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

Bayona Maldonado, L.M.

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.

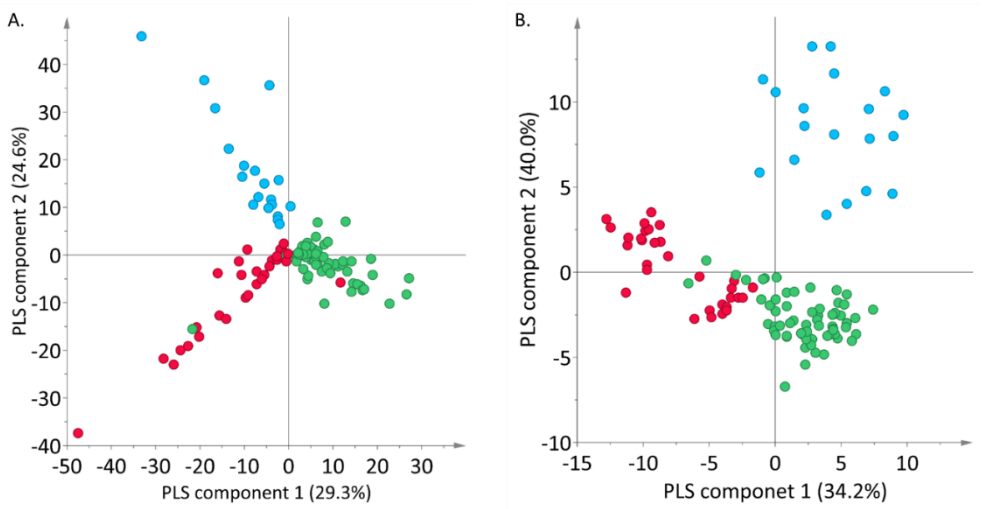

Figure 6.1: First two components of the partial least square-discriminant analysis (PLS-DA) model for the three genetic groups, group 1 (green), group 2 (blue), group 3 (red), of *X. testudinaria* samples A: LC-MS data (R^2 = 0.916, Q^2 = 0.686) and B: NMR data (R^2 =0.907, Q^2 = 0.849)

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

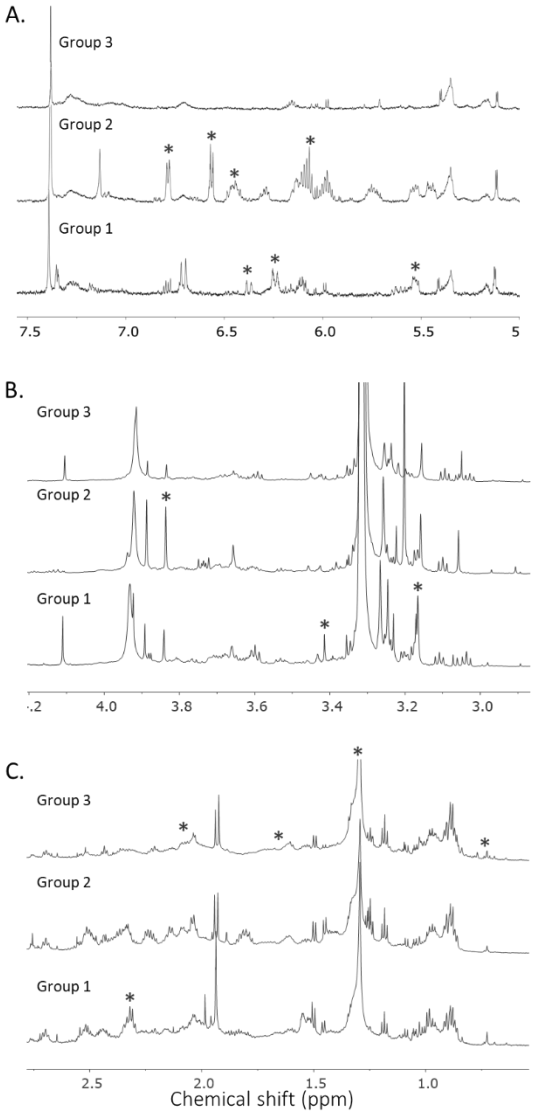

Figure 6.2: Typical ^{1}H-NMR (600MHz in CH_3OH-d_4) spectra of *X. testudinaria* belonging to three different genetic groups. A: δ_H 5.0-7.5, B: δ_H 3.0-4.0 and C: δ_H 0.6-2.6. the characteristic signals for each group are indicated by a red asterisk.

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.

