraction Parameters

Lina M. Bayona [1], Melina Videnova [1] and Young Hae Choi [1]*

[1] Natural Products Laboratory, Institute of Biology, Leiden University, 2333BE Leiden, The Netherlands.

* Corresponding author e-mail: y.choi@chem.leidenuniv.nl

Marine Drugs (2018), 16(10).

https://doi.org/10.3390/md16100393

Abstract

Metabolomics has become an important tool in the search for bioactive compounds from natural sources, with the recent inclusion of marine organisms. Of the several steps performed in metabolomics studies, the extraction process is a crucial step, which has been overlooked for a long time. In the present study, a pressurized liquid extraction system was used to investigate the effect of extraction parameters such as pressure, temperature, number of cycles, and solvent polarity on the chemical diversity of the extract obtained from the marine sponge, *Xestospongia*. For this, a full factorial design (2^4) was performed using a chemical diversity index, which was found to be a suitable tool to determine the efficiency of the extraction process, as the response variable. This index was calculated using a logarithmic transformation of ^{1}H-NMR signals. Three factors (number of cycles, temperature, and solvent polarity) and two interactions were found to affect the chemical diversity of the obtained extracts significantly. Two individual factors (temperature and solvent polarity) were selected for further study on their influence on sponge metabolites using orthogonal partial least square (OPLS) modeling. Based on the results, the groups of compounds that were most influenced by these parameters were determined, and it was concluded that ethanol as extraction solvent together with low temperatures were the conditions that provided a higher chemical diversity in the extract.

Keywords: marine organism, extraction, metabolomic diversity, ^{1}H-NMR, design of experiment

1. Introduction

Drug discovery is an arduous process that requires enormous amounts of time and money and, more importantly, it requires a vast pool of candidates to uncover a lead compound. It has been calculated that between 5000 to 10000 chemical candidates are needed at the beginning of the drug development process in order to obtain one approved drug (Torjesen 2015). Consequently, the search for a sustainable source which can provide a large number of chemical entities has always been a long-term goal for scientists. Natural products, defined as all the small molecules naturally synthesized by a living organism, have undoubtedly been one of the most prolific sources of bioactive chemicals for drug discovery. For instance, of all the drugs that were approved between 1981 and 2014, over 45% were natural products, natural products derivatives, or synthetic drugs that were inspired by natural products (Newman and Cragg 2016).

Information regarding traditional medicine has been collected over centuries from regions such as Sumer, Egypt, and China, and this has led to the use of medicinal plants to treat a wide range of diseases (Cragg et al. 1997). Relatively easy access for collection and the wealth of ethnopharmacological information enabled the development of a great number of drugs from plants during the 19th century. Since the beginning of the 20th century there was a surge in the development of modern biotechnology and advancement of scuba techniques, allowing for other natural sources such as microorganisms and marine organisms to present themselves as a rich alternative source of new bioactive compounds with very different metabolic pools to those found in plants (Blunt et al. 2015, 2017, 2018).

Marine organisms have opened a new era for the discovery of a novel pool of molecules that could be used as drug candidates. Interestingly, many compounds isolated from marine organisms have shown very distinctive chemical characteristics, which could consequently provide a more productive source of chemical entities. For instance, some of these compounds have presented with unique structural features, such as the inclusion of bromine and chlorine atoms by a covalent bond, polyketide compounds, and nonribosomal peptides (Sashidhara et al. 2009; Villa and Gerwick 2010). This makes marine organisms a fascinating source of molecules with a potential to be used as drugs. Nowadays, thousands of new marine natural products with diverse bioactivity are being reported every year. Recently, the FDA approved seven drugs from marine sources for the treatment of different diseases (Gerwick and Moore 2012). However, despite the potential and vast level of chemical diversity found in marine organisms, the number of registered new drugs appears to remain low.

The difficulties for the development of new drugs come from the low natural abundance of marine organisms in the wild and restrictions in sample collection. Additionally, marine organisms carry an extremely low concentration of active compounds, consequently requiring large amounts of samples. For example, in the case of sponges, it is estimated that the active compounds correspond to less than 0.4% of the dry weight of the organism (Belarbi et al. 2003). Moreover, the effect of environmental and ecological factors on the production of active compounds is still unclear (Paul and Puglisi 2004). To mitigate the adverse effects of these issues, every step during the process of investigating marine organisms should be carried out carefully to minimize the loss of compounds. For these reasons, meticulous attention is required for appropriate sample preparation. Preanalytical processing, including extraction, is a particularly important step in marine organisms. Unlike terrestrial samples, marine samples usually contain more inorganic salts and fatty materials due to their environmental conditions. Additionally, the places from which samples are collected are often remote, which leads to long delays between collection and extraction in the laboratory. Therefore, a robust protocol is required to avoid interferences caused by salt, sample storage, and transportation during the sample preparation.

Following sample collection, extraction is performed in order to dissociate metabolites from the matrix. When extracting natural products, it is not only essential to obtain a high yield of targeted specific metabolites but also a qualitative feature: the number of metabolites to be extracted. Particularly in the case of marine organisms; managing to isolate the large variety of metabolites that are present within the matrix in small quantities is of high importance during the extraction process. This has encouraged increased applications of metabolomics studies to marine organisms as part of the recent trends of life science research. In these holistic approaches, it is important to extract a wide range of compounds to fulfill the aim of acquiring a comprehensive overview of the chemical composition of an organism (Viant 2007). Previously, many extraction methods have been suggested for metabolomics study as was reviewed by Mushtaq and colleagues (Mushtaq et al. 2014). Additionally, as an alternative to the conventional extraction methods, a comprehensive extraction method using multiple solvents was suggested (Yuliana et al. 2011). However, no methods have been described to target marine organisms specifically.

During extraction, several parameters can influence the efficiency of the process such as solvent polarity, temperature, pressure, the ratio between the samples and solvents, as well as mechanical assistance. The most suitable choice of solvent is crucial for effective extraction, and the selection of the wrong extraction conditions could have a dramatic effect on the

extraction yield, e.g., the yield of the extraction of sulfur compounds could be diminished by 90% depending on the chosen solvent (Sass-Kiss et al. 1999). The relationship of other factors such as temperature and pressure with the extraction efficiency have been previously ignored, despite their importance. Although it has been mainly used in terrestrial plants, pressurized extraction can increase the efficiency of extraction. This system combines the use of liquid solvents at elevated temperature and pressure, which could allow to reduce extraction time and maximize the contact between sample and solvent and consequently results in improving the extraction efficiency (Ju and Howard 2003; Lee and Lin 2007). Furthermore, protocols for the investigation of marine natural products require some preanalytical steps in order to ensure the reproducibility of the results. These steps include desalting of the crude extracts using solid phase extraction (SPE) cartridges as well as the removal of lipid contents using liquid–liquid partition (Bauvais et al. 2017; Farag et al. 2016; Ivanišević et al. 2011).

In this study, several extraction parameters were investigated on marine organisms, particularly the effect on the metabolic diversity of the extract. A marine sponge was selected as a model organism, as sponges (Porifera) are known to have the highest number of newly reported compounds amongst marine organisms, illustrating the tremendous chemical diversity that this phylum contains (Blunt et al. 2017, 2018). Amongst the sponges, *Xestospongia* is an excellent example from this phylum, as over 350 different compounds have been reported in this sponge, including alkaloids, terpenoids, quinones, steroids, and brominated fatty acids (Zhou et al. 2010). Additionally, the wide distribution of *Xestospongia* in tropical oceans around the world, including the Indo-Pacific and Atlantic oceans, makes it an exciting organism for further studies of its chemical diversity (Swierts et al. 2017).

In this research, several parameters were examined, including the temperature, pressure, number of cycles, and solvent polarity using a pressurized extraction system. A design of experiments (DOE) of a full factorial design was used, and the diversity of chemicals measured was selected as the response variable. For the measurement of chemical diversity, proton nuclear magnetic resonance (^{1}H-NMR) was applied to the extracts and the diversity index was calculated. Based on the results, the most influential parameters and the interaction between the tested parameters were established. For the next approach to investigate the specific effect on the individual groups of metabolites, two of the most influential parameters were selected, and the chemical data obtained by ^{1}H-NMR were correlated with the selected factors by a multivariate data analysis; orthogonal partial least square (OPLS) modeling.

2. Results and Discussion

To study the correlation between the extraction parameters and metabolic diversity, a pressurized extraction system was employed by controlling the following parameters: pressure, temperature, solvent polarity, and cycle number. Of the large number of parameters involved in the extraction process, these four parameters were chosen to evaluate their effect on the chemical diversity of the extract. To identify the most influential factor and the interactions between factors on the resulted chemical diversity, a full factorial design (2^4) was applied to the chemical data.

Solvent polarity was selected as a parameter as it is usually the most influential factor on the yield of the extraction. From the numerous available solvents that are used for the extraction of marine organisms, e.g., *n*-hexane, dichloromethane, ethyl acetate, methanol, ethanol, isopropanol and other mixtures of solvents, ethanol, and dichloromethane were selected, and their ratio was used as variables in DOE (Wright 1998). This allows for an examination of two of the most used solvents in marine organism extraction.

Extraction temperature is well-known for increasing the rate of the diffusion of metabolites from the matrix into the solvent. Here, the low and high levels were set to 30 °C and 70 °C, respectively. The low temperature was chosen as it is close to room temperature and many extractions are performed at this temperature. The high temperature was chosen as it is close to the boiling point of ethanol at atmospheric pressure. Additionally, the speed extraction system uses nitrogen as a gas to increase the pressure. For this reason, the temperature can be increased without the risk of degradation of the sample by oxidation. With regards to the pressure parameter, the use of high pressures during extraction is expected to reduce the extraction time as high pressure forces the solvent into the matrix (Kaufmann and Christen 2002). In this study, the pressure was set at 50 and 100 bars for the low and high levels, respectively. These values were selected based on the technical characteristics of the machine; the lowest working pressure for the extraction system is 50 bar. Furthermore, the high level was set to 100 bars as some reports state the successful use of a pressurized extraction system at this pressure level (Kaufmann and Christen 2002; Wright 1998).

As the last parameter, the number of cycles for the extraction was evaluated. This parameter is a measurement of the ratio between the solvent and sample that should be used to perform the extraction. The metabolites that have a higher affinity for the solvent can saturate the solvent, which prohibits the extraction of other compounds and potentially lowers the diversity of metabolites that can be extracted. In this case, the number of cycles was set to

one or three cycles for the low and high settings, respectively. After each extraction cycle, fresh solvent was added to the sample.

2.1. Chemical Diversity Index

One of the most vital necessities in the study of natural products, including those from marine sources, is providing a wide range of chemicals. Particularly in drug discovery, diversity is the critical issue in a chemical pool. To explore the chemical diversity of natural products, chemical profiling tools like metabolomics are being actively applied to many marine organisms. The quality of metabolomics results is dependent on extraction method. To produce a broad range of metabolites, the extraction process must cover as many metabolites as possible in the target organisms.

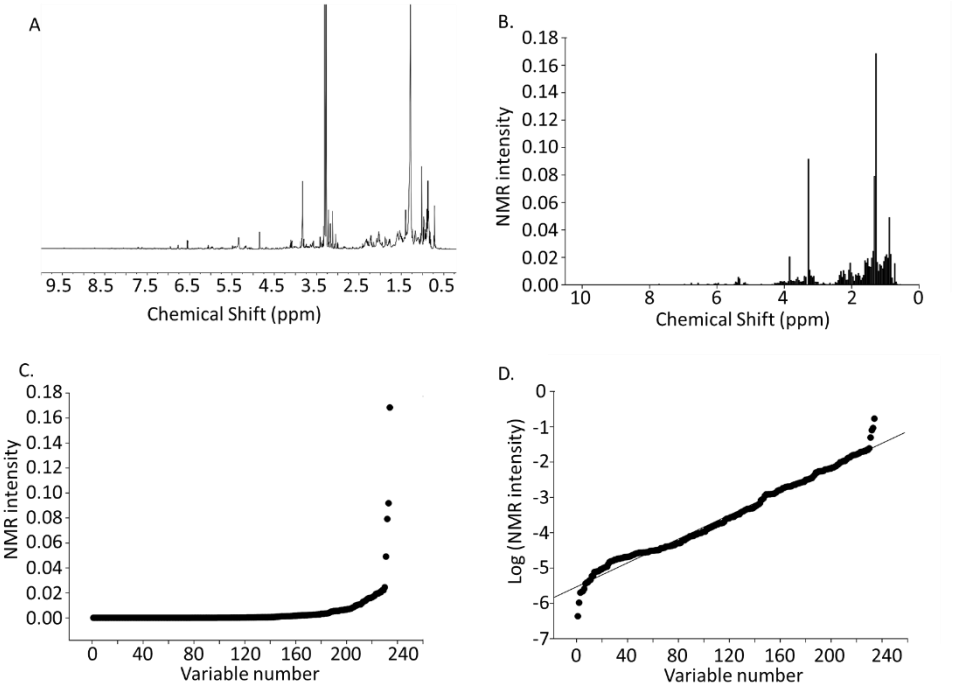


Figure 3.1: Schematic of the data processing for the calculation of the chemical diversity index. A: NMR-processed spectra, B: NMR spectra after bucketing, C: plot of the NMR intensity against the variable number, and D: the plot of the logarithm of the ^{1}H-NMR intensity against the variable number.

To quantify the diverse array of extracted chemicals, an index was developed in this study based on the ^{1}H-NMR analysis of the samples. ^{1}H-NMR can detect various groups of metabolites and all the signals are highly proportional to the molar concentration of the

compounds, which could make it possible to quantitatively compare the signals within the same sample. Therefore, the change in the extraction efficiency of the diverse groups of metabolites in different extraction conditions can be clearly viewed (Kim et al. 2010).

For the calculation of this index (as exemplified in Figure 3.1), firstly, the ^{1}H-NMR spectra (Figure 3.1a) was integrated or bucketed in a certain range, e.g., 0.04 ppm (Figure 3.1b) and the integrated signals were sorted from the highest to the lowest (Figure 3.1c). Some signals, however, were shown to have an extremely high intensity which resulted in an exponential curve (Figure 3.1c). This happened because a few signals, which mainly corresponded to fatty acids, were very high in comparison to others, such as the aromatic region. To solve the problems of the signals, the ^{1}H-NMR data were transformed logarithmically (Figure 3.1d) and the transformed signal intensities fit the 1st order line. The inverse of the 1st order fitting curve was referred to as the chemical diversity index. A higher value of the slope indicates that a specific group of metabolites was extracted preferentially, whilst a slope close to zero indicates that all the metabolites are extracted with the same efficiency. For further evaluation of the extraction factors, the calculated chemical diversity index was used as the response variable in the DOE analysis.

2.2. Evaluation of Extraction Parameters Using A Design of Experiments

To investigate the effects of solvent polarity, temperature, pressure, and the number of cycles (factors) on the resulted chemical diversity (response), a pure screening DOE was applied to the data set. The response (chemical diversity index) was modeled using a constant (β_0), four coefficients of individual parameters (β_0–β_4), and six coefficients of interaction parameters (β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, and β_{34}). As shown in Table 3.1 (the significant values are marked with *), three factors and two interactions between the factors were found to significantly affect the extraction process. Firstly, the number of cycles factor displayed a positive correlation with the extraction process, indicating that more cycles allow for the extraction of more diverse metabolites in the extract. This might be due to the fact that metabolites with low solubility could gradually be extracted by the newly added solvent during each cycle.

The second factor of interest was the polarity of the solvent in the system of ethanol and dichloromethane. An increase in the solvent polarity was found to correlate positively with the diversity of the extracted compounds. It shows that ethanol may be a more suitable solvent for the extraction of a wide range of compounds than mixtures of ethanol and dichloromethane. Although a large percentage of sponge metabolites are lipophilic fatty chemicals, ethanol may be more efficient to extract polar secondary metabolites, which

resulted in higher number of extracted metabolites. The final factors that were selected were temperature and pressure. Interestingly, a higher temperature of the extraction procedure resulted in a decrease of the chemical diversity of the extract. On the other hand, in the case of pressure, it was not found to be significantly related to the extracted chemical diversity of the extraction. In general, the physical properties of liquids are not affected by pressure as much as gas, including solvent power.

Table 3.1. Analysis of the coefficient values of the model using the chemical diversity index as the response variable.

Model Parameter	Regression Coefficients	Coefficient Values
Constant	β_0	2.68×10^{1}
Temperature	β_1	$-2.83 \times 10^{-1*}$
Pressure	β_2	-2.79×10^{-4}
Cycles	β_3	$5.06 \times 10^{-1*}$
Solvent polarity	β_4	$3.27 \times 10^{-1*}$
Temperature-Pressure	β_{12}	1.91×10^{-1}
Temperature-Cycles	β_{13}	$3.60 \times 10^{-1*}$
Temperature-Solvent polarity	β_{14}	1.11×10^{-1}
Pressure-Cycles	β_{23}	$-2.28 \times 10^{-1*}$
Pressure-Solvent polarity	β_{24}	-1.94×10^{-2}
Cycles-Solvent polarity	β_{34}	-2.92×10^{-2}
R^2		6.70×10^{-1}
R^2 adjusted		5.81×10^{-1}
p value		2.20×10^{-6}
Lack of fit		5.72×10^{-2}

* significant at $p < 0.05$.

The interaction between the number of cycles and the temperature in the extraction process also has a significant effect on the outcome of chemical diversity. At low temperature, the extract displayed almost no difference between the chemical diversity depending on the number of cycles. However, at higher temperatures, the diversity decreased dramatically if one cycle was used for extraction. This may be related to the penetration efficiency of the solvent through the matrix. At higher temperatures, lipids from the matrix are extracted more efficiently when compared to an extraction at lower temperatures. Due to this effect, other metabolites are not extracted, and the chemical diversity is reduced. Therefore, at higher temperatures more cycles are required to achieve a similar chemical diversity in the extraction.

A similar effect was found in the interaction between the number of cycles and the pressure of the extraction. At lower pressures, the chemical diversity of the extract was smaller if one cycle was used for the extraction. However, at higher pressures one cycle displayed a higher chemical diversity whilst three cycles showed a decrease in the diversity.

For the next steps, solvent polarity and temperature were selected for further investigation with regards to which groups of metabolites were more influenced by these selected parameters.

2.3. Effect of the Solvent Polarity and Temperature on An Individual Group of Metabolites

To investigate the effect of solvent polarity and temperature in detail, a supervised multivariate data analysis (MVDA), OPLS modeling, was applied to the ^{1}H-NMR data using the percentage of ethanol and temperature as Y-variables. Figure 3.2 shows the effect of the polarity of the extraction solvent on the resulting metabolites, for which the solvents with different ethanol percentages were used at the same temperature (30 °C). In the score plot, clear differences between the samples with different percentages of ethanol were found (Figure 3.2a), and the OPLS model was validated as Q^2 was 0.984 and the cross- validation analysis of variance (CV-ANOVA) test had a p-value < 0.05. The results show that there is a significant change in the chemical composition of the extract, depending on the solvent polarity that is employed.

To analyze the correlation between the signals and the separation according to the solvent composition observed in the OPLS modeling, an S-plot was made (Figure 3.2b). For the extracts that were prepared using higher ratios of dichloromethane, the signals between δ 0.72 and δ 2.24 displayed a higher intensity. Signals in this range correspond to fatty acids and sterols which have been widely reported in *Xestospongia* (Gauvin et al. 2004; Liu et al. 2011; Nguyen et al. 2013)[26–28]. These compounds are very nonpolar and are therefore expected to be extracted more efficiently by dichloromethane when compared with ethanol. In the case of a higher percentage of ethanol for the extract, the intensity of signals between δ 2.5 and δ 4.5 was higher than other signals. Many of the signals within the range may be related to functional groups such as methyl groups bound to oxygen or nitrogen atoms present in alkylpyridine alkaloids and isoquinoline quinones, as previously reported in *Xestospongia* (Amnuoypol et al. 2004; Sakemi et al. 1990).

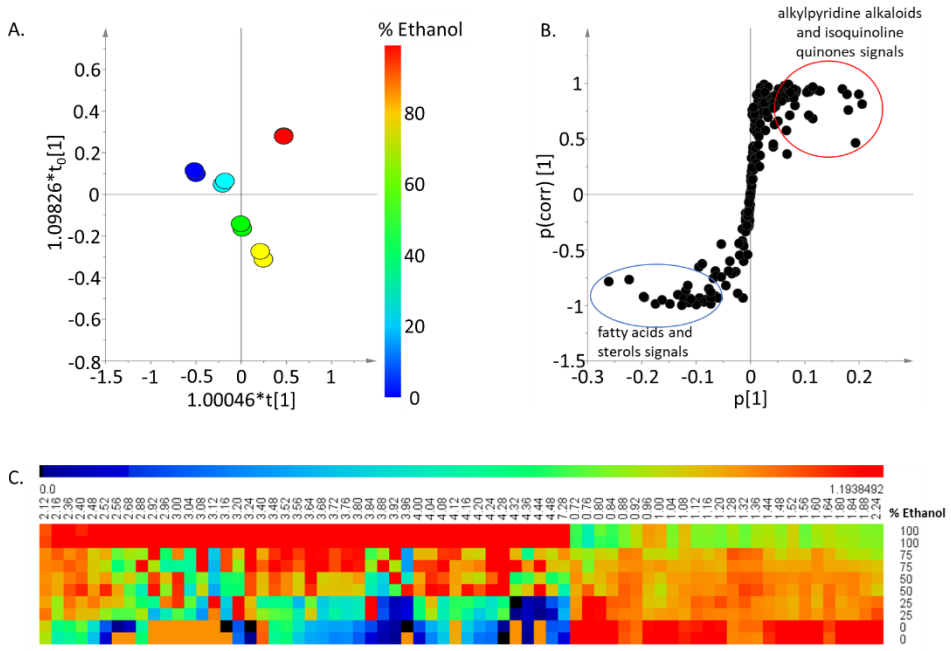


Figure 3.2: A: Orthogonal partial least square modeling using ^{1}H-NMR data of *Xestospongia* extracts combined with solvent polarity (the percentage of ethanol) score plot, B: S-plot, and C: heat map of the relative intensities of the discriminant signal. The extractions were performed at 30 °C. Red circle: Metabolites with higher polarity. Blue circle: metabolites with lower polarity.

The overview of metabolic profile obtained by the solvents at different polarities is shown in the form of a heat map (Figure 3.2c). In this map, signals of a higher chemical shift which were related to an extraction solvent of high polarity displayed intensities that were very close to zero in a low polarity extraction solvent. This indicates that the compounds with high polarity were not extracted at all using solvent mixtures with a high content of dichloromethane. Moreover, signals at lower chemical shifts related to a lower polarity of the extraction solvent are also extracted using solvent mixtures with a higher content of ethanol, but they are extracted less efficiently when compared to the dichloromethane extracts. This explains why the increase in the solvent polarity had a positive effect on the chemical diversity of the extract. Although ethanol has higher affinity for polar compounds, it is also able to extract compounds with very low polarity from the matrix.

2.4. Effect of temperature

Temperature was the second selected factor for the detailed study on the relationship with the individual groups of metabolites. For the temperature effects, 100% ethanol was used as

the extracting solvent and temperatures vary between 30 and 80 °C in 10 °C steps. The score plot of the OPLS modeling showed that the metabolites are greatly influenced by temperature; the chemical composition of extractions between 30 and 50 °C were grouped together as the chemical composition was similar, whilst the extraction between 60 and 80 °C were grouped into a second group (Figure 3.3a). The model was validated, as Q^2 was 0.953 and the CV-ANOVA test had a p-value < 0.05.

To identify the signals that were correlated within each group of temperatures, an S-plot was constructed (Figure 3.3b). For the low temperatures, signals between 0.96 and 3.92 ppm were distinguished as significant. Similarly, in the case of extractions at high temperature, signals in the same region as well as few signals in the aromatic region were correlated. This indicates that the differences in the extraction caused by the temperature do not correspond to a family of compounds as the region of the significant signals is almost the same. Instead, the difference is caused by specific compounds that are extracted differently depending on the temperature.

Lastly, to compare the relative intensities of the discriminant signals, a heat map was produced. Here, it is possible to observe the clear difference between the low temperature (below 50 °C) and high temperature (above 60 °C) extractions. Unlike the solvents, a preferential extraction is not clearly observed. This selectivity of extraction in the compounds depending on the temperature could be related to the increase of the chemical diversity at low temperatures as showed in the first experiment. At low temperatures, all the compounds are extracted with similar efficiency whilst at higher temperatures, a specific group is preferred. In that sense, for analysis that requires a holistic overview of the chemical composition of the samples, lower temperatures for extraction should be preferred.

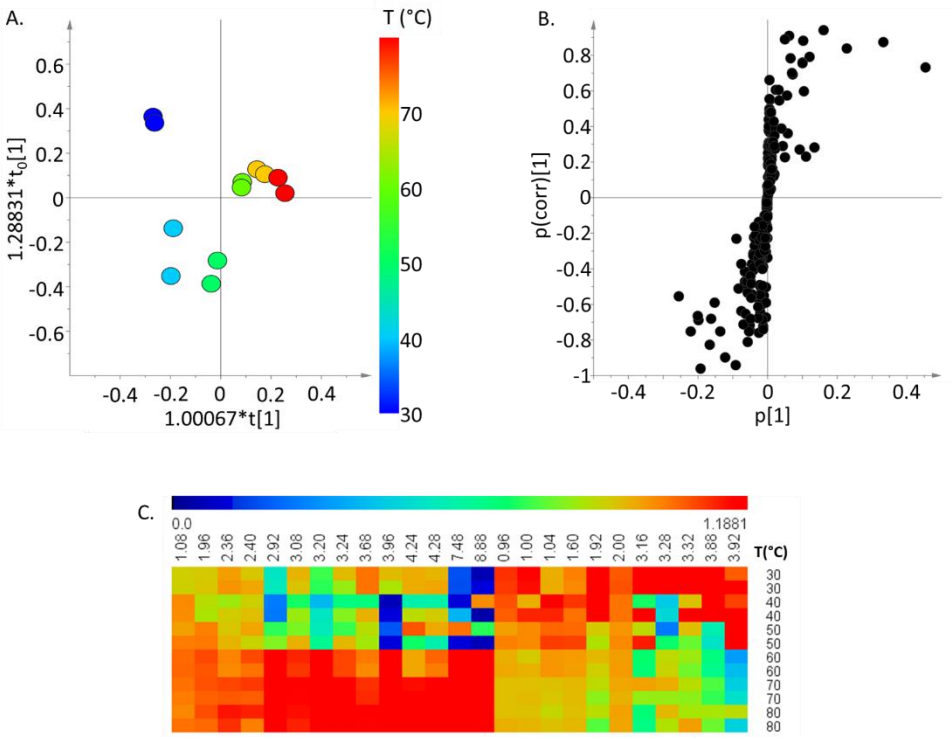


Figure 3.3: Orthogonal partial least square modeling using ^{1}H-NMR data of *Xestospongia* ethanol extracts varying the temperature (A: score plot, B: S-plot) and heat map of the relative intensities of the discriminant signal in C.

3. Conclusions

The extraction process is a crucial step for further detection of metabolites in any natural products study, particularly for metabolomics approaches. The chemical diversity index is a novel tool to quantify the chemical diversity of crude extracts and was successfully used to study the effects of different parameters during the extraction process. From the parameters studied, the number of cycles, solvent polarity, and temperature significantly affected the chemical diversity of the obtained extract. Additionally, the interaction of temperature with the number of cycles and number of cycles with pressure also have a positive and negative effect on the extract, respectively. Lastly, with the second group of experiments, it was possible to confirm that 100% ethanol together with low temperatures were the best conditions to perform the extraction.

4. Materials and Methods

4.1. Sponge Collection

Samples of *Xestospongia* sp. were collected from the inner coral reef in Martinique, in September 2016. The samples were preserved in ethanol for the transportation and were stored at −20 °C. The specimens were identified by Nicole de Voogd (National Museum of Natural History, The Netherlands). The ethanol used for storage was removed by filtration. The samples were frozen under liquid nitrogen and freeze-dried using a Labconco FreeZone 4.5 plus freeze dryer. The dried samples were ground and pooled into a single sample batch (17.9 g dried weight). The same samples were used for all the experiments in the study.

4.2. Experimental Design

Table 3.2. Experimental design for the extraction of *Xestospongia* using pressurized extraction system.

Number of Experiments	X1 (Temperature)	X2 (Pressure)	X3 (Number of Cycles)	X4 (Solvent)
1	−	−	−	−
2	+	−	−	−
3	−	+	−	−
4	+	+	−	−
5	−	−	+	−
6	+	−	+	−
7	−	+	+	−
8	+	+	+	−
9	−	−	−	+
10	+	−	−	+
11	−	+	−	+
12	+	+	−	+
13	−	−	+	+
14	+	−	+	+
15	−	+	+	+
16	+	+	+	+

In this study, a full factorial design with n = 4 was used to identify if the pressure, temperature, number of cycles, and solvent polarity influence the chemical diversity of the extract. Two levels were defined for each variable, − and +. The temperature was changed between 30 and 70 °C and the pressure was changed between 50 bar and 100 bar. Number of cycles was changed between 1 and 3. The solvent was changed between dichloromethane/ethanol

1:1(v/v) and 100% ethanol. The detailed conditions of the experiments are shown in Table 3.2. The experiments were carried out in triplicate.

For the second part of experiments conditions of temperature and solvent polarity varied after pressure was set at 90 bar and the number of cycles was three. For the temperature evaluation, 100% ethanol was used with temperature variation of 30, 40, 50, 60, 70, and 80 °C. For the solvent polarity evaluation, the temperature was set at 30 °C and the solvent proportion of dichloromethane/ethanol was changed between 100% dichloromethane, 75:25 (v/v), 1:1 (v/v), 25:75 (v/v), and 100% ethanol. All the experiments were carried out by duplicate.

4.3. Pressurized Solvent Extraction

Pressurized solvent extraction was performed using a Speed Extractor E-916 (BÜCHI Labortechnik AG, Flawil, Switzerland). Six stainless steel extraction cells (10 mL) were used at a time, placing two circular glass fiber filters on both ends and a metal frit at the bottom. The cells were packed with 9 g of Quartz sand (diameter: 0.3–0.9 mm, BÜCHI Flawil, Switzerland), mixed with the ground samples (150 mg and 200 mg for 1st and 2nd experiments, respectively).

4.4. ^{1}H-NMR Analysis

From the extract obtained from the Speed Extractor system, 1.5 mL was taken to dryness. The residue was redisolved in CH_3OH-d_4 with hexamethyldisiloxane (HMDSO) as the internal standard. The ^{1}H-NMR spectra were measured at 25 °C in an AV-600 MHz NMR spectrometer (Bruker, Karlsruhe, Germany), operating at the ^{1}H-NMR frequency of 600.13 MHz, and equipped with a TCI cryoprobe and Z gradient system. For internal locking, CH_3OH-d_4 was used. A presaturation sequence was used to suppress the residual water signal, using low power selective irradiation at the H_2O frequency during the recycle delay.

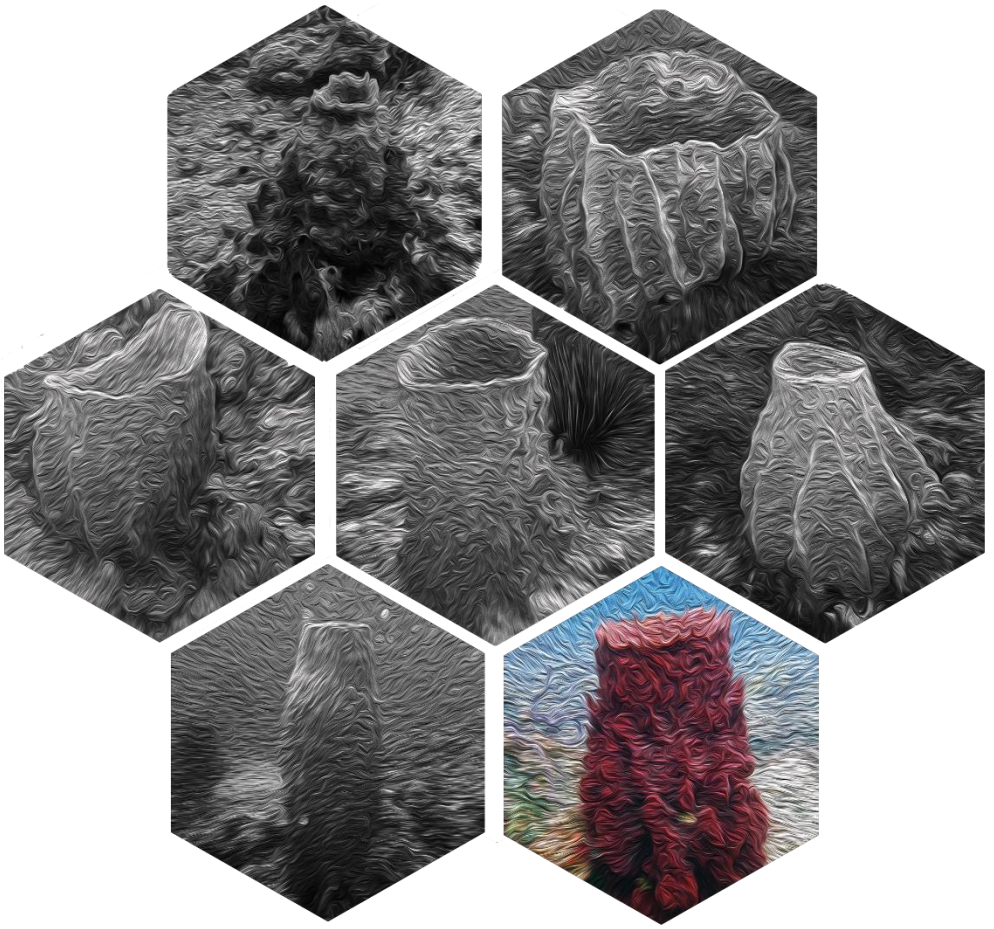
4.5. Data Prepossessing and Statistical Analysis

The resulting spectra were phased, baseline corrected, and calibrated to HMDSO at 0.06 ppm using TOPSPIN V. 3.0 (Bruker Karlsruhe, Germany). The NMR spectra were bucketed using AMIX 3.9.12 (Bruker BioSpin GmbH, Rheinstetten, Germany). Bucket data were obtained by spectra integration at every 0.04 ppm interval from 0.20 to 10.02 ppm. The peak intensity of individual peaks was scaled to the total intensity of the buckets. The regions between 3.32 to 3.28, 4.9 to 4.8, 3.62 to 3.57, and 1.15 to 1.19 ppm were excluded from the analysis because they correspond to solvent residual signals.

The chemical diversity index was calculated using the buckets from the ^{1}H-NMR spectra. The buckets were organized from high to low, and the logarithm of the bucket was plotted against the order number. These plots are straight lines and the inverse of the slope of these lines corresponds to the chemical diversity index that was used as the response variable in the full factorial design. The statistical analysis was performed using MODDE software (v 12.0.1, Sartorius stedim, Goettingen, Germany).

To analyze the effect of changing the value of one separate factor at the time (temperature and solvent) the buckets that were obtained from the NMR spectra were organized in a matrix and multivariate data analysis was performed using SIMCA-P software (v.15.0.2, Sartorius stedim, Goettingen, Germany). Principal component analysis PCA and orthogonal projections to latent structures OPLS were performed.

Acknowledgments: The authors greatly appreciate the help of Nicole de Voogd and Esther van der Ent (Naturalis Biodiversity Center, The Netherlands) for the collection and identification of *Xestospongia* specimens. The specimens were collected during the Madibenthos expeditions (PI Philippe Bouchet) organised by the Museum National d'Histoire Naturelle (MNHN) and the Marine Protected Areas Agency (AAMP), the Regional Directorate for the Environment (DEAL) and the Martinique Water Bureau (ODE) with funding from the European Regional Development Fund (ERDF), the Territorial Collectivity of Martinique (CTM) and Saint-James Plantations and BRED Banque populaire. This work was supported by the COLCIENCIAS (science technology and innovation ministry, Colombian government)



Chapter 4

Influence of the geographical location on the metabolic production of giant barrel sponges (*Xestospongia* spp.) revealed by metabolomics tools

Lina M. Bayona [1], Gemma van Leeuwen [1], Özlem Erol [1], Thomas Swierts [2,3], Esther van der Ent [2,3] Nicole J. de Voogd [2,3], Young Hae Choi [1*]

[1]Natural Products Laboratory, Institute of Biology, Leiden University, Leiden, The Netherlands.

[2]Marine Biodiversity, Naturalis Biodiversity Center, Leiden, The Netherlands

[3] Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands

* Corresponding author e-mail: y.choi@chem.leidenuniv.nl

ACS Omega (2020), 5(21), 12398–12408.

https://doi.org/10.1021/acsomega.0c01151

Abstract

Despite their high therapeutic potential, only a limited number of approved drugs originate from marine natural products. A possible reason for this is their broad metabolic variability related to the environment, which can cause reproducibility issues. Consequently, a further understanding of environmental factors influencing the production of metabolites is required. Giant barrel sponges, *Xestospongia* spp.*,* are a source of many new compounds and are found in a broad geographical range. In this study, the relationship between the metabolome and the geographical location of sponges within the genus *Xestospongia* spp. was investigated. One hundred and thirty-nine specimens of giant barrel sponges (*Xestospongia* spp.) collected in four locations, Martinique, Curaçao, Taiwan, and Tanzania, were studied using a multiplatform metabolomics methodology (NMR and LC-MS). A clear grouping of the collected samples according to their location was shown. Metabolomics analysis revealed that sterols and various fatty acids, including polyoxygenated and brominated derivatives, were related to the difference in location. To explore the relationship between observed metabolic changes and their bioactivity, antibacterial activity was assessed against *Escherichia coli* and *Staphylococcus aureus.* The activity was found to correlate with brominated fatty acids. These were isolated and identified as (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**), Xestospongic acid (**2**), (7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid (**3**) and two previously unreported compounds.

Keywords: Metabolomics, antibacterial, marine sponge, giant barrel sponge, geographic location, brominated fatty acids, *Xestospongia* spp.

1. Introduction

Marine natural products (MNP) have a wide chemical diversity, covering a broader area in the chemical spectrum compared to their terrestrial counterparts (Blunt et al. 2018). The chemical structures of metabolites isolated from marine organisms contain highly characteristic features and many of them have shown diverse bioactivities. In the past decades, the isolation of novel and bioactive molecules from marine organisms has been a hot issue in natural product research resulting, so far, in the development of eight drugs, which have been approved and are currently available for the treatment of cancer, HIV and pain(Altmann 2017; Gerwick and Moore 2012; Newman and Cragg 2004). Despite their potential, the number of approved drugs is low considering the large number of compounds that have been discovered from marine sources. In fact, while more than 1200 new compounds are reported every year, the number of MNP-derived approved drugs has not been increasing at the same rate (Blunt et al. 2016, 2017, 2018).

Although many of the metabolites produced by marine organisms have proved to be active, these compounds are usually produced in very small amounts (Belarbi et al. 2003). During the process of drug development, large quantities of the compound are required to perform all the preclinical and clinical trials that are necessary for a drug to be approved (Gupta 2011). Unfortunately, the large scale harvesting of the organism required for this is not feasible either from an economical or an ecological perspective (Altmann 2017). Moreover, the production of metabolites in marine organisms can change due to environmental factors such as pH, temperature, predation pressure and subsequent changes in symbionts community, making them too unreliable both qualitatively and quantitatively as a natural source of compounds (Viant 2007).

To overcome this, diverse approaches have been suggested, including aqua- and mariculture (Belarbi et al. 2003; Pomponi 1999). Although these techniques have not been used yet for the production of compounds at a commercial scale, it is thought that their implementation could provide sufficient amounts of the compounds to meet the demand for clinical and preclinical trials (Cuevas and Francesch 2009). The successful cultivation of marine organisms, mainly of sponges (de Voogd 2007; Ruiz et al. 2013; Santiago et al. 2019), resulting in the production of higher quantities of active metabolites (Hadas et al. 2005; Page et al. 2005), could guarantee the reliability of the sources, paving the way for their approval for medicinal use. Optimization of growth and production conditions for the cultivation of the organisms requires an understanding of how biotic and environmental factors affect their metabolome. Such a study involving so many variables can benefit from the use of an untargeted approach

that allows the acquisition of the most inclusive picture of the metabolome and then observes how it varies with changing external factors. Metabolomics, defined as comprehensive profiling of all the metabolites produced by an organism, cell, or tissue at a certain point in time can provide the information, which could then be used for guidance on a variety of compounds produced and uncovering the factors associated with their production (Kim et al. 2010).

Among marine organisms, sponges have been considered to be the most prolific in the production of secondary metabolites, most of which have biological activity as proved by their performance in a wide variety of bioassays (Belarbi et al. 2003; Blunt et al. 2018; Mehbub et al. 2014). In particular, giant barrel sponges, which belong to the genus *Xestospongia*, have drawn the attention of the scientific community due to their pharmacological activities and their role in ecosystems (Fiore et al. 2013; Zhou et al. 2010). In ecological systems, their large size allows them to play an essential role in the reef, providing habitat for other organisms and filtering vast amounts of seawater (Diaz and Rützler 2001; Swierts et al. 2018). Therefore, the tight interaction of giant barrel sponges with their environment makes them an interesting model to study the relationship between metabolites and environmental factors. Also, in some locations these sponges have been reported to cover up to 9% of the reef substrate, being more abundant than any other invertebrate (McMurray et al. 2008; Zea 1993). Their chemical composition has been studied, and a wide range of compounds have been isolated including alkaloids, brominated fatty acids, and sterols. Many of these compounds have proved to be bioactive, displaying antibacterial, cytotoxicity, fungicide, and antiretroviral activities (Zhou et al. 2010).

In addition, giant barrel sponges can be found in a wide geographical range: *Xestospongia testudinaria* from the Red Sea to the Indo-Pacific Ocean and Australia, and *Xestospongia muta* in the tropical regions of the Atlantic Ocean. These two species show very similar genetic and morphological markers (Setiawan et al. 2016). Furthermore, recent studies revealed the presence of cryptic species in both ocean basins (Swierts et al. 2017). Interestingly for this study, some of the species present in the Caribbean Sea are genetically much closer to species in the Indo-Pacific than to other species in the same location (Swierts et al. 2013, 2017). These similarities in the cryptic species between locations provides the opportunity to focus on the differences in the metabolome caused by environmental factors.

Geographical location has been identified as one of the most influential factors related to the variation of many sponge metabolites (Page et al. 2005; Rohde et al. 2012; Sacristan-Soriano et al. 2011). However, the results that led to this conclusion were aimed at a few target

metabolites, while the more general effect on the whole metabolome, which requires a holistic approach, has scarcely been studied (Reverter et al. 2018). To study the correlation between the geographical location and metabolic production, 139 specimens of giant barrel sponges (*Xestospongia* spp.), collected in four different geographic locations: Martinique, Curaçao, Taiwan, and Tanzania were studied using a holistic approach. Applying multiplatform metabolomics methodology (nuclear magnetic resonance spectroscopy (NMR) and liquid chromatography-mass spectrometry (LC-MS)), we aimed to investigate the effect of geographical location on the chemical composition of the sponges. Additionally, the correlation between the metabolic changes observed in the samples and their antibacterial activity was evaluated. This proved that the implementation of a metabolomics approach to MNPs can provide relevant information on the conditions required to optimize the production of bioactive compounds. Furthermore, the presence of minor active compounds largely influenced by location-related factors can be revealed using this approach.

2. Results and discussion

The metabolic profile of giant barrel sponge samples collected in four different geographical locations showed clear differences in the chemical composition of the specimens collected in each location. To compare the general metabolic profile of the samples, ^{1}H-NMR and LC-MS were separately applied to the same sample set. These data were further analyzed using an orthogonal partial least-squares discriminant analysis (OPLS-DA) model (Figure 4.1). Both models, ^{1}H-NMR and LC-MS, were validated with a Q^2 value > 0.4 and cross-validation analysis of variance (CV-ANOVA) test $p < 0.05$ (Cai et al. 2012; Zheng et al. 2011).

In fact, for giant barrel sponges *X. muta* and *X. testudinaria*, the composition of sterols (Gauvin et al. 2004) and some brominated fatty acids (Zhou et al. 2010) was previously found to be similar between sponges collected in different oceans. These previous studies showed that despite large geographical separation, giant barrel sponges could share a common metabolic background in qualitative features. In this study, however, a significant separation between the samples collected from different places was observed in the OPLS-DA analysis (Figure 4.1). This result might indicate that the environmental conditions in each location could quantitatively influence the metabolome of the sponges.

The location in which sponges grow involves a number of factors that can influence their development and metabolism, including abiotic factors such as temperature, pH, salinity, or the biotic predatory stress. The effect of the combination of these factors could cause that sponges collected from a specific location produce similar metabolites. Furthermore,

Xestospongia spp. are high microbial abundance (HMA) sponges, and microbial communities have been reported to mainly be affected by geographical location (Swierts et al. 2018). Thus, it is plausible to find differences in the chemical composition of sponges from different locations, as the metabolome corresponds to the holobiont and the metabolites found can either be produced by the sponge, by the microorganisms or they can be the product of the interaction of the sponges with microorganisms (Gerwick and Moore 2012).

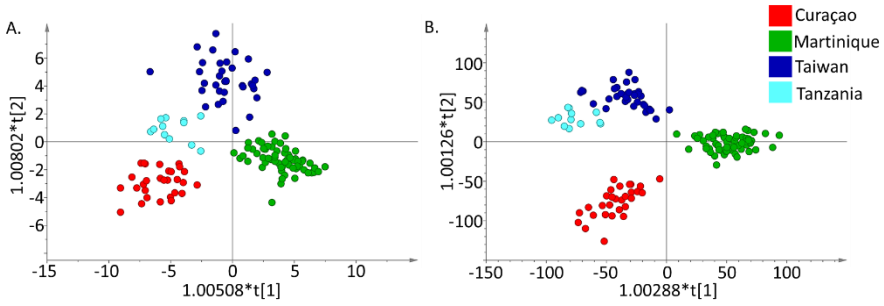


Figure 4.1: First two components of the OPLS-DA analysis based on ^{1}H-NMR (A) and LC-MS (B) of *Xestospongia* spp. samples collected in four locations: Curaçao (Red), Martinique (Green), Taiwan (Dark blue) and Tanzania (Light blue).

The loading plots of the OPLS-DA analysis (NMR and LC-MS data) were analyzed to select the discriminating signals and subsequently identify the corresponding compounds. The characteristic ^{1}H-NMR chemical shifts are shown in a heat map in Figure 4.2, obtained by calculation of the variable importance for the projection (VIP) values. The signals correlated with the samples from Martinique were found mainly in two regions of the spectra. The region between δ_H 0.80 and 1.00 was assigned to methyl groups in sterols. Particularly the singlets in the range of δ_H 0.7-0.8 were assigned to methyls H-18 and H-19 in sterols. Many steroids have been reported in *Xestospongia* spp., including conventional sterols (Kerr et al. 1991), and brominated fatty acids esters (Pham et al. 1999). The aromatic region between δ_H 7.04 and δ_H 7.32 is characteristic of phenolic signals that could correspond to known phenolics of *Xestospongia* such as quinones (Roll et al. 1983), isoquinoline alkaloids (Calcul et al. 2003), and β-carboline alkaloids (Kobayashi et al. 1995). Samples from Curaçao were distinguished by abundant signals in the range of δ_H 2.50-3.80. Signals at downfield of this range (δ_H.3-3.8) correspond to protons attached to oxygen-bearing carbons. These could thus be attributed to hydroxylated polyunsaturated fatty acids, since there are many reports of the isolation of this type of fatty acids from *Xestospongia* spp.(Jiang et al. 2011; Liu et al. 2011; Morinaka et al. 2007). Taiwan samples displayed characteristic signals between δ_H 6.40 and 6.60, which

correspond to double bonds commonly occurring in brominated unsaturated fatty acids. Samples from Tanzania had no distinguishing signals in a specific region of the spectrum, indicating that the changes present in this location do not involve a family of compounds, but rather specific compounds.

The NMR analysis provided a general overview of the metabolic profiles, allowing the detection of families of compounds predominant in each location. However, the congestion of signals in the spectra and the relatively low sensitivity rendered the identification of individual metabolites unfeasible. Thus, LC-MS/quadrupole time of flight (Q-TOF) was used to identify these metabolites, especially the minor ones. As shown in Figure1b, metabolic differences in the samples from each location were as clear as those observed with ^{1}H-NMR. As in the case of ^{1}H-NMR, a VIP plot was also used for the identification of peaks responsible for the separation. However, dereplication of the 50 most relevant peaks obtained from the VIP plot was not successful, because most of the selected MS features could not be identified, or they corresponded to several isomers. Nevertheless, information on a specific metabolites group, brominated fatty acids, was obtained from MS data. Different types of brominated fatty acids were found to be differential features in the samples on each location. Martinique samples showed no bromine-containing signals, while the Curaçao samples were discriminated by their characteristic dibrominated metabolites and the samples from Taiwan and Tanzania by monobrominated ones.

The variation in the chemical composition of the samples observed in this study proves the plasticity of *Xestospongia* spp. in terms of their biosynthesis processes. This could partly explain the great diversity in compounds isolated from this same sponge genus all over the world. Considering that these compounds exhibit a wide range of biological activities, it could be presumed that this metabolic differentiation observed in samples from different locations could be reflected in their bioactivity (Page et al. 2005). To investigate this potential correlation, the antimicrobial activity of *Xestospongia* spp. extracts against a Gram positive (*Staphylococcus aureus*) and a Gram negative (*Escherichia coli*) bacteria was assayed. This particular bioactivity was chosen due to numerous reports of antimicrobial compounds in *Xestospongia* spp. collected throughout the world (Bourguet-Kondracki et al. 1992; He et al. 2015).