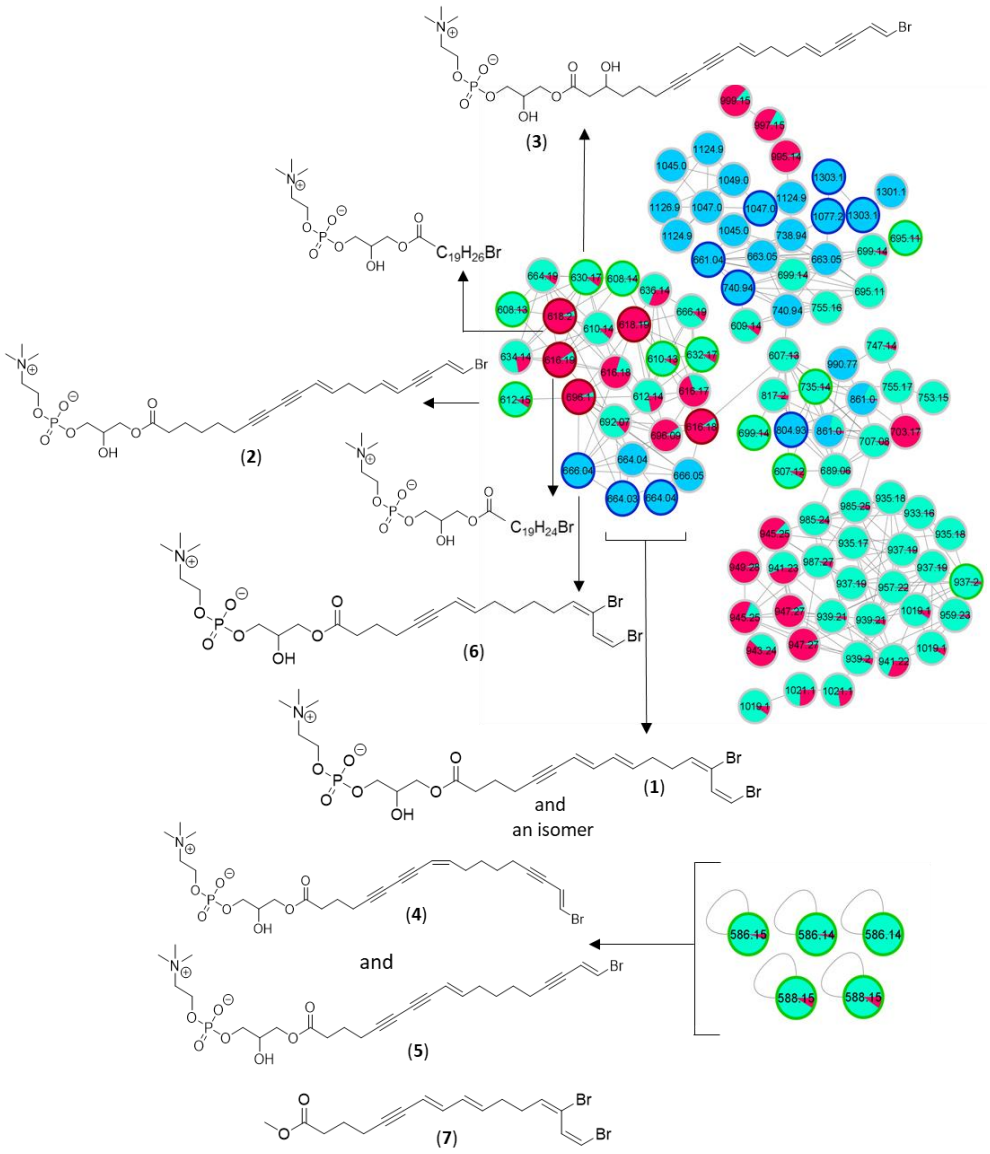

Figure 6.3: Molecular network of *X. testudinaria* samples belonging to three genetic groups showing the structure of the isolated compounds 1-7 related to the separation between groups. Each node is colored as a pie chart showing the abundance of the feature in each genetic group in green (group 1), 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).

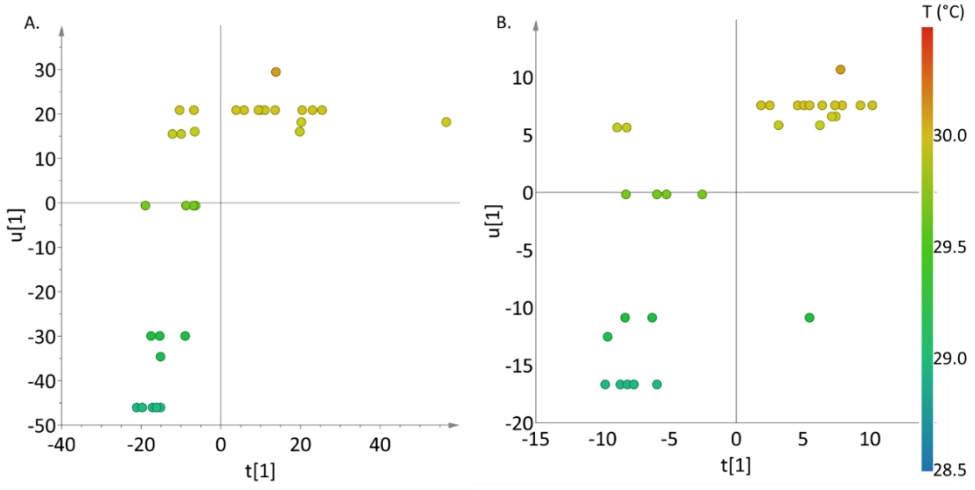

Figure 6.4: First component of the partial least square (PLS) analysis for X variable (t[1]) vs Y variables (u[1])based on A: LC-MS data and B: NMR Data of the samples belonging to genetic group 3 (C6A2) of *X. testudinaria* in response to variations in the sea surface temperature.