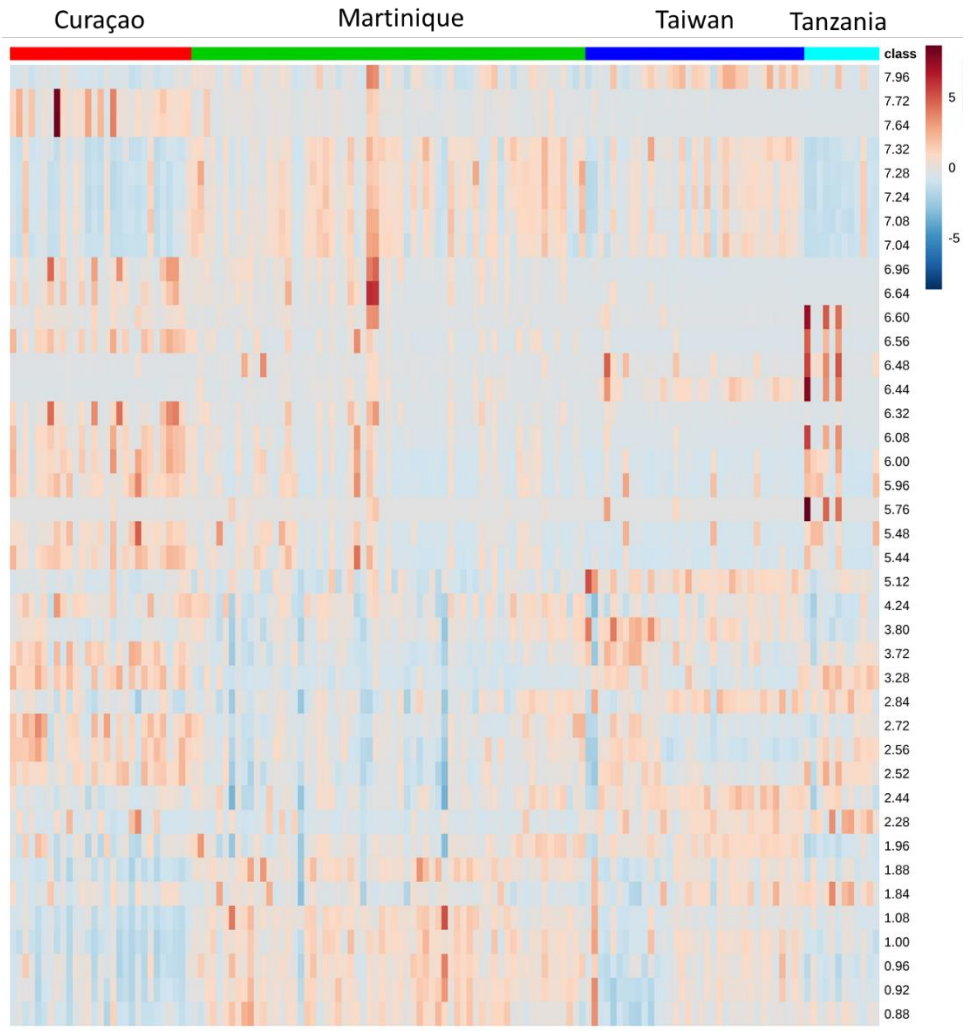


Figure 4.2: Heat map of characteristic signals from the ^{1}H-NMR data obtained from the variable importance for the projection (VIP) plot of orthogonal partial least square discriminant analysis (OPLS-DA).

The result of the activity test showed that some sponge extracts were active against *S. aureus* at a concentration of 512 µg/mL. From the whole sample set, 11.5% of the collected samples displayed activity, although there was a large variation in the activity according to the location. For example, while 20% of the samples collected in Taiwan had antimicrobial activity, none of the samples from Tanzania displayed activity. Although differences in the activity between collection places were observed, the ratio of active and nonactive samples was not significantly related to the collection places (χ^2 (2) = 2.72, p = 0.256). The lack of relation

between these two factors suggests that the production of antibacterial compounds is triggered by a factor occurring within a smaller spatial scale or is driven by genetic variation.

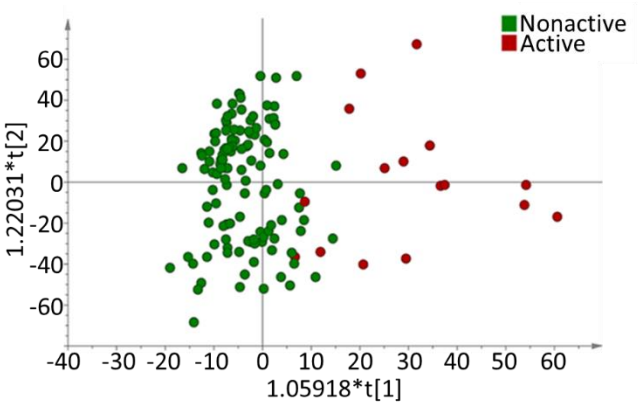


Figure 4.3: orthogonal partial least square discriminant analysis (OPLS-DA) model for the 139 *Xestospongia* spp. samples categorized by their activity against *Staphylococcus aureus* using LC-MS data.

On the other hand, none of the samples showed activity against *E. coli*, a proteobacteria, when tested at a concentration of 512 µg/mL. The lack of activity against *E.coli* can be explained by the fact that proteobacteria are one of the most predominant phyla among the bacterial communities of *Xestospongia* spp.(Fiore et al. 2013; Swierts et al. 2018). Therefore, it is a natural result that they do not produce compounds that could inhibit the growth of these types of bacteria.

To identify the compounds specifically involved in the antibacterial activity against *S. aureus*, an OPLS-DA model was built, grouping the samples as active (showed activity at 512 µg/mL) and nonactive (no activity shown at concentrations of 512 µg/mL) and using both NMR and LC-MS data. The model based on NMR data was not validated and did not reveal differences between the two groups. In this case, overlapping of signals belonging to compounds of the same family or low sensitivity could explain the lack of validation, as the activity must be related to specific compounds. On the other hand, with the LC-MS data, it was possible to separate the samples that displayed activity from the non-active samples as shown in Figure 4.3. Although variation in the chemical composition among the active samples was observed, a list of the masses of potentially active compounds was made using an S-plot, (appendix 1 Table S2). These features, together with the list obtained previously from the OPLS-DA analysis using location as a factor, were used to target the compounds of interest from samples collected in Martinique, Curaçao and Taiwan. The great dispersion observed between the

active samples suggested that although all samples exhibited activity, it was not necessarily due to the same compounds or alternatively, that there was a significant variation in the amount of active compounds present in the samples depending on the location. To clarify this, some of the most active compounds were isolated and tested, and their resulting activity was compared with their occurrence in different locations.

2.1 Isolation and structural elucidation

Ethanolic extracts of samples from Martinique, Curaçao, and Taiwan that were active against *S. aureus* were prepared to isolate potentially active antimicrobial compounds. These extracts were subjected to fractionation with liquid chromatography, using the list of potential active features as a criterion for fraction selection. This led to the isolation of five brominated fatty acid analogues (Figure 4.4): two from Martinique extracts (**1**,**2**), two from Curaçao extracts (**3**,**4**), and one from Taiwan extracts (**5**).

Compound **1** was isolated from a Martinique sample as a white powder. The (+)-HRESIMS spectrum of **1** showed the proton adduct $[M+H]^+$ ions at *m/z* 347.0646 and 349.0631, with relative intensities of 1:1, suggesting the presence of one bromine atom in the molecule. The molecular formula was deduced to be $C_{18}H_{19}BrO_2$. The ^{1}H-NMR (CH_3OH-d_4, 600 MHz) spectrum of the compound showed the presence of two double bonds and the ^{13}C-NMR (CH_3OH-d_4, 150 MHz) showed one carboxylic acid carbon, and the presence of three triple bonds. The molecular formula together with the characteristic NMR signals were dereplicated using the Dictionary of Natural Products. The compound was identified as (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid, which had been previously isolated from *X. muta* collected in Columbus Island, Bahamas, and reported to inhibit the HIV-1 protease with an IC_{50} of 8 µM (Patil et al. 1992).

Compound **2** was also isolated from a Martinique sample as a white powder. The (+)-HRESIMS spectrum of **2** showed the proton adduct $[M+H]^+$ ions at *m/z* 351.0956 and 353.0939, with relative intensities of 1:1. This isotopic pattern suggested the presence of a bromine atom in the molecule. The molecular formula was deduced to be $C_{18}H_{23}BrO_2$. The ^{1}H-NMR (CH_3OH-d_4, 600 MHz) spectroscopic data, showed the presence of two double bonds and ^{13}C-NMR (CH_3OH-d_4, 150 MHz) showed the presence of two triple bonds. A search using the molecular formula and the characteristic NMR signals in the Dictionary of Natural Products, yielded the compound (9*E*,17*E*)-18-bromooctadeca-9,17-dien-7,15-diynoic acid also known as xestospongic acid. This compound had been originally isolated from *Xestospongia sp* samples

collected in Australia as one of the most abundant compounds in the sample accounting for 0.1% of the dry weight material (Quinn and Tucker 1985).

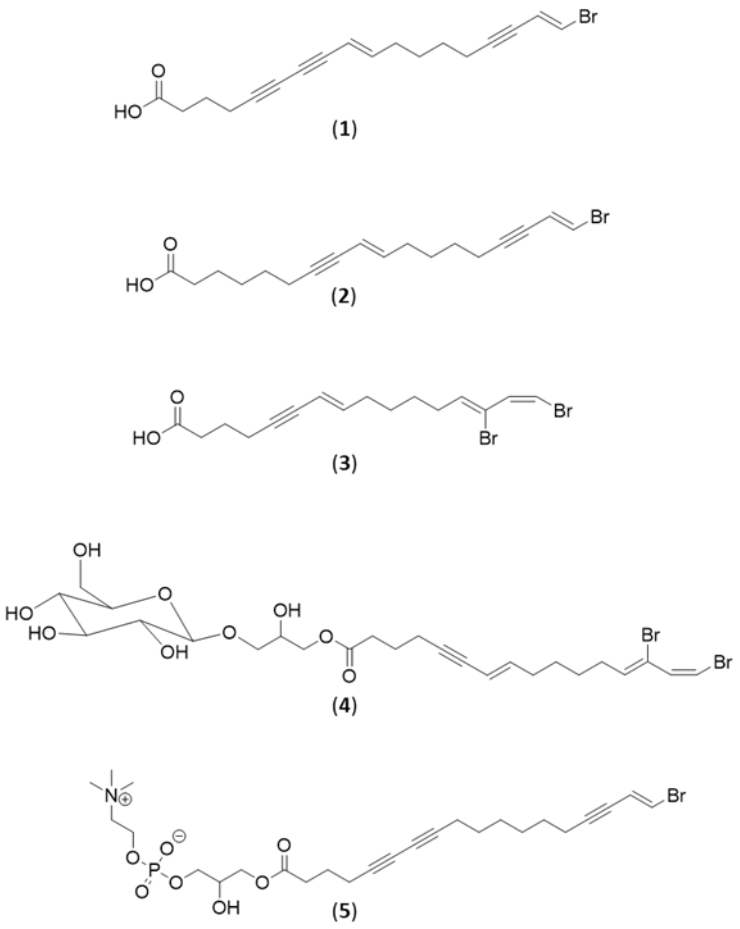


Figure 4.4: Structure of the compounds isolated from the giant barrel sponge (*Xestospongia* spp.). (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**), Xestospongic acid (**2**), (7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid (**3**), compound (**4**) and compound (**5**)

Compound **3** was isolated from a Curaçao sample as a white powder. Its (+)-HRESIMS spectrum showed the proton and sodium adduct $[M+H]^+$ and $[M+Na]^+$ ions at *m/z* 402.9904, 404.9884, and 406.9867, and 424.9727, 426.9706, and 428.9685, respectively, both having a relative intensity of 1:2:1. This isotopic pattern indicates the presence of two bromine atoms in the molecule. The molecular formula was deduced to be $C_{16}H_{20}Br_2O_2$. The ^{1}H-NMR (CH_3OH-d_4, 600

MHz) spectroscopic data, showed the presence of three double bonds in the molecule and the ^{13}C-NMR (CH_3OH-d_4, 150 MHz) showed one carboxylic acid carbon and one triple bond. The molecular weight together with the NMR signal in the Dictionary of Natural Products led to the identification of the compound as (7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid. This compound had been previously reported from *X. muta* collected in Summerland Key, Florida, USA and in Portobelo Bay, Panama (Schmitz and Gopichand 1978; Villegas-Plazas et al. 2019).

Table 4.1: NMR spectroscopic data for compounds **4** and **5**

	Compound 4		Compound 5	
Position	^{13}C-NMR δ, type	^{1}H-NMR δ, (*J* in Hz)	^{13}C-NMR δ, type	^{1}H-NMR δ, (*J* in Hz)
1’	174.9, C	---	174.6, C	---
2’	33.8, CH_2	2.48 t (7.3)	33.7, CH_2	2.49 t (7.6)
3’	25.2, CH_2	1.82 m	24.8, CH_2	1.83 quint (7.2)
4’	19.4, CH_2	2.35 td (7.0,1.8)	19.2, CH_2	2.35 t (7.0)
5’	88.1, C	---	76.8, C	---
6’	80.9, C	---	66.3, C	---
7’	111.3, CH	5.45 dm (15.8)	66.3, C	---
8’	144.0, CH	5.99 dt (15.8, 7.1)	78.2, C[b]	---
9’	33.6, CH_2	2.08 m	19.6, CH_2	2.28 m
10’	29.5, CH_2	1.40m	29.3, CH_2	1.43 m
11’	29.1, CH_2	1.44 m	29.3, CH_2	1.54 m
12’	32.0, CH_2	2.04 m	29.3, CH_2	1.54 m
13’	137.4, CH	6.07 td (7.7,1.5)	29.4, CH_2	1.43 m
14’	114.8, CH	----	19.9, CH_2	2.30 m
15’	132.3, CH	6.78 dm (7.6)	78.3, C[b]	---
16’	113.4, CH	6.56 d (7.6)	93.8, C	---
17’	---	---	119.2, CH	6.24 dt (14.0, 2.3)
18’	---	---	117.9, CH	6.70 d (14.0)
1	66.7, CH_2	4.17 m	67.8, CH_2	3.92 m
2	69.6, CH	4.00 m	69.8, CH	3.99 m
3	71.9, CH_2	3.92 dd (10.5, 5.2), 3.66 m	66.3, CH_2	4.21 dd (11.4, 4.5), 4.14 dd (11.4, 6.2)
1’’	104.7, CH	4.28 d (7.8)	60.4, CH_2	4.31 m
2’’	75.1, CH	3.21 m	67.0, CH_2	3.66 m
3’’	77.9, CH	3.36 bs	---	---
4’’	71.6, CH	3.29 m	---	---
5’’	78.0, CH	3.28 bs	---	---
6’’	62.7, CH_2	3.87 dd (12.1, 1.8) 3.67 m	---	---
N-Me	---	---	54.7, CH_3	3.24 s

[a] NMR spectra were recorded in CH_3OH-d_4, ^{1}H 600 MHz, ^{13}C 150 MHz. [b] these carbons are interchangeable

Compound **4** was also isolated from a Curaçao sample as a white powder. Its (+)-HRESIMS spectrum showed the proton and sodium adducts $[M+H]^+$ and $[M+Na]^+$ ions at *m/z* 639.0777, 641.0761, and 643.0749 and 661.0580, 663.0581, and 665.0563, respectively, both sets of ions with a relative intensity of 1:2:1. This isotopic pattern indicates the presence of two bromine atoms in the molecule. The molecular formula was deduced to be $C_{25}H_{36}Br_2O_9$, which requires 7 degrees of unsaturation. The ^{1}H-NMR and ^{13}C-NMR spectroscopic data (Table 4.1) and heteronuclear single quantum correlation (HSQC) spectrum revealed 10 methylene (δ_H/δ_C 1.40/29.5, 1.44/29.1, 1.82/25.2, 2.04/32.0, 2.08/33.6, 2.35/19.4, 2.48/33.8, 3.67-3.87/62.7, 3.92-3.66/71.9 4.17/66.7), six methine (δ_H/δ_C 3.21/75.1, 3.28/78.0, 3.29/71.6, 3.36/77.9, 4.00/69.6, 4.28/104.7) and five olefinic protons (δ_H/δ_C 5.45/111.3, 5.99/144.0, 6.07/137.4, 6.56/113.4, 6.78/132.3). The signal at 104.7 ppm is very characteristic for a carbon atom joined to two oxygen atoms, which indicates the presence of a sugar moiety in the molecule. In addition, the ^{13}C-NMR spectrum showed four nonprotonated carbons, consisting of one ester carbonyl (δ_C 174.9), two *sp* carbons (δ_C 80.9, 88.1) and one olefinic carbon (δ_C 114.8). The presence of aliphatic signals together with a carbonyl and *sp* and sp^2 carbons indicates that the structure contains an unsaturated fatty acid moiety. Two of the olefinic carbons are shifted upfield, indicating the presence of a substituent that increases the protection over those carbons. This is in agreement with the presence of two bromine atoms observed in the mass spectra and with the lack of any terminal methyl or methylene groups. It was thus possible to establish the attachment of bromine atoms to terminal olefinic carbons at δ_c 114.8 and δ_c 113.4.

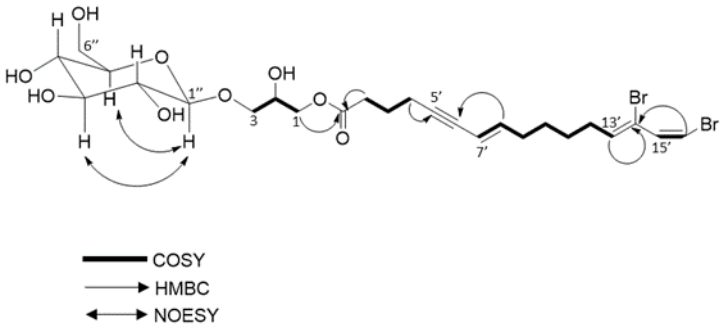


Figure 4.5: Important COSY, HMBC and NOE correlations of compound **4**

Further examination of HMBC and COSY correlations allowed us to establish the full structure of compound **4** (Figure 4.5) as consisting of three moieties: a dibrominated unsaturated fatty acid, a glycerol molecule and a sugar moiety. The brominated fatty acid and the sugar are attached to C1 and C3 of the glycerol molecule respectively. The chemical shift of δ_c 104.7 was

assigned to the anomeric carbon of the sugar, which is attached to C3 of the glycerol moiety through a glycosidic bond. Additionally, NOESY showed correlations between the anomeric proton and those in positions 3'' and 5''. This correlation together with the coupling constants of the anomeric proton (*J* = 7.8 Hz) and protons 3''and 4'' (*J* > 8 Hz) obtained from *J*-Resolved spectra allowed the identification of the sugar moiety as β-glucose. This was also supported by reported ^{13}C-NMR chemical shifts of β-glucose moiety in similar analogues (Fan 1996; Wicke et al. 2000). The identical chemical shift and coupling constants of H-13' indicated that the double bond in position 13'would have the same configuration as that of compound 3 Lastly, the double bond in position 7' was confirmed to have an E configuration with its characteristic coupling constant (*J* = 15.8 Hz), while the terminal double bond was found to have a Z configuration with the coupling constant (*J* = 7.6 Hz) (Schmitz and Gopichand 1978; Villegas-Plazas et al. 2019).

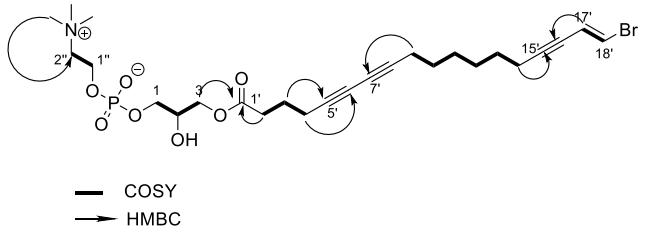


Figure 4.6: COSY and HMBC important correlations of compound **5**

Compound **5** was isolated from a sample from Taiwan as a white powder. The (+)-HRESIMS spectrum of **5** showed the proton adduct $[M+H]^+$ ions at *m/z* 588.1718 and 590.1702. The ions have a relative intensity of 1:1, indicating the presence of a bromine atom in the molecule. The molecular formula was deduced to be $C_{26}H_{39}BrNO_7P$. The ^{1}H-NMR and APT^{13}C-NMR spectroscopic data (Table 4.1) and HSQC correlation revealed the presence of three overlapping methyl groups joined to a nitrogen atom (δ_H/δ_C 3.24/54.7 x 3), 13 methylene (δ_H/δ_C 1.43/29.4, 1.43/29.3, 1.54/29.3 x 2, 1.83/24.8, 2.28/19.6, 2.30/19.9, 2.35/19.2, 2.49/33.7, 3.66/67.0, 3.92/67.8, 4.14-4.21/66.3, 4.31/60.4), one methyne (δ_H/δ_C 3.99/69.8), two olefinic protons (δ_H/δ_C 6.24/119.2, 6.70/117.9), and seven carbons with no protons attached, consisting of one carbonyl ester (δ_C 174.6) and six *sp* carbons (δ_C 66.3 x 2, 76.8, 78.2, 78.3, 93.8). The *sp* carbons indicate the presence of three triple bonds in the molecule. However, some of the δ_C are shifted to downfield, suggesting that two of the triple bonds are conjugated. As for compound **5,** the lack of a terminal methyl or methylene along with the low chemical shift of the olefinic carbon indicates the presence of a terminal olefinic bond attached to a bromine atom. Further examination of HMBC and COSY correlations established

the structure of **5** (Figure 4.6) as consisting of three moieties: a brominated fatty acid, a molecule of glycerol and a molecule of phosphatidylcholine. The presence of a phosphate group can be deduced from analysis of the exact mass of the molecule.

All of the isolated compounds contained one or more triple bonds in their structures, thus they are classified as polyacetylenes. This kind of compound has been reported in a wide range of marine organisms such as algae, corals, mollusks and sponges. In the case of sponges, the genera *Petrosia*, *Callyspongia* and *Xestospongia* are the main sources of polyacetylene compounds, and in some cases they have even been considered to be a chemotaxonomic marker of these genera (Zhou et al. 2015). Although the biosynthetic pathway and ecological function of this kind of compound are still unclear, they have shown a wide range of biological activities. In this study, all the isolated compounds exhibited mild activity against *S. aureus* (**1**: 64 µg/mL, **2** 256 µg/mL, **3** 64 µg/mL, **4** 64 µg/mL, **5** 128 µg/mL). Thus, the inconsistency in the relationship between the activity and location in which the sponges were collected can be explained by the fact that the compounds responsible for the activity might differ in their concentration or their structure in each location.

A comparison of the occurrence of the isolated compounds between the locations showed different patterns for each compound (Figure 4.7). Interestingly, compound **2**, isolated from a sample collected in Martinique, was more abundant in samples from the other three locations. This compound has been previously isolated from *Xestospongia* spp. samples collected in Australia(Quinn and Tucker 1985), the Red Sea(Hirsh et al. 1987), and Mayotte in the coast of Africa (Bourguet-Kondracki et al. 1992). The occurrence of **2** in *Xestospongia* spp. samples collected all over the world indicates that although it is a constitutive metabolite of sponges of the genus *Xestospongia,* the environmental factors prevalent in each location may affect the amount in which this metabolite is produced.

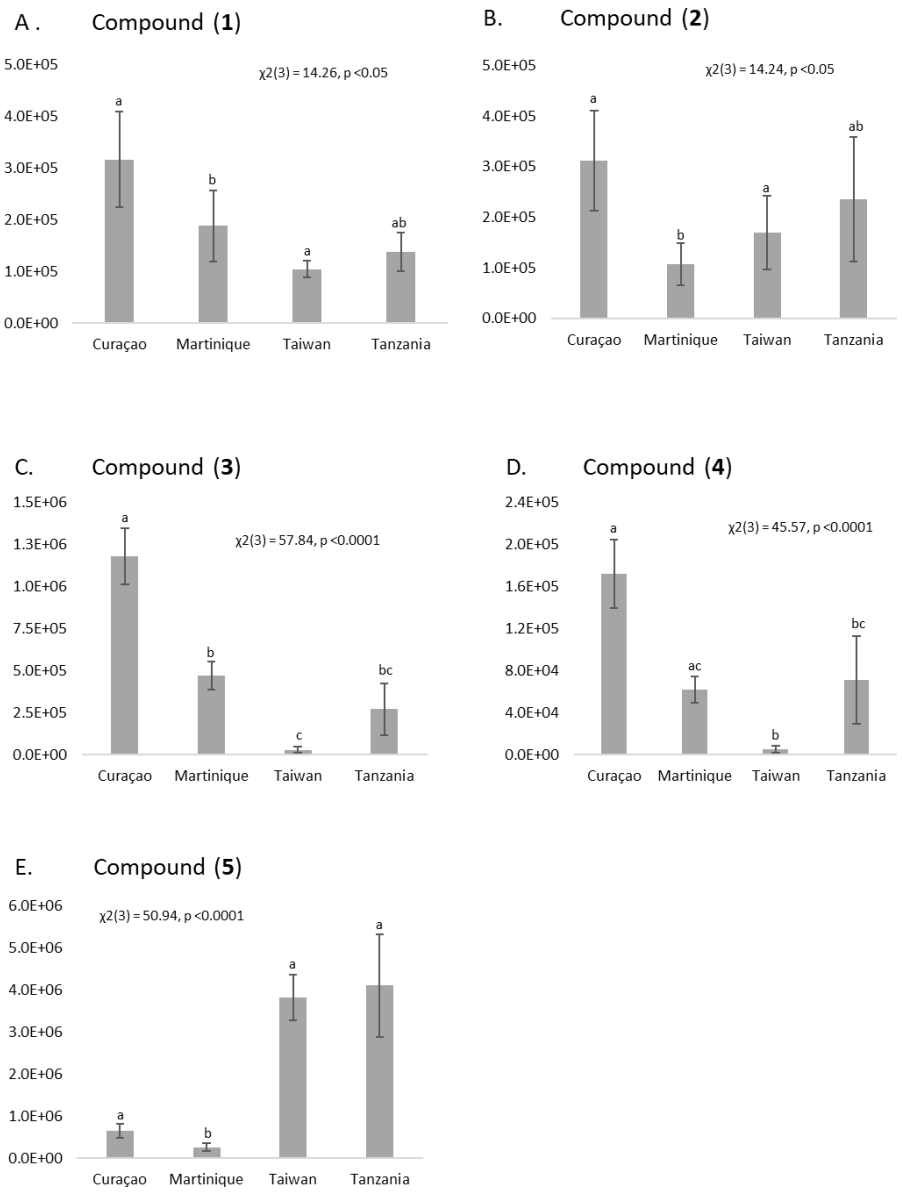


Figure 4.7: Intensity of the buckets of the most intense peak of the mass spectra for compound **1**-**5** in each location. Error bars indicate the standard error. Results of a Kruskal-Wallis Test are shown in each graph. Different letters indicate significant differences in the Post-Hoc test.

Compounds **3** and **4** were more abundant in samples from the Caribbean region, mainly Curaçao. Both compounds have two atoms of bromine in their structures that distinguish them from the other compounds isolated in this study. Compound **3** has been previously isolated

from a sample collected in Florida, USA, and although compound **4** as such, has not been previously reported, but its fatty acid moiety corresponds to compound **3**. Moreover, compound has been previously isolated from a sample collected in Florida and Panamá (Schmitz and Gopichand 1978; Villegas-Plazas et al. 2019). This suggests that these compounds occur higher quantities in sponges located in the Caribbean region and this fact might be used to distinguish the sponges from this region. Lastly, compound **5** is a phospholipid from the phosphatidylcholine group. These compounds are known to be part of the cellular membrane in animals, having not only structural functions but also playing a role in the signaling of metabolic pathways (D'Arrigo and Servi 2010). The variability observed in the amount of **5,** which is more abundant in samples from Taiwan and Tanzania than samples from the Caribbean, suggests that, similarly to what occurs in animal cell membranes, this compound also has more than just a structural role in *Xestospongia* spp. and its production is thus conditioned by the environmental factors related to each location.

3. Experimental section

3.1 Sample collection and extraction

Xestospongia spp. samples were collected in Martinique, Curaçao, Tanzania, and Taiwan and stored in ethanol at -20°C (appendix 1 Table S1). Samples were transported to the Institute of Biology of Leiden University for further analysis. The *Xestospongia* samples were ground and extracted with ethanol and sonicated for 20 min. The extraction was done in triplicate. An aliquot of 1 mL of each extract was dried and used for ^{1}H-NMR analysis. The remaining extracts were dried. The salt from the extracts was removed using C-18 SPE Supelco Supelclean LC-18, (Merck, Darmstadt, Germany) cartridges. For each extract, 50 mg were loaded into the cartridge and eluted with solvents of decreasing polarity, i.e., H_2O (F1), MeOH (F2), and MeOH/DCM (1:1) (F3). The methanol fraction (F2) was used for further LC-MS analysis.

3.2 ^{1}H-NMR Analysis and data preprocessing

The dry extract was resuspended in 1mL of deuterated methanol (CH_3OH-d_4) with hexamethyl disiloxane (HMDSO) as the internal standard. The ^{1}H-NMR spectra were measured at 25 °C in an AV-600 MHz NMR spectrometer (Bruker, Karlsruhe, Germany), operating at the ^{1}H-NMR frequency of 600.13 MHz, and equipped with a TCI cryoprobe and Z gradient system. For internal locking, CH_3OH-d_4 was used. A presaturation sequence was used to suppress the residual water signal, using low power selective irradiation at the H_2O frequency during the recycle delay.

The resulting spectra were phased, baseline corrected, and calibrated to HMDSO at 0.07 ppm using TOPSPIN V. 3.0 (Bruker, Karlsruhe, Germany). The NMR spectra were bucketed using AMIX 3.9.12 (Bruker BioSpin GmbH, Rheinstetten, Germany). Bucket data were obtained by spectra integration at 0.04 ppm intervals from 0.20 to 10.02 ppm. The peak intensity of individual peaks was scaled to the total intensity of the buckets. The regions between 3.32 and 3.28, 4.9 and 4.8, 3.62 and 3.57, and 1.15 and 1.19 ppm were excluded from the analysis because they correspond to solvent residual signals.

3.3 LC-MS analysis and data processing

The methanol fractions obtained from the SPE were dried, and 1 mg was dissolved in ACN/H_2O 1:1 to obtain solutions with a final concentration of 1mg/mL. The fractions were analyzed using an UHPLC-DAD-MS, Thermo Scientific (Dreieich, Germany) UltiMate 3000 system coupled to a Bruker (Bremen, Germany) OTOF-Q II spectrometer with electrospray ionization (ESI). The UHPLC separation was performed on a Phenomenex (Utrecht, The Netherlands), Kinetex, C18 (2.1 x 150 mm, 2.6 µm) using a two-step gradient of 0.1% formic acid in H_2O (A) and 0.1% formic acid in ACN (B), starting at 45% B to 60% in 15 minutes, 60% to 90% in 12.5 min and 90% to 98% B in 2.5 minutes. The flow rate was 0.300 mL/min, and the column temperature was maintained at 40°C. The injection volume was set at 1 µL. The mass spectrometer parameters were set as follow: nebulizer gas 2.0 bar, drying gas 10.0 mL/min, temperature 250°C, capillary voltage 3500 V. The mass spectrometer was operated in positive mode with a scan range of 100 - 1650 m/z and sodium formate was used as a calibrant.

The resulting chromatogram was processed to obtain a matrix for further analysis using Brucker Daltonics Profile Analysis (version 2.1, Bremen, Germany). The spectra were divided into buckets of 1 minute between 1 and 30 minutes and 1 m/z between 100 and 1450 m/z. The buckets were organized in a matrix, and data were filtered to remove those buckets that presented a %CV above 20% in the quality control samples.

3.4 Statistical analysis

The matrixes obtained from the NMR and LC-MS were used to perform multivariate data analysis using SIMCA-P software (v.15.0.2, Umetrics, Umeå, Sweden). Principal component analysis PCA, discriminant analysis of partial least square PLS-DA, and orthogonal partial least square OPLS-DA were performed. For the analysis, data were scaled using united variance scaling (NMR) and pareto scaling (LC-MS), and the models were tested using a permutation test and a cross-validation ANOVA (CV-ANOVA) test. The model was considered valid if CV-

ANOVA showed $p < 0.05$. For the prediction power of the model, Q^2 values above 0.4 were required; otherwise, the model was considered valid but with no prediction power.

A heatmap was created using a data matrix with the top 40 signals of the VIP plot. This matrix was uploaded on the Metaboanalyst R2.0 Web site (http://www.metaboanalyst.ca) (Chong et al. 2019). The dendrogram was obtained by hierarchical cluster analysis using the Euclidean distance and the "Ward" algorithm.

To test if the concentration of each compound differed among locations, the intensity of buckets corresponding to the most intense ion observed in its mass spectra was used. Data were analyzed with IBM SPSS Statistics Version 22 (Armonk, NY, USA) using a Kruskal-Wallis Test. Location was used as a factor, and the number of samples was 139. For the compounds that appeared to be significantly different between locations, a Post-Hoc test was done, and the P values were Bonferroni-corrected.

3.5 Isolation and elucidation

For the isolation of active compounds, extracts of samples from Martinique, Curaçao and Taiwan were prepared as mentioned in section 1. The crude extracts were fractionated using an SPE 20 mL LC-18 Supelco Supelclean (Merck, Darmstadt, Germany) cartridge and eluted using two different methods according to the sample location. The sample from Martinique (1.0 g) was eluted with 100 mL of H_2O, MeOH, and MeOH/DCM (1:1), yielding three fractions: FM1, FM2, and FM3, respectively. The samples from Curaçao (1.9 g) and Taiwan (2.5 g) were eluted using 50 mL of each of the following solvents: 100% H_2O; H_2O/MeOH 8:2, H_2O/MeOH 6:4, H_2O/MeOH 4:6, H_2O/MeOH 2:8, 100% MeOH and MeOH/DCM 1:1. This resulted in fractions of each extract FC1–FC7 for Curaçao samples and FT1–FT7 for Taiwan samples, respectively.

Fraction FM2 (212 mg) was submitted to a size-based separation. The fraction was resuspended in 10 mL of MeOH and injected into a Sepacore flash system (Büchi, Hendrik-Ido-Ambacht, The Netherlands) with a Sephadex LH-20 (Merck KGaA, Darmstadt, Germany) column and a sample loop of 20 mL. Samples were eluted at a flow rate of 2.5 mL/min with MeOH. Fractions were collected automatically every minute and combined into nine FM2.1-FM2.9 fractions based on their TLC profiles. The purification of fractions FM2.7, FM2.9, FC4 and FT4 was performed using an Agilent (Santa Clara, CA, USA) 1200 series system on a Phenomenex (Utrecht, The Netherlands) Luna 5 µm, C-18, 250 mm x 10 mm column and eluted at a flow rate of 3.50 mL/min with different gradients of 0.1% formic acid in H_2O (A) and 0.1% formic acid in MeOH (B). The fractions FM.2.7-FM2.9 (49.87mg) were eluted with a

gradient of 75% B to 80% B in 25 min, 15 min of 80% B, 80% to 100% B in 2 min and 100% B for 5 min. This yielded 2.26 mg of **1** and 1.58 mg of **2.** The fraction FC4 (70.6 mg) was eluted using the following gradient: 72% to 85% B in 52 min, 85% to 100% B in 2 min and 100% B for 10 min. This allowed the isolation of **3** (7.16 mg) and **4** (2.14 mg). The fraction FT5 (94.08 mg) was eluted using the following gradient: 72% to 80% B in 34 min, 80% to 85% B in 16 min, 85% to 100% B in 3 min and 100% B for 3 min. This led to compounds **5** (1.51 mg) and **1**(1.00 mg).

(9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**)

White amorphous powder;^{1}H-NMR (CH_3OH-d_4, 600 MHz) δ_H 1.52 m, 1.82 quint (*J* = 7.3 Hz), 2.18 m, 2.29 m, 2.31 m, 2.38 t (*J* = 7.0 Hz), 5.55 dm (*J* = 15.9 Hz), 6.23 m, 6.27 m, 6.70 d (*J* = 14.0 Hz). ^{13}C-NMR (CH_3OH-d_4, 150 MHz) δ_c 19.5, 19.6, 25.8, 28.6, 28.6, 33.2, 36.9, 65.9, 73.3, 73.9, 80.2, 83.3, 92.8, 109.8, 117.7, 118.9 and 148.3. HRESIMS *m/z* $[M+H]^+$ 347.0646 and 349.0631 (Calcd for $C_{18}H_{20}BrO_2^+$, 347.0647 and 349.0626).

Xestospongic acid (**2**)

White amorphous powder;^{1}H-NMR (CH_3OH-d_4, 600 MHz) δ_H 1.43 m, 1.49 m, 1.51 m, 1.63 quint (*J* = 7.5 Hz), 2.09 m, 2.24 m, 2.26 m, 2.28 m, 5.45 dm (*J* = 15.8 Hz),5.96 dt (*J* = 15.8, 7.2 Hz), 6.22 dt (*J* = 14.0, 2.3 Hz), 6.68 d (*J* = 14.0 Hz). ^{13}C-NMR (CH_3OH-d_4, 150 MHz) δ_C 19.8 x 2, 26.3, 28.9, 29.2, 29.7, 29.8, 33.3, 36.6, 78.3, 80.2, 89.2, 93.6, 111.6, 117.9, 119.2, and 143.3. HRESIMS *m/z* $[M+H]^+$ 351.0956 and 353.0939 (Calcd for $C_{18}H_{24}BrO_2^+$, 351.0960 and 353.0939).

(7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid (**3**)

White amorphous powder;^{1}H-NMR (CH_3OH-d_4, 600 MHz) δ_H,1.40 m, 1.44 m, 1.78 quint (*J* = 7.2 Hz), 2.04 m, 2.08 m, 2.34 dt (*J* = 2.0, 7.0 Hz), 2.40 t (*J* = 7.4 Hz), 5.45 dm (*J* = 15.8 Hz), 5.99 dt (*J* = 15.8, 7.1 Hz), 6.07 td (*J* = 7.7, 1.5 Hz), 6.55 d (*J* = 7.6 Hz), 6.78 dm (*J* = 7.6 Hz). ^{13}C-NMR (CH_3OH-d_4,150 MHz) δ_c 19.4, 25.4, 29.1, 29.5, 32.0, 33.6, 33.9, 80.8, 88.1, 111.3, 113.4, 114.8, 132.3, 137.4, 143.9 and 177.3. HRESIMS *m/z* $[M+H]^+$ 402.9904, 404.9884, 406.9867 (Calcd for $C_{16}H_{21}Br_2O_2^+$, 402.9903, 404.9883, 406.9862) and $[M+Na]^+$ 424.9727, 426.9706, 428.9685 (Calcd for $C_{16}H_{20}Br_2NaO_2^+$, 424.9723, 426.9702, 428.9682).

Compound (**4**)

White amorphous powder;^{1}H-NMR (CH_3OH-d_4, 600 MHz) δ_H in Table 4.1. ^{13}C-NMR (CH_3OH-d_4, 150 MHz) δ_C in Table 4.1. HRESIMS *m/z* $[M+H]^+$ 639.0777, 641.0761, 643.0749 (Calcd for $C_{25}H_{37}Br_2O_9^+$, 639.0804, 641.0784, 643.0763) and , $[M+Na]^+$ 661.0580, 663.0581, 665.0563 (Calcd for $C_{25}H_{36}Br_2NaO_9^+$, 661.0624, 663.0603, 665.0583)

Compound (**5**)

White amorphous powder;^{1}H-NMR (CH_3OH-d_4,600MHz) δ_H in Table 4.1. ^{13}C-NMR (CH_3OH-d_4, 150 MHz) δ_C in Table 4.1. HRESIMS *m/z* $[M+H]^+$ 588.1718, 590.1702 (Calcd for $C_{26}H_{40}BrNO_7P^+$ m/z: 588.1726, 590.1705)

3.6 Antibacterial activity test

Strains used in this study were the Gram-positive bacteria *S. aureus* (CECT976) and Gram-negative bacteria *E. coli* (DH5α,Promega). The strains had been kept at – 80 °C (in 100% glycerol). For their use, the strains were transferred onto Mueller-Hinton agar plates (MHA) (Sigma-Aldrich, Zwijndrecht,The Netherlands) and incubated overnight at 37 °C.

A broth microdilution method was used to determine the minimum inhibitory concentration (MIC) according to the CLSI (Clinical Laboratory Standards Institute) guidelines using 96-wells microtiter plates. The MIC is defined as the lowest concentration of each extract, which completely inhibits bacterial growth. For antimicrobial testing, the extracts were dissolved in 100% DMSO in a concentration of 10 mg/mL. All experiments were performed in triplicate. Ninety microliters of Mueller-Hinton broth (MHB) and 10 μL of the tested extract were added into the first well. Then two-fold serial dilutions of the extracts were prepared by dilution with MHB to achieve a decreasing range of concentrations from 512–16 μg/mL in the microtiter plates. The highest concentration of DMSO after dilution was < 5 %, to avoid affecting the growth of the bacterial strains. From the overnight cultures of the bacterial strains, a single colony was used to inoculate the MHB at 37 °C with agitation (150 rpm). The cultures were then further diluted in MHB and adjusted to a turbidity level of 0.5 McFarland standard solution (approximately 10^6 CFU/mL). Each well was then inoculated with 50 μL of the bacterial solution at a density of 10^6 CFU/mL. Spectinomycin (100 mg/mL) (Sigma-Aldrich) was used as a positive control and 5% dimethyl sulfoxide (DMSO) as a negative control. The inoculated microtiter plates were incubated at 30 °C for 24 h. Bacterial growth was detected by optical density.

Acknowledgment

The authors greatly appreciate the contribution of Dr. Erika G. Wilson in the scientific discussion of this paper. The specimens from Martinique were collected during the Madibenthos expeditions (PI Philippe Bouchet) organised by the Museum National d'Histoire Naturelle (MNHN) and the Marine Protected Areas Agency (AAMP), the Regional Directorate for the Environment (DEAL) and the Martinique Water Bureau (ODE) with funding from the

European Regional Development Fund (ERDF), the Territorial Collectivity of Martinique (CTM) and Saint-James Plantations and BRED Banque populaire. Yusheng Huang of the National Penghu University of Science and Technology, Penghu, Taiwan and Christian Vaterlaus of Marine Cultures, Jambiana Tanzania are thanked for their assistance in the field. This work was supported by the COLCIENCIAS (science technology and innovation ministry, Colombian government) and NWO-VIDI (#16.161.301) and NWO-Aspasia (#105-010.030).

Appendix 1

Table S1: Collection places of *Xestospongia* samples and their geographical information.

Code	Location	Latitude	Longitude
LMB29902	Curaçao	12.10771	-68.94976
LMB30102	Curaçao	12.10771	-68.94976
LMB30302	Curaçao	12.10771	-68.94976
LMB29502	Curaçao	12.10771	-68.94976
LMB31302	Curaçao	12.10771	-68.94976
LMB30902	Curaçao	12.10771	-68.94976
LMB30702	Curaçao	12.10771	-68.94976
LMB30502	Curaçao	12.06500	-68.86027
LMB30402	Curaçao	12.06500	-68.86027
LMB30602	Curaçao	12.06500	-68.86027
LMB30802	Curaçao	12.06500	-68.86027
LMB31002	Curaçao	12.06500	-68.86027
LMB29402	Curaçao	12.12206	-68.96925
LMB31102	Curaçao	12.12206	-68.96925
LMB31202	Curaçao	12.12206	-68.96925
LMB30002	Curaçao	12.12206	-68.96925
LMB29802	Curaçao	12.12206	-68.96925
LMB29702	Curaçao	12.12206	-68.96925
LMB31402	Curaçao	12.32977	-69.15191
LMB29602	Curaçao	12.32977	-69.15191
LMB30202	Curaçao	12.32977	-69.15191
LMB59202	Curaçao	12.39386	-69.15723
LMB59302	Curaçao	12.39386	-69.15723
LMB59402	Curaçao	12.39386	-69.15723
LMB59502	Curaçao	12.39386	-69.15723
LMB59602	Curaçao	11.98617	-68.64665
LMB59702	Curaçao	11.98617	-68.64665
LMB59802	Curaçao	11.98617	-68.64665
LMB59902	Curaçao	11.98617	-68.64665
LMB7202	Martinique	14.53330	-61.08789
LMB7602	Martinique	14.63160	-61.12880
LMB7802	Martinique	14.63160	-61.12880
LMB7402	Martinique	14.44440	-61.03769

Code	Location	Latitude	Longitude
LMB19102	Martinique	14.44440	-61.03769
LMB23702	Martinique	14.44380	-61.04059
LMB7702	Martinique	14.44380	-61.04059
LMB19202	Martinique	14.44380	-61.04059
LMB7102	Martinique	14.44460	-60.89990
LMB7902	Martinique	14.44460	-60.89990
LMB7302	Martinique	14.45510	-60.92529
LMB7502	Martinique	14.45510	-60.92529
LMB25302	Martinique	14.45510	-60.92529
LMB20302	Martinique	14.44800	-60.89970
LMB20602	Martinique	14.44800	-60.89970
LMB18402	Martinique	14.44800	-60.89970
LMB17602	Martinique	14.49630	-60.77170
LMB17402	Martinique	14.44210	-61.03960
LMB19402	Martinique	14.44210	-61.03960
LMB26102	Martinique	14.44210	-61.03960
LMB18902	Martinique	14.44070	-61.02909
LMB24302	Martinique	14.44070	-61.02909
LMB8002	Martinique	14.44070	-61.02909
LMB8102	Martinique	14.44070	-61.02909
LMB24602	Martinique	14.46490	-61.01940
LMB25002	Martinique	14.46490	-61.01940
LMB20002	Martinique	14.46490	-61.01940
LMB19302	Martinique	14.46490	-61.01940
LMB20102	Martinique	14.46490	-61.01940
LMB22602	Martinique	14.46490	-61.01940
LMB23602	Martinique	14.86850	-60.89429
LMB24402	Martinique	14.64620	-60.85079
LMB24002	Martinique	14.91740	-61.14710
LMB22302	Martinique	14.91440	-61.14900
LMB24102	Martinique	14.47600	-61.08590
LMB20202	Martinique	14.47600	-61.08590
LMB17302	Martinique	14.51850	-61.09770
LMB20702	Martinique	14.51850	-61.09770
LMB17702	Martinique	14.84160	-61.22770
LMB18202	Martinique	14.84160	-61.22770

Code	Location	Latitude	Longitude
LMB19802	Martinique	14.84160	-61.22770
LMB24702	Martinique	14.84160	-61.22770
LMB19502	Martinique	14.66950	-61.17509
LMB20802	Martinique	14.65710	-61.15750
LMB17902	Martinique	14.65710	-61.15750
LMB23002	Martinique	14.65710	-61.15750
LMB22402	Martinique	14.65710	-61.15750
LMB24202	Martinique	14.65710	-61.15750
LMB19602	Martinique	14.74070	-61.18029
LMB19902	Martinique	14.86250	-61.20749
LMB20402	Martinique	14.86250	-61.20749
LMB25102	Martinique	14.86250	-61.20749
LMB8902	Martinique	14.57630	-61.05480
LMB26002	Martinique	14.57630	-61.05480
LMB19002	Martinique	14.57630	-61.05480
LMB18702	Martinique	14.45850	-60.96949
LMB25702	Martinique	14.45850	-60.96949
LMB18002	Martinique	14.45850	-60.96949
LMB17502	Martinique	14.46700	-61.03440
LMB20502	Martinique	14.46700	-61.03440
LMB23902	Martinique	14.46700	-61.03440
LMB24802	Martinique	14.46700	-61.03440
LMB19702	Martinique	14.51850	-61.09770
LMB24902	Taiwan	23.5517	119.6412
LMB25202	Taiwan	23.3986	119.3231
LMB4902	Taiwan	23.5725	119.4931
LMB5002	Taiwan	23.5725	119.4931
LMB5102	Taiwan	23.5725	119.4931
LMB5202	Taiwan	23.5517	119.6412
LMB5302	Taiwan	23.5517	119.6412
LMB5402	Taiwan	23.5725	119.4931
LMB5502	Taiwan	23.5517	119.6412
LMB5602	Taiwan	23.3986	119.3231
LMB56102	Taiwan	23.2504	119.6743
LMB56202	Taiwan	23.2504	119.6743
LMB56302	Taiwan	23.5371	119.5443

Code	Location	Latitude	Longitude
LMB56402	Taiwan	23.2442	119.6199
LMB56502	Taiwan	23.2504	119.6743
LMB56602	Taiwan	23.5371	119.5443
LMB56702	Taiwan	23.5371	119.5443
LMB56802	Taiwan	23.5371	119.5443
LMB56902	Taiwan	23.5371	119.5443
LMB57002	Taiwan	23.5371	119.5443
LMB5702	Taiwan	23.3986	119.3231
LMB57102	Taiwan	23.2504	119.6743
LMB5802	Taiwan	23.3986	119.3231
LMB5902	Taiwan	23.5725	119.4931
LMB6002	Taiwan	23.5725	119.4931
LMB6102	Taiwan	23.3986	119.3231
LMB6202	Taiwan	23.5517	119.6412
LMB6302	Taiwan	23.5725	119.4931
LMB6402	Taiwan	23.5517	119.6412
LMB6502	Taiwan	23.5725	119.4931
LMB6602	Taiwan	23.2575	119.6791
LMB6702	Taiwan	23.3986	119.3231
LMB6802	Taiwan	23.5517	119.6412
LMB6902	Taiwan	23.5725	119.4931
LMB7002	Taiwan	23.5517	119.6412
LMB61002	Tanzania	-6.31011	39.58258
LMB61102	Tanzania	-6.31011	39.58258
LMB60302	Tanzania	-6.70895	39.28272
LMB60202	Tanzania	-6.70895	39.28272
LMB60002	Tanzania	-6.70895	39.28272
LMB60602	Tanzania	-6.70895	39.28272
LMB60402	Tanzania	-6.70895	39.28272
LMB60902	Tanzania	-6.70895	39.28272
LMB60102	Tanzania	-6.31011	39.58258
LMB60802	Tanzania	-6.31011	39.58258
LMB60702	Tanzania	-6.31011	39.58258
LMB60502	Tanzania	-6.31011	39.58258

Table S2: List of mass features selected from S-plot of Orthogonal partial least square – discriminant analysis (OPSD-DA) with two classes (active and non-active samples).

m/z	P1 value
749.5	0.0570212
405.5	0.0562273
404.5	0.0546859
425.5	0.0531678
597.5	0.0530375
385.5	0.0521159
642.5	0.0504513
403.5	0.049805
323.5	0.0489471
238.5	0.0483625
370.5	0.0475088
347.5	0.0474848
378.5	0.0472359
296.5	0.0471683
324.5	0.0470832
231.5	0.0470479
377.5	0.0468182
375.5	0.0467828
295.5	0.0467675
348.5	0.0461731
369.5	0.045865
218.5	0.0457138
366.5	0.0456151
351.5	0.0454663
240.5	0.045461
184.5	0.0452214
226.5	0.0445743

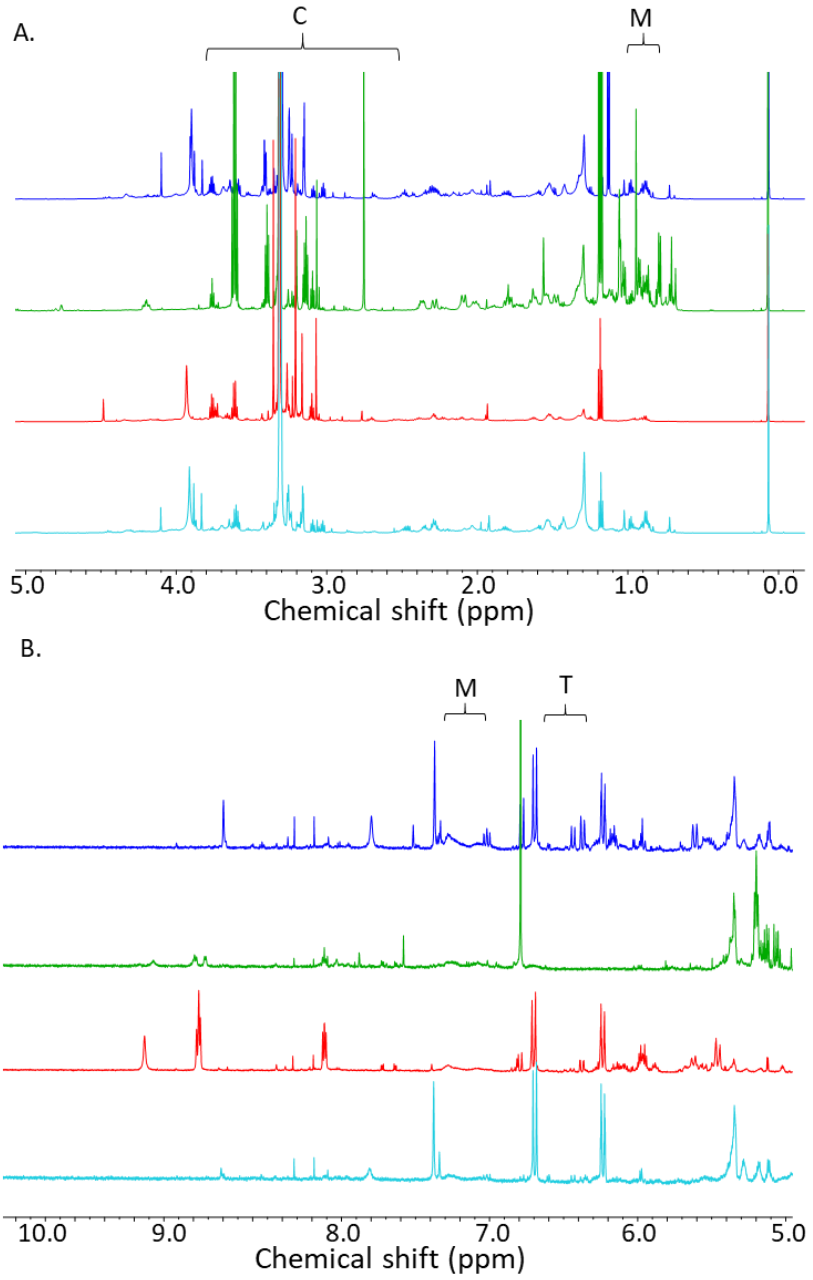


Figure S1: Typical ^{1}H-NMR spectra profile (CH_3OH-d_4, 600 MHz) of *Xestospongia* samples from four locations: Martinique (Green), Curaçao (Red), Taiwan (Dark blue), and Tanzania (Light blue) divided in two regions A: from δ_H 0 - 5 and B: from δ_H 5 - 10. The characteristic ^{1}H-NMR signal ranges of Martinique (M), Curaçao (C), and Taiwan (T) are shown in each spectrum.

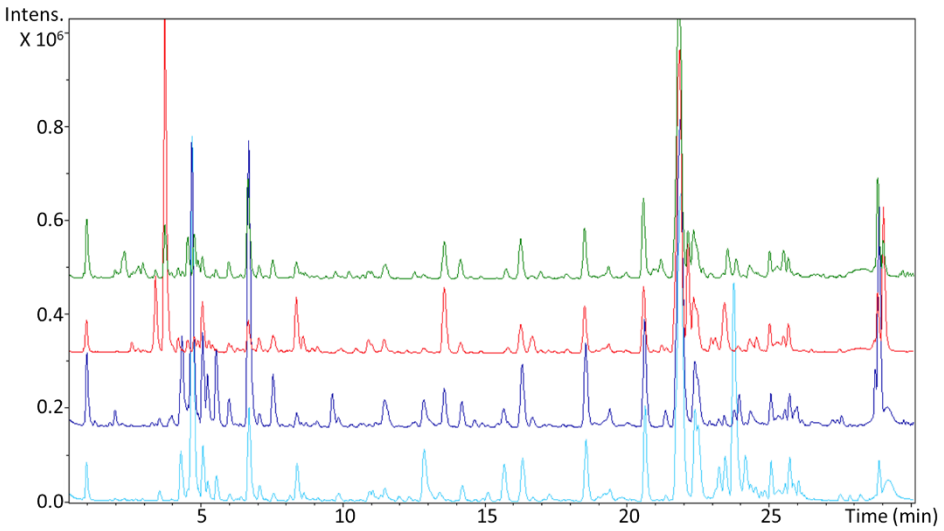


Figure S2: Typical LC-MS chromatographic profiles of *Xestospongia* samples from the four locations: Martinique (Green), Curaçao (Red), Taiwan (Dark blue), and Tanzania (Light blue).

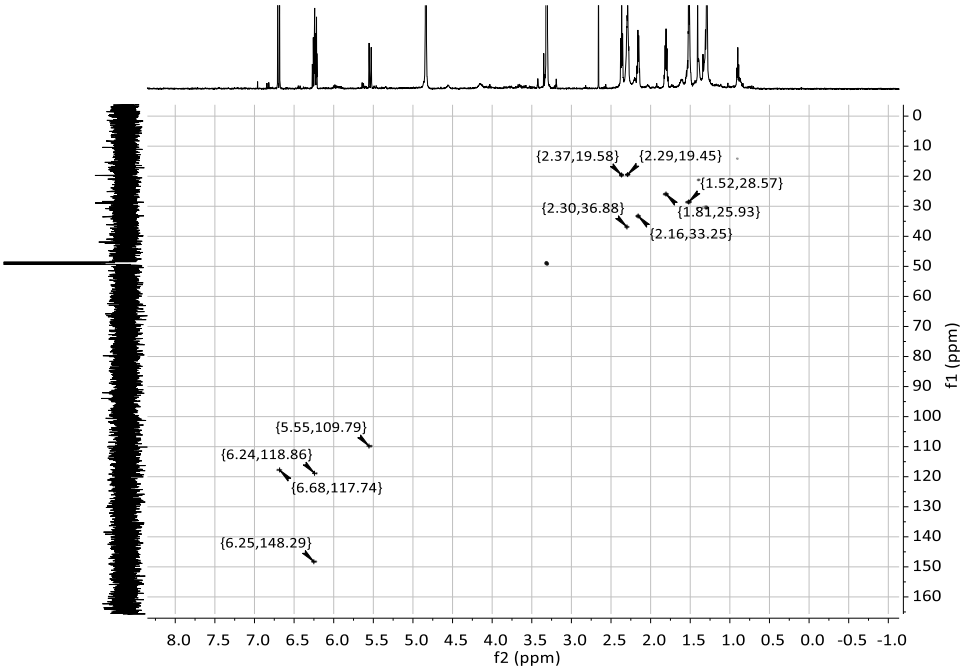


Figure S3: Heteronuclear single quantum coherence (HSQC) spectrum (CH_3OH-d_4, 600 MHz) of (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**).

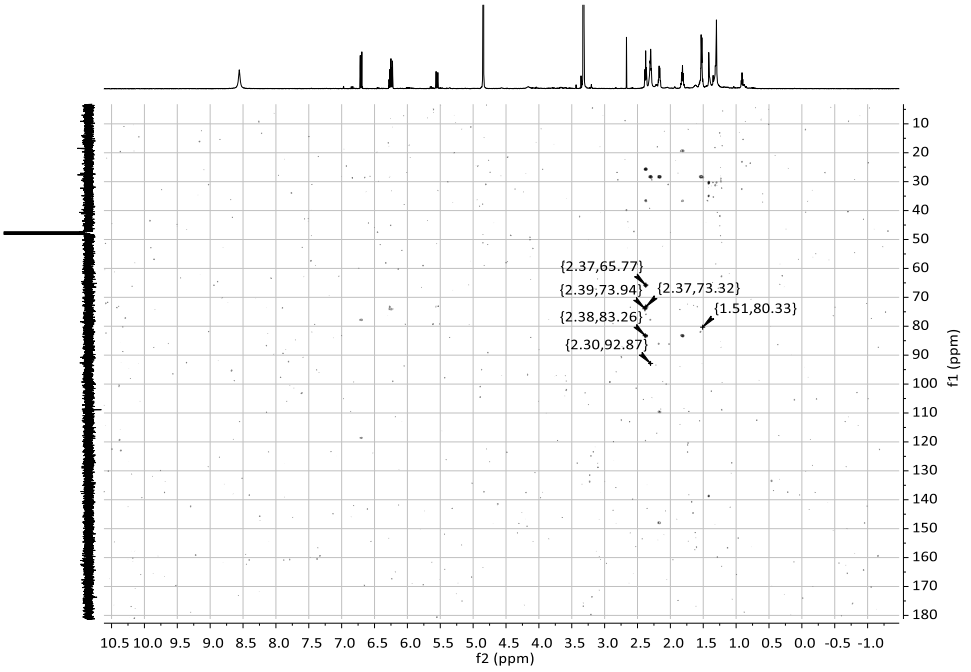


Figure S4: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-d_4, 600 MHz) of (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**).

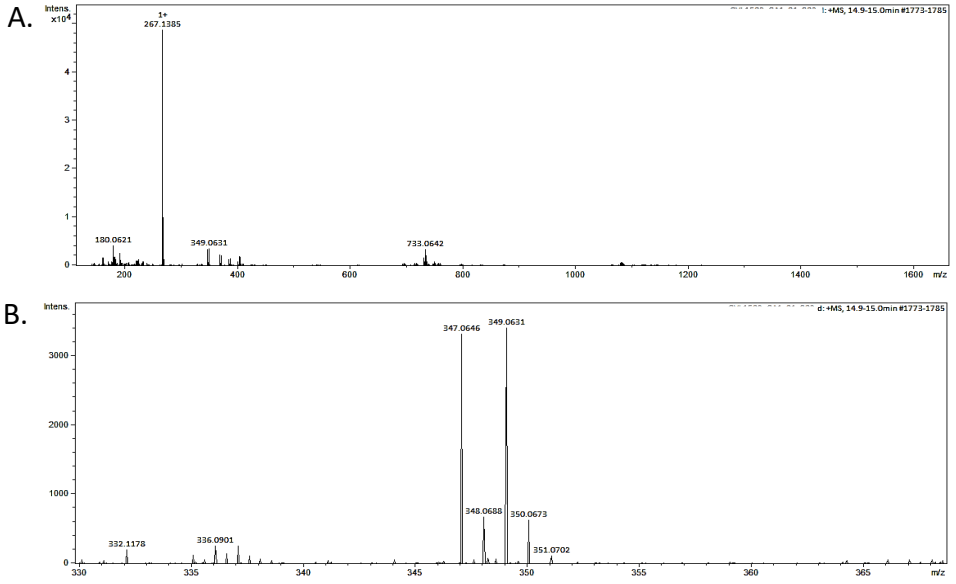


Figure S5: Electrospray ionization-quadrupole-time of flight (ESI-qTOF) MS spectrum in full range (A) and the expanded region around the $[M+H]^+$ ion (B) of (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (**1**).

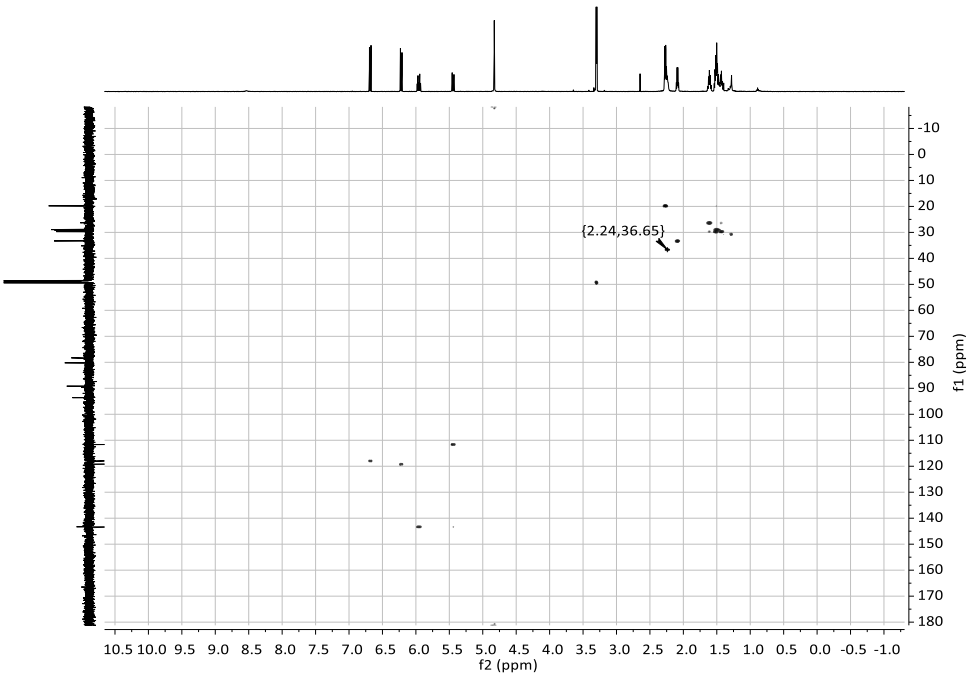


Figure S6: Heteronuclear single quantum coherence (HSQC) spectrum (CH_3OH-d_4, 600 MHz) of Xestospongic acid (**2**).

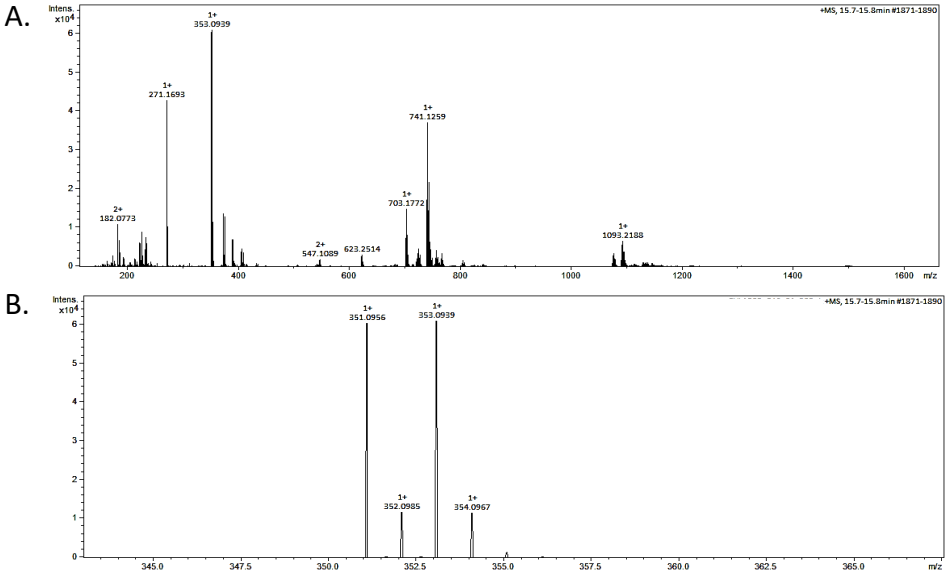


Figure S7: Electrospray ionization-quadrupole-time of flight (ESI-qTOF) MS spectrum in full range (A) and the expanded region around the $[M+H]^+$ ion (B) of Xestospongic acid (**2**).

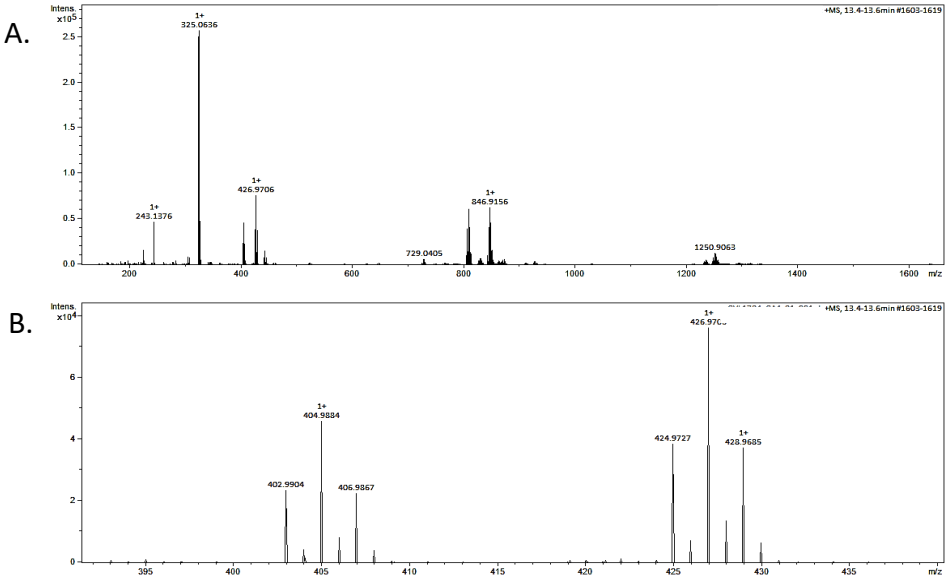


Figure S8: Electrospray ionization-quadrupole-time of flight (ESI-qTOF) MS spectrum in full range (A) and the expanded region around the $[M+H]^+$ and $[M+Na]^+$ ions (B) of (7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid (**3**).

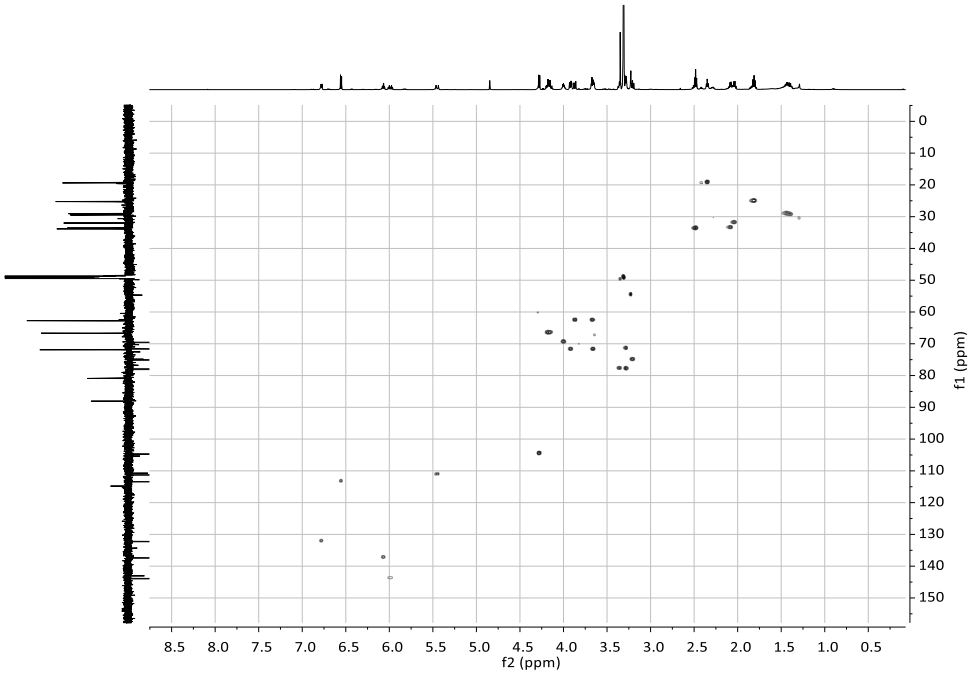


Figure S9: Heteronuclear single quantum coherence (HSQC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**4**).

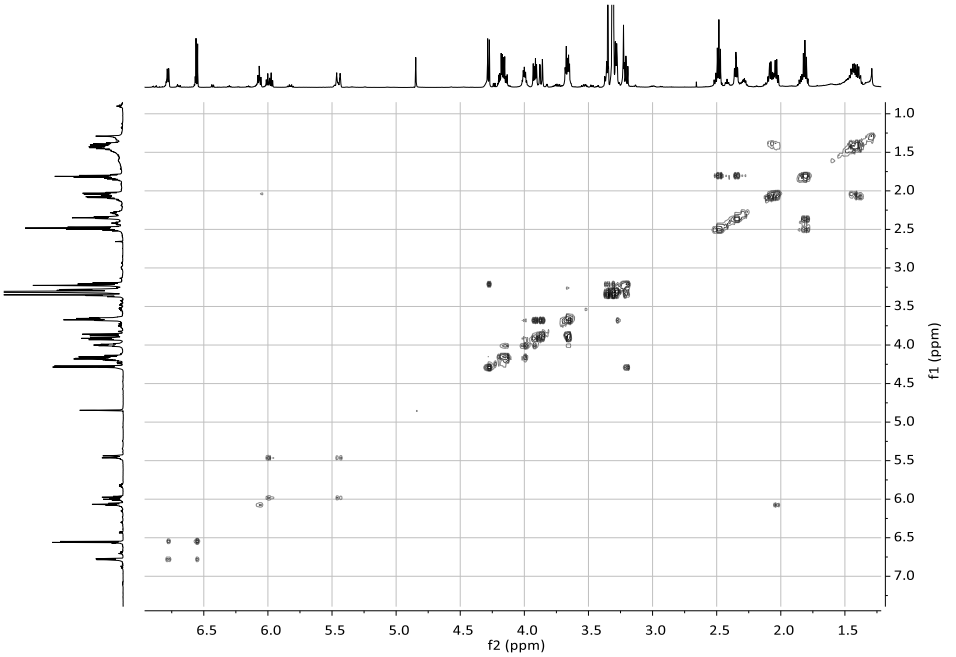


Figure S10: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**4**).

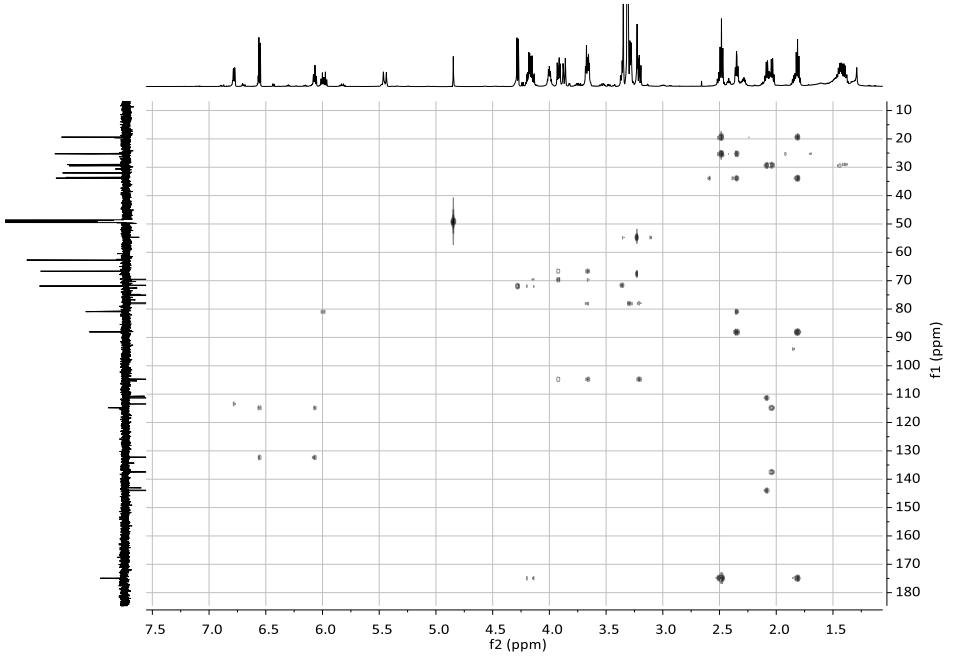


Figure S11: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**4**).

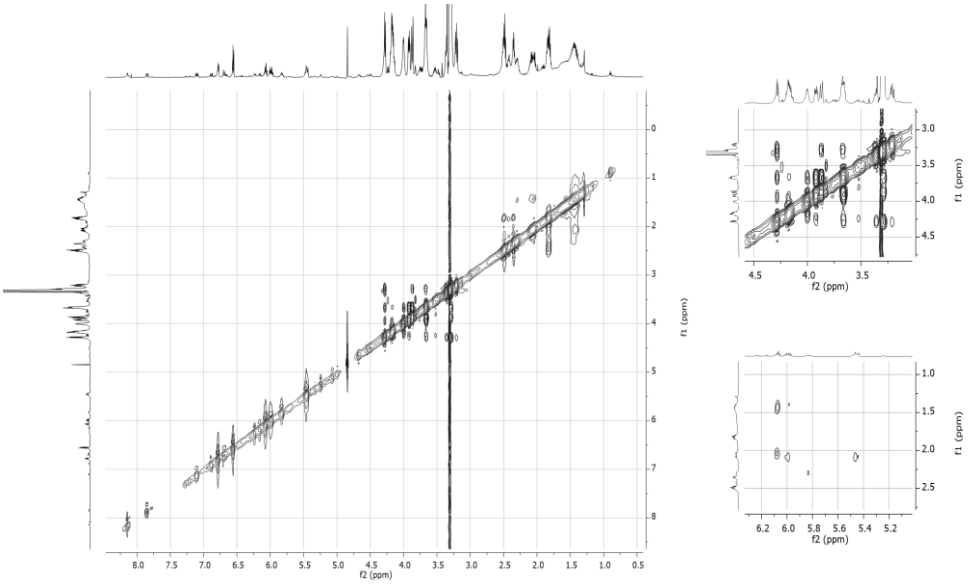


Figure S12: Nuclear Overhauser effect spectroscopy (NOESY) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**4**).

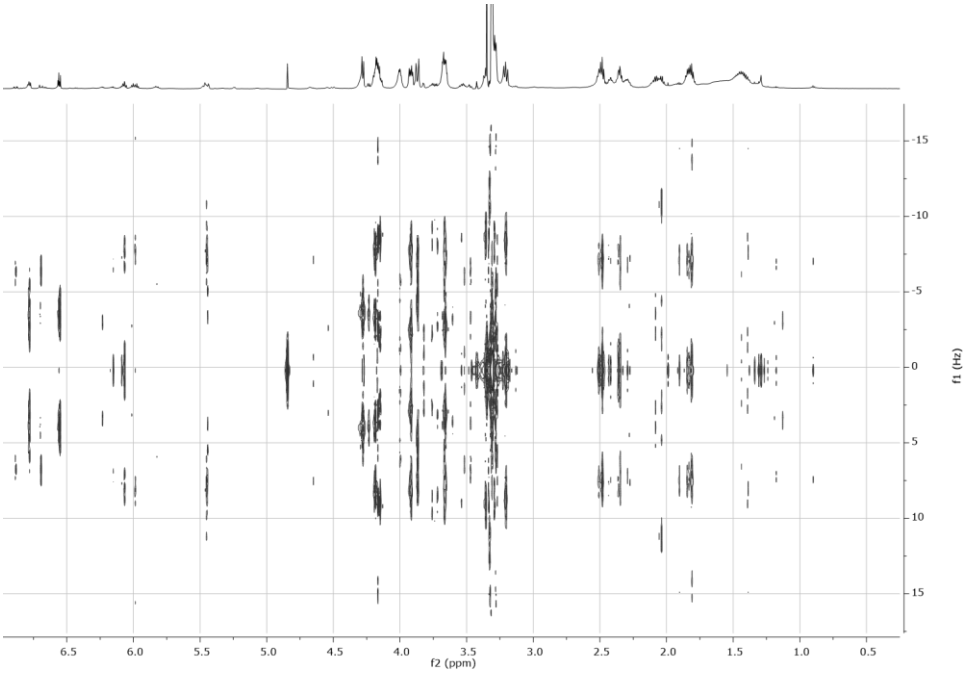


Figure S13: Two-dimensional ^1H-^1H *J*-Resolved spectrum (CH_3OH-d_4, 600 MHz) of Compound (**4**).

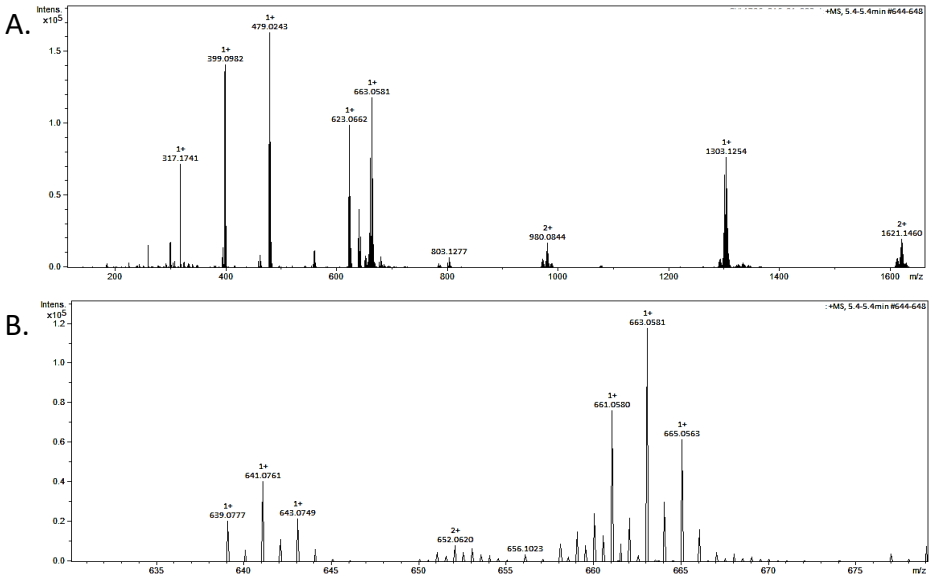


Figure S14: Electrospray ionization-quadrupole-time of flight (ESI-qTOF) MS spectrum in full range (A). and the expanded region around the $[M+H]^+$ and $[M+Na]^+$ ions (B) of Compound (**4**).

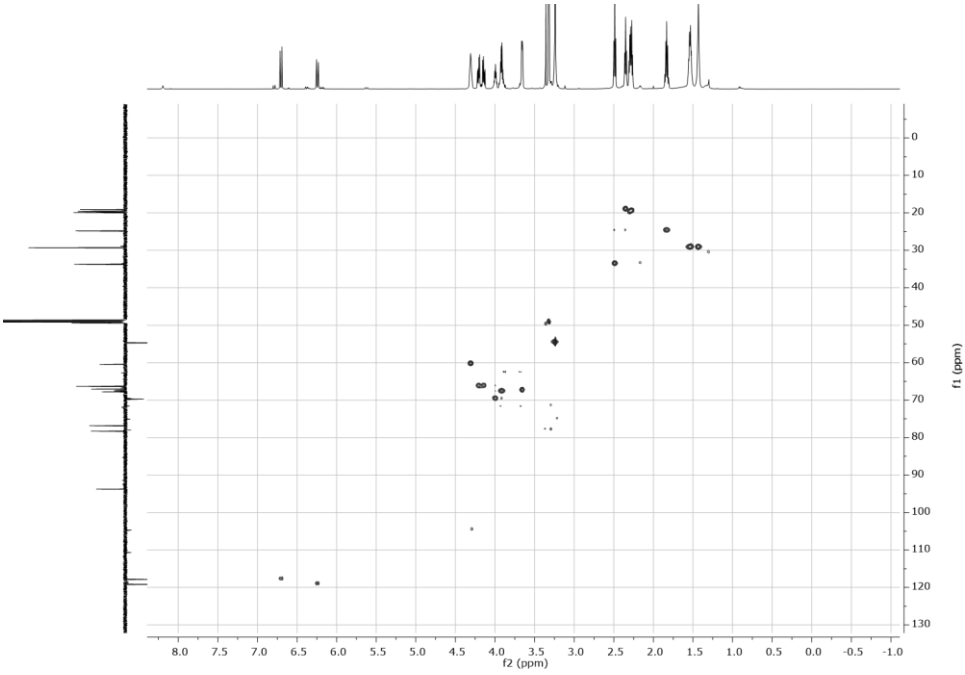


Figure S15: Heteronuclear single quantum coherence (HSQC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**5**).

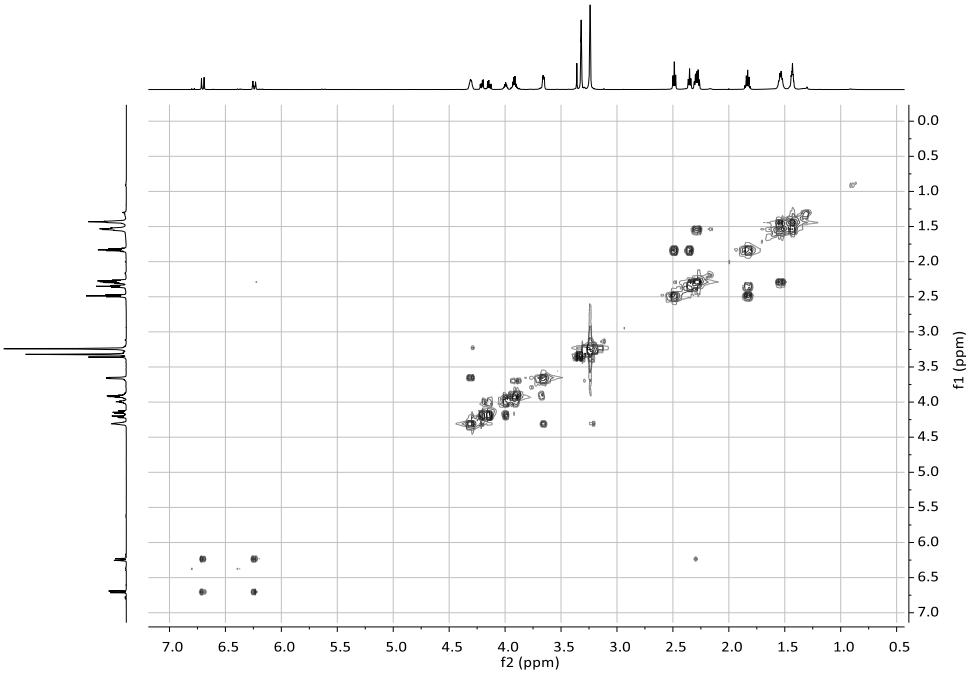


Figure S16: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**5**).

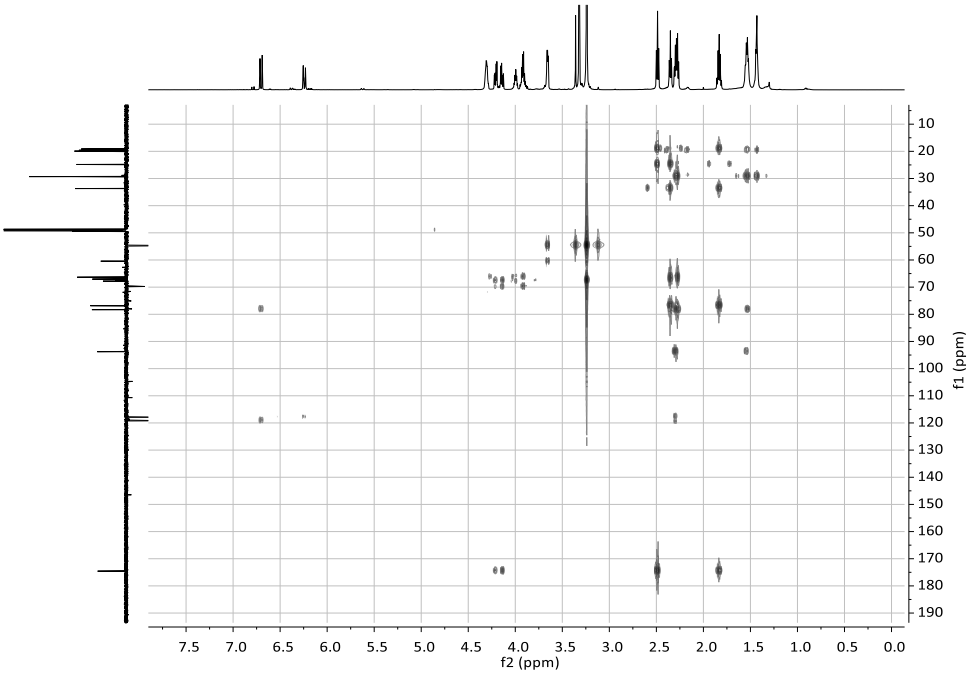


Figure S17: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**5**).

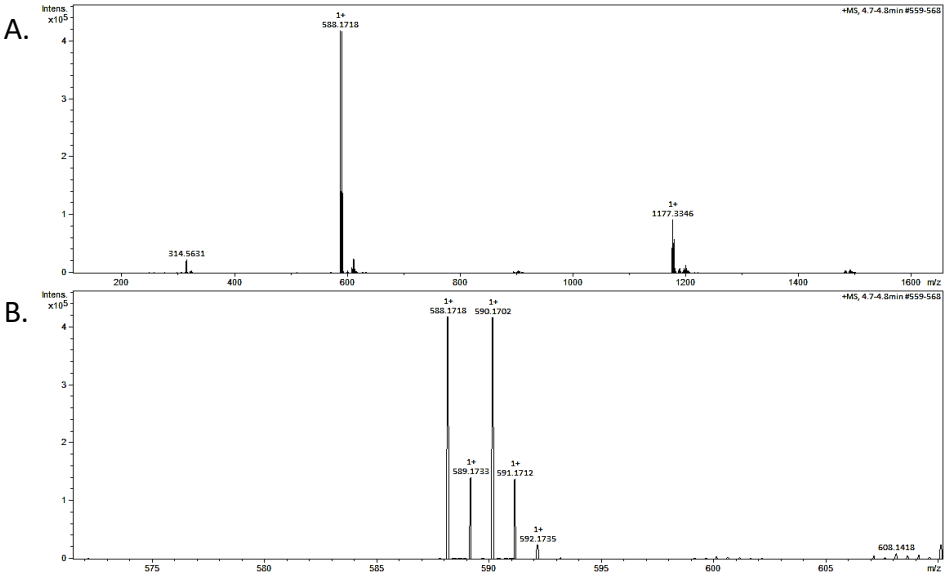
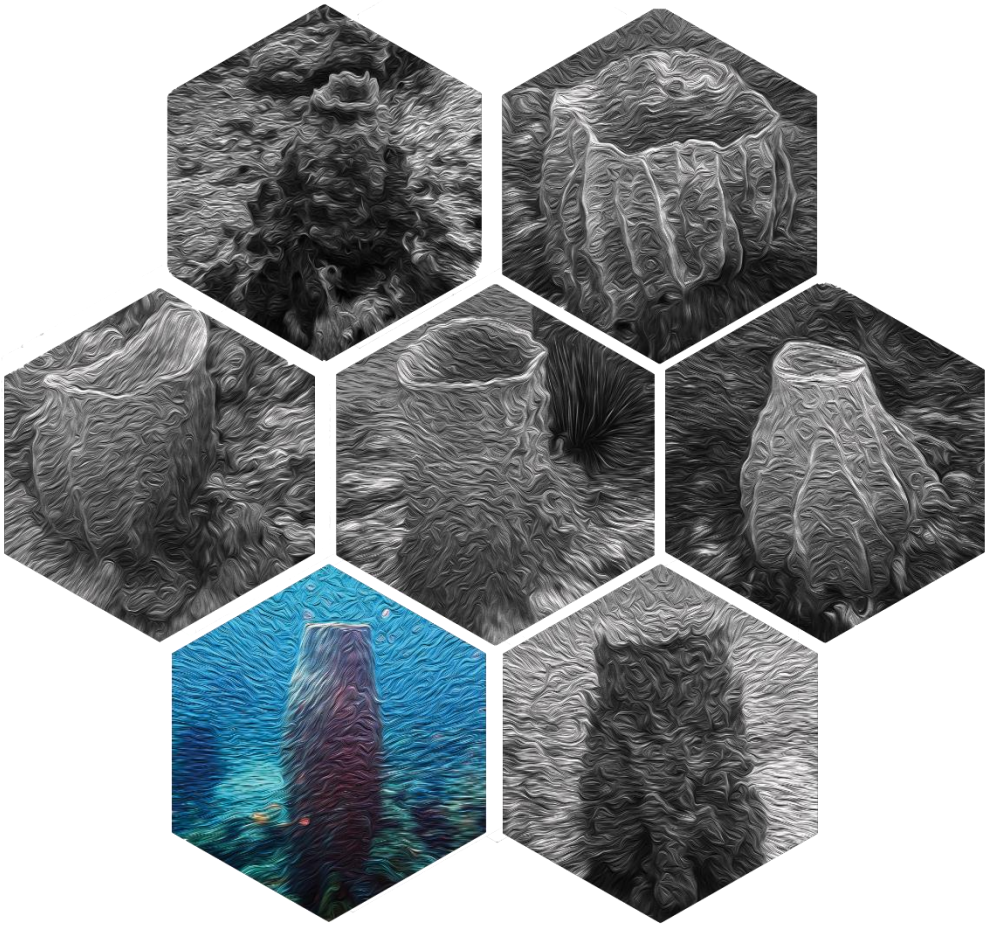


Figure S18: Electrospray ionization-quadrupole-time of flight (ESI-qTOF) MS spectrum in full range (A) and the expanded region around the $[M+H]^+$ ion (B) of Compound (**5**).



Chapter 5

Metabolic variation in Caribbean giant barrel sponges: influence of age and sea-depth

Lina M. Bayona[1], Min-Sun Kim[2], Thomas Swierts[3], Geum-Sook Hwang[4], Nicole J. de Voogd[3,5], Young Hae Choi[1]

[1] Natural Products Laboratory, Institute of Biology, Leiden University, 2333 BE Leiden, The Netherlands.

[2] Food Analysis Center, Korea Food Research Institute, Wanju, Korea.

[3] Naturalis Biodiversity Center, Marine Biodiversity, 2333 CR Leiden, The Netherlands.

[4] Integrated Metabolomics Research Group, Western Seoul Center, Korea Basic Science Institute, Seoul, Korea.

[5] Institute of Environmental Sciences, Leiden University, 2333 CC Leiden, The Netherlands.

Manuscript submitted to marine environmental research

Abstract

The biochemical differentiation of widely distributed long-living marine organisms according to their age or the depth of waters in which they grow is an intriguing topic in marine biology. Especially sessile life forms such as sponges, could be expected to actively regulate biological processes and interactions with their environment through chemical signals in a multidimensional manner. Conventional targeted analysis is too limited to be able to give comprehensive answers to this interesting, yet complex question. In recent years, the development of chemical profiling methods such as metabolomics provided an approach that has encouraged the investigation of the chemical interactions of these organisms. In this study, LC-MS based metabolomics followed by feature based molecular networking (FBMN) was used to explore the effects of both biological and environmental factors on the metabolome of giant barrel sponges, chosen as model organisms as they are distributed throughout a wide range of sea-depths. Samples of giant barrel sponges belonging to two different genetic groups and of different ages, were collected along a significant depth gradient (7-43 m). Using molecular networking, metabolites were deduced mainly for the families of compounds present in the samples. To identify the metabolites influenced by each factor, further multivariate data analyses were performed including the correlations between the metabolome and the considered factors, i.e., age and depth. It was found that depth only caused significant changes in the metabolome of one of the genetic groups. The effect of sponge age, on the other hand, had a significant effect on the metabolome of sponges, independently of their genetic group or depth of growth. Among the identified chemical components of the sponges, the phospholipid profile was found to be most affected by age and depth but each of these factors affected specific phospholipids. As a result, two previously unreported metabolites that were found to be related to depth, compounds **1** and **2**, were chemically elucidated using NMR and MS.

Keyword: *Xestospongia muta*, metabolomics, phospholipids, molecular networking, environmental condition, sponge growth.

1. Introduction

In recent years, coral reefs around the world have experienced substantial changes in their biodiversity as a response to changes in the environment resulting from both natural and anthropogenic phenomena (Hughes et al. 2007; Pandolfi et al. 2003). Global warming causes an increase in sea surface temperatures that leads to massive bleaching events of corals and are the primary driver of recent global reef decline (Hughes et al. 2017). The decay of corals has a profound influence on other organisms in the reef, including sponges and algae, which are taking over the substrate space freed by the coral decline (Bell et al. 2013, 2018; McManus and Polsenberg 2004). Sponges in particular, play a vital role in reef life, not only because they are the most abundant taxa in Caribbean reefs, but also because they are the responsible for the transfer of nutrients along the trophic chain (Diaz and Rützler 2001; Pawlik and McMurray 2020; Rix et al. 2018).

Giant barrel sponges, which belong to the genus *Xestospongia*, have been identified as a important reef player. Due to their large size and barrel-shape appearance, these sponges provide a habitat for many other animals (Swierts, et al. 2018a). In fact, the Caribbean giant barrel sponges can cover up to 9% of the available surface (Zea 1993), and recent studies have shown that the population growth rate of *Xestospongia muta* has been rapidly increasing (McMurray et al. 2015), to the point that they are one of the most abundant organisms in reefs (Loh and Pawlik 2014). They also play an active part in the regulation of the nitrogen and carbon cycles in their habitat (Fiore et al. 2013; Southwell et al. 2008)

Xestospongia muta is characterized as a high microbial abundance (HMA) sponge with a slow growth rate and fluctuating chemical defenses (Gloeckner et al. 2014; Loh and Pawlik 2014; Pawlik et al. 1995). The interaction of these sessile organisms with their environment, as well as their biological processes, must be mediated essentially, through chemical signals. It can be anticipated, thus, that changes in environmental conditions will be reflected at a metabolic level as an alteration of the metabolome (Hay 2009; Paul et al. 2006; Pawlik 2011). As expected, a wide range of metabolites have been reported in *Xestospongia* spp. sponges, most of which have shown significant responses to bioassays, confirming that they might have a role in various biological and physiological processes (Zhou et al. 2010). However, the biotic or abiotic factors driving the production of these compounds remain unclear mainly due to the lack of methods able to deal with their chemical complexity. The traditional single-targeted method has proved to be too limited to dereplicate such chemically complex situations, and it was not till the advent of a more holistic approach such as that provided by metabolomics, that these problems could be successfully revisited.

The use of "omics" platforms to gain insight into the chemical ecology of marine sponges has changed the perspectives for their exploration (Paul et al. 2019). In particular metabolomics, defined as the study of all metabolites present in an organism under specific conditions (Viant 2007), has been very useful for environmental studies (Bundy et al. 2008). Although targeted analyses have revealed significant information on changes in metabolites resulting from environmental factors (Noyer et al. 2011; Page et al. 2005), the use of metabolomics as an unbiased method enables a much broader approach that can lead to the identification of responses associated with a variety of previously unreported compounds.

Giant barrel sponges used in this study as model organisms, have been considered as the "redwood of the reef" due to their unparalleled longevity. For instance, a specimen found in Curaçao was estimated to be up to 2300 years old, positioning *Xestospongia* as one of the oldest animals alive (Van Soest et al. 2012). With the development of a growth model for *Xestospongia*, determining the age of these sponges has become relatively easy (McMurray et al. 2008). However, the effect that age might have on the lifecycle and physiological variation of the sponge is still far from being clearly understood. In particular, age could be expected to affect the metabolome, as has been reported in studies on most other organisms, such as several terrestrial plants (Lee et al. 2019; Yoon et al. 2019), humans (Jové et al. 2014; Yu et al. 2012) and marine sponge associated bacteria (Ng et al. 2013).

Biological development of marine sponges could also very likely be greatly affected by another factor: sea-depth. Marine habitats can vary substantially depending on their distance from the sea surface due to great variations in environmental conditions such as the availability of sunlight, pH, pressure, temperature and presence of predators. Therefore, giant barrel sponges, which can be found at depths ranging from less than 10 to beyond 100 m (Olson and Gao 2013; Van Soest et al. 2014), should unavoidably be influenced by these environmental variations. Several studies on the influence of depth on the microbiome of this sponge suggested that depth gradients could cause great changes in some of the microorganisms associated with the sponge (Morrow et al. 2016; Olson and Gao 2013; Villegas-Plazas et al. 2019). Considering that *X. muta* is an HMA sponge, changes in the microbiome could be expected to have a significant effect on the metabolome of the sponge. Notably, notwithstanding the importance of this relationship in the field of marine ecology, the studies on the effect of depth on the metabolic production of giant barrel sponges are still limited.

Recent studies have shown that the species previously known as *X. muta* is in fact, a species complex that can be separated into three distinct genetic groups (cryptic species) according to their mitochondrial DNA (Swierts et al., 2017). In the present study, samples belonging to

X. muta of two different genetic groups, different ages and depths were collected in Curacao and chemically profiled in a holistic manner to detect possible variations in their metabolome. For this, liquid chromatography coupled with mass spectrometry (LC-MS) was used as an analytical platform, combined with multivariate data analysis (MVDA) and a molecular networking workflow. The high sensitivity of LC-MS, one of its outstanding features, allowed the detection of metabolites present in very low concentrations (Goulitquer et al. 2012). After applying MVDA to the obtained data, a molecular networking workflow allowed it interpretation from a broader perspective.

2. Result and Discussion

Sponges are so far one of the most studied sources of marine natural products (MNP). Among these, giant barrel sponges have been most extensively studied since their widespread distribution and relatively large size facilitates their identification and collection. Their chemical composition includes sterols, brominated fatty acids and terpenoids (Zhou et al. 2010). However, the qualitative and quantitative composition of their metabolome was found to vary depending on environmental conditions (Bayona et al. 2020; Villegas-Plazas et al. 2019). The correlations between changes in the metabolome of giant barrel sponges and two possible driving factors of its variation (age and depth) were investigated using LC-MS, given its capacity to detect very low concentrations of most metabolites. Samples were collected in the Caribbean Sea off the coast of Curaçao. With the recent findings of Swierts and co-workers (Swierts et al. 2017) regarding the existence of three genetic groups among what had been previously considered to be only *X. muta* in mind, the collected samples were studied and classified accordingly. Thus, among the 69 samples collected, 41 belonged to genetic group 7 and 28 to genetic group 8 (Swierts et al. 2017). Considering that these genetic groups could have different responses to biotic or abiotic factors such as age or depth, the interaction between genetic groups and these two factors was also included in the study.

The effect of age, depth and genetic group on the chemical profile (LC-MS data) was firstly tested using a PERMANOVA analysis (Table 5.1). Among the factors evaluated, age and genetic group were found to significantly influence the variation of the sponge metabolome, whereas depth correlation was only marginally significant. The effect of the interaction between the studied factors was also calculated. In this case, no correspondence was found between metabolomic changes related to sample age and depth and/or genetic groups, that is, the changes in the metabolome caused by the aging processes did not appear to be affected by the genetic group nor by the depth of the sample collection. On the other hand, while the effect of depth was marginally significant when considering the whole sample set, it did seem

to vary according to the genetic group. This is supported by the fact that the interaction between the depth of collection and the genetic groups was found to be significant.

Table 5.1: Permutational multivariate analysis of variance (PERMANOVA) testing the effect of depth of sample collection, age, and their genetic group of the samples on the metabolome.

	F-value	R^2	P value
Age	$F_{(1,69)} = 2.45$	0.024	0.014*
Depth	$F_{(1,69)} = 2.04$	0.020	0.050
Genetic group	$F_{(1,69)} = 30.16$	0.300	<0.001*
Age*Depth	$F_{(1,69)} = 0.78$	0.007	0.593
Age* Genetic Group	$F_{(1,69)} = 0.64$	0.006	0.781
Depth* Genetic Group	$F_{(1,69)} = 2.57$	0.026	0.020*
Age*Depth*Genetic group	$F_{(1,69)} = 0.79$	0.008	0.596

*Significant factors confirmed by PERMANOVA.

In view of the significance of their interaction, the effect of depth on each genetic group was analyzed individually. The PERMANOVA results showed the metabolome of genetic group 7 to be largely affected by the depth ($F_{(1,45)} = 3.018$, $R^2 = 0.064$, $p = 0.003$) and this was also confirmed by the results of a canonical principal coordinate analysis (Figure 5.1). In contrast, the metabolome of genetic group 8 was not clearly correlated with the depth factor (PERMANOVA, $F_{(1,27)} = 1.457$, $R^2 = 0.053$, $p = 0.112$). This finding is in accordance with previous studies that identified depth as one factor that could influence the chemical composition of *X. muta* (Villegas-Plazas et al. 2019). However, the dependence of a particular genetic group on changes due to depth has never been studied.

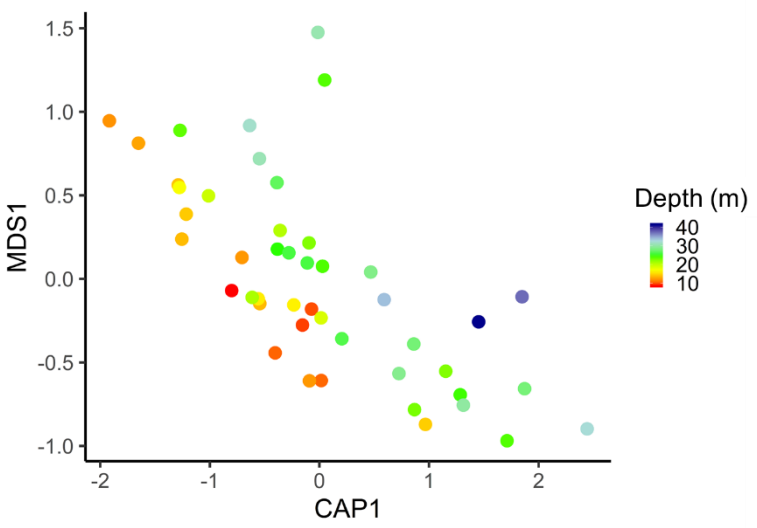


Figure 5.1: Canonical principal coordinate analysis between depth (m) factor and metabolome of the genetic group 7. Constrain canonical variable 1 (CAP 1) vs non constrain variable out of multidimensional scaling (MDS1).

The principal coordinate analysis (PCoA), an unsupervised multivariate data analysis, revealed differences between the metabolomes of the two phylogenetic groups (Figure 5.2a) Samples belonging to group 7 were found to have more unique features than group 8, consisting mostly of secondary metabolites (Figure 5.2b). The higher chemical richness of group 7 can be associated with its higher potential for producing metabolites implying that sponges belonging to group 7 are more likely to produce different secondary metabolites as a response to changes in environmental factors along the depth gradient. Furthermore, considering that *X. muta* is an HMA sponge, the variations observed in its metabolome could be associated with changes in the metabolism of the host itself (sponge), in the metabolism or the composition of its symbionts (microbiome) or changes in the metabolic interaction between the sponge and the symbionts. The alteration of the microbiome of giant barrel sponges in the Caribbean due to environmental factors related to differences in depth has been reported (Morrow et al. 2016; Villegas-Plazas et al. 2019) as was the influence of the phylogenetic group of the host on the composition of the microbiome of *Xestospongia* spp (Swierts 2019; Swierts et al. 2018b). From this perspective, the changes in the metabolome due to depth could also be attributed not only to different responses according to the phylogenetic group but also to differences in their specific symbionts

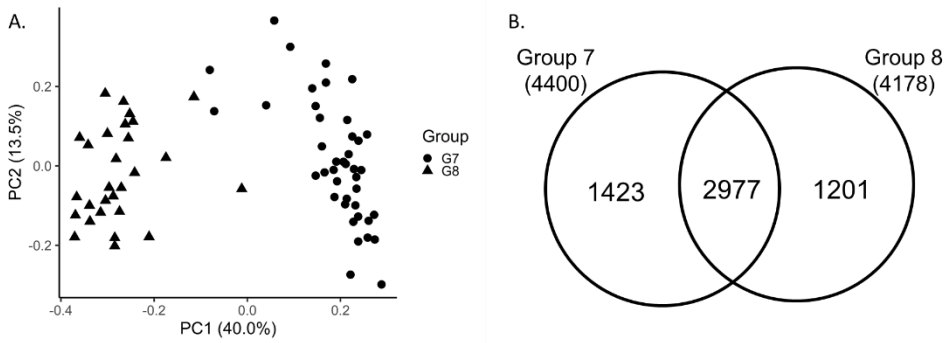


Figure 5.2: A: Principal coordinate analysis of 69 *Xestospongia* spp. samples using LC-MS data, principal coordinate 1 (PC1) Vs principal coordinate 2 (PC2). B: Venn-Diagram of the features obtained from the LC-MS data matrix in the intercept shows features present in both genetic groups.

The LC-MS data obtained from samples belonging to genetic group 7 was analyzed using the molecular networking workflow in the GNPS platform (Nothias et al. 2020; Wang et al. 2016) followed by *in silico* dereplication methods such as network annotation propagator (da Silva et al. 2018) and MolNetEnhancer (Ernst et al. 2019). This resulted in 40 molecular networks that contain 3 or more nodes, 20 of which could be classified into a molecular family according to the ClassyFire chemical ontology (Djoumbou Feunang et al. 2016). Most of the classified networks corresponded to lipids that include both fatty acids and glycerophospholipids or lipophilic compounds such as terpenoids and steroids while others were classified as polyketides and aromatic compounds such as benzoic acid derivates or quinones. To gain insight into the metabolites influenced by the depth gradient, the 15 features that most contributed in each direction of the first axis in the canonical principal coordinate analysis - positive (deep) and negative (shallow) - were selected for further investigation. As is shown in Figure 5.3, most of the features related to samples collected at shallow depths were grouped in one cluster and classified as glycerophospholipids. A closer look at the MS spectra of these compounds revealed that they all contain a phosphatidylcholine moiety. In addition, considering their molecular weight, it was deduced that they only contain one fatty acid bonded to the glycerol unit, implying that they belong specifically to the *lyso*-phosphatidylcholine lipids (*lyso*-PC) family. Lastly, the isotopic pattern showed the presence of bromine atoms in some of the compounds.

Table 5.2: ^{1}H-NMR and ^{13}C-NMR (600 MHz, CH_3OH-d_4) of compounds **1** and **2**

	Compound 1			Compound 2		
	δ_H (ppm)	mult, *J* (Hz)	δ_C (ppm)	δ_H (ppm)	mult, *J* (Hz)	δ_C (ppm)
1'	---	---	174.8		---	175.1
2'	2.48	t, 7.4	33.8	2.39	t, 7.5	34.2
3'	1.81	quint, 7.2	25.2	1.74	quint, 7.4	25.1
4'	2.35	td, 7.0, 1.9	19.4	2.35	qd,7.4, 1.1	30.5
5'	---	---	88.1	5.89	dt, 10.7, 7.5	142.6
6'	---	---	80.9	5.61	brd, 10.7	111.3
7'	5.45	dm, 15.8	111.3		---	85.6
8'	5.99	dt, 15.8, 7.1	144.0		---	93.7[a]
9'	2.08	m	33.6	5.66	dq, 15.8, 1.8	111.3
10'	1.40	m	29.5	6.10	dt, 15.8, 7.1	144.9
11'	1.44	m	29.1	2.17	qd, 7.0, 1.4	33.5
12'	2.04	m	32.0	1.53	m	29.1[b]
13'	6.07	td, 7.7, 1.4	137.4	1.53	m	28.9[b]
14'			114.8	2.30	m	19.8
15'	6.78	dm, 7.6	132.3		---	93.6[a]
16'	6.56	d, 7.6	113.4		---	78.3
17'	---	---	---	6.24	dt, 14.0, 2.3	119.2
18'	---	---	---	6.70	d, 14.0	118.0
1	4.19, 4.12	dd, 11.4, 4.5, dd, 11.4, 6.2	66.3	4.17, 4.12	dd, 11.4, 4.7, dd, 11.4, 6.1	66.2
2	3.98	m	69.8	3.98	m	69.8
3	3.90	m	67.8	3.90	m	67.8
1''	4.30	m	60.4	4.10	m	61.6
2''	3.64	m	67.5	3.24	tm, 4.8	50.9
N-Me	3.23	s	54.7	2.73	s	33.5

[a] Interchangeable carbons, [b] interchangeable carbons

The dereplication of the selected features led to the identification of one brominated polyacetylene fatty acid and a *lyso*-phosphatidylcholine lipid that contains palmitic acid as the acyl moiety. Further studies of the features of the glycerophospholipids network that showed different brominating patterns led to the identification of a previously unreported compound, 1-*O*-((7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoyl)-sn-glycero-3-phosphocholine (Compound **1**). This compound was isolated as a white powder. The full elucidation of this compound was done using both NMR (Table 5.2) and HRMS. Its (+)-HRMS

spectrum showed the proton adduct $[M+H]^+$ ions at *m/z* 642.0835, 644.0836, and 646.0803 corresponding to the molecular formula, $C_{24}H_{38}Br_2NO_7P^+$ (calc. 644.0810, 642.0831, 646.0790); the presence of the two bromine atoms was deduced from the isotopic pattern. The ^{13}C-NMR spectrum showed the presence of a carbonyl carbon at δ_C 174.8 in addition to six sp^2 carbons and 2 *sp* carbons corresponding to three alkenes and one alkyne, respectively, indicating the presence of an acyl moiety. The carbons attached to heteroatoms (oxygen and nitrogen) in the region between δ_C 54 and 70 suggested the presence of one glycerol and one phosphatidylcholine moiety. The two bromine atoms observed in the mass spectra were found to be bonded to carbons involved in double bonds as shown by the shift in the chemical shift of these carbons. The full structure of compound **1** was determined using the COSY and HMBC spectra. The brominated fatty acid and phosphatidylcholine were found to be bonded to C1 and C3 of the glycerol molecule, respectively. Using the coupling constant, the stereochemistry of the double bonds was established as 7*E*,15*Z.* The stereochemistry of the double bond in C'-13 was determined to be *E* due to the correlation observed in the NOESY between protons in position 12' and 15'.

Lyso-phosphatidylcholines *(lyso-PC)* account for only 0.3 to 4% of the total lipid content in sponge extracts (Genin et al. 2008), coinciding with studies conducted in animal cell membranes (Torkhovskaya et al. 2007). Lysophospholipids have been recognized as playing a role beyond being part of the membrane structure as molecular signaling molecules (Birgbauer and Chun 2006; D'Arrigo and Servi 2010; Torkhovskaya et al. 2007). These types of molecules have been related to different mechanism of cell proliferation , such as the induction of growing factors synthesis or increasing the sensitivity of cells to growth factors (Torkhovskaya et al. 2007). In addition, similar molecules have been reported to be involved in the self-recognition and activation of immune systems though pro-inflammation signaling in sponges and corals (Müller and Müller 2003; Quinn et al. 2016). Therefore, the increase in the production of *lyso*-PC in shallow depth sponges could be related to a defense mechanism developed as a response to higher predatory stress in shallower waters (Chanas and Pawlik 1997). A similar phenomenon has been observed in sponges of the genus *Oscarella* that displayed a seasonal variation of two lysophospholipids with an increased production that coincided with the embryogenesis and the larval development period (Ivanisevic et al. 2011). This shows that these types of molecules can perform a variety of roles in the development of the sponge. However, the variation in their structures makes it difficult to establish specific roles for each of them. Lastly, even though lysophospholipids in general, and *lyso*-PC in particular, are universally present among metazoans, some chemical features in the acyl chain present in compounds from *Xestospongia*, such as bromine atoms and acetylenic groups, are

characteristic of these specific organisms (Zhou et al. 2010; Zhou et al. 2015). The fatty acids from *Xestospongia* have been reported to have a wide range of biological activities, suggesting that it is not only the lysophospholipid that has a biological role in this case, but the free fatty acid which could *per se* play an additional role in the sponge.

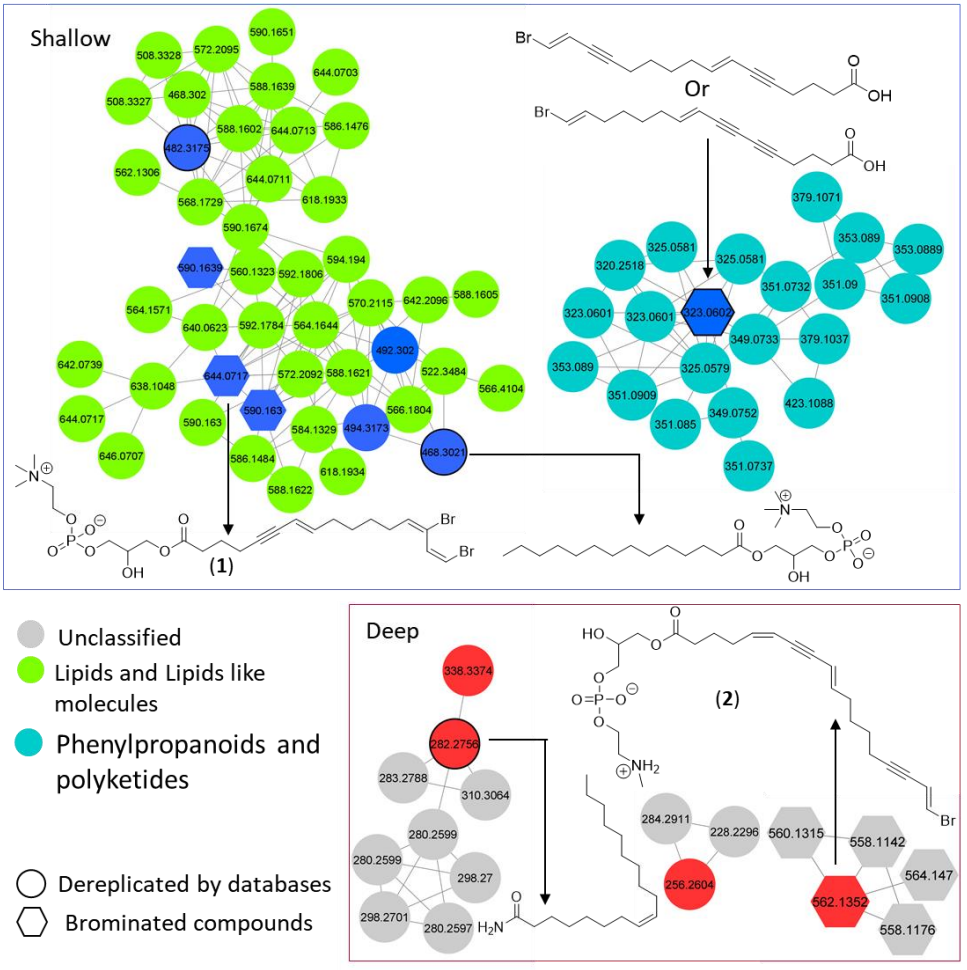


Figure 5.3: Molecular network showing the clusters of the features identified to be changing due to depth. The nodes corresponding to important features obtained from the canonical analysis of principal coordinate are colored in blue (shallow) and red (deep)

For samples collected at increasing depth, the features that were identified as important were not grouped in one single cluster. Instead, the significant compounds were found in small clusters within the molecular networking analysis, as shown in Figure 5.3. This indicates that the environmental changes along the depth gradient do not cause a change in the production

of a specific group of molecules, but rather, that the changes observed in the metabolome are compound specific. One of the reasons why the environmental changes along the depth gradient produce a response involving specific compounds could be that not all environmental and biotic factor variations are directly related to depth. While abiotic factors such as pressure and temperature are related to depth, other factors such as predatory stress can vary along the depth gradient. This can cause heterogenicity in the compounds whose production is altered at larger depths. Among the compounds that were identified as associated with greater depths it was possible to isolate another previously unreported compound, 1-*O*-((5*Z*,9*E*,17*E*)-18-bromooctadeca-5,9,17-trien-7,15-diynoyl)-sn-glycero-3-*N*-Methylethanolamine (compound **2**).

Compound **2** was also isolated as a white powder. Its (+)-QTOF-ESI-MS spectrum showed the proton adduct $[M+H]^+$ ions at *m/z* 560.1394 and 562.1379, with a relative intensity of1:1, indicating the presence of one bromine atom in the molecule. Consequently, the molecular formula was established as $C_{24}H_{36}BrNO_7P$ (calc. 560.1413 and 562.1392). Its NMR spectroscopic data (Table 5.2) was used to characterize the structure of this compound. The carbonyl carbon at δ_C 175.1 together with the *sp* and sp^2 carbons confirmed the presence of an acyl chain in the molecule. In addition, the absence of a terminal methyl or methylene group together with the unusual chemical shift of the carbons in terminal double bond at δ_c 119.2 and 118.0 confirmed the presence of a bromine atom in this position. The full structure of compound **2** was determined using the COSY and HMBC. The compound was determined to be formed by three moieties, a brominated unsaturated fatty acid, a glycerol molecule and *N*-methylethanolamine**.** The brominated fatty acid and *N*-methylethanolamine are bonded to C1 and C3 of the glycerol molecule, respectively. Using the coupling constant, the stereochemistry of the double bounds was established as 5*Z*,11*E*,17*E*.

Similar to the compounds found in shallower waters, compound **2** also belongs to the lysophospholipids family. However, the presence of a *N*-methylethanolamine moiety is very unusual. Although ethanolamine and choline are moieties commonly found in animal cells, *N*-methylethanolamine has been mostly reported to be present in the membrane of several by microorganisms (Dahal and Kim 2017; Goldfine and Ellis 1963; Schubotz et al. 2011). This suggests that compound **2** is produced totally or partially by an associated microorganism. As was discussed previously, the *Xestospongia* microbiome experiences changes due to depth. The fact that one of the metabolites associated with changes in depth contains a moiety produced by microorganisms is an example of the way in which changes in the microbiome can be reflected in changes in the metabolome of the holobiont.

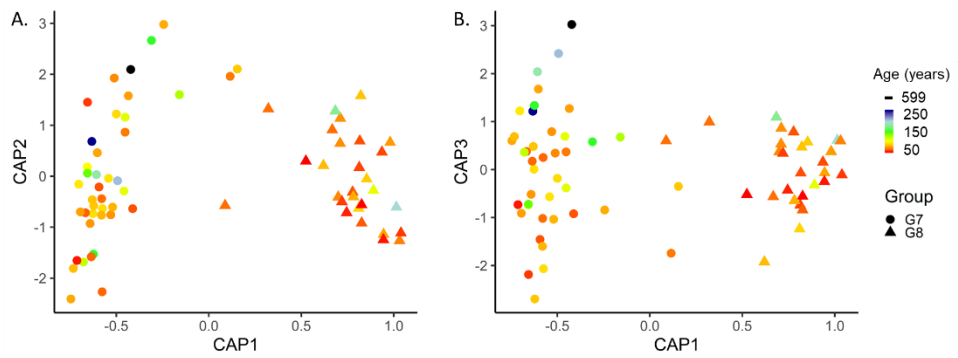


Figure 5.4: Canonical principal coordinate analysis using age (years), depth (m) and genetic group as factors for samples of genetic group 7 and 8. A: CAP 1 vs 2 and B: CAP 1 vs 3.

In the second part of the study the effect of age on the metabolome of the sponge was investigated (Figure 5.4). The metabolome of sponges proved to be affected by age, providing an important insight into their biological and chemical processes. From a biological perspective, this could imply that in sponges, the biological stage of development of a sponge is reflected by the metabolome as observed in other organisms (Yoon et al. 2019; Yu et al. 2012). For sponges in general, and particularly in the case of giant barrel sponges, the finding that the metabolome can be an indicator for the development of the sponge is important, as very little is known about the biological development of these sponges. Moreover, these findings suggest that external factors could also affect the metabolome, as older sponges have had to adapt over time to more variations in the conditions of the water column, for example in the microbiome, the temperature or type and amount of nutrients. Therefore, it is possible that over the years, older sponges have, for example, recruited specific symbionts into their metabolome that afforded an adaptive advantage (Ribes et al. 2016; Turon et al. 2018).

The fact that the interaction between age and depth does not affect the metabolome is in agreement with the lack of any influence of depth on growth rate of *X. muta* (McMurray et al. 2008). Similarly, the interaction between age and genetic groups showed no significant effect on the metabolome, possibly because growth rates have been calculated for *X. muta,* but not for the individual genetic groups. Each genetic group could have different growth rates, as suggested by previous reports of differences in growth rate of *X. muta* and its sister species *X. testudinaria* (McGrath et al. 2018).

Considering that the effect of age was reflected in the metabolome, the second aim of this research was to identify those metabolites most affected by this process. The canonical principal coordinated analysis revealed the effect of age mainly CAP 2 and 3 as can be observed

in Figure 5.4. The features that were found to contribute the most to this axis were selected as characteristic metabolites associated with age. Subsequently, a molecular network was built using the GNPS platform followed by *in silico* tools to increase the annotation of the features. This showed that most of the clusters in which the nodes corresponding to the compounds related to age were located were not classified. However, a closer look into the MS spectra of the selected compounds allowed us to draw some conclusions about the chemical structures of these compounds.

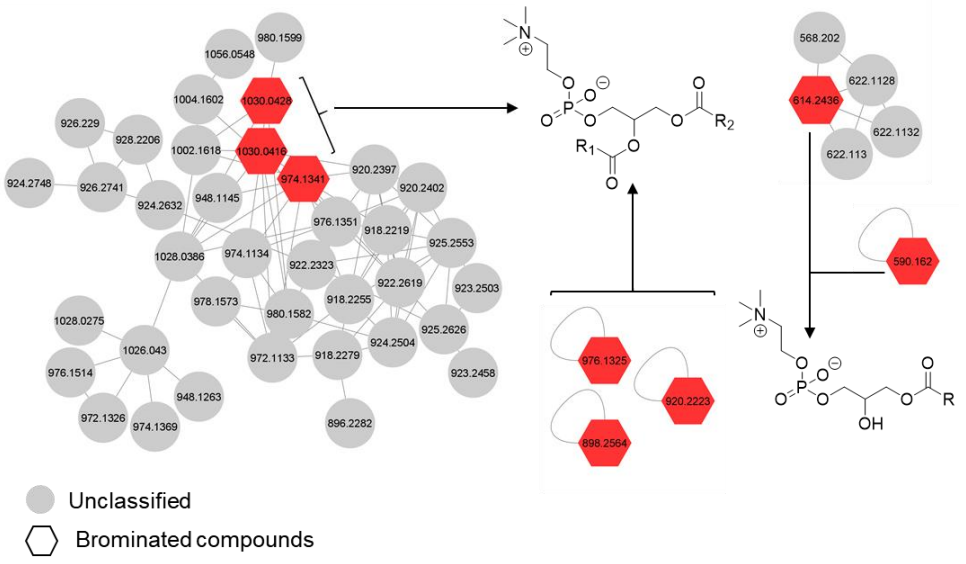


Figure 5.5: Molecular network showing the clusters of the features identified to be changing depending on the age of the sponge. The nodes corresponding to features related to older sponges obtained from the canonical analysis of principal coordinate are colored in and red.

As shown in Figure 5.5, most of the compounds that were identified to be related to older sponges displayed *m/z* values between 800 and 1000. This, together with their fragmentation pattern, indicated that the structure of these compounds corresponded to glycerophospholipids with one phosphatidylcholine and two acyl chains joined to the glycerol moiety. Further, the isotopic pattern of these compounds revealed the presence of 1 to 5 bromine atoms. A possible explanation for their presence could be that were cellular division increased in older sponges these compounds could also be increased as they are a constitutive element in membranes. Considering that it has been reported that the growth rate of giant barrel sponges decreases with the age of the sponge (McMurray et al. 2008), it is possible that the increment in the cellular division associated with these membrane lipids is related to higher cell turnover rates rather than a growing state (Alexander et al. 2014). Cell turnover

rates have proven to be particularly fast in sponges compared to other animals and it is believed to be a strategy to prevent cellular damage (De Goeij et al. 2009). Another alternative could be that these molecules act as a sort of reservoir of the bioactive polybrominated fatty acids that are bonded to the glycerol. Polybrominated fatty acids have been reported to exhibit a wide range of biological activities, such as antiviral, antimicrobial and cytotoxic activity. Therefore, besides a mere functional role, the glycerophospholipids could act as a bank of these fatty acids that are involved in defense mechanisms and can be released in the event of situations that require their participation. However, further research in this field is needed in order to prove this hypothesis.

3. Conclusions

The metabolome of giant barrel sponges belonging to two genetic groups revealed their different response to changes along a depth gradient. Genetic group 7 showed a higher content of lysophosphatidylcholine lipids at shallower depths. These compounds have been reported to play several biological roles in animal cells indicating that environmental conditions such as light, temperature and predatory stress present in shallower waters might trigger the production of this family of compounds. Samples collected at greater depths did not reveal the overexpression of any specific group of compounds, but one of the compounds that was found to be related to these conditions contains a *N*- methylethoanolamine unit that has been reported only in bacteria, suggesting that it is the result of the interaction between the metabolism of the sponges and their associated microorganisms. Apart from depth, age was also found to affect the metabolome of giant barrel sponges with older specimens having increased amounts of glycerophospholipids. This is an important contribution as it is the first time that age has been evaluated as factor for change in the metabolome of these sponges and could be indicative of an increase in cell replication in older sponges.

4. Materials and Methods

4.1 Sample collection

Giant barrel sponges (*Xestospongia* spp) were collected by SCUBA diving in Curaçao from multiple locations with different depths around the island (appendix 2 Table S1). Collected sponge samples were immediately stored on site in 98% ethanol (w/w) at -20 °C. The samples were identified by DNA sequencing, using the I3-M11 and partition of the CO1 mitochondrial gene according to Swierts and co-workers (Swierts et al. 2013). The DNA analysis allowed the samples to be identified as belonging to three main genetic classes for giant barrel sponges in the Caribbean (groups 7 and 8) (Swierts et al. 2017). Although a few samples could be

identified as belonging to group 9, these were not included in the study as there was not a sufficient quantity to carry out statistical analysis. The age of the specimens was determined through on-field measurements according to previous work (McMurray et al., 2008). The depth at which each specimen was collected was also recorded and that value was used for the analyses.

4.2 Sample preparation for metabolomic analysis

The *Xestospongia* spp. samples were ground and extracted with ethanol and sonicated for 20 min three times. The 3 extracts were combined and desalted on 500mg C-18 SPE cartridges of 45 µm particle size (Supelco Supelclean, Bellefonte, PA, USA). For this, 50 mg of each extract were loaded onto the cartridge and eluted with water (F1), methanol (F2) and dichloromethane/methanol 1:1 (F3). The methanol fractions were analyzed by LC-MS.

4.3 Liquid chromatography- Mass spectrometry (LC-MS) analysis

The methanol fractions obtained were taken to dryness with a Centrivap vacuum concentrator (Labconco, Kansas City, MO, USA) and the obtained residue was redissolved in 50% methanol (v/v) to a final concentration of 1 mg/mL. Samples were injected in randomized sequence within each genetic group. The analysis was carried out using an UHPLC-DAD-ESI-MS system consisting of a UPLC Acquity I-Class (Waters, Milford, MA, USA) hyphenated to a Bruker Impact HD MS spectrometer (Bremen, Germany) with electrospray ionization (ESI). The separation was performed on a Waters C18 (2.1 x 100 mm, 2.1 µm) column and eluted with gradient of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) of 10% to 100% B in 30 min, and 100% B for 5 min; 100% to 10% B in 3 minutes and 2 minutes equilibration at 10% B. The flow rate was 0.3 mL/min, column temperature was 40 °C and injection volume was 2 µL. The mass spectrometer parameters were set as follows: 1.5 bar of nebulizer gas, 6.0 L/min of drying gas, drying gas was 350 °C, and capillary voltage was 4000 V. The mass spectrometer was operated in positive mode in a range of 50 to 1200 m/z.

4.4 Statistical analysis

Data files obtained from the LC–MS analyses were converted to mzXML format using Brucker Daltonics DataAnalysis (version 4.1, Bremen, Germany). LC-MS data were processed using MZMine2 (Pluskal et al. 2010). To build the feature matrix, the mass detection was performed using centroid data. The noise level was set at 10000 for MS and 100 for MS/MS. The chromatograms were built using the ADAP chromatogram builder (Myers et al. 2017) with a minimum number of scans of 3, group intensity threshold 1000, minimum highest intensity 10000 and m/z tolerance of 0.05. Chromatograms were deconvoluted using a baseline cutoff

algorithm with the following parameters: minimum peak height of 10000, peak width range of 0.02−1.00 min, and baseline level of 1000. Chromatograms were deisotoped using an isotopic peak grouper algorithm with an *m/z* tolerance of 0.05 and retention time tolerance of 0.1 min. The features of each sample were aligned using a join alignment algorithm with the following parameters: 0.05 *m/z* tolerance and 0.1 retention time tolerance. Using these parameters, three matrixes were created, one including all samples and one matrix with samples of each genetic group.

All the statistical analyses were performed using RStudio 1.2.1335 (RStudio team,2018). The PERMANOVA (Permutational analysis of variance) analyses were performed using adonis() function in the 'vegan' package (Oksanen J., 2018). This was applied to different age, sea-depth and genetic group samples, for which 9999 permutation and Bray-Curtis as dissimilarity measurement were used to investigate mutual interactions between metabolites and the factors. Canonical principal coordinate analyses were performed for the data matrixes obtained from LC-MS analysis and the factors found to be significant in the PERMANOVA capscale() function with Bray-Courtois dissimilarity as a distance parameter was used. Variation in the metabolome of genetic groups was assessed with a principal coordinate analysis (PCoA) using the cmdscale() function and the Bray-Courtois dissimilarity matrix as an input.

4.5 Molecular networking and dereplication

Feature base molecular networking (FBMN) workflow (Nothias et al. 2020) was carried out using the GNPS online platform (http://gnps.ucsd.edu) (Wang et al. 2016). Quantitative data matrix and MS/MS obtained from MZmine2 are shown in section 4.4. Files were uploaded to GNPS platform and the ions in the region between +/- 17 Da around the precursor *m/z* were removed to filter the data. Additionally, only the top six fragment ions of the MS/MS data were considered for further use to reduce data size. The mass tolerance for both precursor and the MS/MS fragment ions were set to 0.03 Da. With these tolerance and filter parameters, a molecular network was constructed, for which minimum cosine value for the edges and the matched fragments were set at 0.7 and 6, respectively. In addition, edges between two nodes were only included when both nodes were in common to their respective top 10 matching nodes. The maximum size of a molecular family was set to 100. The generated spectra in the network were then dereplicated using GNPS spectral libraries (Horai et al. 2010; Wang et al. 2016), and visualized using Cytoscape software (Shannon et al. 2003).To improve the annotation of the features in the MNs Network Annotation Propagation (NAP) an *in silico* tool was used (da Silva et al. 2018). The following NAP parameters were employed for the

annotation: exact mass error for database search, 10 ppm; $[M + H]^+$ as adduct type:, cosine = 0.7 to select inside a cluster; 10 maximum candidate structures in the graph. The result from both FBMN and NAP were introduced in MolNetEnhancer workflow (https://ccms-ucsd.github.io/GNPSDocumentation/molnetenhancer/) (Ernst et al. 2019)to classify the features into chemical classes according to ClassyFire chemical ontology (Djoumbou Feunang et al. 2016).The results of the molecular networks for the matrixes with all the samples and the samples from genetic group 7 can be found in https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=cd2ffac11ce24c32ad6b8f84ce85240e and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=48bd32ff37cc45a98575a78121bb33d2, respectively.

4.6 Isolation and identification of compounds

Two compounds were isolated from two of the ethanol extracts of the sponge samples collected at different depths. The crude extracts were fractionated using 20-mL SPE LC-18 cartridges (Supelco) eluted successively with 100 mL each of 100% water, 80% 60%, 40% 20% methanol/water mixtures, 100 % methanol and methanol-dichloromethane (1:1, v/v). The SPE fractionation yielded seven fractions labelled FS1-FS7 and FD1-FD7 for sample1 and 2, respectively. Using LC-MS profiling, the target compounds, corresponding to the top 15 most contributing features (negative and positive) to the CAP 1 from the canonical principal coordinate analysis (appendix 2 Table S2), were identified in FS5 and FD5. Successive semi-preparative HPLC using an Agilent 1200 series system (Santa Clara, CA, USA) and a Luna 5 µm, C-18, 250 mm x 10 mm column (Phenomenex, Torrance, CA, USA) eluted at a flow rate of 3.5 mL/min allowed the isolation of compound **1** from FS5 and compound **2** from FD5.

The structure of these compounds was elucidated by NMR analyses at 25°C with an AV-600 MHz NMR spectrometer (Bruker, Karlsruhe, Germany), operating at the ^{1}H-NMR frequency of 600.13 MHz, equipped with a TCI cryoprobe and Z gradient system. Deuterated methanol was used as the internal lock. Their 1-D ^{1}H-NMR and ^{13}C attached proton test (APT) spectra were recorded initially and 2-D ^{1}H-NMR techniques, COSY, HSQC, and HMBC were used to confirm the connectivity of the atoms in compound **1** and **2**

Appendix 2

Table S1: Information of depth, age, genetic group and geographical location of *Xestospongia* the samples collected in Curaçao

Code	Depth	Age	Genetic group
LMB360	28.7	28	Group 7
LMB362	16.2	82	Group 7
LMB366	9.7	18	Group 7
LMB367	8.6	37	Group 7
LMB368	23.0	60	Group 7
LMB369	26.0	25	Group 7
LMB370	29.8	33	Group 7
LMB374	27.9	40	Group 7
LMB375	28.0	54	Group 7
LMB376	20.6	47	Group 7
LMB377	13.8	141	Group 7
LMB379	14.6	599	Group 7
LMB380	23.7	9	Group 7
LMB381	13.7	36	Group 7
LMB382	30.8	47	Group 7
LMB383	22.6	125	Group 7
LMB384	9.4	18	Group 7
LMB385	12.1	43	Group 7
LMB386	12.3	62	Group 7
LMB387	24.1	63	Group 7
LMB388	24.7	52	Group 7
LMB389	37.3	28	Group 7
LMB390	31.8	40	Group 7
LMB391	32.4	44	Group 7
LMB393	24.0	53	Group 7
LMB394	12.8	23	Group 7
LMB396	26.3	42	Group 7
LMB397	21.0	209	Group 7
LMB398	22.6	13	Group 7
LMB399	16.0	21	Group 7
LMB400	12.3	65	Group 7
LMB402	10.5	266	Group 7
LMB403	14.4	50	Group 7
LMB405	10.4	179	Group 7
LMB408	18.8	72	Group 7
LMB409	24.4	104	Group 7
LMB410	24.0	141	Group 7
LMB411	17.1	96	Group 7
LMB412	34.0	85	Group 7
LMB414	26.9	61	Group 7
LMB415	14.2	23	Group 7
LMB416	14.8	10	Group 8
LMB418	24.1	8	Group 8
LMB419	24.1	15	Group 8
LMB422	19.8	25	Group 8
LMB423	15.4	17	Group 8
LMB425	20.7	25	Group 8
LMB426	24.0	25	Group 8
LMB427	28.7	171	Group 8
LMB428	32.2	45	Group 8
LMB429	22.7	35	Group 8
LMB430	18.8	9	Group 8
LMB432	34.2	49	Group 8
LMB433	14.8	18	Group 8
LMB434	9.1	32	Group 8
LMB436	18.1	34	Group 8
LMB438	9.4	195	Group 8
LMB440	17.3	41	Group 8
LMB441	33.8	20	Group 8
LMB442	15.2	16	Group 8
LMB444	22.3	52	Group 8
LMB445	15.9	13	Group 8
LMB446	17.7	82	Group 8
LMB448	17.2	58	Group 8
LMB452	21.0	14	Group 8
LMB453	10.0	30	Group 8
LMB455	19.4	8	Group 8
LMB456	16.1	38	Group 8
LMB457	19.5	55	Group 8

Table S2: Top 15 most contributing features (negative and positive) to the CAP 1 from the canonical principal coordinate analysis

m/z value	Retention time (min)	CAP 1 value
590.1722	14.9	0.066929
1030.038	24.0	0.072301
576.4033	19.3	0.074658
664.454	19.2	0.077708
616.4698	30.2	0.077844
892.1917	25.1	0.084319
562.1453	15.0	0.085022
620.4288	19.3	0.08547
751.5224	33.9	0.086833
338.3374	29.3	0.118655
896.218	25.5	0.121921
793.4967	30.8	0.126926
256.2603	23.5	0.147761
593.3947	30.2	0.182476
282.2756	24.3	0.707957
590.1629	15.4	-0.17504
496.3693	19.5	-0.16223
644.0717	14.1	-0.146
494.3173	16.3	-0.13164
468.3021	14.5	-0.11018
482.3175	15.0	-0.08795
482.3174	16.5	-0.07844
323.0601	19.8	-0.07703
590.1638	14.8	-0.07477
492.3019	15.2	-0.06393
663.0504	15.5	-0.06306
510.3483	18.6	-0.05543
721.4981	33.6	-0.05309
482.3174	15.5	-0.05242

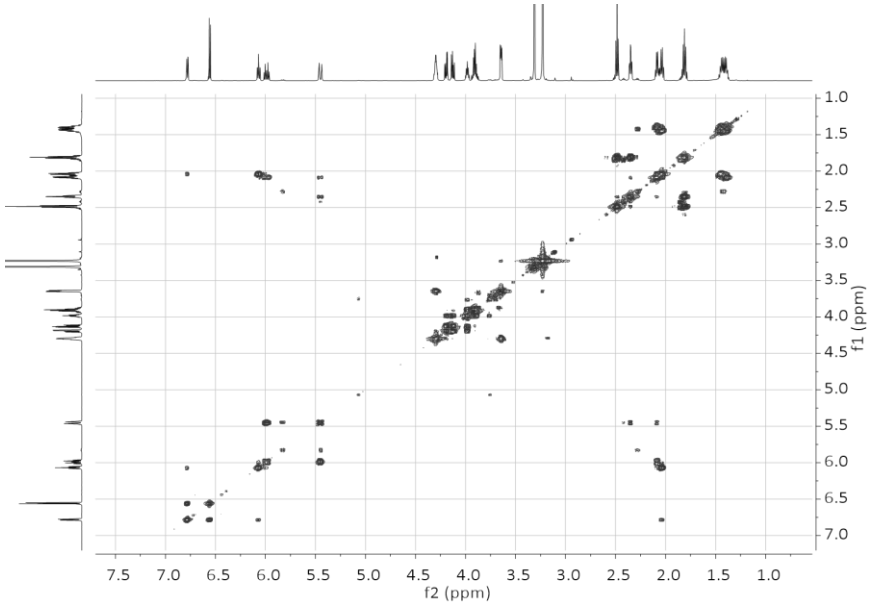


Figure S1: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**1**).

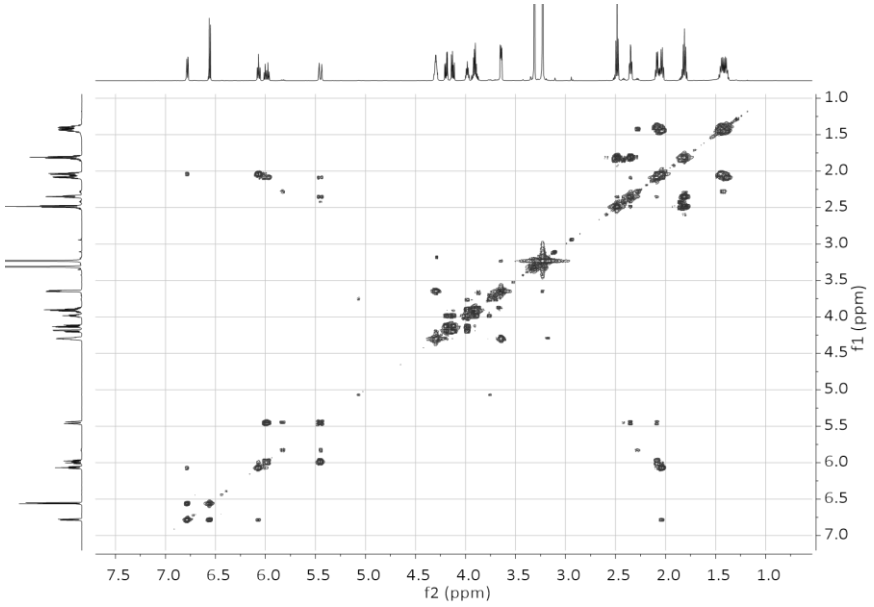


Figure S2: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**1**).

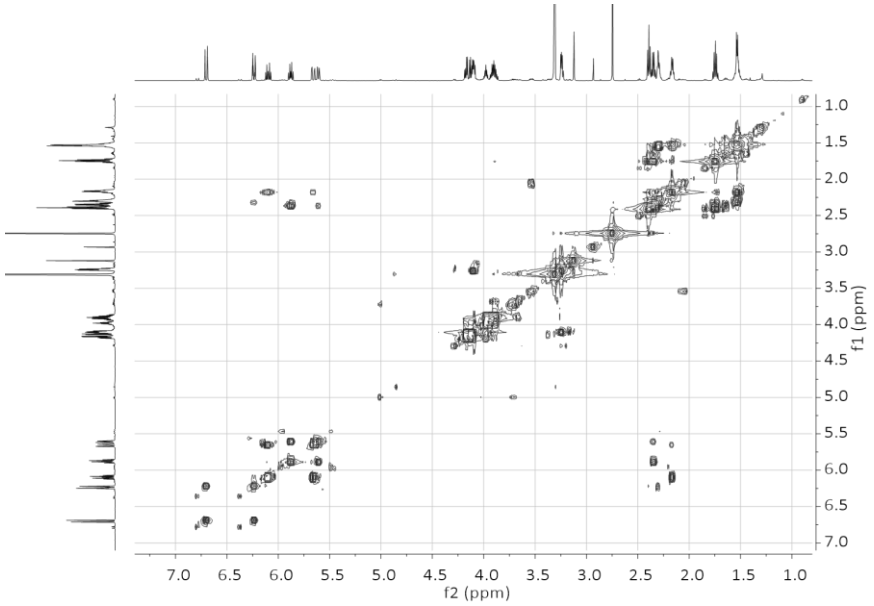


Figure S3: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**2**).

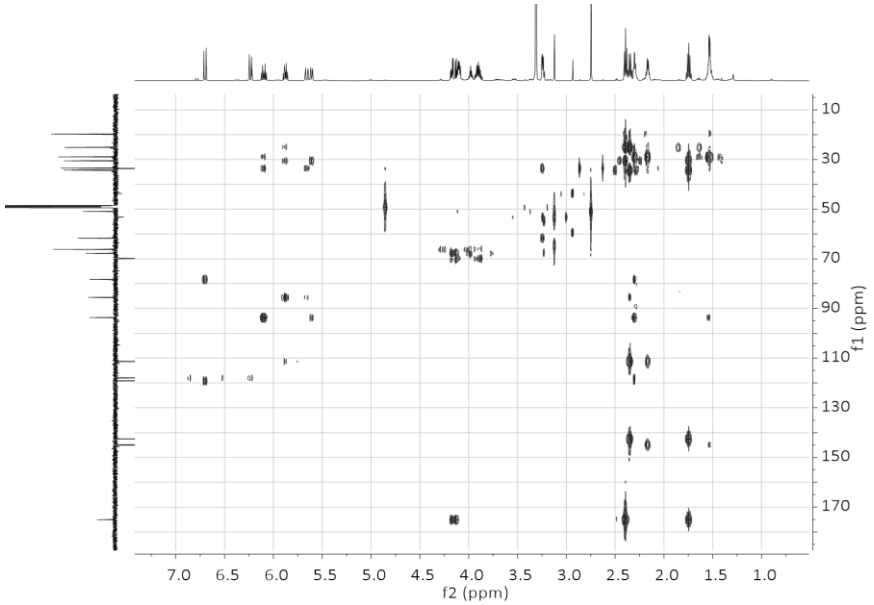
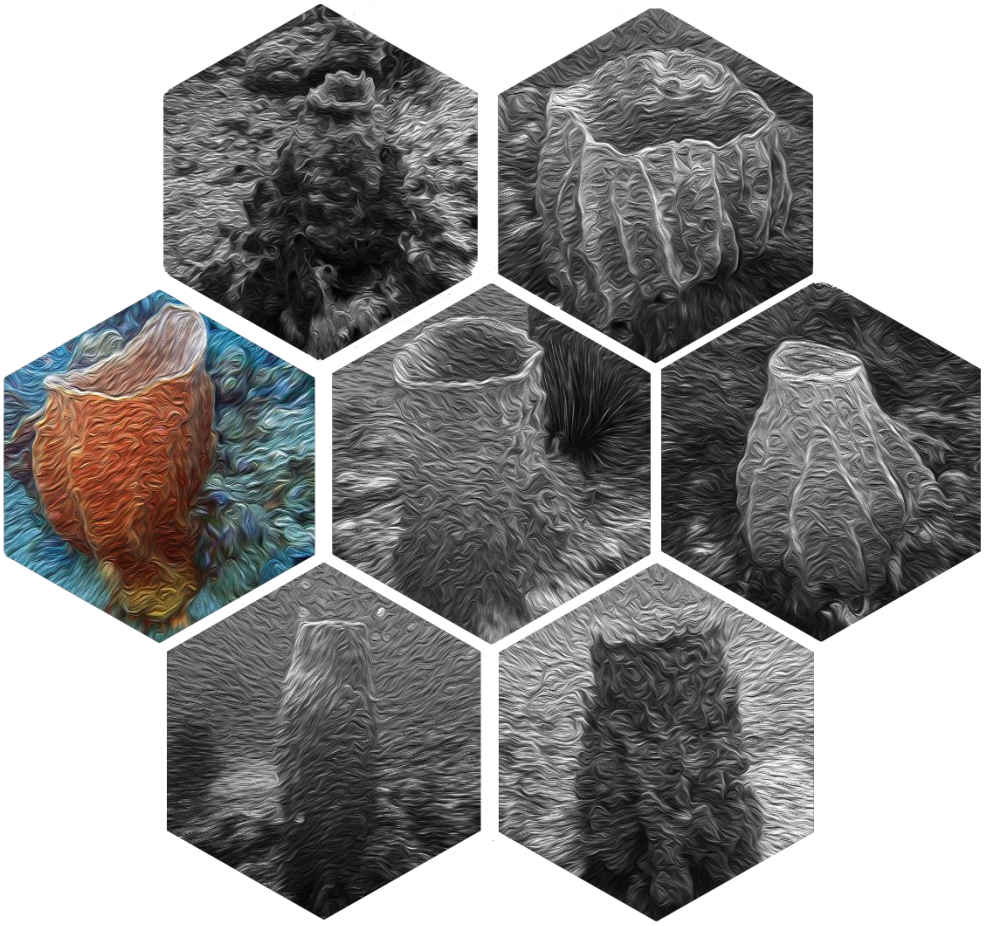


Figure S4: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-d_4, 600 MHz) of Compound (**2**).



Chapter 6

Study of the lipid profile of three different genetic groups of the Indo-Pacific giant barrel sponge and possible implications in their response to environmental conditions

Lina M. Bayona[1], Esther van der Ent[2,3], Rohani Ambo-Rappe[4], Young Hae Choi[1], Nicole J. de Voogd[2,3*]

[1]Institute of Biology, Natural Products Laboratory, Leiden University, Sylviusweg 72, 2333 BE, Leiden, The Netherlands.

[2]Naturalis Biodiversity Center, Marine Biodiversity, Darwinweg 2, 2333 CR, Leiden, The Netherlands.

[3]Institute of Environmental Sciences, Leiden University, Einsteinweg 2, 2333 CC, Leiden, The Netherlands.

[4] Faculty of Marine Science and Fisheries, Department of Marine Science, Hasanuddin University, Makassar, Indonesia.

Manuscript in preparation

Abstract

The presence of cryptic species in sponges has been a recurrent obstacle for their classification, due to the plasticity in their morphology and the fact that some genetic markers fail to give enough resolution at the species or even genus level. Therefore, their chemical characterization could provide supplementary information, contributing to reduce the misclassification of these organisms. Another advantage of counting on a deeper knowledge of the metabolome is the possibility of predicting the influence of various environmental factors such as pH and temperature on closely related sponges. Understanding the metabolic responses to several factors can be hampered by their chemical complexity, inevitably requiring a multiplatform workflow to obtain an accurate, comprehensive picture. In this study, the chemical variation among three genetic groups of the giant barrel sponge (*Xestospongia testudinaria*) collected in the Spermonde Archipelago (SW Sulawesi, Indonesia) was investigated. The samples were analyzed using NMR and LC-MS based metabolomics and the differences between the genetic groups and their individual responses to certain environmental conditions was further determined using multivariate data analysis. The genetic groups exhibited highly characteristic chemical profiles, particularly related to brominated unsaturated fatty acids derivatives, reflecting a chemical variation caused by the genetic background of these species. The study of these brominated unsaturated fatty acids derivatives allowed the isolation and elucidation of the structures of six *lyso*-phospholipids (**1-6**) and a fatty acid methyl ester (**7**). Among the identified phospholipids, five compounds were found to have previously unreported structures (**1-5**). A further study of the relationship between differences in the genetic groups and their lipid profile and possible implications in the response to changes in two environmental conditions that are very relevant for marine organisms, seawater pH and temperature was also conducted. Results showed that all groups were unaffected by the range of pH values studied, while only one of the genetic groups showed significant changes in their metabolic profile as a response to variations in temperature. The difference observed in this case was the presence of higher amounts of saturated fatty acids and membrane lipids at higher sea surface temperature (SST).

Key word: *Xestospongia testudinaria*, cryptic species, metabolomics, environmental changes, genetic groups

1. Introduction

Sponges are ubiquitous in reef ecosystems around the world (Van Soest et al. 2012). Many species have been proved to play diverse roles in the reefs related to phenomena such as bioerosion and stabilization of the reef substrate, nutrient cycling (nitrogen, silicon, and carbon), water filtering, and participation in the interactions among other organisms in the reef (Bell 2008; Wulff 2006, 2016). These ecological interactions stem from their function as habitats to many different organisms, including microorganisms, fish and invertebrates, which results in the production of a wide range of secondary metabolites. Although the importance of sponges in the sustainability of reef systems has not been as extensively studied as for corals, sponges are unarguably relevant players in the development and conservation of these ecosystems (Bell and Carballo 2008; Diaz and Rützler 2001). Moreover, sponges have proved to be more resilient than corals to certain environmental changes such as the increase in the ocean temperature and acidification, opening the possibility of sponge-dominated reefs in the future (Bell et al. 2013, 2018)

Giant barrel sponges are one of the most prominent and widely distributed sponges in coral reefs around the globe (Diaz and Rützler 2001). They are characterized by their size, longevity and high abundance in some reefs (McMurray et al. 2008, 2010; Zea 1993). Giant barrel sponges can belong to any of three accepted species from the genus: *Xestospongia bergquistia, Xestospongia*: *Xestospongia muta*, and *Xestospongia testudinaria* (Van Soest et al. 2012). Recent studies, however, have shown that the originally denominated *X. muta* and *X. testudinaria* actually consist of a number of different species (Bell et al. 2014; Setiawan et al. 2016; Swierts et al. 2013, 2017). According to Swierts and co-workers (Swierts et al. 2017) individuals previously classified as *X. muta* and *X. testudinaria* could in fact be classified into nine genetic groups based on haplotypes of two markers: cytochrome oxidase I (CO1) and adenosine triphosphate synthase subunit 6 (ATP6). Certain groups are only found in the tropical Atlantic region (groups 7, 8, and 9) whereas others (groups 1, 2, and 3) inhabit the Indo-Pacific region. However, this coappearance does not result in separate monophyletic clades, as one giant barrel sponge could be more closely related to another sponge in a different ocean than to the ones present in the same location (Swierts et al. 2017).

More than 300 compounds have been identified so far from *Xestospongia* sponges (Zhou et al. 2010). Out of these, metabolites reported for giant barrel sponges (*X. muta* and *X. testudinaria*), have mainly consist of sterols and brominated fatty acids. The sterol profile of these sponges has been considered to be useful to discriminate among chemotypes. Kerr and Kelly-Borges (1993) proposed that three *X. muta* chemotypes had distinguishable sterol

profiles. The observed qualitative and quantitative differences might be associated with the three main genetic groups found in the tropical Atlantic. Similarly, the different patterns observed in the sterol composition of *Xestospongia* samples collected in the Indo-Pacific (Gauvin et al. 2004) could be related to the different genetic groups of giant barrel sponges known to be present in this region. However, while sterol profiles show some differences, other types of metabolites have not been studied yet and it is likely that overall differences in these profiles could be connected to cryptic species of *X. testudinaria.* In general, there is plenty evidence of the relationship between chemical patterns and genetic divergence as reported for the sponges of the genus *Oscarella* and the zoanthid *Parazoanthus axinellae* (Boury-Esnault et al. 2013; Cachet et al. 2015; Ivanišević et al. 2011).

To gain insight into the chemical divergences among living organisms, including sponges, it is necessary to count on a profiling method that can deal with complex mixtures. Metabolomics, defined as the study of all the metabolites present in a sample under a set of conditions (Viant 2007) has been developed precisely to handle the complexity of the metabolome. It can offer an untargeted overview of the metabolome of the sponges, which can in turn reveal differences in the production of compounds or families of compounds that might have been overlooked in targeted studies. For this, the analytical platforms used must meet several requirements: cover a broad range of detectable compounds with high resolution and sensitivity, a reliable identification capacity and data robustness. Mass spectrometry- and nuclear magnetic resonance-based metabolomics tools have been applied in many metabolomics studies and are now the most popular techniques for data collection. In general, NMR is more often used in cases where a general overview is required, while MS-based methods are used in studies that demand a higher sensitivity.

Although metabolomics studies were initially focused on terrestrial organisms, marine organism metabolomics has become increasingly popular in this field. It has been used, for example, to identify changes in the metabolome caused by environmental conditions (Chapter 2, this thesis). In particular, the implementation of this approach allowed the identification of changes related to environmental conditions such as depth, season and geographical location in several sponge genera such as *Xestospongia* (Bayona et al. 2020; Villegas-Plazas et al. 2019), *Haliclona* (Reverter et al. 2018)*, Aplysina* (Reverter et al. 2016) and *Crambe* (Ternon et al. 2017). These changes in the metabolome can be associated with biotic factors, such as predatory stress and modifications in the microbiome, or abiotic factors, such as pH, temperature, salinity, oxygen levels and nutrient concentration (Januar et al. 2015; Reverter et al. 2016, 2018). The significance of this type of studies is enormous if considering that a

greater understanding of how modifications in the environment of sponges affect their production of secondary metabolites could lead to much needed insight into the resilience of these organisms to massive global events such as ocean acidification and global warming.

The presence of cryptic species in sponges has been a recurrent problem in the classification of these organisms. Therefore, the addition of new markers, such as a chemical profile, could contribute to their accurate classification. Moreover, understanding how environmental conditions can affect giant barrels sponges is particularly important due to the role they play in reef ecosystems, including their interaction with both micro and macro organisms and their participation in nutrient cycles. Lastly, it is important to establish if the different genetic groups respond differently to variations in their environment, as this could have important implications for giant barrel sponge diversity and survival in reefs around the world.

In order to investigate these issues, both NMR and LC-MS based metabolomics was implemented to explore the differences in chemical production between three genetic groups of *X. testudinaria*. While NMR provided a general overview of the chemical profile of each genetic group, LC-MS based metabolomics allowed the identification of compounds that were present in low concentrations. The metabolomics analyses were additionally used to establish whether two different environmental conditions, seawater pH and surface sea temperature (STT), could cause changes in the metabolome of these sponges and if so, whether these changes differed among genetic groups. For this, 112 samples of giant barrel sponges were collected in the Spermonde Archipelago, SW Sulawesi, Indonesia and their genetic groups were determined according to previous reports (Swierts et al. 2017).

2. Results and discussion

Giant barrel sponges (*X. testudinaria*) inhabiting the Indo-Pacific region can be divided into at least three different genetic groups (Swierts et al. 2017). The sponge samples collected at different sites in the Spermonde Archipelago (SW Sulawesi, Indonesia) were assigned to haplotypes using mitochondrial CO1 and ATP6 markers (in total 112 samples). Samples were found to belong to group 1 (C2A1, n=62), group 2 (C4A3, n=19) and group 3 (C6A2, n=31). The samples were then chemically profiled using NMR and LC-MS analytical platforms to define their metabolome. The classification of the genetic groups was then used as a category to perform a partial least square– discriminant analysis (PLS-DA) using NMR and LC-MS data. As shown in Figure 6.1, in both cases the three genetic groups can be distinguished by their metabolic profiles. The models were validated using a cross validation ANOVA test, exhibiting p values < 0.05.

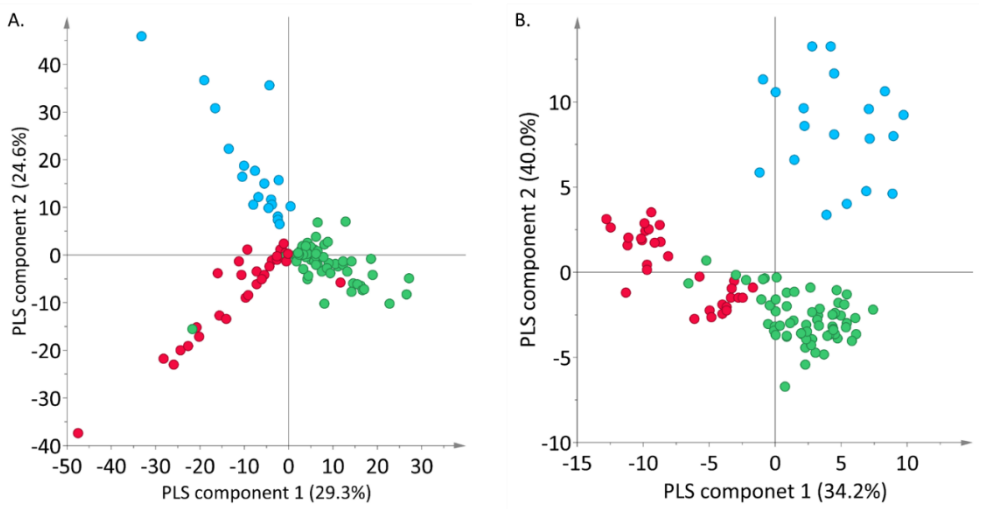


Figure 6.1: First two components of the partial least square-discriminant analysis (PLS-DA) model for the three genetic groups, group 1 (green), group 2 (blue), group 3 (red), of *X. testudinaria* samples A: LC-MS data (R^2 = 0.916, Q^2 = 0.686) and B: NMR data (R^2 =0.907, Q^2 = 0.849)

Working with the NMR data, the 10 most conspicuous signals among those altered in each group as observed in the biplot, were selected for further analysis. Figure 6.2 shows the characteristic signals found for each genetic group. Group 1 and 2 exhibited several characteristic signals in the olefinic region between δ_H 5.0 and 7.0 which correspond to double bonds, presumably of unsaturated fatty acids. This is in agreement with reports of the presence of a wide range of brominated unsaturated fatty acids (Zhou et al. 2010) in giant barrel sponges. This was additionally confirmed by the presence of the signals corresponding to double bonds shifted downfield around δ_H 7.0. On the contrary, group 3 spectra show no characteristic signals in this region (δ_H 5.0 and 7.0) as shown in Figure 6.2, and actually shows very few signals in general. This indicates that brominated unsaturated fatty acids might be present in very small amounts in samples from this genetic group.

Similarly, some signals in the region between δ_H 3.0 - 4.0 differentiate group 2 and 3 from group 1. Some of them correspond to methyl groups bonded to a heteroatom of methyl esters of fatty acids, methoxylated fatty acids or the methyl group of choline present in phospholipids (Bayona et al. 2020; Brantley et al. 1995; Jiang et al. 2011; Quinn and Tucker 1991). Although some higher ^{1}H-NMR signals in this region were observed in both group 2 and 3 spectra, the individual signals differ, indicating that while they produce similar kinds of compounds, mainly brominated fatty acids, the specific compounds produced by each genetic group are different. The aliphatic region between δ_H 0.5 – 2.5 also showed some differences among the genetic

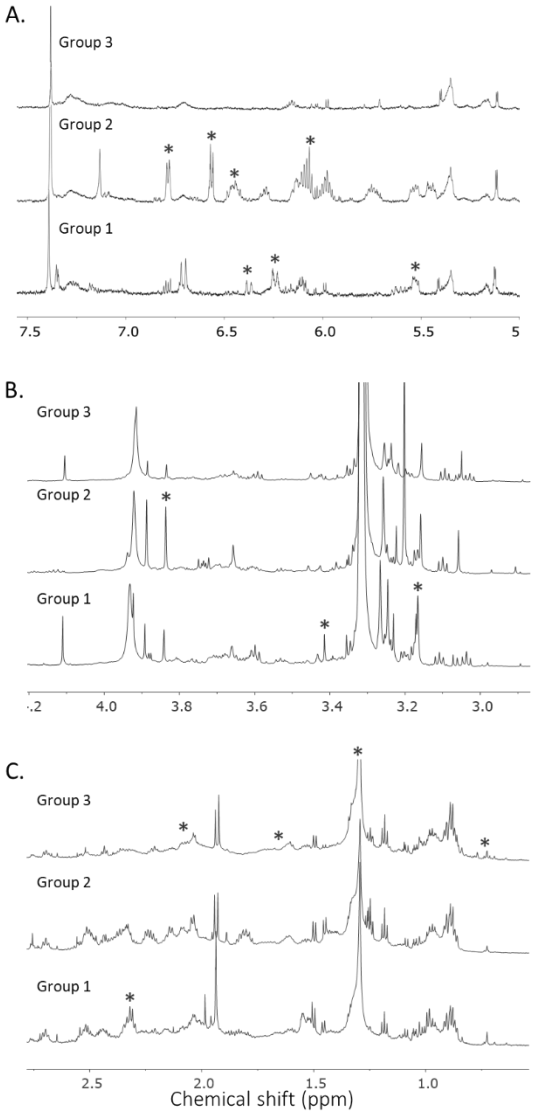


Figure 6.2: Typical ^{1}H-NMR (600MHz in CH_3OH-d_4) spectra of *X. testudinaria* belonging to three different genetic groups. A: δ_H 5.0-7.5, B: δ_H 3.0-4.0 and C: δ_H 0.6-2.6. the characteristic signals for each group are indicated by a red asterisk.

groups. Group 3 was found to have many increased signals which could correspond to lineal saturated fatty acids and in particular the signals between δ_H 0.7 -0.8 could indicate the presence of the methyl groups of fatty acids as well as H-18 and H-19 of sterols (Carballeira and Maldonado 1988; Kerr et al. 1991). After obtaining the general chemical profile of each

genetic group using NMR- based analysis, LC_MS analysis was done to obtain additional information on specific compounds characteristic for each group

The LC-MS data provided several features characteristic for each group. In this case the top 75 features in the VIP plot of the first two projections of the model were selected for further analysis (appendix 3 Table S1 and S2). After looking into the MS and MS/MS spectra of these features, it was observed that most of them corresponded to brominated compounds, confirming the presence of brominated fatty acids in the samples as suggested by the NMR spectra. Interestingly, for each group, several features had the same *m/z* value but different retention times, implying the presence of several isomers. This is consistent with studies that report several isomers of unsaturated brominated fatty acids in giant barrel sponges (Jiang et al. 2011; Liu et al. 2011; Zhou et al. 2010). The analysis of isotopic patterns of the selected features revealed the presence of two bromine atoms in most of those related to group 2, one Br atom in group 3, while group 1 compounds had both one and two bromine atoms. This implies that the presence of bromine might not be a conclusive discriminant feature between groups.

As a next step, a molecular network using the GNPS workflow (Nothias et al. 2020) was built to establish structural relationships between the discriminant features find in the PLS-DA analysis. In total, 69 networks with three or more nodes were built. Out of the nodes, 3 were present only in group 1 and 20 only in group 2. Most of the nodes (n=1948) were present in all the genetic groups, indicating that differences in the metabolome between genetic groups is quantitative rather than qualitative. The annotation of the discriminant features by dereplication using the MS/MS spectra was unsuccessful. Therefore, to identify the distinctive compounds of each genetic group, compounds were isolated from extracts of samples from different groups, targeting those with *m/z* values from the PLS-DA analysis and the nodes that were grouped into one network, which suggests these compounds were structurally related.

In total, 7 compounds were isolated from three extracts using successive HPLC separations and LC-MS analysis to confirm the presence of target compound in the fractions obtained. In the MS/MS analysis, compounds **1** to **6** showed a fragment with an m/z value of 184.07, suggesting the presence of a phosphatidylcholine moiety in these molecules. In addition, compounds **1** to **6** exhibited the same pattern in their ^{1}H-NMR spectra in the δ_H 3.2 - 4.3 region (Table 6.1). The singlet signal at δ_H 3.23 that integrates for 9 protons and the signals at 4.29 and 3.64 confirm the presence of a choline moiety in these molecules. Moreover, the signals at δ_H 3.91, 3.98 and the diasterotopic protons at δ_H 4.12 and 4.18 confirmed the presence of a glycerol moiety (Shin et al. 1999). Therefore, compounds **1** to **6** correspond to *lyso*-

phosphatidylcholines and differing only in their fatty acid chain. To determine the structure of the fatty acids, the molecular mass of the fatty acid moiety was dereplicated and the ^{1}H-NMR signals were compared with those reported in literature. This allowed the fatty acid moieties of compounds **1**,**2**,**4**,**5** and **6** to be identified as (7*E*,9*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,9,13,15-tetraen-5-ynoic acid (Ichiba et al. 1993), (11*E*,15*E*,19*E*)-20-bromoicosa-11,15,19-trien-7,9,17-triynoic acid (Taniguchi et al. 2008), (9*Z*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (Fusetani et al. 1993), (9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoic acid (Fusetani et al. 1993; Zhou et al. 2011), and (7*E*,13*E*,15*Z*)-14,16-dibromohexadeca-7,13,15-trien-5-ynoic acid (Schmitz and Gopichand 1978). No match was found for the molecular mass of compound **3**, but the comparison of its ^{1}H-NMR spectra with that of compound **2** (Table 6.1) showed that the only difference between these compounds was the substitution of one of the protons in position 3' by a hydroxyl group, causing the chemical shift of proton 3' to shift to a lower field at δ_H 4.03 and protons in position 2' to become diasterotopic with δ_H 2.45 and 2.55. Lastly, compound **7** was established to be the methyl ester of the fatty acid in compound **1** by comparison of their ^{1}H-NMR spectra and the presence of a singlet signal with δ_H 3.67 corresponding to the methyl group connected to an oxygen atom.

Table 6.1: ^{1}H-NMR (600 MHz, CH_3OH-d_4) of compounds **1**-**7**

	1		**2**		**3**		**4**		**5**		**6**		**7**	
	δH	***J* (Hz)**	**δH**	***J* (Hz)**	**δH**	**mult, *J* (Hz)**	**δH**	***J* (Hz)**	**δH**	***J* (Hz)**	**δH**	***J* (Hz)**	**δH**	***J* (Hz)**
2'	2.49	t, 7.4	2.38	t, 7.4	2.45, 2.55	dd, 15.1, 8.7, dd, 15.1, 4.2	2.49	t, 7.3	2.49	t, 7.3	2.48	t, 7.4	2.44	t, 7.4
3'	1.82	quint, 7.2	1.65	quint, 7.4	4.03	m	1.86	quint, 7.2	1.84	quint, 7.3	1.81	quint, 7.2	1.81	quint, 7.2
4'	2.39	td, 7.0, 2.0	1.45	m	1.60	m	2.44	t, 6.5	2.41	t, 7.1	2.35	td, 7.0, 1.9	2.38	td, 7.0, 2.2
5'	---	---	1.55	quint, 7.0	1.70	m	---	---	---	---	---	---	---	---
6'	---	---	2.34	t, 7.2	2.37	t, 6.5	---	---	---	---	---	---	---	---
7'	5.52	dm, 15.5	---	---	---	---	---	---	---	---	5.45	dm, 15.8	5.52	dm, 15.6
8'	6.45	dd, 15.5, 10.7	---	---	---	---	---	---	---	---	5.99	dt, 15.8, 7.1	6.45	dd, 15.6, 10.7
9'	6.12	dd, 15.1, 11.0	---	---	---	---	5.52	brd, 10.8	5.55	brd, 15.8	2.08	q, 7.1	6.11	brdd, 15.1, 10.6
10'	5.74	dt, 15.1, 7.1	---	---	---	---	6.10	dt, 10.8, 7.5	6.27	dt, 15.8, 7.1	1.40	m	5.74	dt, 15.1, 7.0
11'	2.23	q, 7.2	5.57	brd, 15.8	5.57	brd, 15.8	2.34	m	2.16	m	1.44	m	2.23	q, 7.2
12'	2.13	m	6.23	dm, 15.8	6.23	dm, 15.8	1.54	m	1.51	m	2.04	q, 7.6	2.13	m
13'	6.06	td, 7.6, 1.4	2.24	m	2.24	m	1.54	m	1.51	m	6.06	td, 7.7, 1.4	6.06	td, 7.6, 1.5
14'	---	---	2.24	m	2.24	m	2.31	m	2.28	m	---	---	---	---
15'	6.78	dq, 7.6, 1.3	6.15	dt, 15.8, 6.6	6.14	m	---	---	---	---	6.78	dq, 7.6, 1.3	6.77	dq, 7.6, 1.4
16'	6.56	d, 7.6	5.63	dd, 15.9, 1.8	5.63	brd, 15.7	---	---	---	---	6.56	d, 7.6	6.56	d, 7.6
17'	---	---	---	---	---	---	6.24	dt, 14.0, 2.3	6.23	dt, 14.0, 2.3	---	---	---	---
18'	---	---	---	---	---	---	6.70	d, 14.0	6.69	d, 14.0	---	---	---	---
19'	---	---	6.37	dd,14.0, 2.2	6.37	dd, 13.9, 2.2	---	---	---	---	---	---	---	---
20'	---	---	6.79	d, 14.0	6.79	dd, 13.9	---	---	---	---	---	---	---	---
1	4.13, 4.19	dd, 11.4, 6.2, dd, 11.4, 4.5	4.12, 4.18	dd, 11.4, 6.1, dd, 11.4, 4.6	4.16, 4.21	dd, 11.4, 6.0, dd, 11.4, 4.7	4.14, 4.20	dd, 11.4, 6.2, dd, 11.4, 4.5	4.13, 4.20	dd, 11.4, 6.1, dd, 11.4, 4.5	4.13, 4.19	dd, 11.4, 6.2 dd, 11.4, 4.5	---	---
2	3.98	m	3.98	m	3.98	m	3.99	m	3.98	m	3.98	m	---	---
3	3.91	m	3.90	m	3.91	m	3.91	m	3.91	m	3.90	m	---	---
1''	4.30	m	4.29	m	4.29	m	4.29	m	4.29	m	4.29	m	---	---
2''	3.65	m	3.64	m	3.64	m	3.64	m	3.64	m	3.64	m	---	---
***N*-Me**	3.23	s	3.23	s	3.23	s	3.23	s	3.23	s	3.23	s	---	---
***O*-Me**	---	---	---	---	---	---	---	---	---	---	---	---	3.67	s

Additionally, from the molecular networking analysis as shown in Figure 6.3 compounds **1**-**3** and **6** are connected in the same network therefore it is possible to deduce that they are structurally related. In addition, the pie chart in each node represents the prevalence of each compound among the samples of each genetic group. In agreement with the results obtained from the PLS-DA, the nodes that correspond to compounds **2**-**5** are mainly present in group 1 samples. Likewise, compounds **1** and **6** are mainly present in group 2 samples. For group 3, some signals that were discriminant in the LC-MS analysis appear in the cluster together with compounds **1**-**3** and **6**. However, it was not possible to isolate the compounds corresponding to these nodes due to their low concentration in the extracts. Moreover, besides the general overview provided by the molecular network analysis, the presence of the unidentified features in the same molecular network as the isolated compounds indicated that these compounds might have some structural similarities. Thus, it was possible to infer that the features correspond to *lyso* -phosphatidylcholine lipids. This, added to the dereplication of the molecular mass corresponding to the fatty acid moiety, allowed to propose putative structures for the discriminant compound of genetic group 3. Using this information, the structure for the compound corresponding to the node 616.19 was a *lyso*-phospholipid with a phosphatidylcholine moiety with the (11*E*,15*E*,19*E*)-20-bromoicosa-11,15,19-trien-9,17-diynoic acid (appendix 3 Figure S9a) as the fatty acid chain as previously reported for *Xestospongia testudinaria* (Akiyama et al. 2013). In the same way, the node with m/z 618.20 could correspond to an analog with one unsaturation less in the fatty acid chain (appendix Figure S9b). These compounds differ from compounds from genetic group 1 in the absence of the acetylene group between carbons 5 and 6. Ultimately, the compounds found to be related to the differences between genetic groups were *lyso*- phospholipids of the choline type. The fatty acid moiety corresponded to long chain fatty acids with lengths ranging from 16 to 20 carbons with different degrees of unsaturation and bromination substitution.

In the past, differences in the chemical composition of marine organisms have been used for a taxonomic classification supporting other data (Cachet et al. 2015). The separation of the three genetic groups of giant barrel sponges based on their chemical composition endorses the presence of cryptic species in what has been classified as *X. testudinaria* according to Swierts and co-workers (Swierts et al. 2013, 2017). One important aspect of the compounds related to the differentiation of the genetic group is that the fatty acid units correspond to brominated polyacetylene fatty acids, which have been reported to be biochemical markers for *Xestospongia* (Erpenbeck and van Soest 2006). Usually, these compounds are biosynthesized from C_{14}-C_{18} saturated fatty acids, although in marine organism they can also come from the polyketide pathway (Minto and Blacklock 2008). In the case of chains longer

than C_{18}, like the C_{20} acyl moieties observed in characteristic compounds from group 1 and 3, an extra elongation step is needed (Leonard et al. 2004). This might imply the activation of different metabolic pathways related to the synthesis of these longer fatty acid chains in specimens belonging to group 1 and 3 compared with those belonging to group 2. Moreover, the separation between the groups was found to be related to *lyso*-phospholipids of the phosphatidylcholine (PC) kind. In general, these types of compounds have been identified as molecular signals that mediate several processes in animal cells (Birgbauer and Chun 2006; Torkhovskaya et al. 2007). In sponges, *lyso*-phospholipids have been isolated from several genera and responded positively to different bioassays (Jang et al. 2012; Lin et al. 2015; Shin et al. 1999; Zhao et al. 2003). However, information about the role of these lipids in the sponges is still limited. Studies conducted on the Mediterranean sponge *Oscarella tuberculata* showed the seasonal variation of the levels of two *lyso*-phospholipids which could be related to the reproductive cycle of the sponge (Ivanisevic et al. 2011). Similarly, two sponges of the genus *Haliclona* showed an increase in the levels of *lyso*-PAF (Platelet activating factor) type of compounds between April and May (Reverter et al. 2018). In *Xestospongia* sponges *lyso*-phospholipids have also been related to changes in environmental conditions experienced by the sponges at different depths (Chapter 5, this thesis). The fact that the family of compounds that was found to be a discriminating factor between the genetic groups has been related also to environmental conditions and the life cycle of sponges could indicate two things. In the first place, considering that specimens belonging to different genetic groups experience similar environmental conditions, the differences in the chemical composition could be the result of different responses to the environmental conditions between the groups. Secondly, that changes in the composition of *lyso*-phospholipids could be related to the reproductive cycle of the sponge, similarly to *O. tuberculate*, suggesting that the reproductive cycle of each genetic group occurs at different time lapses. Further experiments will be needed to confirm this hypothesis.

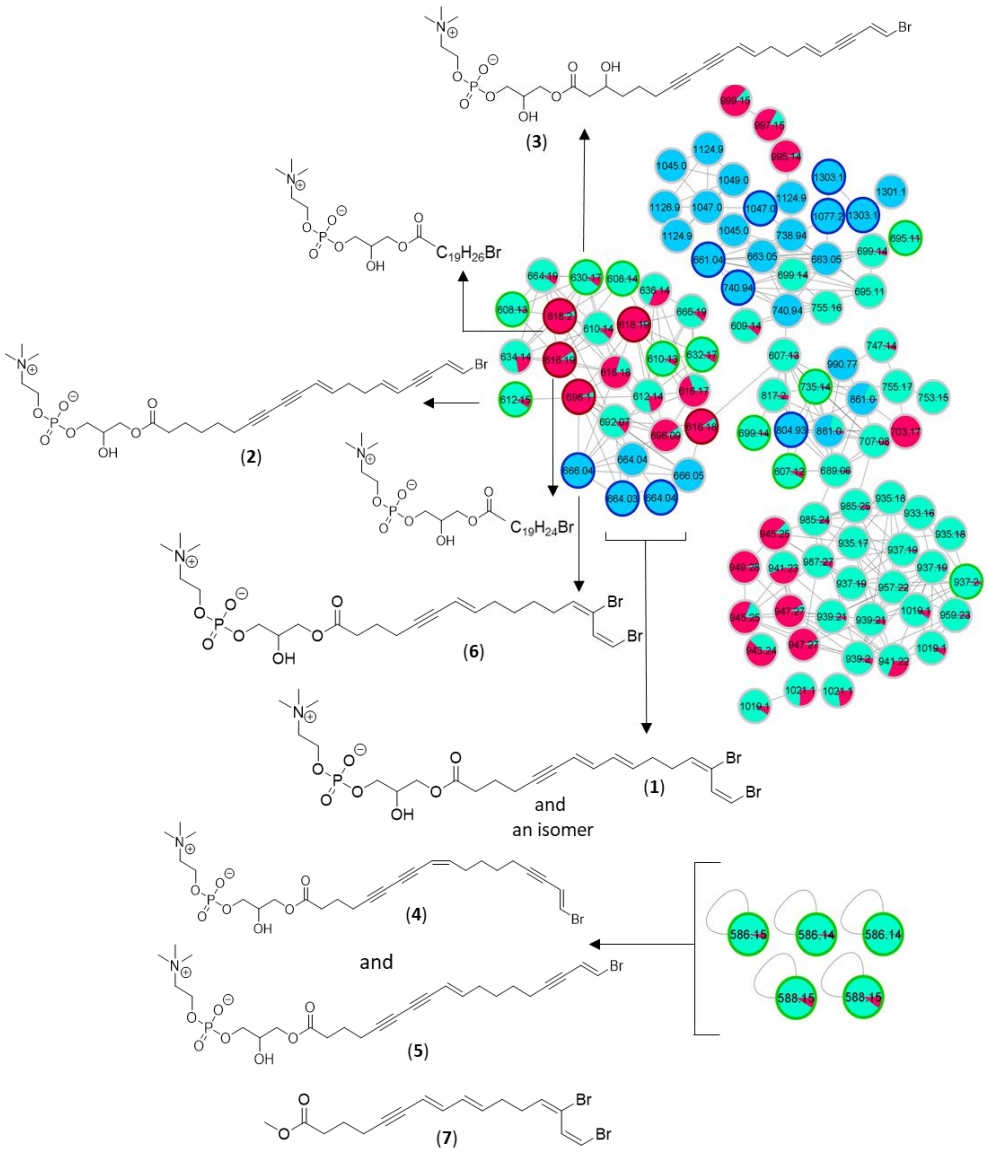


Figure 6.3: Molecular network of *X. testudinaria* samples belonging to three genetic groups showing the structure of the isolated compounds 1-7 related to the separation between groups. Each node is colored as a pie chart showing the abundance of the feature in each genetic group in green (group 1), blue (group 2) and red (group 3).The nodes with bold borders represent the features that were selected from the partial least square-discriminant analysis.

In order to determine if environmental variables, seawater pH and sea surface temperature, could induce changes in the metabolome of sponges from different genetic groups, PLS

models (LC-MS and NMR data) for each genetic group were constructed using the values of pH and temperature as Y variables. The pH values varied in the range of 8.02 - 8.19 for group 1, 8.06 - 8.17 for group 2 and 8.02 - 8.17 for group 3. None of the models showed any significant correlation, with p-values above 0.05 in CV-ANOVA test. For SST measurements, group 1 ranged between 28.7 to 30.2 °C, and group 2 and 3 between 29.0 and 30.1 °C. The models for groups 1 and 2 exhibited *p*-values above 0.05 for the CV-ANOVA test and Q^2 below 0.2. They were therefore considered non-valid. In contrast, the PLS models using group 3 samples and temperature as the Y variable showed p-values below 0.05 for the CV-ANOVA test and R^2 and Q^2 above 0.6 and 0.4 respectively, proving the validity of the models (Figure 6.4).

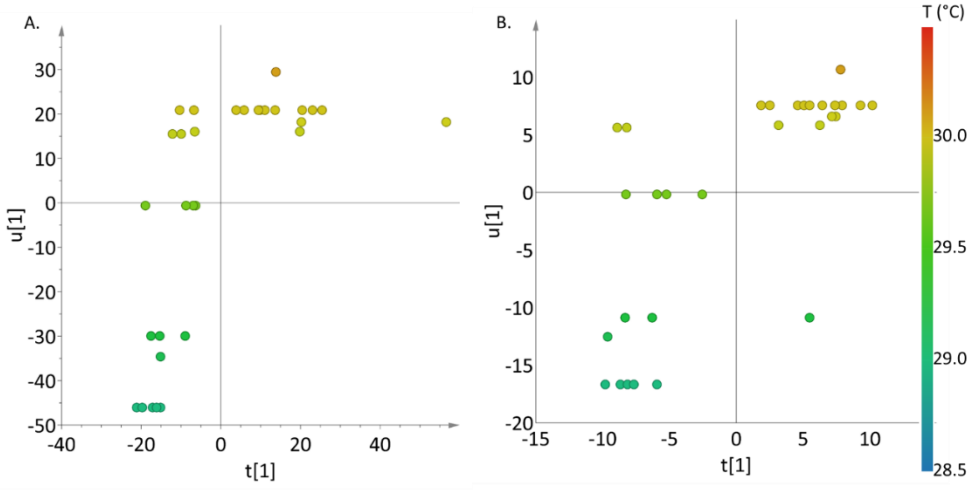


Figure 6.4: First component of the partial least square (PLS) analysis for X variable (t[1]) vs Y variables (u[1])based on A: LC-MS data and B: NMR Data of the samples belonging to genetic group 3 (C6A2) of *X. testudinaria* in response to variations in the sea surface temperature.

The metabolome of an organism is highly adaptive and can respond to modifications in the environmental conditions to which the organism is exposed (Bundy et al. 2008). In our study on the effect of the pH on the metabolome of the three genetic groups of giant barrel sponges, changes in their metabolome were not significant. This apparent lack of a visible change in their metabolome could have several explanations. In the first place, the different water pH at the sampling locations was only between 0.11-0.17 below normal pH values, which is way below the estimated decrease of 0.3-0.5 pH units projected for 2100 and that many studies use as reference to observe changes in sponge health (Bell et al. 2018). Another important aspect that could explain the known resilience of sponges to ocean acidification is that most

of them have silicon and organic skeletons that are far less susceptible to decrease in pH than organisms with calcareous skeletons such as corals (Bell et al. 2013). This apparent resistance is also supported by reports of the lack of significant differences in the concentration of major metabolites in some species submitted to changes in pH (Duckworth et al. 2012). However, the threat posed by more severe changes in pH conditions should not be underestimated, as *Xestospongia* spp. could experience changes due to modifications in their metabolism in order to adapt to the new environmental conditions.

On the other hand, variation in the SST only caused a significant change in the metabolome of group 3. This proves that even when exposed to very similar environmental conditions the three genetic groups can have different metabolic responses. This supports the separation according to their metabolome shown earlier in this study. To gain insight into the metabolites that differed in abundance due to the increase in temperature, the top 25 and top 50 VIP signals for the NMR and LC-MS models respectively were selected (appendix 3 Figure 10 and Table S3). The NMR signals that were related to higher temperatures were mainly found in the aliphatic region between δ_H 0.68-1.60 and δ_H 2.36., corresponding to aliphatic chain linear and branched fatty acids. The lack of discriminant signals in the olefinic regions indicates that these fatty acids are mainly saturated. Therefore, it is possible to infer that the changes caused by increases in the temperature are associated with the primary metabolism. In addition, the features selected from the LC-MS analysis showed some patterns that coincided with the information obtained from NMR. Most of the features selected showed a fragment of *m/z* 184.07 in their MS/MS spectra, indicating once again the presence of a phosphatidylcholine moiety in the molecules. However, in this case the retention times (between 23 and 27 min) and the high *m/z* values (above 1000) suggested the presence of phospholipids with two fatty acid chains joined to the glycerol moiety together with the PC moiety. These lipids have a structural role in cells as they are major components of the cell membrane (van Meer et al. 2008). The increase in membrane lipids could be related to higher growth rates at these higher temperatures. This is consistent with the higher growth rate reported in the summer months for *X. muta* in the Caribbean (McMurray et al. 2008) and the almost 50 % increase of the *Xestospongia* populations in some locations in the Caribbean between 2000 and 2006 (McMurray et al. 2010) where the average water temperature has been constantly increasing (Change, 2020). In contrast, previous studies have shown that exposure to temperatures above 30 °C caused an increase in the expression level of the heat shock protein *hsp*70 (López-legentil et al. 2008). However, signs of necrotic tissue were only observed at 40 °C, which is almost 9 degrees higher that the temperatures recorded in this study.

3. Conclusions

Genetic groups of giant barrel sponges can be distinguished by their chemical profile. Particularly, the *lyso*-phospholipid composition has been observed to vary between the genetic groups. In sponges, this kind of compound has been reported to be related to the changes in environmental conditions and to the life cycle of the sponges. This suggests that life cycles differ among genetic groups, which is an indication of reproductive isolation and could thus confirm that they are indeed separate species. It could also be presumed that genetic groups respond differently when exposed to similar environmental conditions. The evaluation of the response of the three genetic groups to differences in sea water temperature confirmed this possibility, since only genetic group 3 samples showed changes in their metabolome related to temperature change. The metabolic variation observed in these sponges consisted in an increase in membrane phospholipids which could be associated with an increase in growth rate at higher temperatures.

4. Materials and methods

4.1 Sample collection and genetic group classification

The research area is located in the Spermonde archipelago, just off the coast of Makassar, SW Sulawesi, Indonesia. The archipelago has an area of 1800 km^2 and consists of 160 fringing reefs, barrier reefs and patch reefs. Four ecological reef zones have been identified based on cross-shelf distribution of various reef taxa, bathymetry, geomorphology and distance offshores and the zones differ in abiotic parameters. Giant barrel sponges (*X. testudinaria*) were collected by SCUBA diving from multiple sites up to 60 km from the shore in April-May 2018. Water temperatures and pH values were recorded using a Hanna YSI1030 datalogger in each collection site at the moment of collection (appendix 3, Table S4). Sponge samples were stored in 96% ethanol (w/w) immediately upon collection and preserved at -20 °C. For molecular analyses, DNA was extracted using Qiagen Blood & Tissue kit, following the manufacturers protocol for spin column extractions. The samples were identified by DNA sequencing of the I3-M11 partition of the mitochondrial CO1 gene and the ATP6 gene following the protocol established by Swierts and co-workers (Swierts et al. 2013). For the CO1 gene, the primers C1-J2165 (5'-GAAGTTTATATTTTAATTTTACCDGG-3') and C1-Npor 2760 (5'-TCTAGGTAATCCAGCTAAACC-3') were used to amplify a fragment of 544 base pairs (bp). Amplification was performed in a 25 µL total reaction volume with 15.5 µL sterile water, 5 dNTPs (2.5 mM), 2.5 µL coralload buffer (Qiagen), 0.4 µL of each primer (10 µM), 0.25 µL taq polymerase (Qiagen) and 1 µL DNA template (20 ng μL^{-1}). For the ATP6 gene, the primers

ATP6porF (5'-GTAGTCCAGGATAATTTAGG-3') and ATP6porR (5'-GTTAATAGACAAAATACATAAGCCTG-3') were used to amplify a product of 445 bp. Amplification was performed in a 25 µL total reaction volume with 14 µL sterile water, 5 µL dNTPs (2.5 mM), 2.5 µL coralload buffer (Qiagen), 1.5 µL BSA (Promega), 0.4 µL (10 µM) of each primer, 0.25 µL taq polymerase (Qiagen) and 1 µL DNA template (20 ng μL^{-1}). For both genes, a PCR protocol consisting in an initial denaturing step (95 °C for 5 min), followed by 35 cycles of denaturing (95 °C for 30 s), annealing (42 °C for 45 s) and extension (68 °C for 1.00 min), and a final extension step (72 °C for 10 min) executed in a T100 thermal cycler (Bio-Rad) was implemented. The DNA analysis enabled the classification of the samples in mitochondrial haplotypes and subsequently in one of three genetic groups reported for giant barrel sponges in the Indo-Pacific region (Swierts et al. 2017).

4.2 Sample preparation for metabolomic analysis

The sponge samples were ground and sonicated with ethanol for 20 min by triplicate. One milliliter of the resulting extract was dried and used for ^{1}H-NMR analysis. The remaining extracts were dried using a Centrivap concentrator (Labconco, Kansas City, MO, USA). Resulting samples were desalted using Supelclean C-18 SPE,500 mg, 45 µm, cartridges (Supelco, Bellefonte, PA, USA). Fifty mg of each extract were loaded onto the cartridge and eluted sequentially with water (F1), methanol (F2) and dichloromethane/methanol 1:1 (F3). The methanol fraction was used for LC-MS analysis.

4.3 ^{1}H-NMR Analysis and data processing

The dry extract was resuspended in 1mL of deuterated methanol (CH_3OH-d_4) with hexamethyl disiloxane (HMDSO) as the internal standard. The ^{1}H-NMR spectra were measured at 25 °C in an AV-600 MHz NMR spectrometer (Bruker, Karlsruhe, Germany), operating at the ^{1}H-NMR frequency of 600.13 MHz, and equipped with a TCI cryoprobe and Z gradient system. Deuterated methanol was used as an internal lock. A presaturation sequence was used to suppress the residual water signal, using low power selective irradiation at the H_2O frequency during the recycle delay.

The resulting spectra were phased, baseline corrected and calibrated to HMDSO at 0.07 ppm using TOPSPIN V. 3.0 (Bruker Karlsruhe, Germany). The NMR spectra were bucketed using AMIX 3.9.12 (Bruker BioSpin GmbH, Rheinstetten, Germany). Bucket data were obtained by spectra integration at 0.04 ppm intervals from 0.20 to 10.02 ppm. The peak intensity of individual peaks was scaled to the total intensity of the buckets. The regions between 4.9 to

4.8, 3.32 to 3.28, 3.36 to 3.35, 3.63 to 3.59, and 1.16 to 1.20 ppm were excluded from the analysis because of residual signals of water, methanol and ethanol.

4.4 LC-MS analysis

The methanol fractions obtained from the SPE were dried using a Centrivap concentrator (Labconco, Kansas City, MO, USA). One milligram of the dried extract was dissolved in MeOH to obtain solutions with a final concentration of 1mg/mL. The fractions were analyzed using a UHPLC-DAD UltiMate 3000 system (Thermo Scientific Waltham, USA) coupled to a Micro OTOF-Q II spectrometer with electrospray ionization (ESI) (Bruker, Bremen, Germany). Samples (1 µL) were separated on a Kinetex C18, 2.1 x 150 mm, 2.6 µm column (Phenomenex (Torrance, USA), eluted with a gradient of 0.1% formic acid in H_2O (A) and 0.1% formic acid in ACN (B), starting at 45% to 60% B in 15 min, 60% to 90% B in 12.5 min and 90% to 98% B in 2.5 min. The flow rate was 0.300 mL/min, column temperature was 40 °C. The mass spectrometer parameters were as follows: nebulizer gas 2.0 bar, drying gas 10.0 mL/min, temperature 250 °C, capillary voltage 4000 V. The mass spectrometer was operated in positive mode with a scan range of 100 - 1650 m/z and sodium formate was used as a calibrant. For MS/MS fragmentation, the 10 most intense ions were selected and fragmented following the parameters used by Garg and Co-workers (Garg et al. 2015).

Data files obtained from the LC–MS were converted to mzXML format using Brucker Daltonics DataAnalysis (version 4.1, Bremen, Germany). The LC-MS data were processed using MZMine2 (Pluskal et al. 2010) to build the feature matrix using the following parameters: the mass detection that was performed with centroid data using a noise level of 1000 for MS and 10 for MS/MS spectra. The chromatograms were processed using the tool ADAP chromatogram builder (Myers et al. 2017) with a minimum number of scans of 3, group intensity threshold 100, minimum highest intensity 100 and m/z tolerance 0.05. The chromatogram deconvolution was performed using a baseline cutoff algorithm with the following parameters: minimum peak height of 100, peak duration range of 0.01–0.50 min, and baseline level of 100. Chromatograms isotopic peaks grouper tool was used to remove the isotopes with an m/z tolerance of 0.05 and retention time tolerance of 0.1 min. The features of each sample were aligned using the join alignment tool with the following parameters: 0.05 m/z tolerance, 75% m/z weight, 0.1 retention time tolerance and 25% of retention time weight.

4.5 Statistical analysis

The matrixes obtained from the NMR and LC-MS were used to perform multivariate data analysis using SIMCA-P software (v.15.0.2, Umetrics, Umeå, Sweden). Discriminant analysis of

partial least square PLS-DA analysis was performed using the genetic group as a category. In addition, PLS models for each class were done using the pH and temperature recorded at the collection site as a response variable. For all the analysis, data were scaled using united variance scaling and the models were tested using a permutation test and a cross-validation ANOVA (CV-ANOVA) test. The model was considered valid if CV-ANOVA showed $p < 0.05$. For the prediction power of the model, Q^2 values above 0.4 were required; otherwise, the model was considered valid but with no prediction power.

For the PLS model of samples belonging to genetic group 3 and temperature as a response variable a heatmap was created using the top 25 signals of the VIP plot in the online platform Metaboanalyst R4.0 Web site (http://www.metaboanalyst.ca) (Chong et al. 2018). The signals were scaled using UV scaling.

4.6 Molecular networking and dereplication

The Feature-Based Molecular Networking (FBMN) workflow (Nothias et al. 2020) on GNPS (https://gnps.ucsd.edu, (Wang et al. 2016)) was used to create a molecular network. Firstly, LC-MS data were processed using MZMINE2 as shown in section 4.4. The features matrix and the MS/MS spectra were exported to GNPS for FBMN analysis. To build the MN, all the MS/MS spectra were filtered by selecting a window of +/- 50 Da of the top 6 fragment ions. The precursor ion mass and MS/MS fragment ion tolerances were set at 0.03 Da. Using the parameters mentioned above, an MN was created filtering the edges to have a cosine score above 0.7 and more than six matched peaks. Edges between two nodes were kept if they were common to both their top 10 most similar nodes. The maximum size of a cluster was set to 100, resulting in the removal of nodes with lowest cosine scores from the clusters until the size of the cluster was below this value. For dereplication, the MS/MS spectra in the network were then screened against GNPS spectral libraries (Horai et al. 2010; Wang et al. 2016). For a match to be accepted a score above 0.7 and at least 6 matched peaks were required. The molecular networks were visualized using Cytoscape software(Shannon et al. 2003) and displayed as a pie chart where each node was colored to show the prevalence of the feature in each genetic group. The FBMN analysis can be found in https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=031fb987b93b4466ab452714a87399e5

4.6 Isolation and identification of compounds

The target compounds were isolated from extracts of genetic groups 1, 2 and 3. The crude extracts were fractionated using 20 mL Supelclean LC-18 SPE, 5 g C18 modified silica,45 µm cartridges (Supelco, Bellefonte, PA, USA) and eluted successively with 100 mL of the following

solvents: 100% water 80%, 60%, 40%, 20% water/methanol mixtures,100% methanol and methanol-dichloromethane (1:1). This resulted in seven fractions of each extract. Using LC-MS profiling, the target compounds with the *m/z* values selected from the top 75 features of the VIP plot for the first two projections of the PLS model (appendix 3 Table S1 and S2) were found to be the fractions eluted with 80% methanol. The fractions were purified by semi-preparative HPLC on an Agilent 1200 series system (Santa Clara, CA, USA) with a Luna C-18, 250 mm x 10 mm, 5 µ column (Phenomenex, Torrance, CA, USA) and eluted at a flow rate of 4.5 mL/min with mobile phases of water: acetonitrile: 0.1% . Successive HPLC separations led to the isolation of compound **1** (9.45 mg), **2** (0.69 mg), **3** (0.50 mg), **4** (0.50 mg), **5** (0.45 mg), **6** (3.70 mg), and **7** (12.16 mg).

1-*O*-((7*E*,9*E*,13*Z*,15*Z*)-14,16-dibromohexadeca-7,9,13,15-tetraen-5-ynoyl)-sn-glycero-3-phosphocholine (**1**)

^{1}H-NMR (CH_3OH-d_4 600 MHz): see Table 6.1. ^{13}C-NMR (CH_3OH-d_4 150 MHz): 1'- δ_c 174.3, 8'- δ_c 141.6, 13'- δ_c 136.3, 10'- δ_c 135.8, 15'- δ_c 131.9, 9'- δ_c 131.6, 14'- δ_c 114.9, 16'- δ_c 113.1, 7'- δ_c 110.9, 5'- δ_c 91.4, 6'- δ_c 81.1, 2- δ_c 69.4, 3- δ_c 67.5, 2''- δ_c 67.1, 1- δ_c 65.9, 1''- δ_c 60.1, N-Me- δ_c 54.3, 2'- δ_c 33.5, 11'- δ_c 32.3, 12'- δ_c 31.6, 3'- δ_c 24.8, 4'- δ_c 19.2. (+)-QTOF-ESI-MS: 640.0658, 642.0646 and 644.0617 ratio 1:2:1 (calc. for $C_{24}H_{37}Br_2NO_7P^+$ 640.0669, 642.0648 and 644.0628)

1-*O*-((11*E*,15*E*,19*E*)-20-bromoicosa-11,15,19-trien-7,9,17-triynoyl)-sn-glycero-3-phosphocholine (**2**)

^{1}H-NMR (CH_3OH-$d_4$600 MHz): see Table 6.1. (+)-QTOF-ESI-MS 612.1709 and 614.1695 ratio 1:1 (calc. for $C_{28}H_{40}BrNO_7P^+$ 612.1726 and 614.1705)

1-*O*-((11*E*,15*E*,19*E*)-20-bromo-3-hydroxyicosa-11,15,19-trien-7,9,17-triynyl)-sn-glycero-3-phosphocholine (**3**)

^{1}H-NMR (CH_3OH-d_4 600 MHz): see Table 6.1. (+)-QTOF-ESI-MS 628.1659 and 630.1646 ratio 1:1 (calc. for $C_{28}H_{40}BrNO_8P^+$ 628.1675 and 630.1717)

1-*O*-((9*Z*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoyl)-sn-glycero-3-phosphocholine (**4**)

^{1}H-NMR (CH_3OH-d_4 600 MHz): see Table 6.1. (+)-QTOF-ESI-MS:586.1549 and 588.1528 ratio 1:1 (calc. for $C_{26}H_{38}BrNO_7P^+$ 586.1569 and 588.1549).

1-*O*-((9*E*,17*E*)-18-bromooctadeca-9,17-dien-5,7,15-triynoyl)-sn-glycero-3-phosphocholine (**5**)

^{1}H-NMR (CH_3OH-d_4 600 MHz): see Table 6.1. (+)-QTOF-ESI-MS:586.1549 and 588.1528 ratio 1:1 (calc. for $C_{26}H_{38}BrNO_7P^+$ 586.1569 and 588.1549).

1-*O*-((7E,13Z,15Z)-14,16-dibromohexadeca-7,13,15-trien-5-ynoyl)-sn-glycero-3-phosphocholine (**6**)

^{1}H-NMR (CH_3OH-d_4 600 MHz): see Table 6.1. (+)-QTOF-ESI-MS: 642.0840, 644.0835 and 646.0820 ratio 1:2:1 (calc. for $C_{24}H_{39}Br_2NO_7P^+$ 642.0831, 644.0810 and 646.0844)

(7*E*,9*E*,13*Z*,15*Z*)-14,16-dibromohexadeca-7,9,13,15-tetraen-5-ynoyl methyl ester (**7**)

^{1}H-NMR (CH_3OH-d_4, 600 MHz): see Table 6.1. (+)-QTOF-ESI-MS: 414.9915, 416.9896 and 418.9886 ratio 1:2:1 (calc. for $C_{17}H_{21}Br_2O_2^+$ 414.9908, 416.988 and 418.9867)

Acknowledgements

Research permits were issued by the Indonesian State Ministry for Research and Technology (RISTEK). This work was funded by NWO-VIDI with project number 16.161.301. We thank our colleagues from Universitas Hasanuddin and especially Prof. Dr. Jamaluddin Jompa, and their students for their invaluable support in arranging the fieldwork permits and assistance in the field. We also acknowledge Willem Renema, Jelle Rienstra, and Niels van der Windt. This work was supported by the COLCIENCIAS (Science Technology and Innovation Ministry, Colombia) and NWO-VIDI (#16.161.301) and NWO-Aspasia (#105-010.030).

Appendix 3

Table S1: Top 75 signal of the first component VIP plot of the PLS-DA analysis of *X. testudinaria* samples belonging to three genetic groups using LC-MS data

m/z	Retention time	VIP[1]*	*m/z*	Retention time	VIP[1]*
588.16	4.7	2.09318	938.66	23.2	2.77203
590.17	4.7	2.08789	674.13	9.6	2.16151
588.15	4.0	2.35568	672.12	9.7	2.11477
588.16	4.4	2.63546	1197.30	4.4	2.1403
964.56	1.2	2.10357	607.13	6.2	2.45603
735.14	29.8	2.70261	605.13	6.2	2.46268
752.41	1.2	2.72962	721.13	27.7	3.21544
586.15	4.0	2.28357	751.12	29.8	3.00339
586.15	4.4	2.53782	630.17	4.6	2.63229
940.66	23.2	2.75481	938.65	23.3	3.0488
610.18	4.6	2.34049	588.15	3.4	2.08988
645.50	29.7	2.09861	810.53	24.5	2.23966
788.55	24.5	2.30964	538.36	3.4	2.18995
874.41	24.0	2.14452	1193.29	4.5	2.23874
596.21	6.5	2.26261	588.16	4.3	2.54668
1211.41	6.5	2.27461	911.70	25.2	2.23194
674.13	8.6	2.14436	610.16	4.7	2.69581
594.20	6.0	2.28554	911.70	18.3	2.11285
592.19	6.0	2.12763	1089.20	16.4	2.48043
676.13	9.6	2.07007	1089.20	15.4	2.59916
699.15	27.7	3.23793	630.16	4.1	2.83636
937.20	25.6	2.10716	628.16	4.1	2.74647
586.15	4.2	2.1273	586.15	4.0	2.48301
630.16	4.1	2.66084	610.14	4.0	2.268
628.16	4.1	2.41828	632.17	4.6	2.69112
695.11	14.9	2.11639	916.61	22.9	3.09761
612.15	4.8	2.26212	696.11	8.6	2.23477
1083.15	14.9	2.47709	607.13	5.7	2.12681
594.21	6.5	2.23755	605.13	5.7	2.17294
578.32	1.8	2.12092	695.11	14.4	2.13175
616.19	6.5	2.3304	608.14	4.5	2.6129
618.19	6.5	2.5494	608.13	4.3	2.64691
550.38	24.6	2.1514	676.40	17.1	2.54801

m/z	Retention time	VIP[1]*	*m/z*	Retention time	VIP[1]*
618.21	7.3	2.23137	608.14	4.1	2.45087
674.13	9.6	2.16111	1213.43	7.3	2.10409
726.01	7.9	2.10814	676.13	9.6	2.15733
616.18	6.0	2.24119	657.20	8.6	2.22464
591.40	21.6	2.3688			

* VIP[1]: Variable importance for the projection first component

Table S2: Top 75 signal of the second component VIP plot of the PLS-DA analysis of *X. testudinaria*. samples belonging to three genetic groups using LC-MS data* VIP[2]: Variable importance for the projection first component

m/z	Retention time	VIP[2]*	*m/z*	Retention time	VIP[2]*
752.41	1.2	2.20794	674.13	9.6	2.08333
940.66	23.2	2.06675	672.12	9.7	2.06709
642.06	3.4	2.4677	721.13	27.7	2.3887
1303.13	5.4	2.1446	751.12	29.8	2.23959
1077.27	5.4	2.41362	938.65	23.3	2.27429
1305.11	3.4	2.6313	1303.13	5.3	2.11841
642.06	3.1	2.40725	1295.11	4.8	2.20371
804.94	12.0	2.99503	808.97	13.4	2.53666
661.04	4.8	2.45363	1047.02	23.4	2.07055
1309.14	3.8	2.37067	806.97	13.4	2.53065
1305.11	3.1	2.24859	1295.11	3.4	2.12065
639.06	4.7	2.5052	562.13	2.2	2.60339
642.06	2.9	2.81295	560.14	2.2	2.56245
723.97	4.9	2.03006	1248.88	12.0	3.39673
620.08	4.2	2.14593	1254.92	13.4	3.08805
596.21	6.5	2.23898	661.04	4.4	2.32665
1211.41	6.5	2.19332	665.36	22.7	2.20932
674.13	8.6	2.07801	1246.88	12.0	3.34085
594.20	6.0	2.25269	1305.14	3.8	2.04323
596.20	6.0	2.05741	1252.92	13.4	2.91128
592.19	6.0	2.14163	666.05	3.4	2.55258
699.15	27.7	2.40978	630.16	4.1	2.12604
803.70	25.9	2.26157	628.16	4.1	2.05887
594.21	6.5	2.21745	916.61	22.9	2.31731
616.19	6.5	2.38487	718.97	6.8	2.09963
618.19	6.5	2.48158	644.07	3.4	2.3353
618.21	7.3	2.19406	676.07	1.3	2.74533
674.13	9.6	2.08334	879.70	25.9	2.35779
726.01	7.9	2.05387	696.11	8.6	2.22391
616.18	6.0	2.29315	664.04	3.4	2.64816
591.40	21.6	2.28312	721.97	4.1	2.05229
938.66	23.2	2.06628	664.04	3.1	2.55934
821.71	25.9	2.2906	586.48	11.8	2.42506
962.75	15.7	2.54729	667.40	16.7	2.48662

m/z	Retention time	VIP[2]*	*m/z*	Retention time	VIP[2]*
960.76	15.7	2.37819	676.13	9.6	2.05909
964.75	15.7	2.60114	564.15	2.8	2.11724
738.95	6.8	2.1847	562.15	2.8	2.31877
740.95	6.8	2.08022			

* VIP[2]: Variable importance for the projection first component

Table S3: Top 50 signal of the second component VIP plot of the PLS analysis of *X. testudinaria* samples belonging to genetic group 3 using LC-MS data and temperature as Y variable

m/z	Retention time	VIP[2]*	*m/z*	Retention time	VIP[2]*
564.29	1.6	2.57124	718.32	28.5	1.6936
819.53	27.5	2.23721	676.52	28.2	1.6872
874.41	24.0	2.11865	682.32	26.7	1.68442
819.53	27.4	2.08892	1005.59	5.1	1.67339
734.60	27.4	1.94871	855.41	27.4	1.66868
856.81	24.7	1.93427	762.53	22.5	1.65869
874.41	24.0	1.89624	734.60	25.5	1.65022
657.45	24.8	1.85235	1013.65	7.7	1.64919
819.54	25.3	1.83444	597.34	16.3	1.64213
819.53	26.1	1.83334	680.32	26.8	1.64175
789.54	24.5	1.83003	676.52	28.3	1.63734
810.53	24.5	1.80888	876.78	21.5	1.62349
819.53	27.1	1.77191	676.52	28.3	1.62174
1009.62	6.8	1.7504	971.49	28.7	1.6208
734.60	26.8	1.7488	555.23	22.5	1.61671
991.67	7.7	1.74737	800.60	27.6	1.60319
819.54	25.7	1.74667	717.47	26.0	1.60042
819.54	26.5	1.73756	1060.75	22.1	1.59893
716.32	28.5	1.72553	805.52	25.9	1.59777
637.33	16.3	1.7229	805.52	25.9	1.59536
1121.59	7.1	1.71588	790.56	25.3	1.59155
602.42	25.4	1.71123	819.53	27.9	1.591
676.53	28.3	1.70894	578.32	1.8	1.5908
819.54	25.2	1.70425	600.40	29.0	1.58355
637.33	16.3	1.69665	1015.52	27.3	1.57692

* VIP[2]: Variable importance for the projection first component

Table S4: Collection places, temperature, pH and Haplotypes information of *X. testudinaria* samples

Sample	Genetic	Location	Temperature (°C)	pH	Barcoding CO1	Barcoding ATP6
XID26202	Group 1	Kapoposang	28.78	8.19	C2	A1
XID28202	Group 1	Kapoposang	28.78	8.19	C2	A1
XID30602	Group 1	Lankai	29.79	8.17	C2	A1
XID20202	Group 1	Lankai	29.79	8.17	C2	A1
XID18902	Group 1	Lankai	29.79	8.17	C2	A1
XID25302	Group 1	Pulau Badi	30.2	8.17	C2	A1
XID31102	Group 1	Pulau Badi	30.2	8.17	C2	A1
XID27002	Group 1	Pulau Badi	30.2	8.17	C2	A1
XID17002	Group 1	Pulau Badi	30.2	8.17	C2	A1
XID26402	Group 1	Kudingareng Keke	29.89	8.15	C2	A1
XID29102	Group 1	Polewali	29.9	8.06	C2	A1
XID30502	Group 1	Polewali	29.9	8.06	C2	A1
XID29602	Group 1	Polewali	29.9	8.06	C2	A1
XID24702	Group 1	Lumulumu	30.13	8.12	C2	A1
XID22702	Group 1	Lumulumu	30.13	8.12	C2	A1
XID18402	Group 1	Karanrang	29.94	8.07	C2	A1
XID30802	Group 1	Samalona	30.08	8.16	C2	A1
XID22902	Group 1	Samalona	30.08	8.16	C2	A1
XID16402	Group 1	Samalona	30.08	8.16	C2	A1
XID16902	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID28302	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID27202	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID26102	Group 1	Barang Lompo	29.22	8.14	C2	A1

Sample	Genetic	Location	Temperature (°C)	pH	Barcoding CO1	Barcoding ATP6
XID19402	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID25702	Group 1	Barang Caddi	30.11	8.17	C2	A1
XID25402	Group 1	Barang Caddi	30.11	8.17	C2	A1
XID20002	Group 1	Barang Caddi	30.11	8.17	C2	A1
XID18602	Group 1	Kapoposang	28.78	8.19	C2	A1
XID25802	Group 1	LaeLae	28.98	8.02	C2	A1
XID20502	Group 1	LaeLae	28.98	8.02	C2	A1
XID17402	Group 1	LaeLae	28.98	8.02	C2	A1
XID25902	Group 1	Bone Lola	29.91	8.16	C2	A1
XID20102	Group 1	Bone Lola	29.91	8.16	C2	A1
XID17202	Group 1	Bone Lola	29.91	8.16	C2	A1
XID23002	Group 1	Lankadea	29.66	8.17	C2	A1
XID28102	Group 1	Lankadea	29.66	8.17	C2	A1
XID19302	Group 1	Lankadea	29.66	8.17	C2	A1
XID17502	Group 1	Lankadea	29.66	8.17	C2	A1
XID22502	Group 1	Padjenekang	29.15	8.16	C2	A1
XID22202	Group 1	Padjenekang	29.15	8.16	C2	A1
XID17902	Group 1	Padjenekang	29.15	8.16	C2	A1
XID19202	Group 1	Kapoposang	28.78	8.19	C2	A1
XID24602	Group 1	Kapoposang	28.78	8.19	C2	A1
XID23802	Group 1	Kapoposang	28.78	8.19	C2	A1
XID24202	Group 1	Kapoposang	28.78	8.19	C2	A1
XID21502	Group 1	Kudingareng Keke	29.42	8.06	C2	A1
XID21702	Group 1	Kudingareng Keke	29.42	8.06	C2	A1

Sample	Genetic	Location	Temperature (°C)	pH	Barcoding CO1	Barcoding ATP6
XID17602	Group 1	Kudingareng Keke	29.42	8.06	C2	A1
XID19102	Group 1	Samalona	29.34	8.17	C2	A1
XID20302	Group 1	Samalona	29.34	8.17	C2	A1
XID18702	Group 1	Samalona	29.34	8.17	C2	A1
XID21802	Group 1	Samalona	29.34	8.17	C2	A1
XID24802	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID25502	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID16102	Group 1	Barang Lompo	29.22	8.14	C2	A1
XID29402	Group 1	Lankai	29.01	8.17	C2	A1
XID21902	Group 1	Lankai	29.01	8.17	C2	A1
XID21602	Group 1	Lankai	29.01	8.17	C2	A1
XID29802	Group 1	Bone Baku	28.72	8.13	C2	A1
XID22802	Group 1	Bone Baku	28.72	8.13	C2	A1
XID31502	Group 1	Bone Baku	28.72	8.13	C2	A1
XID20602	Group 1	Bone Baku	28.72	8.13	C2	A1
XID18502	Group 2	Lankai	29.79	8.17	C4	A3
XID29302	Group 2	Lankai	29.79	8.17	C4	A3
XID21202	Group 2	Polewali	29.9	8.06	C4	A3
XID24902	Group 2	Polewali	29.9	8.06	C4	A3
XID23902	Group 2	Polewali	29.9	8.06	C4	A3
XID22402	Group 2	Polewali	29.9	8.06	C4	A3
XID20902	Group 2	Karanrang	29.94	8.07	C4	A3
XID17302	Group 2	Karanrang	29.94	8.07	C4	A3
XID23702	Group 2	Karanrang	29.94	8.07	C4	A3

Sample	Genetic	Location	Temperature (°C)	pH	Barcoding CO1	Barcoding ATP6
XID28802	Group 2	Karanrang	29.94	8.07	C4	A3
XID26702	Group 2	Karanrang	29.94	8.07	C4	A3
XID25002	Group 2	Karanrang	29.94	8.07	C4	A3
XID18002	Group 2	Samalona	30.08	8.16	C4	A3
XID24302	Group 2	Barang Lompo	29.22	8.14	C4	A3
XID30102	Group 2	Barang Lompo	29.22	8.14	C4	A3
XID26602	Group 2	Barang Lompo	29.22	8.14	C4	A3
XID22002	Group 2	Barang Lompo	29.22	8.14	C4	A3
XID27602	Group 2	Barang Lompo	29.22	8.14	C4	A3
XID23202	Group 2	Lankai	29.01	8.17	C4	A3
XID18302	Group 3	Polewali	29.9	8.06	C6	A2
XID19502	Group 3	Polewali	29.9	8.06	C6	A2
XID28902	Group 3	Karanrang	29.94	8.07	C6	A2
XID16302	Group 3	Karanrang	29.94	8.07	C6	A2
XID30202	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID26802	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID28402	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID19902	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID22302	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID19002	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID16802	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID17102	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID26302	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID16002	Group 3	Barang Baringan	29.98	8.1	C6	A2

Sample	Genetic	Location	Temperature (°C)	pH	Barcoding CO1	Barcoding ATP6
XID20802	Group 3	Barang Baringan	29.98	8.1	C6	A2
XID27402	Group 3	Barang Lompo	29.22	8.14	C6	A2
XID24402	Group 3	Barang Caddi	30.11	8.17	C6	A2
XID22102	Group 3	LaeLae	28.98	8.02	C6	A2
XID31002	Group 3	LaeLae	28.98	8.02	C6	A2
XID23402	Group 3	LaeLae	28.98	8.02	C6	A2
XID17702	Group 3	LaeLae	28.98	8.02	C6	A2
XID27102	Group 3	LaeLae	28.98	8.02	C6	A2
XID18802	Group 3	Bone Lola	29.91	8.16	C6	A2
XID31302	Group 3	Bone Lola	29.91	8.16	C6	A2
XID27802	Group 3	Lankadea	29.66	8.17	C6	A2
XID30302	Group 3	Lankadea	29.66	8.17	C6	A2
XID24502	Group 3	Lankadea	29.66	8.17	C6	A2
XID20402	Group 3	Lankadea	29.66	8.17	C6	A2
XID19702	Group 3	Padjenekang	29.15	8.16	C6	A2
XID24102	Group 3	Barang Lompo	29.22	8.14	C6	A2
XID31202	Group 3	Barang Lompo	29.22	8.14	C6	A2

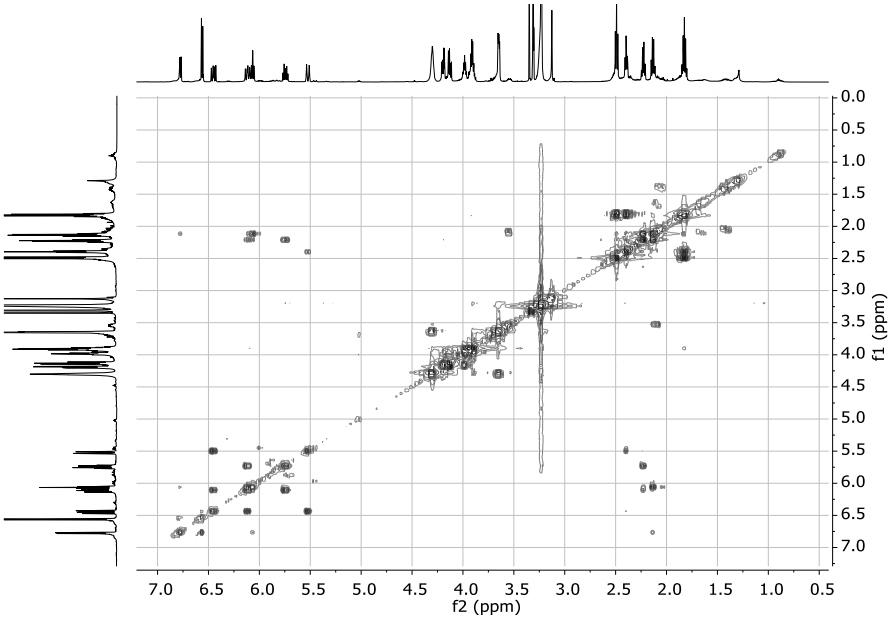


Figure S1: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-*d*$_4$, 600 MHz) of Compound **1**.

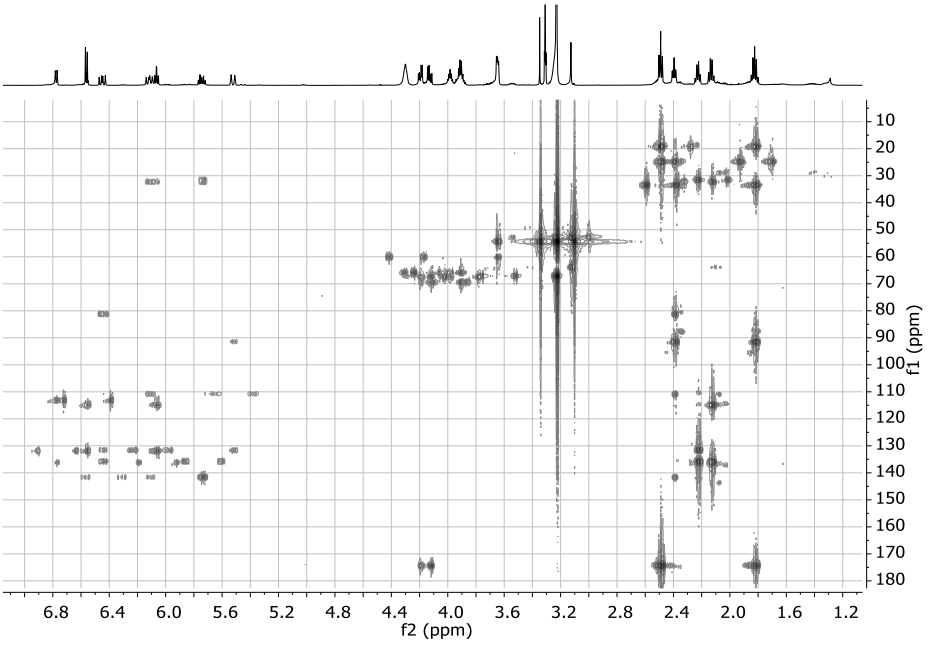


Figure S2: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-*d*$_4$, 600 MHz) of Compound **1**.

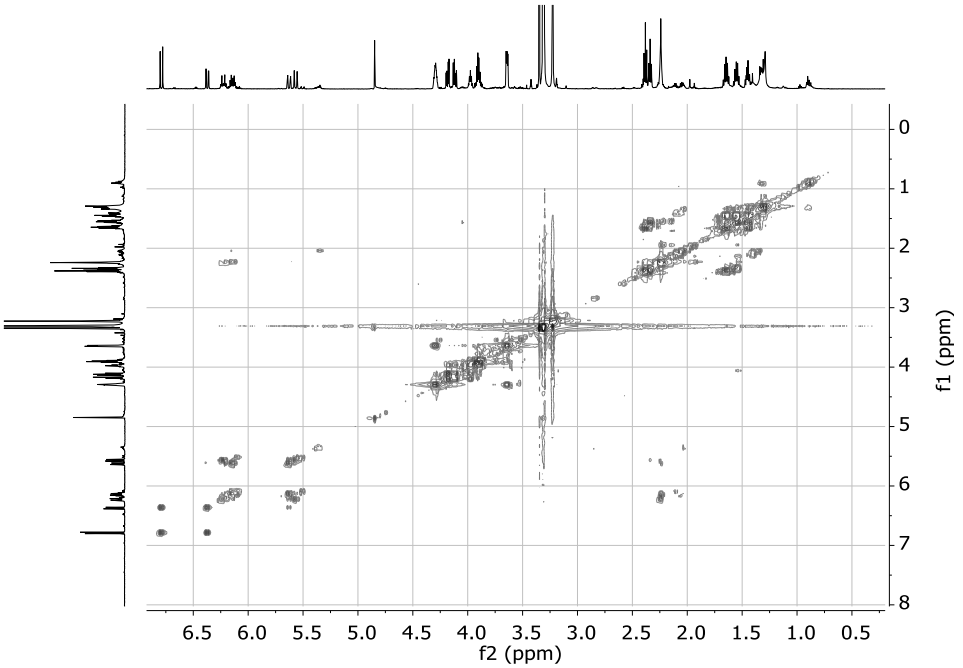


Figure S3: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **2**.

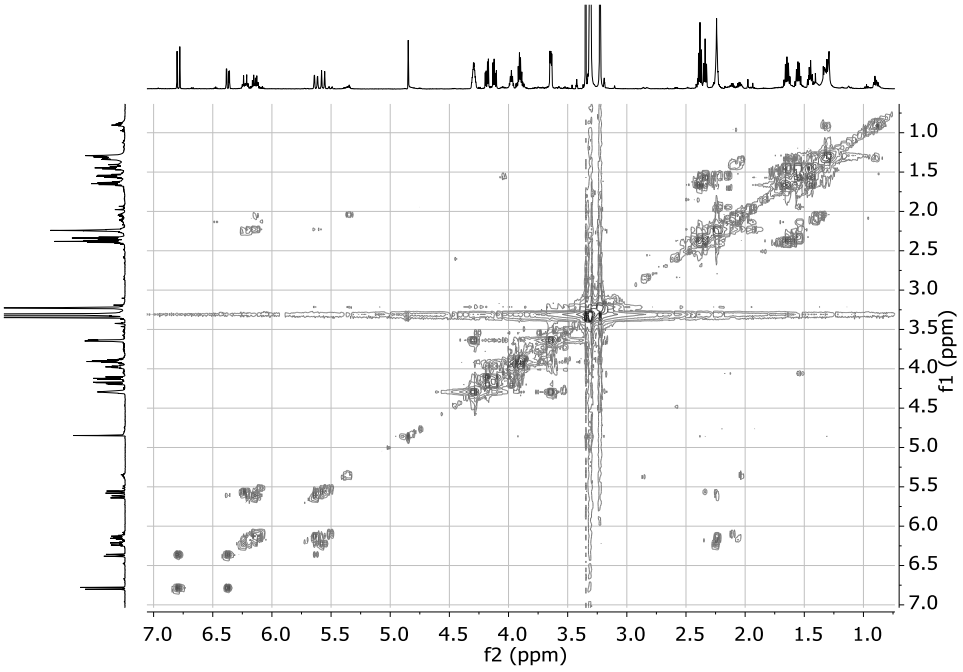


Figure S4: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **3**.

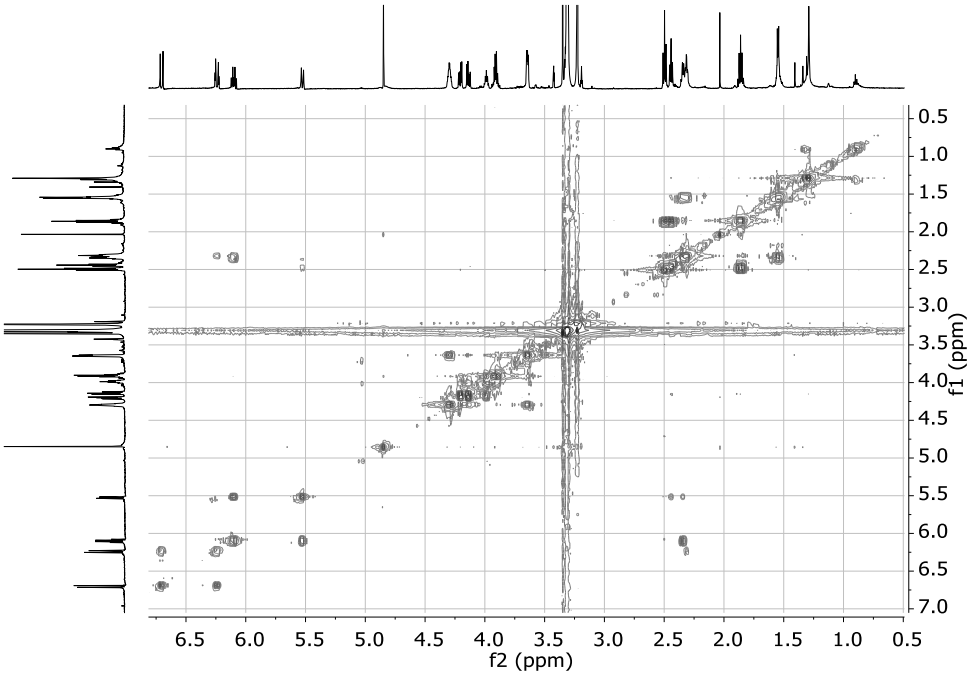


Figure S5: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **4**.

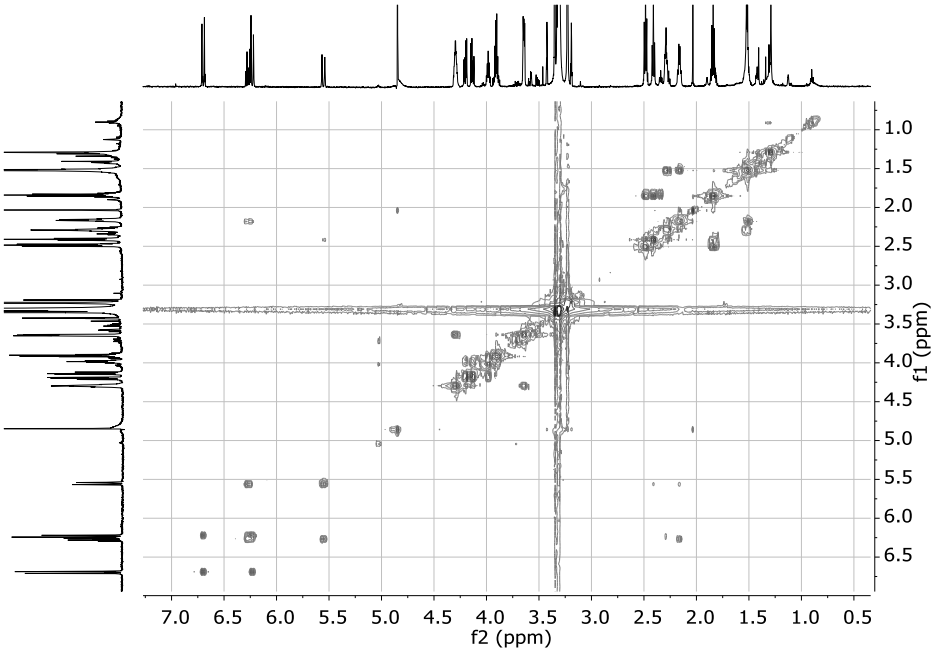


Figure S6: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **5**.

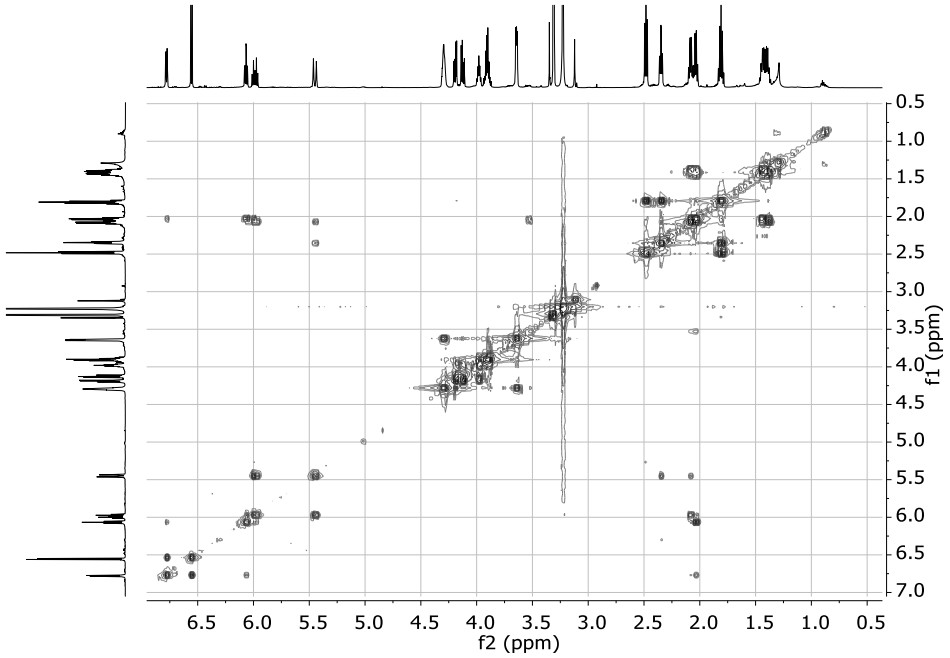


Figure S7: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **6**.

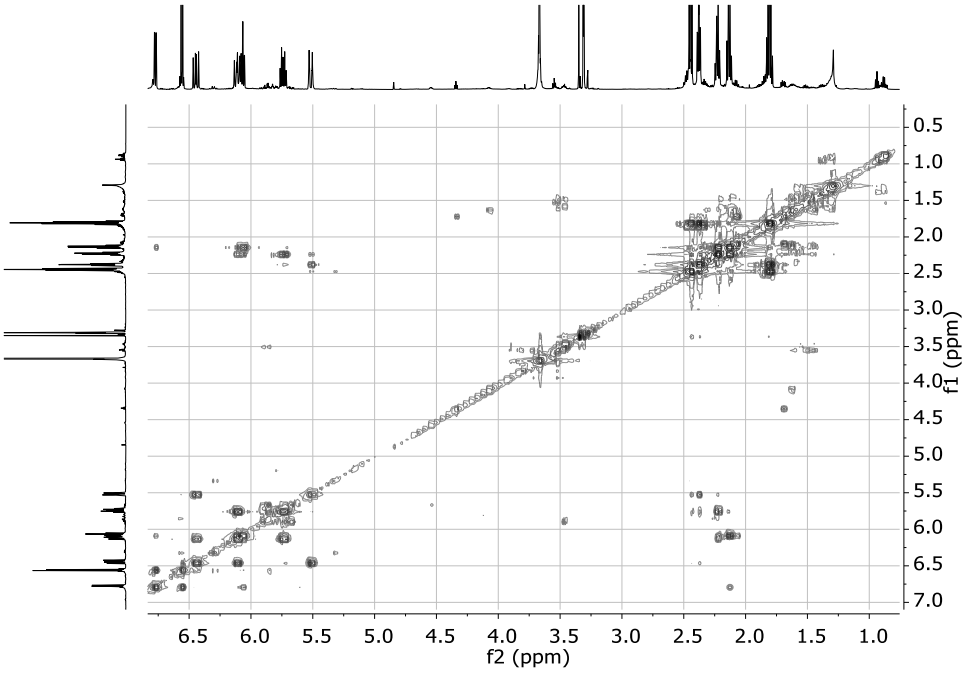


Figure S8: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **7**.

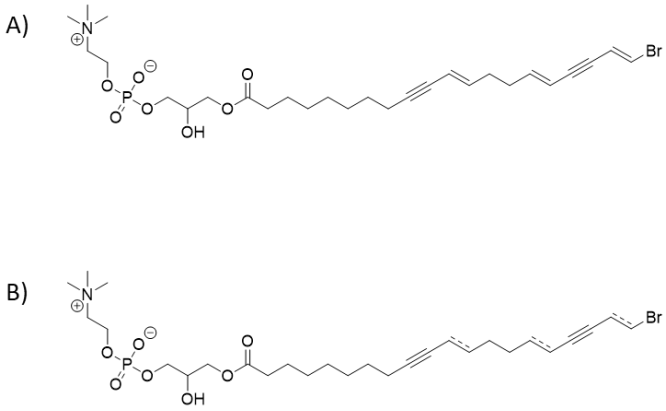


Figure S9: Putative structures characteristic for Group 3 A: m/z 618.20 and B: m/z 616.19

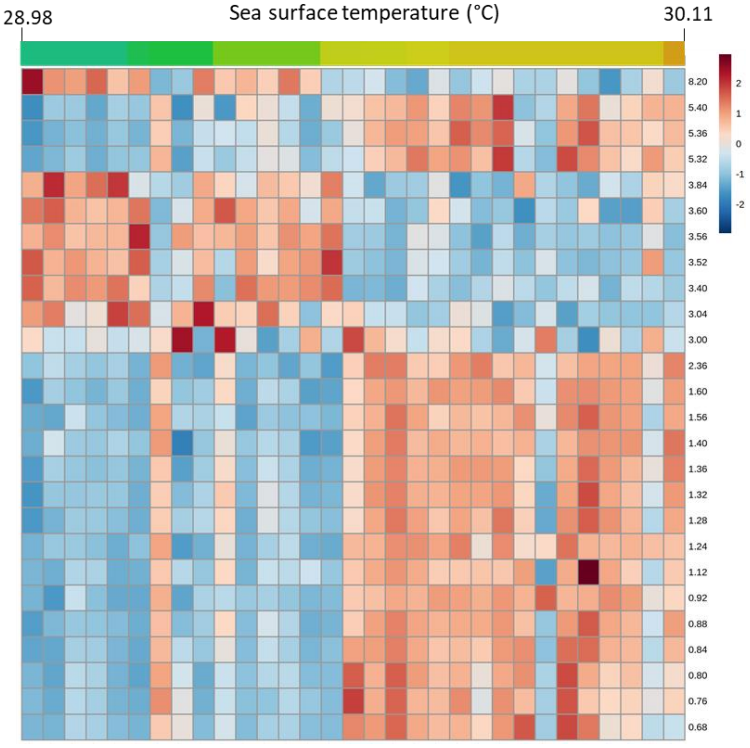
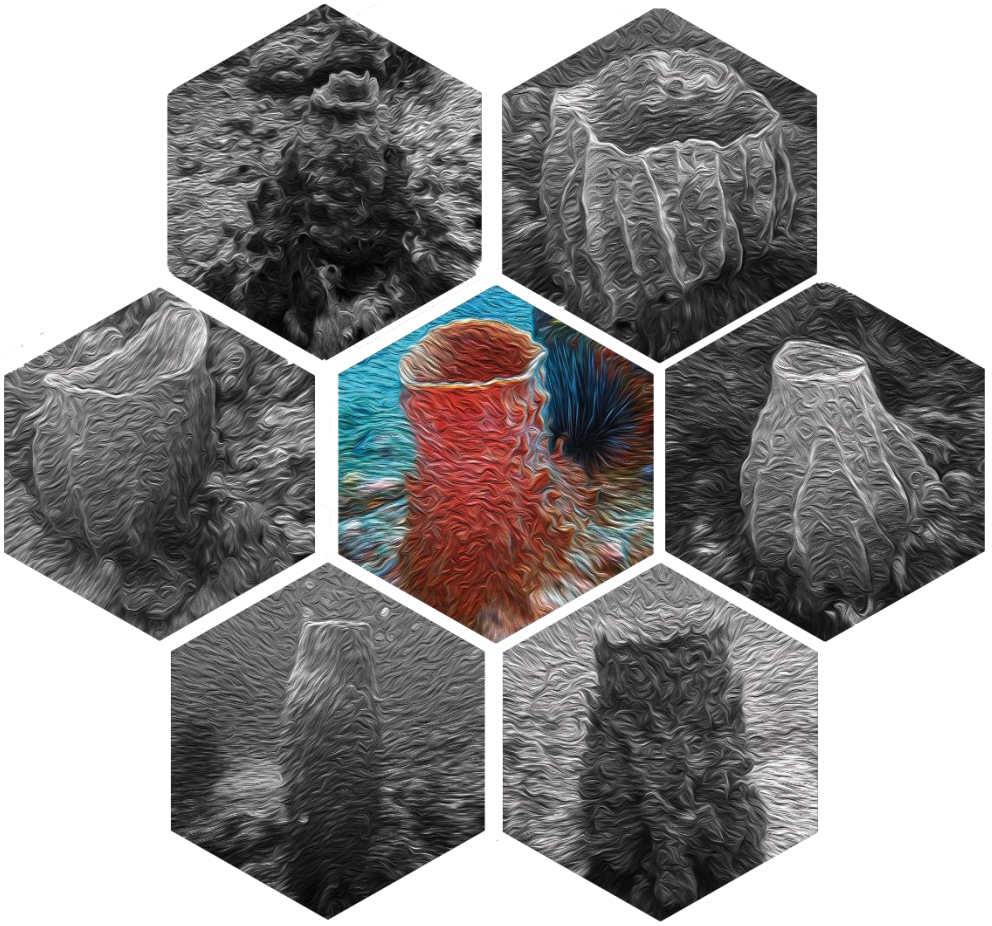


Figure S10: Heat map of characteristic signals from ^{1}H-NMR data obtained from the variable importance for the projection (VIP) plot of partial least-squares (PLS) of samples belonging to genetic group 3 using as Y variable the temperature.



Chapter 7

Concluding remarks and future perspectives

Optimization of Extraction Parameters: obtaining a real metabolome

Since the first report on marine natural products (MNP), marine organisms have been a valuable source of bioactive chemicals due mainly to their chemical diversity and significant bioactivities (Carroll et al. 2019). The exploration of the chemical diversity of marine organisms has increased significantly thanks to the promising results achieved with the implementation of metabolomics, an approach that is based on unbiased chemical profiling that provides a holistic view of their chemical diversity as well as biological activities and ecological functions (Goulitquer et al. 2012; Paul et al. 2019). This profiling technique has allowed the discovery of new compounds that exhibit positive results in different bioassays. But even further, as shown in Chapter 2, the information obtained from the metabolomics studies could be used to unravel the putative ecological functions of many metabolites. However, the great chemical diversity exhibited by MNPs is challenging even for the metabolomics approach. The first and perhaps most difficult challenge encountered is the selection of the extraction method of the samples. Considering that the goal of metabolomics is to study the whole metabolome, it is quite clear that no extraction method could possibly deliver such an extract. The variation in polarity and hydro/lipophilicity of the different metabolites added to their interaction with different matrixes and extreme concentration range make it impossible to extract all metabolites with one single solvent. Notwithstanding this, it is possible at least to optimize conditions related to the extraction process itself that can increase its efficiency in terms of yield and number of metabolites obtained. The influence of factors such as temperature, number of extraction cycles and matrix composition on the chemical diversity of extracts is unsurprising and has been extensively studied (Heavisides et al. 2018; Johnson et al. 2017). Chapter 3 describes the study of extraction parameters such as solvent polarity, temperature, pressure and number of extraction cycles, using a pressure assisted extraction system followed by a Design of Experiment (DOE) analysis. The results showed that the combination of polar solvents like ethanol together with lower extraction temperatures yielded an extract of a giant barrel sponge with the highest chemical diversity. Although these studies provide the first steps towards the development of a protocol for extensive extraction methods, the design of a universal extraction method, particularly for marine sessile organisms, is still far from being a reality. Actually, the understanding of the complex relationship of marine organisms such as corals, algae, and sponges with their symbionts and the debate about which organism really produces many of the MNP could benefit from new extraction strategies which could provide extracts of the host and their symbionts separately.

Redwoods of the reef: Giant barrel sponge age

Marine sponges have been widely studied due to their chemical diversity and the ecological roles as part of many marine ecosystems. Among sponges, giant barrel sponges have drawn particular attention as they are conspicuous organisms in the reefs. In the Caribbean Sea, a growth model of these sponges has shown that they can live for thousands of years, being thus, among the most long-living animals on the planet. In Chapter 5, the study of the influence of the age of sponges on their metabolome showed that the level of phosphatidylcholine phospholipids is higher in older sponges. Moreover, many fatty acids in these phospholipids were found to contain bromine, which have been widely reported in giant barrel sponges. The increase of these lipids, as a constitutive part of the cell membrane, might be related to an increase in the cellular division of the sponge cells. In addition, the higher level of brominated fatty acids in the phospholipids could be related to a chemical defense of the sponge since this kind of compound has been reported to be antiviral, antibacterial and/or cytotoxic (Zhou et al. 2010). From this perspective, these specific fatty acids could play a role in the defense mechanism since giant barrel sponges might use membrane lipids as a storage of these metabolites that can be released by hydrolysis when the sponge is attacked by a predator or a pathogen. This would balance the metabolic cost of these brominated fatty acids (Thoms and Schupp 2007).

The accurate determination of the age of sponges is a difficult task because they lack characteristic markers linked to their age. This difficulty has been circumvented by developing growth models. The Caribbean giant barrel sponge (*Xestospongia muta*) was the first species of sponges for which the age was determined using a growth model (McMurray et al. 2008). Therefore, the influence of the age on the metabolome of sponges shown in Chapter 5 could contribute to the identification of markers of sponge age. Recently, growth models have been developed for other sponge species, which are not as long-lived as *X. muta* (McGrath et al. 2018; Olinger et al. 2019). Studies about whether the age of these sponges can also influence their metabolome could extend the use of chemical markers in the determination of their age. Furthermore, better quantitative information of the relationship between age and chemical production could have implications in the field of drug discovery, as some of the active metabolites may be produced only at certain stages of development (age), a well-known factor in some plants (Yoon et al. 2019). Lastly, the knowledge of age and its effect on sponges could contribute to conservation efforts, such as the creation of artificial reefs. Changes in the metabolome of sponges due to age could affect interactions between sponges and other

animals in the reef, causing differences in the biodiversity between natural and artificial reefs (Perkol-Finkel et al. 2006).

Cryptic species and metabolomics

As a whole, the identification and classification of marine organisms, particularly sponges, still remain an exceedingly difficult task largely because of the ambiguity of their morphologic characteristics which eventually leads to misclassifications and underestimation of the ocean biodiversity. Although the introduction of genetic markers has made it possible to discover cryptic species in many marine taxa, including sponges, the accurate taxonomical classification lags behind (Xavier et al. 2010). Therefore, the possibility of finding cryptic species (two or more different species which are classified as a single species) is quite high. In fact, conserving biodiversity cannot be completed without accurate taxonomical information as well as the knowledge about community stability and population size of the members of an ecosystem (Bickford et al. 2007). Likewise, the commercial use and biotechnological applications of marine sponges, such as the discovery of new bioactive compounds, could be undermined by the utilization of erroneously classified organisms.

The conventional classification of giant barrel sponges species in the past was reviewed in recent studies, and as a result *X. muta* and *X. testudinaria* were divided into nine putative species (genetic groups) (Swierts 2019). Therefore, the differentiation of some of these species from a metabolic perspective was comprehensively studied in Chapter 6, and to a lesser extent in Chapter 5. Chapter 6 shows that the genetic groups present in one location in the Indo-Pacific can be clearly distinguished by their metabolic fingerprint. The difference between genetic groups was mainly attributed to *lyso*-phospholipids that contain phosphatidylcholine. The level of this kind of compound in sponges has been reported to be related to environmental factors as will be discussed in the next section, as well as the reproductive cycle. Most sponges are dioecious, and giant barrel sponges reproduce by mass spawning events in which eggs and sperm are released by different individuals into the environment, a process which thus relies upon the synchronicity of the release of gamete cells. However, spawning events of giant barrel sponges have been observed in different seasons both in the Caribbean and in the Indo-Pacific (Fromont and Bergquist 1994; McMurray et al. 2008; Ritson-Williams et al. 2005) and systematic studies of *X. muta* spawning events also showed some variability at a seasonal, lunar, and temporal scale (Neely and Butler 2020). In this sense, quantitative and qualitative changes in *lyso*- phospholipids contents could indicate that the genetic groups of giant barrel sponges were at different periods in their reproductive

cycle which could result in reproductive isolation of the genetic groups. This would agree with the presence of separate species of giant barrel sponges.

In addition, the differences in the chemical production among the genetic groups of giant barrel sponges can also change the way they interact with other organisms. Previous reports have established that giant barrel sponges display variability in their chemical defenses, implying that the metabolome of sponges of the same population can differ substantially according to their exposure to predators (Chanas and Pawlik 1997; Loh and Pawlik 2014). These observations could be also be related to differences in the metabolic composition of each genetic group since individuals of one or more genetic groups may have a different chemical component in their defense system. This was particular the case in the Caribbean fire sponge, *Tedania ignis,* where the differences in starfish predation of individuals that were supposed to belong to the same species prompted genetic and morphological studies that resulted in the distinction between *T. ignis* and the new species *Tedania klausi*. Lastly, the presence of cryptic species could exacerbate the problem of supply for the discovery of new metabolites as a source of new drugs. For solutions such as mariculture, the selection of the right species for the culture is crucial to secure the production of the active metabolites that are needed. Different species could produce the compound in smaller amounts or eventually produce none at all, resulting in a dead end for this biotechnological approach to get a sustainable supply.

Environmental factors can change the metabolome

As filter feeding organisms, marine sponges have developed an elaborate system to interact with their environment. Therefore, changes in abiotic factors such as light, temperature, pH, salinity, and nutrient concentration can influence the production of secondary metabolites in marine sponges. In Chapters 4, 5 and 6 the effect of different environmental conditions was evaluated to determine their effect on the metabolome of giant barrel sponges. From a global perspective, Chapter 4 shows the differences in the chemical profile of giant barrel sponges collected in two locations in the Caribbean Sea and two in the Indo-Pacific region. The location can be perceived as a complex factor, as multiple conditions such as temperature, pH, predatory stress, water currents, and nutrients might differ. Therefore, the metabolic production observed in each location is the result of the interaction between all these conditions. At the same time, establishing general trends is difficult, as shown for example by the activity of some of the tested extracts against *Staphylococcus aureus*. No correlation was found between the locations and activity of the samples indicating that the production of

active compounds might be triggered either by environmental conditions occurring at a smaller geographical scale or by biotic factors.

In Chapter 5, another complex factor, depth, was evaluated with sponges collected in the Caribbean Sea. It was determined that of the two genetic groups studied, only sponges belonging to genetic group 7 showed a change in their metabolome along the depth gradient. Depth is also considered a complex factor due to variations in conditions experienced by the sponges at different depths. Further aspects related to environment are reported in Chapter 6 which describes the effects of pH and sea surface temperature (SST) of *Xestospongia testudinaria* in Indonesia. Once again, among the three studied genetic groups (groups 1, 2 and 3) only the metabolome of group 1 displayed changes due to differences in SST, while none showed significant variations in the metabolome as a result of differences in pH. Interestingly, in both studies the response of different genetic groups to similar gradients in environmental conditions was different. Even though the classification into nine genetic groups could imply that there are at least nine different species of giant barrel sponges (Swierts 2019), these species are still very closely related. The fact that they respond differently to changes in their environment could indicate that they have different levels of adaptability when exposed to similar conditions. Among others, this could have implications on the survival of these sponges, particularly nowadays, as ocean environments are experiencing extreme changes related to global events such as global warming and ocean acidification. From a conservational perspective, this could mean that one or more of the genetic groups of giant barrel sponges could be less fit to endure the new environmental conditions and these groups could decline or even disappear. This prospect is more worrying if we take into account that the different genetic groups of giant barrel sponges were discovered only a few years ago, and similar cases could be occurring for several other sponge species. Consequently, the lack of adaptation of some of these cryptic species could result in a largely unnoticed loss of biodiversity.

Another aspect that is noteworthy is that the microbiomes of sponges in general, and giant barrel sponges in particular, have been reported to change their composition depending on environmental factors such as geographical location, depth, temperature and season. It is reasonable to anticipate that these changes in the microbiome will be reflected in changes in the metabolome, since the analyzed extracts correspond to the holobiont. However, it is also possible that modifications in the metabolome of the holobiont are not strictly related to the composition of the microbiome but can also respond to the activation of different metabolic pathways in the microorganism associated with the sponge. Extensive studies into the

interactions between sponges and their microsymbionts are therefore needed to gain insight into the production of metabolites and its regulation.

Future perspectives

The use of metabolomics to explore the ecological relationships of marine organisms and as an incentive for the discovery of new molecules from marine sources has provided a holistic perspective to these fields. This interest has been enhanced by the continuous development of technical and bioinformatic tools for metabolomics analysis that has allowed a deeper understanding of the metabolism of marine organism. At present, metabolomics provides only a snapshot, an instantaneous image of the metabolism of an organism at the moment of its collection. However, natural processes are extremely dynamic, and even more so in marine environments, therefore a method that continually monitors the metabolome in small time lapses would provide valuable information about the metabolism of marine organism and the ecological interactions mediated through chemical compounds, which is especially significant in sessile organisms such as sponges. In addition, in the last few years, the debate about which organism produces the secondary metabolites commonly isolated from marine invertebrates is favoring the associated microorganisms. This has increased the interest in the metabolites produced by microorganisms. Yet, while they have proved to be a prolific source of new compounds, in many cases the chemical space occupied by the metabolites isolated from associated microorganisms is very different from the one occupied by metabolites from the holobiont. This thesis explores the changes related to environmental conditions in the metabolome of giant barrel sponges as holobionts. Studies conducted on the composition of the microbiome in similar conditions have also shown that they are similarly affected (Lesser et al. 2016; Montalvo and Hill 2011; Swierts et al. 2018; Villegas-Plazas et al. 2019). However, the intersection between these situations has not really been evaluated. Understanding the correlation between the changes in the microbiome and the metabolome would be a good starting point to unravel the complexity of the networks that connect the sponge cells with the sponge microsymbionts through chemical compounds. The development of technologies such as mass spectrometry imaging and single cell metabolomics, constitute a step in that direction, showing that specific metabolites are accumulated in the sponge cells while others are mainly in the microsymbionts. However, the question of the relocation of compounds that can be used as precursors from the microsymbiont to the sponge cells or vice-versa remains unknown.

References

Adnani, N., Chevrette, M. G., Adibhatla, S. N., Zhang, F., Yu, Q., Braun, D. R., et al. (2017). Coculture of Marine Invertebrate-Associated Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. *ACS Chemical Biology*, 12(12), 3093–3102.

Adpressa, D. A., & Loesgen, S. (2016). Bioprospecting Chemical Diversity and Bioactivity in a Marine Derived *Aspergillus terreus*. *Chemistry & Biodiversity*, 13(2), 253–259.

Akiyama, T., Takada, K., Oikawa, T., Matsuura, N., Ise, Y., Okada, S., & Matsunaga, S. (2013). Stimulators of adipogenesis from the marine sponge *Xestospongia testudinaria*. *Tetrahedron*, *69*(32), 6560–6564.

Alexander, B. E., Liebrand, K., Osinga, R., van der Geest, H. G., Admiraal, W., Cleutjens, J. P. M., et al. (2014). Cell Turnover and Detritus Production in Marine Sponges from Tropical and Temperate Benthic Ecosystems. *PLOS ONE*, 9(10), e109486.

Ali, K., Iqbal, M., Yuliana, N. D., Lee, Y.-J., Park, S., Han, S., et al. (2013). Identification of bioactive metabolites against adenosine A1 receptor using NMR-based metabolomics. *Metabolomics*, 9(4), 778–785.

Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., et al. (2016). Integration of Molecular Networking and *In-Silico* MS/MS Fragmentation for Natural Products Dereplication. *Analytical Chemistry*, 88(6), 3317–3323.

Allwood, J. W., & Goodacre, R. (2010). An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochemical Analysis*, 21(1), 33–47.

Alsufyani, T., Weiss, A., & Wichard, T. (2017). Time Course Exo-Metabolomic Profiling in the Green Marine Macroalga *Ulva* (Chlorophyta) for Identification of Growth Phase-Dependent Biomarkers. *Marine drugs*, 15(1), 14.

Altmann, K.-H. (2017). Drugs from the Oceans: Marine Natural Products as Leads for Drug Discovery. *CHIMIA International Journal for Chemistry*, 71(10), 646–652.

Amnuoypol, S., Suwanborirux, K., Pummangura, S., Kubo, A., Tanaka, C., & Saito, N. (2004). Chemistry of Renieramycins. Part 5. Structure Elucidation of Renieramycin-Type Derivatives O, Q, R, and S from Thai Marine Sponge *Xestospongia* Species Pretreated with Potassium Cyanide. *Journal of Natural Products*, 67(6), 1023–1028.

Anderson, P. E., Mahle, D. A., Doom, T. E., Reo, N. V., DelRaso, N. J., & Raymer, M. L. (2011). Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. *Metabolomics*, 7(2), 179–190.

Antcliffe, J. B., Callow, R. H. T., & Brasier, M. D. (2014). Giving the early fossil record of sponges a squeeze. *Biological Reviews*, 89(4), 972–1004.

AntiBase, Wiley-VCH. (n.d.). https://application.wiley-vch.de/stmdata/antibase.php. Accessed 30 October 2019

Barra, L., Barac, P., König, G. M., Crüsemann, M., & Dickschat, J. S. (2017). Volatiles from the fungal microbiome of the marine sponge *Callyspongia* cf. *flammea*. *Organic & Biomolecular Chemistry*, 15(35), 7411–7421.

Bauvais, C., Bonneau, N., Blond, A., Pérez, T., Bourguet-Kondracki, M.-L., & Zirah, S. (2017). Furanoterpene Diversity and Variability in the Marine Sponge *Spongia officinalis*, from Untargeted LC–MS/MS Metabolomic Profiling to Furanolactam Derivatives. *Metabolites*, 7(2), 27.

Bayet-Robert, M., Lim, S., Barthomeuf, C., & Morvan, D. (2010). Biochemical disorders induced by cytotoxic marine natural products in breast cancer cells as revealed by proton NMR spectroscopy-based metabolomics. *Biochemical Pharmacology*, 80(8), 1170–1179.

Bayona, L. M., van Leeuwen, G., Erol, Ö., Swierts, T., van der Ent, E., de Voogd, N. J., & Choi, Y. H. (2020). Influence of Geographical Location on the Metabolic Production of Giant Barrel Sponges (*Xestospongia* spp.) Revealed by Metabolomics Tools. ACS Omega, 5(21), 12398–12408.

Bayona, L. M., Verpoorte, R., Klinkhamer, P. G. L., & Choi, Y. H. (2019). Thin-Layer Chromatography | Metabolomics. In P. Worsfold, C. Poole, A. Townshend, & M. Miró (Eds.), *Encyclopedia of Analytical Science* (Third Edition) (pp. 59–75). Oxford: Academic Press.

Bayona, M. L., Videnova, M., & Choi, H. Y. (2018). Increasing Metabolic Diversity in Marine Sponges Extracts by Controlling Extraction Parameters. *Marine Drugs*, 16(10), 393.

Becerro, M. A., Thacker, R. W., Turon, X., Uriz, M. J., & Paul, V. J. (2003). Biogeography of sponge chemical ecology: comparisons of tropical and temperate defenses. *Oecologia*, 135(1), 91–101.

Belarbi, E. H., Contreras Gómez, A., Chisti, Y., García Camacho, F., & Molina Grima, E. (2003). Producing drugs from marine sponges. *Biotechnology Advances*, 21(7), 585–598.

Belghit, I., Rasinger, J. D., Heesch, S., Biancarosa, I., Liland, N., Torstensen, B., et al. (2017). In-depth metabolic profiling of marine macroalgae confirms strong biochemical differences between brown, red and green algae. *Algal Research*, 26, 240–249.

Bell, James J., Biggerstaff, A., Bates, T., Bennett, H., Marlow, J., McGrath, E., & Shaffer, M. (2017). Sponge monitoring: Moving beyond diversity and abundance measures. *Ecological Indicators*, 78, 470–488.

Bell, J. J., Bennett, H. M., Rovellini, A., & Webster, N. S. (2018). Sponges to Be Winners under Near-Future Climate Scenarios. *BioScience*, *68*(12), 955–968.

Bell, James J., & Carballo, J. L. (2008). Patterns of sponge biodiversity and abundance across different biogeographic regions. *Marine Biology*, 155(6), 563–570.

Bell, J. J., Davy, S. K., Jones, T., Taylor, M. W., & Webster, N. S. (2013). Could some coral reefs become sponge reefs as our climate changes? *Global Change Biology*, 19(9), 2613–2624.

Bell, J. J., Rovellini, A., Davy, S. K., Taylor, M. W., Fulton, E. A., Dunn, M. R., et al. (2018). Climate change alterations to ecosystem dominance: how might sponge-dominated reefs function? *Ecology*, 99(9), 1920–1931.

Bell, J.J., & Smith, D. (2004). Ecology of sponge assemblages (Porifera) in the Wakatobi region, south-east Sulawesi, Indonesia: Richness and abundance. *Journal of the Marine Biological Association of the United Kingdom*, 84(3), 581–591.

Bell, J. J., Smith, D., Hannan, D., Haris, A., Jompa, J., & Thomas, L. (2014). Resilience to Disturbance Despite Limited Dispersal and Self-Recruitment in Tropical Barrel Sponges: Implications for Conservation and Management. *PLOS ONE*, 9(3), e91635.

Bell, James J. (2008). The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*, 79(3), 341–353.

Bergquist, P. R. (1978). Sponges. Hutchinson of London.

Betancur, L. A., Naranjo-Gaybor, S. J., Vinchira-Villarraga, D. M., Moreno-Sarmiento, N. C., Maldonado, L. A., Suarez-Moreno, Z. R., et al. (2017). Marine Actinobacteria as a source of compounds for phytopathogen control: An integrative metabolic-profiling / bioactivity and taxonomical approach. *PLOS ONE*, 12(2), e0170148.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., et al. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148–155.

Bingol, K., Bruschweiler-Li, L., Li, D.-W., & Brüschweiler, R. (2014). Customized Metabolomics Database for the Analysis of NMR ^{1}H–^{1}H TOCSY and ^{13}C–^{1}H HSQC-TOCSY Spectra of Complex Mixtures. *Analytical Chemistry*, 86(11), 5494–5501.

Birgbauer, E., & Chun, J. (2006). New developments in the biological functions of lysophospholipids. *Cellular and Molecular Life Sciences CMLS*, 63(23), 2695–2701.

Blunt, J. W., Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., & Prinsep, M. R. (2018). Marine natural products. *Natural Product Reports*, 35(1), 8–53.

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., & Prinsep, M. R. (2017). Marine natural products. Natural Product Reports, 34(3), 235–294.

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., & Prinsep, M. R. (2016). Marine natural products. *Natural Product Reports*, 33(3), 382–431.

Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., & Prinsep, M. R. (2015). Marine natural products. *Natural Product Reports*, 32(2), 116–211.

Blunt, J., Munro, M., & Upjohn, M. (2012). The Role of Databases in Marine Natural Products Research. In E. Fattorusso, W. H. Gerwick, & O. Taglialatela-Scafati (Eds.), *Handbook of Marine Natural Products* (pp. 389–421). Dordrecht: Springer Netherlands.

Boroujerdi, A. F. B., Vizcaino, M. I., Meyers, A., Pollock, E. C., Huynh, S. L., Schock, T. B., et al. (2009). NMR-Based Microbial Metabolomics and the Temperature-Dependent Coral Pathogen *Vibrio coralliilyticus*. *Environmental Science & Technology*, 43(20), 7658–7664.

Borras, E., Aksenov, A. A., Baird, M., Novick, B., Schivo, M., Zamuruyev, K. O., et al. (2017). Exhaled breath condensate methods adapted from human studies using longitudinal metabolomics for predicting early health alterations in dolphins. *Analytical and Bioanalytical Chemistry*, 409(28), 6523–6536.

Bose, U., Hewavitharana, A. K., Ng, Y. K., Shaw, P. N., Fuerst, J. A., & Hodson, M. P. (2015). LC-MS-Based Metabolomics Study of Marine Bacterial Secondary Metabolite and Antibiotic Production in *Salinispora arenicola*. *Marine Drugs*, 13(1), 249-266.

Boudreau, P. D., Monroe, E. A., Mehrotra, S., Desfor, S., Korobeynikov, A., Sherman, D. H., et al. (2015). Expanding the Described Metabolome of the Marine Cyanobacterium *Moorea producens* JHB through Orthogonal Natural Products Workflows. *PLOS ONE*, 10(7), e0133297.

Bourguet-Kondracki, M. L., Rakotoarisoa, M. T., Martin, M. T., & Guyot, M. (1992). Bioactive bromopolyacetylenes from the marine sponge *Xestospongia testudinaria*. *Tetrahedron Letters*, 33(2), 225–226.

Boury-Esnault, N., Lavrov, D. V., Ruiz, C. A., & Pérez, T. (2013). The Integrative Taxonomic Approach Applied to Porifera: A Case Study of the *Homoscleromorpha*. *Integrative and Comparative Biology*, *53*(3), 416–427.

Brantley, S. E., Molinski, T. F., Preston, C. M., & DeLong, E. F. (1995). Brominated acetylenic fatty acids from *Xestospongia* sp., a marine sponge-bacteria association. *Tetrahedron*, *51*(28), 7667–7672.

Bundy, J. G., Davey, M. P., & Viant, M. R. (2008). Environmental metabolomics: a critical review and future perspectives. *Metabolomics*, 5(1), 3-21.

Cachet, N., Genta-Jouve, G., Ivanisevic, J., Chevaldonné, P., Sinniger, F., Culioli, G., et al. (2015). Metabolomic profiling reveals deep chemical divergence between two morphotypes of the zoanthid *Parazoanthus axinellae*. *Scientific Reports*, 5, 8282.

Cai, H.-L., Li, H.-D., Yan, X.-Z., Sun, B., Zhang, Q., Yan, M., et al. (2012). Metabolomic Analysis of Biochemical Changes in the Plasma and Urine of First-Episode Neuroleptic-Naïve Schizophrenia Patients after Treatment with Risperidone. *Journal of Proteome Research*, 11(8), 4338–4350.

Calcul, L., Longeon, A., Mourabit, A. A., Guyot, M., & Bourguet-Kondracki, M.-L. (2003). Novel alkaloids of the aaptamine class from an Indonesian marine sponge of the genus *Xestospongia*. *Tetrahedron*, 59(34), 6539–6544.

Cappello, T., Fernandes, D., Maisano, M., Casano, A., Bonastre, M., Bebianno, M. J., et al. (2017). Sex steroids and metabolic responses in mussels *Mytilus galloprovincialis* exposed to drospirenone. *Ecotoxicology and Environmental Safety*, 143, 166–172.

Cappello, T., Giannetto, A., Parrino, V., De Marco, G., Mauceri, A., & Maisano, M. (2018). Food safety using NMR-based metabolomics: Assessment of the Atlantic bluefin tuna, *Thunnus thynnus*, from the Mediterranean Sea. *Food and Chemical Toxicology*, 115, 391–397.

Carballeira, N. M., & Maldonado, L. (1988). The phospholipid fatty acids of the marine sponge *Xestospongia muta*. *Lipids*, *23*(7), 682–684.

Carr, M. H., Neigel, J. E., Estes, J. A., Andelman, S., Warner, R. R., & Largier, J. L. (2003). Comparing marine and terrestrial ecosystems: Implications for the design of coastal marine reserves. *Ecological Applications*, 13(1), 90–107.

Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., & Prinsep, M. R. (2019). Marine natural products. *Natural Product Reports*, 36(1), 122–173.

Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., & Prinsep, M. R. (2020). Marine natural products. *Natural Product Reports*, 37(2), 175–223.

Chanas, B., & Pawlik, J. R. (1997). Variability in the chemical defense of the Caribbean reef sponge *Xestospongia muta*. In Lessios HA, Macintyre IG (eds) *Proceedings of the 8th international coral reef symposium* (Vol. 2, pp. 1363–1368). Smithsonian Tropical Research Institute, Balboa.

Change, N. G. C. (2020). Global Surface Temperature | NASA Global Climate Change. *Climate Change: Vital Signs of the Planet*. https://climate.nasa.gov/vital-signs/global-temperature. Accessed 28 September 2020

ChemSpider, RSC http://www.chemspider.com/. Accessed 30 October 2019

Chen, S., Zhang, C., Xiong, Y., Tian, X., Liu, C., Jeevithan, E., & Wu, W. (2015). A GC-MS-based metabolomics investigation on scallop (*Chlamys farreri*) during semi-anhydrous living-preservation. Innovative *Food Science & Emerging Technologies*, 31, 185–195.

Cheng, K., Müllner, E., Moazzami, A. A., Carlberg, H., Brännäs, E., & Pickova, J. (2017). Metabolomics Approach To Evaluate a Baltic Sea Sourced Diet for Cultured Arctic Char (*Salvelinus alpinus L.*). *Journal of Agricultural and Food Chemistry*, 65(24), 5083–5090.

Chong, J., Yamamoto, M., & Xia, J. (2019). MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. Metabolites, 9(3), 57.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, 46(W1), W486-W949.

Clendinen, C. S., Lee-McMullen, B., Williams, C. M., Stupp, G. S., Vandenborne, K., Hahn, D. A., et al. (2014). ^{13}C NMR metabolomics: applications at natural abundance. *Analytical chemistry*, 86(18), 9242–9250.

Cragg, G. M., Newman, D. J., & Snader, K. M. (1997). Natural Products in Drug Discovery and Development. *Journal of Natural Products*, 60(1), 52–60.

Cuevas, C., & Francesch, A. (2009). Development of Yondelis® (trabectedin, ET-743). A semisynthetic process solves the supply problem. *Natural Product Reports*, 26(3), 322–337.

Cutignano, A., Nuzzo, G., Ianora, A., Luongo, E., Romano, G., Gallo, C., et al. (2015). Development and Application of a Novel SPE-Method for Bioassay-Guided Fractionation of Marine Extracts. *Marine Drugs*, 13(9), 5736-5749.

D'Arrigo, P., & Servi, S. (2010). Synthesis of Lysophospholipids. *Molecules*, 15(3), 1354–1377.

da Silva, R. R., Wang, M., Nothias, L.-F., van der Hooft, J. J. J., Caraballo-Rodríguez, A. M., Fox, E., et al. (2018). Propagating annotations of molecular networks using in silico fragmentation. *PLOS Computational Biology*, 14(4), e1006089.

Dahal, R. H., & Kim, J. (2017). Rhodanobacter humi sp. nov., an acid-tolerant and alkalitolerant gammaproteobacterium isolated from forest soil. *International Journal of Systematic and Evolutionary Microbiology*, 67(5), 1185–1190.

de Alencar, D. B., Diniz, J. C., Rocha, S. A. S., dos Santos Pires-Cavalcante, K. M., Freitas, J. O., Nagano, C. S., et al. (2017). Chemical composition of volatile compounds in two red seaweeds, *Pterocladiella capillacea* and *Osmundaria obtusiloba*, using static headspace gas chromatography mass spectrometry. *Journal of Applied Phycology*, 29(3), 1571–1576.

de Goeij, J. M., van Oevelen, D., Vermeij, M. J. A., Osinga, R., Middelburg, J. J., de Goeij, A. F. P. M., & Admiraal, W. (2013). Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. Science, 342(6154), 108-110.

de Goeij, J. M., De Kluijver, A., Van Duyl, F. C., Vacelet, J., Wijffels, R. H., De Goeij, A. F. P. M., et al. (2009). Cell kinetics of the marine sponge *Halisarca caerulea* reveal rapid cell turnover and shedding. *Journal of Experimental Biology*, 212(23), 3892-3900.

de Voogd, N. J. (2007). The mariculture potential of the Indonesian reef-dwelling sponge *Callyspongia* (Euplacella) *biru*: Growth, survival and bioactive compounds. *Aquaculture*, 262(1), 54–64.

De Vos, L., Rützler, K., Boury-Esnault, N., Donadey, C., & Vacelet, J. (1991). Atlas of sponge morphology. Smithsonian Institution Press.

Diaz, M. C., & Rützler, K. (2001). Sponges: An essential component of Caribbean coral reefs. *Bulletin of Marine Science*, 62(2), 535–546.

Ding, Y. C., Pang, M. L., Liang, Z.-X., Goh, K. K., Glukhov, E., Gerwick, H. W., & Tan, T. L. (2018). MS/MS-Based Molecular Networking Approach for the Detection of Aplysiatoxin-Related Compounds in Environmental Marine Cyanobacteria. *Marine Drugs*, 16(12), 505.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., et al. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of cheminformatics*, 8, 61–61.

DNP 28.2, Dictionary of Natural Products 28.2, CRC. http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml. Accessed 20 May 2020

Duckworth AR, West L, Vansach T, Stubler A, & Hardt M. (2012). Effects of water temperature and pH on growth and metabolite biosynthesis of coral reef sponges. *Marine Ecology Progress Series*, *462*, 67–77.

Dunn, W. B., & Ellis, David. I. (2005). Metabolomics: Current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry*, 24(4), 285–294.

Ebada, S. S., & Proksch, P. (2012). The Chemistry of Marine Sponges(∗). *Handbook of Marine Natural Products*, 191–293.

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

Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., et al. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, 9(7):123

Ernst, M., Kang, B. K., Caraballo-Rodríguez, M. A., Nothias, L.-F., Wandy, J., Chen, C., et al. (2019). MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites*, 16;9(7):144.

Erpenbeck, D., & van Soest, R. W. M. (2006). Status and Perspective of Sponge Chemosystematics. *Marine Biotechnology*,9,2.

Esquenazi, E., Coates, C., Simmons, L., Gonzalez, D., Gerwick, W. H., & Dorrestein, P. C. (2008). Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. *Molecular BioSystems*, 4(6), 562–570.

Esquenazi, E., Yang, Y.-L., Watrous, J., Gerwick, W. H., & Dorrestein, P. C. (2009). Imaging mass spectrometry of natural products. *Natural Product Reports*, 26(12), 1521–1534.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S., & Thomas, T. (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences*, 109(27) E1878-E1887.

Fan, T. W.-M. (1996). Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 28(2), 161–219.

Farag, M. A., Porzel, A., Al-Hammady, M. A., Hegazy, M.-E. F., Meyer, A., Mohamed, T. A., et al. (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(4), 1274–1287.

Favre, L., Ortalo-Magné, A., Greff, S., Pérez, T., Thomas, O. P., Martin, J.-C., & Culioli, G. (2017). Discrimination of Four Marine Biofilm-Forming Bacteria by LC–MS Metabolomics and Influence of Culture Parameters. *Journal of Proteome Research*, 16(5), 1962–1975.

Fernández-Varela, R., Tomasi, G., & Christensen, J. H. (2015). An untargeted gas chromatography mass spectrometry metabolomics platform for marine polychaetes. *Journal of Chromatography A,* 1384, 133–141.

Fiore, C. L., Baker, D. M., & Lesser, M. P. (2013). Nitrogen Biogeochemistry in the Caribbean Sponge, *Xestospongia muta*: A Source or Sink of Dissolved Inorganic Nitrogen? *PLOS ONE*, 8(8), e72961.

Fiore, C. L., Jarett, J. K., & Lesser, M. P. (2013). Symbiotic prokaryotic communities from different populations of the giant barrel sponge, *Xestospongia muta*. *MicrobiologyOpen*, 2(6), 938–952.

Floros, D. J., Jensen, P. R., Dorrestein, P. C., & Koyama, N. (2016). A metabolomics guided exploration of marine natural product chemical space. *Metabolomics*, 12, 145.

Forcisi, S., Moritz, F., Kanawati, B., Tziotis, D., Lehmann, R., & Schmitt-Kopplin, P. (2013). Liquid chromatography–mass spectrometry in metabolomics research: Mass analyzers in ultra high pressure liquid chromatography coupling. *State-of-the art of (UHP)LC-MS(-MS) techniques and their practical application*, 1292, 51–65.

Forner, D., Berrué, F., Correa, H., Duncan, K., & Kerr, R. G. (2013). Chemical dereplication of marine actinomycetes by liquid chromatography–high resolution mass spectrometry profiling and statistical analysis. *Analytica Chimica Acta*, 805, 70–79.

Forshed, J., Schuppe-Koistinen, I., & Jacobsson, S. P. (2003). Peak alignment of NMR signals by means of a genetic algorithm. *Analytica Chimica Acta*, 487(2), 189–199.

Fromont, J. (1991). Descriptions of species of the *Petrosida* (Porifera:Demospongiae) occurring in the tropical waters of the Great Barrier Reef. *The Beagle, Records of the Northern Territory Museum of Arts and Sciences*, 8(1), 73–96.

Fromont, J., & Bergquist, P. R. (1994). Reproductive biology of three sponge species of the genus *Xestospongia* (Porifera: Demospongiae: Petrosida) from the Great Barrier Reef. *Coral Reefs*, 13(2), 119–126.

Fusetani, N., Li, H., Tamura, K., & Matsunaga, S. (1993). Antifungal Brominated C18 Acetylenic Acids from the Marine Sponge, *Petrosia volcano* Hoshino. *Tetrahedron*, *49*(6), 1203–1210.

Garg, N., Kapono, C. A., Lim, Y. W., Koyama, N., Vermeij, M. J. A., Conrad, D., et al. (2015). Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. *Special Issue: MS 1960 to Now*, *377*, 719–727.

Gauvin, A., Smadja, J., Aknin, M., & Gaydou, E. M. (2004). Sterol composition and chemotaxonomic considerations in relation to sponges of the genus *Xestospongia*. *Biochemical Systematics and Ecology*, 32(5), 469–476.

Gauvin, A., Smadja, J., Aknin, M., & Gaydou, E. M. (2004). Sterol composition and chemotaxonomic considerations in relation to sponges of the genus *Xestospongia*. Biochemical Systematics and Ecology, 32(5), 469–476.

Gazave, E., Lapébie, P., Ereskovsky, A. V., Vacelet, J., Renard, E., Cárdenas, P., & Borchiellini, C. (2012). No longer Demospongiae: *Homoscleromorpha* formal nomination as a fourth class of Porifera. In M. Maldonado, X. Turon, M. Becerro, & M. Jesús Uriz (Eds.), *Ancient Animals, New Challenges: Developments in Sponge Research* (pp. 3–10). Dordrecht: Springer Netherlands.

Genin, E., Wielgosz-Collin, G., Njinkoué, J.-M., Velosaotsy, N. E., Kornprobst, J.-M., Gouygou, J.-P., et al. (2008). New trends in phospholipid class composition of marine sponges. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 150(4), 427–431.

Gerwick, W. H. (2017). The Face of a Molecule. *Journal of Natural Products*, 80(9), 2583–2588.

Gerwick, W., & Moore, B. (2012). Lessons from the Past and Charting the Future of Marine Natural Products Drug Discovery and Chemical Biology. *Chemistry & biology*, 19(1), 85–98.

Gidman, E. A., Jones, M. L. M., Bussell, J. A., Malham, S. K., Reynolds, B., Seed, R., et al. (2007). A methodology for screening haemolymph of intertidal mussels, *Mytilus edulis*, using FT-IR spectroscopy as a tool for environmental assesment. *Metabolomics*, 3(4), 465–473.

Gloeckner, V., Wehrl, M., Moitinho-Silva, L., Gernert, C., Schupp, P., Pawlik, J. R., et al. (2014). The HMA-LMA Dichotomy Revisited: an Electron Microscopical Survey of 56 Sponge Species. *The Biological Bulletin*, 227(1), 78–88.

Goldfine, H., & Ellis, M. E. (1963). N-methyl groups in bacterial lipids. *Journal of Bacteriology*, 87, 8–15.

Gomes, N. G. M., Dasari, R., Chandra, S., Kiss, R., & Kornienko, A. (2016). Marine Invertebrate Metabolites with Anticancer Activities: Solutions to the "Supply Problem." *Marine Drugs*, 14(5), 98.

Goulitquer, S., Potin, P., & Tonon, T. (2012). Mass spectrometry-based metabolomics to elucidate functions in marine organisms and ecosystems. *Marine drugs*, 10(4), 849–880.

Grand-Guillaume Perrenoud, A., Guillarme, D., Boccard, J., Veuthey, J.-L., Barron, D., & Moco, S. (2016). Ultra-high performance supercritical fluid chromatography coupled with quadrupole-time-of-flight mass spectrometry as a performing tool for bioactive analysis. *Journal of Chromatography A*, 1450, 101–111.

Greff, S., Aires, T., Serrão, E. A., Engelen, A. H., Thomas, O. P., & Pérez, T. (2017). The interaction between the proliferating macroalga *Asparagopsis taxiformis* and the coral *Astroides calycularis* induces changes in microbiome and metabolomic fingerprints. *Scientific Reports*, 7, 42625.

Guijas, C., Montenegro-Burke, J. R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., et al. (2018). METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Analytical chemistry*, 90(5), 3156–3164.

Gupta, S. K. (2011). Drug Discovery and Clinical Research. Jaypee Brothers Medical Publishers.

Hadas, E., Shpigel, M., & Ilan, M. (2005). Sea ranching of the marine sponge *Negombata magnifica* (Demospongiae, Latrunculiidae) as a first step for latrunculin B mass production. *Aquaculture*, 244(1), 159–169.

Hammerman, N. M., & García-Hernández, J. E. (2017). The sponge *Xestospongia muta* offers shelter to the stony coral *Madracis auretenra* (Northwest Puerto Rico). *Marine Biodiversity*, 47(1), 57–58.

Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery*, 14, 111-129.

Hay, M. E. (2009). Marine chemical ecology: chemical signals and cues structure marine populations, communities, and ecosystems. *Annual review of marine science*, 1, 193–212.

He, F., Mai, L. H., Longeon, A., Copp, B. R., Loaëc, N., Bescond, A., et al. (2015). Novel Adociaquinone Derivatives from the Indonesian Sponge *Xestospongia* sp. *Marine drugs*, 13(5), 2617–2628.

He, Q., Sun, R., Liu, H., Geng, Z., Chen, D., Li, Y., et al. (2014). NMR-Based Metabolomic Analysis of Spatial Variation in Soft Corals. *Marine Drugs*, 12(4), 1876-1890.

Heavisides, E., Rouger, C., Reichel, F. A., Ulrich, C., Wenzel-Storjohann, A., Sebens, S., & Tasdemir, D. (2018). Seasonal Variations in the Metabolome and Bioactivity Profile of *Fucus vesiculosus* Extracted by an Optimised, Pressurised Liquid Extraction Protocol. *Marine Drugs*, 16(12), 503.

Hentschel, U., Usher, K. M., & Taylor, M. W. (2006). Marine sponges as microbial fermenters. *FEMS Microbiology Ecology*, 55(2), 167–177.

Hirayama, A., Wakayama, M., & Soga, T. (2014). Metabolome analysis based on capillary electrophoresis-mass spectrometry. *TrAC Trends in Analytical Chemistry*, 61, 215–222.

Hirsh, S., Carmely, S., & Kashman, Y. (1987). Brominated unsaturated acids from the marine sponge *Xestospongia* sp. *Tetrahedron*, 43(14), 3257–3261.

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703–714.

Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C. B., Brandão, S. N., et al. (2020, June 29). World Register of Marine Species (WoRMS). WoRMS Editorial Board. http://www.marinespecies.org. Accessed 29 June 2020

Hou, Y., Braun, D. R., Michel, C. R., Klassen, J. L., Adnani, N., Wyche, T. P., & Bugni, T. S. (2012). Microbial Strain Prioritization Using Metabolomics Tools for the Discovery of Natural Products. *Analytical Chemistry*, 84(10), 4277–4283.

Houssen, W. E., & Jaspars, M. (2012). Isolation of Marine Natural Products. In S. D. Sarker & L. Nahar (Eds.), Natural Products Isolation (pp. 367–392). Totowa, NJ: Humana Press.

Hughes, Terence P., Rodrigues, M. J., Bellwood, D. R., Ceccarelli, D., Hoegh-Guldberg, O., McCook, L., et al. (2007). Phase Shifts, Herbivory, and the Resilience of Coral Reefs to Climate Change. *Current Biology*, 17(4), 360–365.

Hughes, Terry P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., et al. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543, 373-377.

Ichiba, T., Scheuer, P. J., & Kelly-Borges, M. (1993). Sponge-Derived Polyunsaturated C16 Di- and Tribromocarboxylic Acids. *Helvetica Chimica Acta*, *76*(8), 2814–2816.

Ivanisevic, J., Pérez, T., Ereskovsky, A. V., Barnathan, G., & Thomas, O. P. (2011). Lysophospholipids in the Mediterranean Sponge *Oscarella tuberculata*: Seasonal Variability and Putative Biological Role. *Journal of Chemical Ecology*, 37, 537.

Ivanišević, J., Thomas, O. P., Lejeusne, C., Chevaldonné, P., & Pérez, T. (2011). Metabolic fingerprinting as an indicator of biodiversity: towards understanding inter-specific relationships among *Homoscleromorpha* sponges. *Metabolomics*, 7(2), 289–304.

Izquierdo-García, J. L., Villa, P., Kyriazis, A., del Puerto-Nevado, L., Pérez-Rial, S., Rodriguez, I., et al. (2011). Descriptive review of current NMR-based metabolomic data analysis packages. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 59(3), 263–270.

Jang, K. H., Lee, Y., Sim, C. J., Oh, K.-B., & Shin, J. (2012). Bioactive lipids from the sponge *Spirastrella abata*. *Bioorganic & Medicinal Chemistry Letters*, *22*(2), 1078–1081.

Januar, H. I., Chasanah, E., M. Tapiolas, D., A. Motti, C., H. Liptrot, C., & D. Wright, A. (2015). Influence of Anthropogenic Pressures on the Bioactivity Potential of Sponges and Soft Corals in the Coral Reef Environment. *Squalen Bulletin of Marine and Fisheries Postharvest and Biotechnology;* 10(2), 51-59.

Jerković, I., Marijanović, Z., Roje, M., Kuś, P. M., Jokić, S., & Čož-Rakovac, R. (2018). Phytochemical study of the headspace volatile organic compounds of fresh algae and seagrass from the Adriatic Sea (single point collection). *PLOS ONE*, 13(5), e0196462.

Jiang, W., Liu, D., Deng, Z., de Voogd, N. J., Proksch, P., & Lin, W. (2011). Brominated polyunsaturated lipids and their stereochemistry from the Chinese marine sponge *Xestospongia testudinaria*. *Tetrahedron*, 67(1), 58–68.

Johnson, W. M., Kido Soule, M. C., & Kujawinski, E. B. (2017). Extraction efficiency and quantification of dissolved metabolites in targeted marine metabolomics. *Limnology and Oceanography: Methods*, 15(4), 417–428.

Jové, M., Portero-Otín, M., Naudí, A., Ferrer, I., & Pamplona, R. (2014). Metabolomics of Human Brain Aging and Age-Related Neurodegenerative Diseases. *Journal of Neuropathology & Experimental Neurology*, 73(7), 640–657.

Ju, Z. Y., & Howard, L. R. (2003). Effects of Solvent and Temperature on Pressurized Liquid Extraction of Anthocyanins and Total Phenolics from Dried Red Grape Skin. *Journal of Agricultural and Food Chemistry*, 51(18), 5207–5213.

Kaufmann, B., & Christen, P. (2002). Recent extraction techniques for natural products: microwave-assisted extraction and pressurised solvent extraction. *Phytochemical Analysis*, 13(2), 105–113.

Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology*, 7(3), 296–307.

Kerr, R. G., & Kelly-Borges, M. (1993). Biochemical and morphological heterogeneity in the Caribbean sponge *Xestospongia muta* (petrosida: Petrosiidae). *In Sponges in Time and Space: Biology, Chemistry, Paleontology: proceedings of the 4th International Porifera Congress* Amsterdam, Netherlands, 19-23 (pp. 65–73).

Kerr, R. G., Kerr, S. L., Pettit, G. R., Herald, D. L., Groy, T. L., & Djerassi, C. (1991). Sterols of marine invertebrates. 63. Isolation and structure elucidation of sutinasterol, the major sterol of the marine sponge *Xestospongia* sp. *The Journal of Organic Chemistry*, 56(1), 58–62.

Keyzers, R. A., & Davies-Coleman, M. T. (2005). Anti-inflammatory metabolites from marine sponges. *Chemical Society Reviews*, 34(4), 355–365.

Kim, H. K., Choi, Y. H., & Verpoorte, R. (2010). NMR-based metabolomic analysis of plants. *Nature Protocols*, 5(3), 536–549.

Kleigrewe, K., Almaliti, J., Tian, I. Y., Kinnel, R. B., Korobeynikov, A., Monroe, E. A., et al. (2015). Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products from Cyanobacteria. *Journal of Natural Products*, 78(7), 1671–1682.

Kobayashi, M., Chen, Y.-J., Aoki, S., In, Y., Ishida, T., & Kitagawa, I. (1995). Four new β-carboline alkaloids isolated from two Okinawan marine sponges of *Xestospongia* sp. and Haliclona sp.). *Tetrahedron*, 51(13), 3727–3736.

Kopka, J. (2006). Current challenges and developments in GC–MS based metabolite profiling technology. *BioPerspectives—From basic research to industrial production*, 124(1), 312–322.

Laport, M. S., Santos, O. C. S., & Muricy, G. (2009). Marine Sponges: Potential Sources of New Antimicrobial Drugs. *Current Pharmaceutical Biotechnology*, 10(1), 86–105.

Lee, H.-J., Jeong, J., Alves, A. C., Han, S.-T., In, G., Kim, E.-H., et al. (2019). Metabolomic understanding of intrinsic physiology in Panax ginseng during whole growing seasons. *Journal of Ginseng Research*, 43(4), 654–665.

Lee, M.-H., & Lin, C.-C. (2007). Comparison of techniques for extraction of isoflavones from the root of *Radix Puerariae*: Ultrasonic and pressurized solvent extractions. *Food Chemistry*, 105(1), 223–228.

Lei, Z., Huhman, D. V., & Sumner, L. W. (2011). Mass Spectrometry Strategies in Metabolomics. *Journal of Biological Chemistry*, 286(29), 25435–25442.

Leonard, A. E., Pereira, S. L., Sprecher, H., & Huang, Y.-S. (2004). Elongation of long-chain fatty acids. *Progress in Lipid Research*, *43*(1), 36–54.

Lesser, M. P., Fiore, C., Slattery, M., & Zaneveld, J. (2016). Climate change stressors destabilize the microbiome of the Caribbean barrel sponge, *Xestospongia muta*. *Journal of Experimental Marine Biology and Ecology*, 475, 11–18.

Li, C.-W., Chen, J.-Y., & Hua, T.-E. (1998). Precambrian Sponges with Cellular Structures. *Science*, 279(5352), 879-882.

Liaw, C.-C., Chen, P.-C., Shih, C.-J., Tseng, S.-P., Lai, Y.-M., Hsu, C.-H., et al. (2015). Vitroprocines, new antibiotics against *Acinetobacter baumannii*, discovered from marine Vibrio sp. QWI-06 using mass-spectrometry-based metabolomics approach. *Scientific Reports*, 5, 12856.

Liland, K. H. (2011). Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends in Analytical Chemistry*, 30(6), 827–841.

Lin, K., Yang, P., Yang, H., Liu, A.-H., Yao, L.-G., Guo, Y.-W., & Mao, S.-C. (2015). Lysophospholipids from the Guangxi Sponge *Spirastrella purpurea*. *Lipids*, *50*(7), 697–703.

Lindon, J. C., & Nicholson, J. K. (2008). Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics. *Annual Review of Analytical Chemistry*, 1(1), 45–69.

Liu, D., Xu, J., Jiang, W., Deng, Z., de Voogd, N. J., Proksch, P., & Lin, W. (2011). Xestospongienols A–L, Brominated Acetylenic Acids from the Chinese Marine Sponge *Xestospongia testudinaria*. *Helvetica Chimica Acta*, 94(9), 1600–1607.

Loh, T.-L., & Pawlik, J. R. (2014). Chemical defenses and resource trade-offs structure sponge communities on Caribbean coral reefs. *Proceedings of the National Academy of Sciences*, 111(11) 4151-4156.

López-Legentil, S., Song, B., Mcmurray, S. E., & Pawlik, J. R. (2008). Bleaching and stress in coral reef ecosystems: hsp70 expression by the giant barrel sponge *Xestospongia muta*. *Molecular Ecology*, *17*(7), 1840–1849.

Lu, J., Feng, J., Cai, S., & Chen, Z. (2017). Metabolomic responses of *Haliotis diversicolor* to organotin compounds. *Chemosphere*, 168, 860–869.

Ludwig, C., & Viant, M. R. (2010). Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. *Phytochemical Analysis*, 21(1), 22–32.

Luzzatto-Knaan, T., Garg, N., Wang, M., Glukhov, E., Peng, Y., Ackermann, G., et al. (2017). Digitizing mass spectrometry data to explore the chemical diversity and distribution of marine cyanobacteria and algae. *eLife*, 6, e24214.

Macintyre, L., Zhang, T., Viegelmann, C., Martinez, J. I., Cheng, C., Dowdells, C., et al. (2014). Metabolomic Tools for Secondary Metabolite Discovery from Marine Microbial Symbionts. *Marine Drugs*, 12(6), 3416-3448.

Maldonado, M., Ribes, M., & van Duyl, F. C. (2012). Chapter three - Nutrient Fluxes Through Sponges: Biology, Budgets, and Ecological Implications. In M. A. Becerro, M. J. Uriz, M. Maldonado, & X. Turon (Eds.), *Advances in Marine Biology* (Vol. 62, pp. 113–182). Academic Press.

MarinLit. A database of the marine natural products literature. RSC. http://pubs.rsc.org/marinlit/introduction. Accessed 30 October 2019

Markley, J. L., Brüschweiler, R., Edison, A. S., Eghbalnia, H. R., Powers, R., Raftery, D., & Wishart, D. S. (2017). The future of NMR-based metabolomics. *Analytical biotechnology*, 43, 34–40.

McGrath, E. C., Woods, L., Jompa, J., Haris, A., & Bell, J. J. (2018). Growth and longevity in giant barrel sponges: Redwoods of the reef or Pines in the Indo-Pacific? *Scientific Reports*, 8, 15317.

McManus, J. W., & Polsenberg, J. F. (2004). Coral–algal phase shifts on coral reefs: Ecological and environmental aspects. Regime shifts in the ocean. *Reconciling observations and theory*, 60(2), 263–279.

McMurray S.E., Pawlik JR, & Finelli CM. (2014). Trait-mediated ecosystem impacts: how morphology and size affect pumping rates of the Caribbean giant barrel sponge. *Aquatic Biology*, 23(1), 1–13.

McMurray, S.E., Blum, J. E., & Pawlik, J. R. (2008). Redwood of the reef: growth and age of the giant barrel sponge *Xestospongia* muta in the Florida Keys. *Marine Biology*, 155(2), 159–171.

McMurray, S.E., Finelli, C. M., & Pawlik, J. R. (2015). Population dynamics of giant barrel sponges on Florida coral reefs. *Journal of Experimental Marine Biology and Ecology*, 473, 73–80.

McMurray, S.E., Henkel, T. P., & Pawlik, J. R. (2010). Demographics of increasing populations of the giant barrel sponge *Xestospongia* muta in the Florida Keys. *Ecology*, 91(2), 560–570.

Mehbub, F. M., Lei, J., Franco, C., & Zhang, W. (2014). Marine Sponge Derived Natural Products between 2001 and 2010: Trends and Opportunities for Discovery of Bioactives. *Marine Drugs*, 12(8), 4539–4577.

Melis, R., Sanna, R., Braca, A., Bonaglini, E., Cappuccinelli, R., Slawski, H., et al. (2017). Molecular details on gilthead sea bream (Sparus aurata) sensitivity to low water temperatures from ^{1}H NMR metabolomics. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 204, 129–136.

Minto, R. E., & Blacklock, B. J. (2008). Biosynthesis and function of polyacetylenes and allied natural products. *Progress in Lipid Research*, *47*(4), 233–306.

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G. L., Cerrano, C., et al. (2017). The sponge microbiome project. GigaScience, 6(10), gix077.

Montalvo, N. F., & Hill, R. T. (2011). Sponge-Associated Bacteria Are Strictly Maintained in Two Closely Related but Geographically Distant Sponge Hosts. *Applied and Environmental Microbiology*, 77(20), 7207–7216.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? PLOS Biology, 9(8), e1001127.

Morinaka, B. I., Skepper, C. K., & Molinski, T. F. (2007). Ene-yne Tetrahydrofurans from the Sponge *Xestospongia muta*. Exploiting a Weak CD Effect for Assignment of Configuration. *Organic Letters*, 9(10), 1975–1978.

Morrow, K. M., Fiore, C. L., & Lesser, M. P. (2016). Environmental drivers of microbial community shifts in the giant barrel sponge, *Xestospongia muta*, over a shallow to mesophotic depth gradient. *Environmental Microbiology*, 18(6), 2025–2038.

Müller, W. E. G., & Müller, I. M. (2003). Origin of the Metazoan Immune System: Identification of the Molecules and Their Functions in Sponges. *Integrative and Comparative Biology*, 43(2), 281–292.

Mushtaq, M. Y., Choi, Y. H., Verpoorte, R., & Wilson, E. G. (2014). Extraction for Metabolomics: Access to The Metabolome. *Phytochemical Analysis*, 25(4), 291–306.

Myers, O. D., Sumner, S. J., Li, S., Barnes, S., & Du, X. (2017). One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. *Analytical Chemistry*, 89(17), 8696–8703.

Naman, C. B., Leber, C. A., & Gerwick, W. H. (2017). Chapter 5 - Modern Natural Products Drug Discovery and Its Relevance to Biodiversity Conservation. In I. Kurtböke (Ed.), *Microbial Resources* (pp. 103–120). Academic Press.

Neely, K. L., & Butler, C. B. (2020). Seasonal, lunar, and diel patterns in spawning by the giant barrel sponge, *Xestospongia muta*. *Coral Reefs*. Doi: 10.1007/s00338-020-02009-2

Newman, D. J., & Cragg, G. M. (2004). Advanced Preclinical and Clinical Trials of Natural Products and Related Compounds from Marine Sources. *Current Medicinal Chemistry*, 11(13), 1693–1713.

Newman, D. J., & Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, 79(3), 629–661.

Newman, D., & Cragg, G. (2016). Drugs and Drug Candidates from Marine Sources: An Assessment of the Current "State of Play". *Planta medica*, 82 9–10, 775–89.

Ng, Y. K., Hewavitharana, A. K., Webb, R., Shaw, P. N., & Fuerst, J. A. (2013). Developmental cycle and pharmaceutically relevant compounds of *Salinispora actinobacteria* isolated from Great Barrier Reef marine sponges. *Applied Microbiology and Biotechnology*, 97(7), 3097–3108.

Nguyen, X. C., Longeon, A., Pham, V. C., Urvois, F., Bressy, C., Trinh, T. T. V., et al. (2013). Antifouling 26,27-Cyclosterols from the Vietnamese Marine Sponge *Xestospongia testudinaria*. *Journal of Natural Products*, 76(7), 1313–1318.

NIST 20 https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:msms/ Accessed 29 July 2020

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., et al. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nature Methods*, 17(9), 905–908.

Noyer, C., Thomas, O. P., & Becerro, M. A. (2011). Patterns of Chemical Diversity in the Mediterranean Sponge *Spongia lamella*. *PLOS ONE*, 6(6), e20844.

Olinger, L. K., Scott, A. R., McMurray, S. E., & Pawlik, J. R. (2019). Growth estimates of Caribbean reef sponges on a shipwreck using 3D photogrammetry. *Scientific Reports*, 9(1), 18398.

Olsen, E. K., Søderholm, K. L., Isaksson, J., Andersen, J. H., & Hansen, E. (2016). Metabolomic profiling reveals the N-Acyl-Taurine geodiataurine in extracts from the marine sponge *Geodia macandrewii* (Bowerbank). *Journal of Natural Products*, 79(5), 1285–1291.

Olson, J. B., & Gao, X. (2013). Characterizing the bacterial associates of three Caribbean sponges along a gradient from shallow to mesophotic depths. *FEMS Microbiology Ecology*, 85(1), 74–84.

Oppong-Danquah, E., Parrot, D., Blümel, M., Labes, A., & Tasdemir, D. (2018). Molecular Networking-Based Metabolome and Bioactivity Analyses of Marine-Adapted Fungi Co-cultivated With Phytopathogens. *Frontiers in microbiology*, 9, 2072–2072.

Page, M. J., Northcote, P. T., Webb, V. L., Mackey, S., & Handley, S. J. (2005). Aquaculture trials for the production of biologically active metabolites in the New Zealand sponge *Mycale hentscheli* (Demospongiae: Poecilosclerida). *Aquaculture*, 250(1), 256–269.

Page, M., West, L., Northcote, P., Battershill, C., & Kelly, M. (2005). Spatial and Temporal Variability of Cytotoxic Metabolites in Populations of the New Zealand Sponge *Mycale hentscheli*. *Journal of Chemical Ecology*, 31(5), 1161–1174.

Palumbi, S. R., Sandifer, P. A., Allan, J. D., Beck, M. W., Fautin, D. G., Fogarty, M. J., et al. (2009). Managing for ocean biodiversity to sustain marine ecosystem services. *Frontiers in Ecology and the Environment*, 7(4), 204–211.

Pandolfi, J. M., Bradbury, R. H., Sala, E., Hughes, T. P., Bjorndal, K. A., Cooke, R. G., et al. (2003). Global Trajectories of the Long-Term Decline of Coral Reef Ecosystems. *Science*, 301(5635), 955-958.

Papale, M., Rizzo, C., Fani, R., Bertolino, M., Costa, G., Paytuví-Gallart, A., et al. (2020). Exploring the Diversity and Metabolic Profiles of Bacterial Communities Associated With Antarctic Sponges (Terra Nova Bay, Ross Sea). *Frontiers in Ecology and Evolution*, 8, 268.

Papazian, S., Parrot, D., Burýšková, B., Weinberger, F., & Tasdemir, D. (2019). Surface chemical defence of the eelgrass *Zostera marina* against microbial foulers. *Scientific Reports*, 9, 3323.

Parrot, D., Papazian, S., Foil, D., & Tasdemir, D. (2018). Imaging the Unimaginable: Desorption Electrospray Ionization – Imaging Mass Spectrometry (DESI-IMS) in Natural Product Research. *Planta Medica*, 84(09/10), 584–593.

Parsons, H. M., Ludwig, C., Günther, U. L., & Viant, M. R. (2007). Improved classification accuracy in 1- and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics*, 8, 234.

Patil, A. D., Kokke, W. C., Cochran, S., Francis, T. A., Tomszek, T., & Westley, J. W. (1992). Brominated Polyacetylenic Acids from the Marine Sponge *Xestospongia muta*: Inhibitors of HIV Protease. *Journal of Natural Products*, 55(9), 1170–1177.

Paul, V. J., & Puglisi, M. P. (2004). Chemical mediation of interactions among marine organisms. *Natural Product Reports*, 21(1), 189–209.

Paul, V. J., & Ritson-Williams, R. (2008). Marine chemical ecology. *Natural Product Reports*, 25(4), 662–695.

Paul, V. J., Freeman, C. J., & Agarwal, V. (2019). Chemical Ecology of Marine Sponges: New Opportunities through "-Omics." *Integrative and Comparative Biology*, 59(4), 765-776.

Paul, V. J., Puglisi, M. P., & Ritson-Williams, R. (2006). Marine chemical ecology. *Natural Product Reports*, 23(2), 153–180.

Pawlik, J. R. (2011). The Chemical Ecology of Sponges on Caribbean Reefs: Natural Products Shape Natural Systems. *BioScience*, 61(11), 888–898.

Pawlik, J. R., & McMurray, S. E. (2020). The Emerging Ecological and Biogeochemical Importance of Sponges on Coral Reefs. *Annual Review of Marine Science*, 12(1), 315-337.

Pawlik, J. R., Chanas, B., Toonen, R. J., & Fenical, W. (1995). Defenses of Caribbean sponges against predatory reef fish. I. Chemical deterrency. *Marine Ecology Progress Series*, 127, 183–194.

Perdicaris, S., Vlachogianni, T., & Valavanidis, A. (2013). Bioactive Natural Substances from Marine Sponges: New Developments and Prospects for Future Pharmaceuticals. *Natural Products Chemistry & Research*, 1(3), 1000114.

Pereira, D. M., Vinholes, J., de Pinho, P. G., Valentão, P., Mouga, T., Teixeira, N., & Andrade, P. B. (2012). A gas chromatography–mass spectrometry multi-target method for the simultaneous analysis of three classes of metabolites in marine organisms. *Talanta*, 100, 391–400.

Perkol-Finkel, S., Shashar, N., & Benayahu, Y. (2006). Can artificial reefs mimic natural reef communities? The roles of structural features and age. *Marine Environmental Research*, 61(2), 121–135.

Pham, N. B., Butler, M. S., Hooper, J. N. A., Moni, R. W., & Quinn, R. J. (1999). Isolation of Xestosterol Esters of Brominated Acetylenic Fatty Acids from the Marine Sponge *Xestospongia testudinaria*. *Journal of Natural Products*, 62(10), 1439–1442.

Pilatti, F. K., Ramlov, F., Schmidt, E. C., Costa, C., Oliveira, E. R. de, Bauer, C. M., et al. (2017). Metabolomics of *Ulva lactuca* Linnaeus (Chlorophyta) exposed to oil fuels: Fourier transform infrared spectroscopy and multivariate analysis as tools for metabolic fingerprint. *Marine Pollution Bulletin*, 114(2), 831–836.

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11, 395.

Polónia, A. R. M., Cleary, D. F. R., Freitas, R., Gomes, N. C. M., & de Voogd, N. J. (2017). Archaeal and bacterial communities of *Xestospongia testudinaria* and sediment differ in diversity, composition and predicted function in an Indonesian coral reef environment. *Journal of Sea Research*, 119, 37–53.

Pomponi, S. A. (1999). The bioprocess–technological potential of the sea. Biotechnological *Aspects of Marine Sponges*, 70(1), 5–13.

Proksch, P., Putz, A., Ortlepp, S., Kjer, J., & Bayer, M. (2010). Bioactive natural products from marine sponges and fungal endophytes. *Phytochemistry Reviews*, 9(4), 475–489.

PubChem, NIH https://pubchem.ncbi.nlm.nih.gov/. Accessed 30 October 2019

Quinn, R. A., Nothias, L.-F., Vining, O., Meehan, M., Esquenazi, E., & Dorrestein, P. C. (2017). Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends in Pharmacological Sciences*, 38(2), 143–154.

Quinn, R. A., Vermeij, M. J. A., Hartmann, A. C., Galtier d'Auriac, I., Benler, S., Haas, A., et al. (2016). Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20160469.

Quinn, R. J., & Tucker, D. J. (1991). Further Acetylenic Acids from the Marine Sponge *Xestospongia testudinaria*. *Journal of Natural Products*, *54*(1), 290–294.

Quinn, R. J., & Tucker, D. J. (1985). A brominated bisacetylenic acid from the marine sponge *Xestospongia testudinaria*. *Tetrahedron Letters*, 26(13), 1671–1672.

Reverter, M., Perez, T., Ereskovsky, A. V., & Banaigs, B. (2016). Secondary Metabolome Variability and Inducible Chemical Defenses in the Mediterranean Sponge *Aplysina cavernicola*. *Journal of Chemical Ecology*, 42(1), 60–70.

Reverter, M., Tribalat, M.-A., Pérez, T., & Thomas, O. P. (2018). Metabolome variability for two Mediterranean sponge species of the genus *Haliclona*: specificity, time, and space. *Metabolomics*, 14, 114.

Ribes, M., Calvo, E., Movilla, J., Logares, R., Coma, R., & Pelejero, C. (2016). Restructuring of the sponge microbiome favors tolerance to ocean acidification. *Environmental Microbiology Reports*, 8(4), 536–544.

Ritson-Williams, R., Becerro, M. A., & Paul, V. J. (2005). Spawning of the giant barrel sponge *Xestospongia muta* in Belize. *Coral Reefs*, 24(1), 160–160.

Rix, L., de Goeij, J. M., van Oevelen, D., Struck, U., Al-Horani, F. A., Wild, C., & Naumann, M. S. (2018). Reef sponges facilitate the transfer of coral-derived organic matter to their associated fauna via the sponge loop. *Marine Ecology Progress Series*, 589, 85–96.

Rohde, S., Gochfeld, D. J., Ankisetty, S., Avula, B., Schupp, P. J., & Slattery, M. (2012). Spatial Variability in Secondary Metabolites of the Indo-Pacific Sponge *Stylissa massa*. *Journal of Chemical Ecology*, 38(5), 463–475.

Roll, D. M., Scheuer, P. J., Matsumoto, G. K., & Clardy, J. (1983). Halenaquinone, a pentacyclic polyketide from a marine sponge. *Journal of the American Chemical Society*, 105(19), 6177–6178.

Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., et al. (2016). OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nature Methods*, 13(9), 741–748.

Ruiz, C., Valderrama, K., Zea, S., & Castellanos, L. (2013). Mariculture and Natural Production of the Antitumoural (+)-Discodermolide by the Caribbean Marine Sponge *Discodermia dissoluta*. *Marine Biotechnology*, 15(5), 571–583.

Sacristan-Soriano, O., Banaigs, B., & Becerro, M. A. (2011). Relevant Spatial Scales of Chemical Variation in *Aplysina aerophoba*. *Marine Drugs*, 9(12), 2499–2513.

Sagar, S., Kaur, M., & Minneman, K. P. (2010). Antiviral Lead Compounds from Marine Sponges. *Marine Drugs*, 8(10), 2619-2638.

Saito, K., & Matsuda, F. (2010). Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology*, 61(1), 463–489.

Sakemi, S., Totton, L. E., & Sun, H. H. (1990). Xestamines A, B, and C, Three New Long-Chain Methoxylamine Pyridines from the Sponge *Xestospongia wiedenmayeri*. *Journal of Natural Products*, 53(4), 995–999.

Salvatore, M. M., Nicoletti, R., Salvatore, F., Naviglio, D., & Andolfi, A. (2018). GC–MS approaches for the screening of metabolites produced by marine-derived *Aspergillus*. *Marine Chemistry*, 206, 19–33.

Santiago, V. S., Manzano, G. G., Yu, C. C., Aliño, P. M., & Salvador-Reyes, L. A. (2019). Mariculture potential of renieramycin-producing Philippine blue sponge *Xestospongia* sp. (Porifera: Haplosclerida). *Aquaculture*, 502, 356–364.

Sashidhara, K. V., White, K. N., & Crews, P. (2009). A Selective Account of Effective Paradigms and Significant Outcomes in the Discovery of Inspirational Marine Natural Products. *Journal of Natural Products*, 72(3), 588–603.

Sass-Kiss, A., Simandi, B., Gao, Y., Boross, F., & Vamos-Falusi, Z. (1999). Study on the pilot-scale extraction of onion oleoresin using supercritical CO2. *Journal of the Science of Food and Agriculture*, 76(3), 320–326.

Schmitz, F. J., & Gopichand, Y. (1978). (7E, 13ξ, 15Z)-14,16-dibromo-7,13,15-hexadecatrien-5-ynoic acid. A novel dibromo acetylenic acid from the marine sponge *Xestospongia muta*. *Tetrahedron Letters*, 19(39), 3637–3640.

Schock, T. B., Duke, J., Goodson, A., Weldon, D., Brunson, J., Leffler, J. W., & Bearden, D. W. (2013). Evaluation of Pacific White Shrimp (*Litopenaeus vannamei*) Health during a Superintensive Aquaculture Growout Using NMR-Based Metabolomics. *PLOS ONE*, 8(3), e59521.

Schubotz, F., Lipp, J. S., Elvert, M., & Hinrichs, K.-U. (2011). Stable carbon isotopic compositions of intact polar lipids reveal complex carbon flow patterns among hydrocarbon degrading microbial communities at the Chapopote asphalt volcano. *Geochimica et Cosmochimica Acta*, 75(16), 4399–4415.

SciFindern is your key to unlocking R&D productivity. CAS. https://www.cas.org/products/scifinder. Accessed 20 May 2020

Setiawan, E., de Voogd, N. J., Swierts, T., Hooper, J. N. A., Wörheide, G., & Erpenbeck, D. (2016). MtDNA diversity of the Indonesian giant barrel sponge *Xestospongia testudinaria* (Porifera: Haplosclerida) – implications from partial cytochrome oxidase 1 sequences. *Journal of the Marine Biological Association of the United Kingdom*, 96(2), 323–332.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498–2504.

Shi, Y., Pan, C., Auckloo, B. N., Chen, X., Chen, C.-T. A., Wang, K., et al. (2017). Stress-driven discovery of a cryptic antibiotic produced by Streptomyces sp. WU20 from Kueishantao hydrothermal vent with an integrated metabolomics strategy. *Applied Microbiology and Biotechnology*, 101(4), 1395–1408.

Shin, B. A., Kim, Y. R., Lee, I.-S., Sung, C. K., Hong, J., Sim, C. J., et al. (1999). Lyso-PAF Analogues and Lysophosphatidylcholines from the Marine Sponge *Spirastrella abata* as Inhibitors of Cholesterol Biosynthesis. *Journal of Natural Products*, *62*(11), 1554–1557.

Shulaev, V., & Isaac, G. (2018). Supercritical fluid chromatography coupled to mass spectrometry – A metabolomics perspective. *Journal of Chromatography B*, 1092, 499–505.

Shulaev, V., Cortes, D., Miller, G., & Mittler, R. (2008). Metabolomics for plant stress response. *Physiologia Plantarum*, 132(2), 199–208.

Sipkema, D., Franssen, M. C. R., Osinga, R., Tramper, J., & Wijffels, R. H. (2005). Marine Sponges as Pharmacy. *Marine Biotechnology*, 7(3), 142-162.

Sipkema, D., Osinga, R., Schatton, W., Mendola, D., Tramper, J., & Wijffels, R. H. (2005). Large-scale production of pharmaceuticals by marine sponges: Sea, cell, or synthesis? *Biotechnology and Bioengineering*, 90(2), 201–222.

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, 78(3), 779–787.

Snelgrove, P. V. R. (2016). An Ocean of Discovery: Biodiversity Beyond the Census of Marine Life. *Planta Medica*, 82(09/10), 790–799.

Sogin, E. M., Anderson, P., Williams, P., Chen, C.-S., & Gates, R. D. (2014a). Application of 1H-NMR Metabolomic Profiling for Reef-Building Corals. *PLOS ONE*, 9(10), e111274.

Sogin, E. M., Putnam, H. M., Anderson, P. E., & Gates, R. D. (2016). Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, *Pocillopora damicornis*. *Metabolomics*, 12, 71.

Southwell, M. W., Weisz, J. B., Martens, C. S., & Lindquist, N. (2008). In situ fluxes of dissolved inorganic nitrogen from the sponge community on Conch Reef, Key Largo, Florida. *Limnology and Oceanography*, 53(3), 986–996.

SpecInfo-Consortium Member-NMR, IR and MS, Wiley. https://www.wiley.com/en-us/SpecInfo+Consortium+Member+NMR%2C+IR+and+MS-p-9780471251415. Accessed 20 May 2020

Stowe, S. D., Richards, J. J., Tucker, A. T., Thompson, R., Melander, C., & Cavanagh, J. (2011). Anti-Biofilm Compounds Derived from Marine Sponges. *Marine Drugs*, 9(10), 2010–2035.

Sumner, L. W., Mendes, P., & Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Plant Metabolomics*, 62(6), 817–836.

Swierts, T, Cleary, D. F. R., & de Voogd, N. J. (2018). Prokaryotic communities of Indo-Pacific giant barrel sponges are more strongly influenced by geography than host phylogeny. *FEMS Microbiology Ecology*, 94(12),fiy194.

Swierts, T. (2019). Diversity in the globally intertwined giant barrel sponge species complex. Leiden University, The Netherlands. Retrieved from isbn:9789461829887

Swierts, T., Huang, Y. M., & de Voogd, N. J. (2018). The giant barrel sponge facilitates the recovery of coral fragments after a tropical storm in Taiwan. *Coral Reefs*, 37(3), 675–675.

Swierts, T., Peijnenburg, K. T. C. A., de Leeuw, C., Cleary, D. F. R., Hörnlein, C., Setiawan, E., et al. (2013). Lock, Stock and Two Different Barrels: Comparing the Genetic Composition of Morphotypes of the Indo-Pacific Sponge *Xestospongia testudinaria*. *PLOS ONE*, 8(9), e74396.

Swierts, Thomas, Peijnenburg, K. T. C. A., de Leeuw, C. A., Breeuwer, J. A. J., Cleary, D. F. R., & de Voogd, N. J. (2017). Globally intertwined evolutionary history of giant barrel sponges. *Coral Reefs*, 36(3), 933–945.

Taniguchi, M., Uchio, Y., Yasumoto, K., Kusumi, T., & Ooi, T. (2008). Brominated Unsaturated Fatty Acids from Marine Sponge Collected in Papua New Guinea. *Chemical and Pharmaceutical Bulletin*, *56*(3), 378–382.

Ternon, E., Perino, E., Manconi, R., Pronzato, R., & Thomas, O. P. (2017). How Environmental Factors Affect the Production of Guanidine Alkaloids by the Mediterranean Sponge *Crambe crambe*. *Marine Drugs*, 15(6), 181.

Thoms, C., & Schupp, P. J. (2007). Chemical defense strategies in sponges: a review. In Porifera Research: Biodiversity, Innovation and Sustainability (pp. 627–637). Rio de Janeiro, RJ, Brasil.

Torjesen, I. (2015). Drug development: the journey of a medicine from lab to shelf. Pharmaceutical Journal. https://www.pharmaceutical-journal.com/test-tomorrows-pharmacist/tomorrows-pharmacist/drug-development-the-journey-of-a-medicine-from-lab-to-shelf/20068196.article. Accessed 21 October 2020

Torkhovskaya, T. I., Ipatova, O. M., Zakharova, T. S., Kochetova, M. M., & Khalilov, E. M. (2007). Lysophospholipid receptors in cell signaling. *Biochemistry (Moscow)*, 72(2), 125–131.

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, 12, 523-526.

Turon, M., Cáliz, J., Garate, L., Casamayor, E. O., & Uriz, M. J. (2018). Showcasing the role of seawater in bacteria recruitment and microbiome stability in sponges. *Scientific Reports*, 8, 15201.

van der Hooft, J. J. J., Wandy, J., Barrett, M. P., Burgess, K. E. V., & Rogers, S. (2016). Topic modeling for untargeted substructure exploration in metabolomics. *Proceedings of the National Academy of Sciences*, 113(48), 13738-13743.

van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology*, *9*(2), 112–124.

Van Soest, R. W. M., Boury-Esnault, N., Hooper, J. N. A., Rützler, K., De Voogd, N. J., Alvarez, B., et al. (2020a). World Porifera Database. *Xestospongia muta* (Schmidt, 1870). http://www.marinespecies.org/aphia.php?p=taxdetails&id=166894. Accessed 29 June 2020

Van Soest, R. W. M., Boury-Esnault, N., Hooper, J. N. A., Rützler, K., De Voogd, N. J., Alvarez, B., et al. (2020b). World Porifera Database. *Xestospongia testudinaria* (Lamarck, 1815). http://www.marinespecies.org/aphia.php?p=taxdetails&id=166902. Accessed 29 June 2020

Van Soest, R. W. M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., De Voogd, N. J., et al. (2012). Global Diversity of Sponges (Porifera). *PLOS ONE*, 7(4), e35105.

Van Soest, R. W. M., Meesters, E. H. W. G., & Becking, L. E. (2014). Deep-water sponges (Porifera) from Bonaire and Klein Curaçao, Southern Caribbean. *Zootaxa*, 3878 (5), 401–443.

Verpoorte, R., Choi, Y. H., & Kim, H. K. (2007). NMR-based metabolomics at work in phytochemistry. *Phytochemistry Reviews*, 6(1), 3–14.

Viant, M. R. (2007). Metabolomics of aquatic organisms. *Marine Ecology Progress Series*, 332, 301–306.

Viegelmann, C., Margassery, M. L., Kennedy, J., Zhang, T., O'Brien, C., O'Gara, F., et al. (2014). Metabolomic Profiling and Genomic Study of a Marine Sponge-Associated *Streptomyces* sp. *Marine Drugs*, 12(6), 3323-3351.

Villa, F. A., & Gerwick, L. (2010). Marine natural product drug discovery: Leads for treatment of inflammation, cancer, infections, and neurological disorders. *Immunopharmacology and Immunotoxicology*, 32(2), 228–237.

Villegas-Plazas, M., Wos-Oxley, M. L., Sanchez, J. A., Pieper, D. H., Thomas, O. P., & Junca, H. (2019). Variations in Microbial Diversity and Metabolite Profiles of the Tropical Marine Sponge *Xestospongia muta* with Season and Depth. *Microbial Ecology*, 78(1), 243–256.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34, 828–837.

Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., et al. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences*, 109(26), E1743-E1752.

Webster, N. S., & Taylor, M. W. (2012). Marine sponges and their microbial symbionts: love and other relationships. *Environmental Microbiology*, 14(2), 335–346.

Weckwerth, W. (2003). Metabolomics in Systems Biology. *Annual Review of Plant Biology*, 54(1), 669–689.

Wicke, C., Hüners, M., Wray, V., Nimtz, M., Bilitewski, U., & Lang, S. (2000). Production and Structure Elucidation of Glycoglycerolipids from a Marine Sponge-Associated *Microbacterium Species*. *Journal of Natural Products*, 63(5), 621–626.

Winnikoff, J. R., Glukhov, E., Watrous, J., Dorrestein, P. C., & Gerwick, W. H. (2014). Quantitative molecular networking to profile marine cyanobacterial metabolomes. *The Journal of antibiotic*s, 67(1), 105–112.

Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics*, 1(1), 92–107.

Wright, A. E. (1998). Isolation of Marine Natural Products. In R. J. P. Cannell (Ed.), *Natural Products Isolation* (pp. 365–408). Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59259-256-2_13

Wu, H., & Wang, W.-X. (2010). NMR-based metabolomic studies on the toxicological effects of cadmium and copper on green mussels *Perna viridis*. *Aquatic Toxicology*, 100(4), 339–345.

Wu, W., Daszykowski, M., Walczak, B., Sweatman, B. C., Connor, S. C., Haselden, J. N., et al. (2006). Peak Alignment of Urine NMR Spectra Using Fuzzy Warping. *Journal of Chemical Information and Modeling*, 46(2), 863–875.

Wulff, J.L. (2016). Sponge Contributions to the Geology and Biology of Reefs: Past, Present, and Future. In D. K. Hubbard, C. S. Rogers, J. H. Lipps, & Jr. Stanley George D. (Eds.), *Coral Reefs at the Crossroads* (pp. 103–126). Dordrecht: Springer Netherlands.

Wulff, J. L. (2006). Ecological interactions of marine sponges. *Canadian Journal of Zoology*, 84(2), 146–166.

Xavier, J. R., Rachello-Dolmen, P. G., Parra-Velandia, F., Schönberg, C. H. L., Breeuwer, J. A. J., & van Soest, R. W. M. (2010). Molecular evidence of cryptic speciation in the “cosmopolitan” excavating sponge *Cliona celata* (Porifera, Clionaidae). *Molecular Phylogenetics and Evolution,* 56(1), 13–20.

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., et al. (2013). Molecular Networking as a Dereplication Strategy. *Journal of Natural Products*, 76(9), 1686–1699.

Ye, J., Zhou, F., Al-Kareef, A. M. Q., & Wang, H. (2015). Anticancer agents from marine sponges. *Journal of Asian Natural Products Research*, 17(1), 64–88.

Yoon, D., Choi, B.-R., Ma, S., Lee, W. J., Jo, I.-H., Lee, Y.-S., et al. (2019). Metabolomics for Age Discrimination of Ginseng Using a Multiplex Approach to HR-MAS NMR Spectroscopy, UPLC–QTOF/MS, and GC × GC–TOF/MS. *Molecules*, 24(13), 2381.

Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., et al. (2012). Human serum metabolic profiles are age dependent. *Aging Cell*, 11(6), 960–967.

Yuliana, N. D., Khatib, A., Verpoorte, R., & Choi, Y. H. (2011). Comprehensive Extraction Method Integrated with NMR Metabolomics: A New Bioactivity Screening Method for Plants, Adenosine A1 Receptor Binding Compounds in Orthosiphon stamineus Benth. *Analytical Chemistry*, 83(17), 6902–6906.

Zea, S. (1993). Cover of sponges and other sessile organisms in rocky and coral reef habitats of Santa Marta, Colombian Caribbean Sea. *Caribbean Journal of Science*, 29, 75–88.

Zhang, C., Idelbayev, Y., Roberts, N., Tao, Y., Nannapaneni, Y., Duggan, B. M., et al. (2017). Small Molecule Accurate Recognition Technology (SMART) to Enhance Natural Products Research. *Scientific Reports*, 7, 14243.

Zhang, F., Jonas, L., Lin, H., & Hill, R. T. (2019). Microbially mediated nutrient cycles in marine sponges. *FEMS Microbiology Ecology*, 95(11), fiz155.

Zhang, W., Hankemeier, T., & Ramautar, R. (2017). Next-generation capillary electrophoresis–mass spectrometry approaches in metabolomics. *Analytical biotechnology*, 43, 1–7.

Zhang, Y., Wu, H., Wei, L., Xie, Z., & Guan, B. (2017). Effects of hypoxia in the gills of the Manila clam Ruditapes philippinarum using NMR-based metabolomics. *Marine Pollution Bulletin*, 114(1), 84–89.

Zhao, Q., Mansoor, T. A., Hong, J., Lee, C.-O., Im, K. S., Lee, D. S., & Jung, J. H. (2003). New Lysophosphatidylcholines and Monoglycerides from the Marine Sponge *Stelletta* sp. *Journal of Natural Products*, *66*(5), 725–728.

Zheng, X., Su, M., Pei, L., Zhang, T., Ma, X., Qiu, Y., et al. (2011). Metabolic Signature of Pregnant Women with Neural Tube Defects in Offspring. *Journal of Proteome Research*, 10(10), 4845–4854.

Zhou, X., Lu, Y., Lin, X., Yang, B., Yang, X., & Liu, Y. (2011). Brominated aliphatic hydrocarbons and sterols from the sponge *Xestospongia testudinaria* with their bioactivities. *Chemistry and Physics of Lipids*, *164*(7), 703–706.

Zhou, X., Xu, T., Yang, X.-W., Huang, R., Yang, B., Tang, L., & Liu, Y. (2010). Chemical and Biological Aspects of Marine Sponges of the Genus *Xestospongia*. *Chemistry & Biodiversity*, 7(9), 2201–2227.

Zhou, Z.-F., Menna, M., Cai, Y.-S., & Guo, Y.-W. (2015). Polyacetylenes of Marine Origin: Chemistry and Bioactivity. *Chemical Reviews*, 115(3), 1543–1596.

Zotti, M., De Pascali, S. A., Del Coco, L., Migoni, D., Carrozzo, L., Mancinelli, G., & Fanizzi, F. P. (2016). 1H NMR metabolomic profiling of the blue crab (*Callinectes sapidus*) from the Adriatic Sea (SE Italy): A comparison with warty crab (*Eriphia verrucosa*), and edible crab (*Cancer pagurus*). *Food Chemistry*, 196, 601–609.

Summary

Marine sponges are important members of reef ecosystems, as they play ecological roles that are essential for the health of the reef. From a chemical perspective, sponges have been widely studied, showing great chemical diversity with biological activity, although this chemical diversity can often not be linked to for instance variation in environmental conditions experienced by sponges. Metabolomics, with its holistic overview of the metabolites present in samples, has provided new tools to gain insight into the function of metabolites in the intricate relationship between sponges and their environment. Among sponges, giant barrel sponges have stood out for their predominance in many reefs, their longevity and the fact that they are widely spread across oceans. In some reefs, giant barrel sponges can cover up to 13% of the available benthic substrate, surpassing those of corals. They can live for more than 2000 years and can be found from New Caledonia to the Caribbean with a presence throughout the Red Sea, the east African coast and the Indo-Pacific region. It has also been shown that what was originally classified as three species (*Xestospongia bergquistia* (confined to the northern Great Barrier Reef), *Xestospongia muta* and *Xestospongia testudinaria*) is actually a species complex that contains at least nine cryptic species. The presence of cryptic species is prevalent in sponge taxa and is one of the main difficulties in the classification of these animals. All these features make giant barrels sponges a good model organism to understand the influence that environmental, biological and genetic factors might have on the metabolic production of marine sponges.

As part of the recognition of metabolomics as an approach that enables the comprehensive study of the metabolism of marine sponges and marine organisms in general, a literature review of the methods and applications of metabolomics in this field was done and included in the thesis as Chapter 2. The first step in a metabolomics study is the extraction method, which is crucial for obtaining an extract that accurately represents the metabolites present in the sample. The influence of extraction conditions (pressure, temperature, solvent polarity, and number of cycles) on the chemical diversity of *Xestospongia* spp. extracts was studied using a pressurized extraction system. It was found that temperature, solvent polarity and number of extraction cycles influenced the chemical diversity of the extracts and that under the set experimental conditions, the extraction with 100% ethanol at lower temperatures provided an extract that best represented the chemical diversity of the sponge (Chapter 3).

Continuing with the exploration of the metabolome of giant barrel sponges, differences in the metabolome of the sponge related to variations in environmental conditions were

investigated using LC-MS and NMR based metabolomics (chapters 4, 5, and 6). Firstly, considering the worldwide distribution of these sponges, the metabolic composition of sponges collected in four locations around the world (Martinique, Curaçao, Taiwan and Tanzania) was compared. Results showed that samples were clearly distinguished according to their location. However, no correlation was found between the antibacterial activity against *Staphylococcus aureus* and *Escherichia coli* exhibited by some samples and their location. This finding indicates that the production of active compounds could be affected by environmental conditions occurring at a smaller scale or that other aspects, such as genetic factors, play an important role in this area. Additionally, using the sample set collected in Curaçao, the effect of sea depth (7-43 m) on the metabolome was evaluated, showing that environmental changes along a depth gradient only had a significant effect on one of the genetic groups (putative species) of *Xestospongia* spp. studied. Lastly, the effect of sea surface temperature (SST) and pH was evaluated on a sample set collected in the Spermonde archipelago, SW Sulawesi (Indonesia). Again, in this case, only one of the genetic groups studied in this region displayed a change in the metabolome related to SST. Moreover, the pH of the seawater did not affect any of the genetic groups, suggesting that these sponges exhibit some resilience when faced with ocean acidification scenarios. In all the cases, when environmental changes altered the metabolome of any of the sponges, the modification was mostly related to lipid type compounds including brominated, hydroxylated and/or polyacetylenic fatty acids that exist either in their free form or as an acyl chain in glycerophospholipids.

Another condition that was considered to be potentially reflected in the metabolome of the sponges is the stage of development throughout their lifespan. Sponges lack age-related markers and *X. muta* was the first sponge for which it was possible to establish the age of the specimen. Using the age of giant barrel sponges as a factor and LC-MS based metabolomics together with molecular networking, it was possible to establish that age indeed has an effect on the metabolome of the sponges (Chapter 5). Moreover, the discriminant signals related to older sponges coincided with phosphatidylcholine glycerophospholipids containing two acyl groups. These types of lipids are well known to be membrane lipids, suggesting a higher rate of cellular division in older sponges. However, additional ecological roles of these compounds cannot be disregarded.

The presence of cryptic species in giant barrel sponges raises the question of whether these putative species could also be distinguished based on their metabolome. In this case samples collected in Indonesia were classified into one of the three genetic groups reported to be present in this region. Using LC-MS and NMR-based metabolomics together with molecular

networking it was possible to observe differences in the chemical production between the genetic groups (Chapter 6). Interestingly, the compounds found to be responsible for the separation were *lyso*-phospholipids which have also been reported to vary in sponges as a response to changes in environmental conditions. This indicates that these genetically closely related sponges might be reacting differently to similar environmental conditions.

In conclusion, this thesis showed how environmental, genetic and biological factors individually or as a result of their interaction can influence the metabolome of giant barrel sponges. This is only the first step towards a better understanding of the role that chemical compounds play in interactions between sponges and the surrounding environment. The advances in this field will rely, among others, on the development of technologies that would allow real time measurement of the metabolome and the study of the interaction between the metabolome and other omics, like the microbiome.

Samenvatting

Zeesponzen zijn belangrijke leden van rifecosystemen, omdat ze ecologische rollen vervullen die essentieel zijn voor de gezondheid van het rif. Vanuit chemisch perspectief zijn sponzen uitgebreid bestudeerd en vertonen ze een grote chemische diversiteit met veel biologische activiteit, hoewel deze chemische diversiteit vaak niet in verband kan worden gebracht met bijvoorbeeld variatie in omgevingscondities die sponzen ervaren. Metabolomics, met zijn holistische overzicht van de metabolieten die in monsters aanwezig zijn, heeft nieuwe tools opgeleverd om inzicht te krijgen in de functie van metabolieten in de ingewikkelde relatie tussen sponzen en hun omgeving. Onder sponzen bleek de reuzenbekerspons (*Xestospongia bergquistia, X. muta en X. testudinaria*), zich te onderscheiden door hun grote aantallen in veel riffen, hun lange levensduur en het feit dat ze wijdverspreid zijn over de oceanen. In sommige riffen kan de reuzenbekerspons tot 13% van het beschikbare benthische substraat bedekken, wat meer is dan koralen. Ze kunnen meer dan 2000 jaar leven en zijn te vinden van Nieuw-Caledonië tot het Caribisch gebied met een aanwezigheid in de hele Rode Zee, de Oost-Afrikaanse kust en de Indo-Pacifische-regio. Er is ook aangetoond dat wat oorspronkelijk als drie verschillende soorten werd geclassificeerd (*X. bergquistia* beperkt tot het noordelijke Groot Barrièrerif; *X. muta* in het Caribische gebied en *X. testudinaria* is wijdverspreid in het Indo-Pacifisch gebied) eigenlijk een soortencomplex is dat ten minste negen cryptische soorten bevat. De aanwezigheid van cryptische soorten komt veel voor in sponstaxa en is een van de grootste problemen bij de classificatie van deze dieren. Al deze kenmerken maken de reuzenbekerspons een goed modelorganisme om de invloed te begrijpen die omgevings-, biologische en genetische factoren kunnen hebben op de metabolische productie van zeesponzen.

Als onderdeel van de onderschrijving van metabolomics als een methodiek die de studie van het metabolisme van zeesponzen en mariene organismen in het algemeen mogelijk maakt, is een literatuuroverzicht van de methoden en toepassingen van metabolomics op dit gebied opgenomen in dit proefschrift in Hoofdstuk 2. De eerste stap in een metabolomics-onderzoek is de extractiemethode, die cruciaal is voor het verkrijgen van een extract dat nauwkeurig de metabolieten in het monster weergeeft. De invloed van extractieomstandigheden (druk, temperatuur, polariteit van het oplosmiddel en aantal cycli) op de chemische diversiteit van extracten van *Xestospongia* spp. werd bestudeerd met behulp van een extractiesysteem onder druk. Het bleek dat temperatuur, oplosmiddelpolariteit en aantal extractiecycli de chemische diversiteit van de extracten beïnvloedden en dat onder de vastgestelde experimentele

omstandigheden de extractie met 100% ethanol bij lagere temperaturen een extract opleverde dat het beste de chemische diversiteit van de spons presenteerde (Hoofdstuk 3).

Voortbouwend op de verkenning van het metaboloom van *Xestospongia* spp. werden verschillen in het metaboloom van de spons gerelateerd aan variaties in omgevingscondities, onderzocht met behulp van LC-MS en NMR-gebaseerde metabolomics (hoofdstukken 4, 5 en 6). Ten eerste werd, gezien de wereldwijde distributie van deze sponzen, de metabolische samenstelling van sponzen die verzameld zijn op vier locaties over de wereld (Martinique, Curaçao, Taiwan en Tanzania) vergeleken. De resultaten toonden aan dat monsters duidelijk konden worden onderscheiden op basis van hun locatie. Er werd echter geen correlatie gevonden tussen enerzijds de antibacteriële activiteit tegen *Staphylococcus aureus* en *Escherichia coli* die door sommige monsters werd vertoond en anderzijds de geografische locatie. Deze bevinding geeft aan dat de productie van actieve verbindingen kan worden beïnvloed door omgevingsomstandigheden die op kleinere schaal voorkomen of dat andere aspecten, zoals genetische factoren, een belangrijke rol spelen op dit gebied. Hiernaast werd met behulp van de op Curaçao verzamelde set monsters het effect van zeediepte (7-43 m) op het metaboloom geëvalueerd, wat aantoont dat omgevingsveranderingen langs een dieptegradiënt alleen een significant effect hadden op één van de genetische groepen (vermoedelijke soorten) van *Xestospongia* spp. die in dit onderzoek zijn bestudeerd. Ten slotte werd het effect van de temperatuur van het zeeoppervlak (SST) en de pH geëvalueerd op een monster set verzameld in de Spermonde-archipel, SW Sulawesi (Indonesië). In dit geval vertoonde nogmaals slechts één van de genetische groepen die in deze regio werden bestudeerd een verandering in het metaboloom gerelateerd aan SST. Bovendien had de pH van het zeewater geen invloed op de genetische groepen, wat suggereert dat deze sponzen enige veerkracht vertonen wanneer ze worden geconfronteerd met scenario's zoals oceaanverzuring. In alle gevallen, wanneer omgevingsveranderingen het metaboloom van een van de sponzen veranderden, was de wijziging meestal gerelateerd aan lipide-achtige verbindingen, waaronder gebromeerde, gehydroxyleerde en/ of polyacetyleenvetzuren die in hun vrije vorm of als een acylketen in glycerofosfolipiden voorkomen.

Een andere conditie waarvan werd aangenomen dat deze mogelijk tot uiting kwam in het metaboloom van de sponzen, zijn de verschillende ontwikkelingsstadia gedurende de gehele levensduur. Sponzen missen leeftijdsgebonden markers en *X. muta* was de eerste spons waarvoor het mogelijk was om de leeftijd van het specimen vast te stellen. Door de leeftijd van reuzenbekersponzen als factor te gebruiken en van LC-MS gebaseerde metabolomics met moleculaire netwerken gebruik te maken, was het mogelijk om vast te stellen dat leeftijd inderdaad een effect heeft op het metaboloom van de sponzen (Hoofdstuk 5). Bovendien

vielen de onderscheidende signalen met betrekking tot oudere sponzen samen met fosfatidylcholine glycerofosfolipiden die twee acylgroepen bevatten. Van dit soort lipiden is bekend dat het membraanlipiden zijn, wat duidt op een hogere celdeling bij oudere sponzen. Bijkomende ecologische rollen van deze verbindingen kunnen echter niet worden genegeerd.

De aanwezigheid van cryptische soorten binnen de reuzenbekersponzen roept de vraag op of deze vermeende soorten ook kunnen worden onderscheiden op basis van hun metaboloom. In dit geval werden monsters verzameld in Indonesië ingedeeld in één van de drie genetische groepen die naar verluidt in deze regio aanwezig zijn. Met behulp van LC-MS en NMR-gebaseerde metabolomics samen met moleculaire netwerken was het mogelijk om verschillen in de chemische productie tussen de genetische groepen waar te nemen (Hoofdstuk 6). Interessant is dat de verbindingen waarvan werd vastgesteld dat ze verantwoordelijk waren voor deze distinctie lyso-fosfolipiden waren waarvan ook is gerapporteerd dat ze variëren in sponzen als een reactie op veranderingen in omgevingsomstandigheden. Dit geeft aan dat deze genetisch nauw verwante sponzen mogelijk anders reageren op vergelijkbare omgevingsomstandigheden.

Concluderend laat dit proefschrift zien hoe omgevingsfactoren, genetische en biologische factoren, individueel of als gevolg van hun interactie, het metaboloom van van de reuzenbekersponzen kunnen beïnvloeden. Dit is slechts de eerste stap naar een beter begrip van de rol die chemische verbindingen spelen in interacties tussen sponzen en de omgeving. Vooruitgang op dit gebied zal onder meer afhangen van de ontwikkeling van technologieën die het mogelijk maken om het metaboloom in real time te meten en de studie van de interactie tussen het metaboloom en andere omics, zoals het microbioom.

Resumen

Las esponjas marinas son miembros importantes de los ecosistemas marinos con distintos roles ecológicos que las hacen esenciales para la salud de los arrecifes. Desde un punto de vista químico, las esponjas han sido ampliamente estudiadas mostrando una gran diversidad química y actividad biológica que no siempre ha podido relacionarse con cambios ambientales experimentados por las mismas. La metabolómica con su visión global de todos los metabolitos presentes en una muestra ha proporcionado nuevas herramientas para entender la función que estos tienen en la relación entre las esponjas y sus hábitats. Entre las esponjas, las denominadas esponjas barril gigantes se han destacado por varias razones, incluyendo: su predominancia en muchos arrecifes, su longevidad y el hecho que están ampliamente distribuidas en los océanos alrededor del mundo. En algunos arrecifes, las esponjas barril gigantes pueden cubrir hasta el 13% de sustrato disponible superando incluso a los corales. Estas pueden vivir por más de 2000 años y se les encuentra desde Nueva Caledonia hasta el Caribe, incluyendo el Mar Rojo, la costa africana oriental y la región Indo-Pacífica. Mas aún, se ha demostrado que lo que originalmente fue clasificado como tres especies diferentes (*Xestospongia bergquistia* (confinada a la parte norte de la gran barrera de coral), *Xestospongia muta* y *Xestospongia testudinaria*) es realmente un complejo de especies que contiene al menos nueve especies crípticas. La presencia de especies crípticas ha sido recurrente en filo porífera y es una de las principales dificultades en la clasificación de estos animales. Todas las características mencionadas anteriormente hacen de las esponjas barril gigantes un excelente organismo modelo para poder entender la influencia que pueden llegar a tener factores ambientales, biológicos y genéticos en la producción de metabolitos por parte de las esponjas marinas.

El análisis metabolómico fue utilizado como una herramienta para el estudio exhaustivo del metabolismo de las esponjas marinas y de los organismos marinos en general. El capítulo 2 de esta tesis incluye una revisión bibliográfica de los métodos y aplicaciones de la metabolómica en este campo. El primer paso en un estudio metabolómico es la selección del método de extracción. Por esta razón, se utilizó un sistema extracción presurizado para estudiar la influencia de que tienen las condiciones de extracción (presión, temperatura, polaridad del solvente y numero de ciclos de extracción) en la diversidad química de extractos de *Xestospongia* spp. Bajo las condiciones estudiadas se encontró que las extracciones usando etanol al 100% a bajas temperaturas dan como resultado el extracto que mejor representa la diversidad química de la esponja (Capítulo 3).

Continuando con la exploración metabolómica de las esponjas barril gigantes se investigaron las variaciones en el metaboloma de las esponjas en relación con las condiciones ambientales en las que se encontraban utilizando como metodología cromatografía líquida acoplada a espectrometría de masas (LC-MS por sus siglas en inglés) y resonancia magnética nuclear (RMN) (Capítulos 4, 5 y 6). Teniendo en cuenta la distribución global de estas esponjas, se realizó una comparación de la composición metabólica de esponjas recolectadas en cuatro locaciones alrededor del mundo (Martinica, Curazao, Taiwán y Tanzania). Los resultados mostraron que es posible distinguir las muestras de acuerdo con su lugar de origen, sin embargo, no se pudo observar una correlación entre la actividad antibacterial observada contra *Staphylococcus aureus* y *Escherichia coli* y el lugar de origen de la muestra lo cual indica que la producción de compuestos bioactivos puede afectarse por condiciones ambientales menor escala o incluso estar relacionado con otras causas como factores genéticos. Adicionalmente, se estudió el efecto que la profundidad (7-43 m) tiene en el metaboloma de las esponjas usando un conjunto de muestras recolectadas en Curazao. Este estudio demostró que el cambio en el ambiente a lo largo del gradiente de profundidad solo tiene efectos significativos en el metaboloma de uno de los grupos genéticos (posibles especies) de *Xestospongia* spp. estudiados. Finalmente, se evaluó el efecto de la temperatura de la superficie del mar (SST por sus siglas en inglés) y el pH del agua en un conjunto de esponjas recolectadas en el archipiélago de Spermode en Sulawesi (Indonesia). En este caso nuevamente solo uno de los grupos genéticos presentes en esta región mostro cambios en el metaboloma relacionados con la SST. Adicionalmente, el pH del agua no mostró tener efecto alguno en ninguno de los grupos genéticos analizados, sugiriendo que estas esponjas presentan cierto grado de resiliencia cuando se ven enfrentadas a escenarios de acidificación de los océanos. En todos los casos presentados anteriormente, cuando las condiciones ambientales causaron una alteración en el metaboloma de estas esponjas, los cambios estuvieron relacionados principalmente con compuestos de tipo lipídico que incluyeron ácidos grasos bromados, hidroxilados y/o con grupos alquino, los cuales pueden estar en forma libre o como cadenas unidas a glicerofosfolípidos.

Otro de los factores que podría causar cambios en el metaboloma de estas esponjas es su etapa de desarrollo en su ciclo de vida. Las esponjas carecen de marcadores que permitan determinar su edad y *X. muta* fue la primera especie para la cual fue posible determinar la edad de los individuos. Usando la edad calculada para las esponjas barril gigantes y utilizando la metodología de redes moleculares basadas en análisis de LC-MS se determinó que la edad de las esponjas puede causar cambios en su metaboloma (Capítulo 5). Más aún las señales características de esponjas más viejas corresponden a glicerofosfolípidos con dos ácidos

grasos unidos. Este tipo de compuestos es bien conocido por ser parte de la membrana celular, lo cual sugeriría una mayor tasa de división celular en las esponjas más viejas, aunque otros roles ecológicos no pueden ser descartarse.

La presencia de especies crípticas en las esponjas barril gigantes plantea la posibilidad de distinguir estas especies a través de sus metabolomas. Para determinar la veracidad de esta hipótesis las muestras recolectadas en Indonesia fueron clasificadas en uno de los tres grupos genéticos reportados en esta región. En el capítulo 6 se muestran las diferencias observadas en el metaboloma de estos tres grupos basados en análisis de LC-MS y RMN combinados con redes moleculares. Los compuestos responsables por la diferenciación entre los grupos genéticos fueron *lyso*-fosfolípidos, de los cuales se ha reportado que pueden cambiar dependiendo de las condiciones ambientales. Esto indicaría que dichos grupos genéticamente muy similares pueden tener diferentes respuestas metabólicas al encontrarse en ambientes similares.

En conclusión, esta tesis mostró como factores ambientales, genéticos y biológicos, independientemente o como resultado de la interacción entre ellos, pueden influenciar el metaboloma de las esponjas barril gigantes. Este es solo un primer paso hacia la comprensión del papel que los metabolitos juegan en la interacción de las esponjas con su entorno. Los avances en este campo dependerán, entre otros, del desarrollo de tecnologías que permitan mediciones del metaboloma en tiempo real y del estudio de las interacciones entre el metaboloma y otras ómicas como el microbioma.

Curriculum Vitae

Lina Maria Bayona Maldonado was born in Bogotá, Colombia, on the 6th of June 1990. She started her bachelor in Chemistry in Colombian National University in 2007. During her bachelor Prof. Dr. Freddy Ramos offered her the opportunity of an internship working with marine cyanobacteria that resulted in a newfound love for marine natural products. After finishing her bachelor in 2012 she started her Master in Chemistry in the same university studying the potential of marine cyanobacteria for the discovery of new cytotoxic compounds. After finishing her master, she started working as a lecturer in the Sabana University (Chia, Colombia), teaching organic chemistry and biochemistry. Teaching was one of the most enriching experiences she ever had and to continue this path she decided to move forward in her academic career and apply for a Ph.D. position. In 2016, she received a scholarship from the Colombian government that allowed her to start her Ph.D. under the supervision of Prof. Dr. Peter Klinkhamer, Prof. Dr. Nicole de Voogd, and Dr. Young Choi to continue the study of marine natural products, this time from a metabolomics perspective. Her work concerning the changes in the metabolome of giant barrel sponges under diverse environmental conditions is described in this thesis.

List of publications

Research papers

Bayona, L.M., Kim, M-S., Swierts, T., Hwang, G.S., de Voogd, N.J., & Choi, Y. H. (2021) Metabolic variation in Caribbean giant barrel sponges: influence of age and sea-depth. *Submitted*

Bayona, L. M., de Voogd, N.J., & Choi, Y.H. (2021) Recent progress in marine organism metabolomics: towards a better understanding of the role of metabolites. *Submitted*

Bayona, L. M., van Leeuwen, G., Erol, Ö., Swierts, T., van der Ent, E., de Voogd, N. J., & Choi, Y. H. (2020). Influence of Geographical Location on the Metabolic Production of Giant Barrel Sponges (*Xestospongia* spp.) Revealed by Metabolomics Tools. ACS Omega, 5(21), 12398–12408. https://doi.org/10.1021/acsomega.0c01151

Bayona, L. M., Videnova, M., & Choi, Y. H. (2018). Increasing Metabolic Diversity in Marine Sponges Extracts by Controlling Extraction Parameters. Marine Drugs, 16(10). https://doi.org/10.3390/md16100393

Liu, X., Ahlgren, S., Korthout, H. A. A. J., Salomé-Abarca, L. F., **Bayona, L. M.**, Verpoorte, R., & Choi, Y. H. (2018). Broad range chemical profiling of natural deep eutectic solvent extracts using a high performance thin layer chromatography–based method. Journal of Chromatography A, 1532, 198–207. https://doi.org/10.1016/j.chroma.2017.12.009

Book Chapter

Bayona, L. M., Verpoorte, R., Klinkhamer, P. G. L., & Choi, Y. H. (2019). Thin-Layer Chromatography | Metabolomics. In P. Worsfold, C. Poole, A. Townshend, & M. Miró (Eds.), Encyclopedia of Analytical Science (Third Edition) (pp. 59–75). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-409547-2.14400-2