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

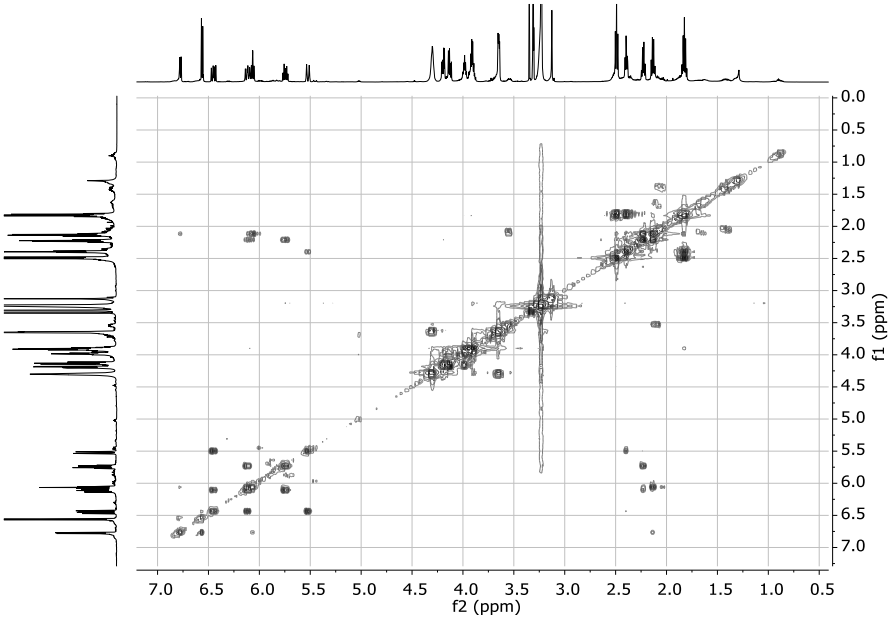

Figure S1: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-*d*$_4$, 600 MHz) of Compound **1**.

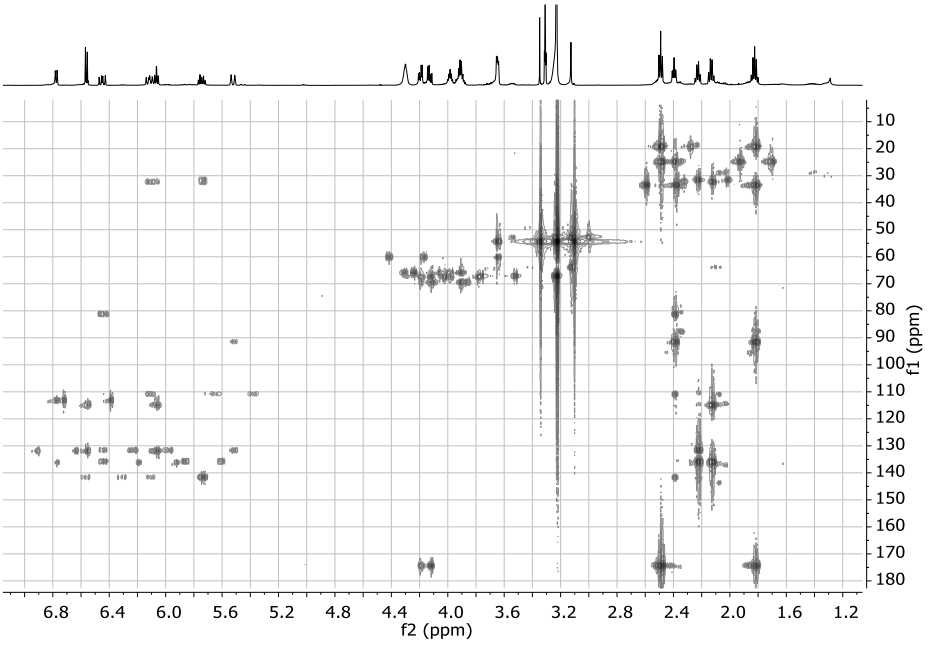

Figure S2: Heteronuclear multiple bond correlation (HMBC) spectrum (CH_3OH-*d*$_4$, 600 MHz) of Compound **1**.

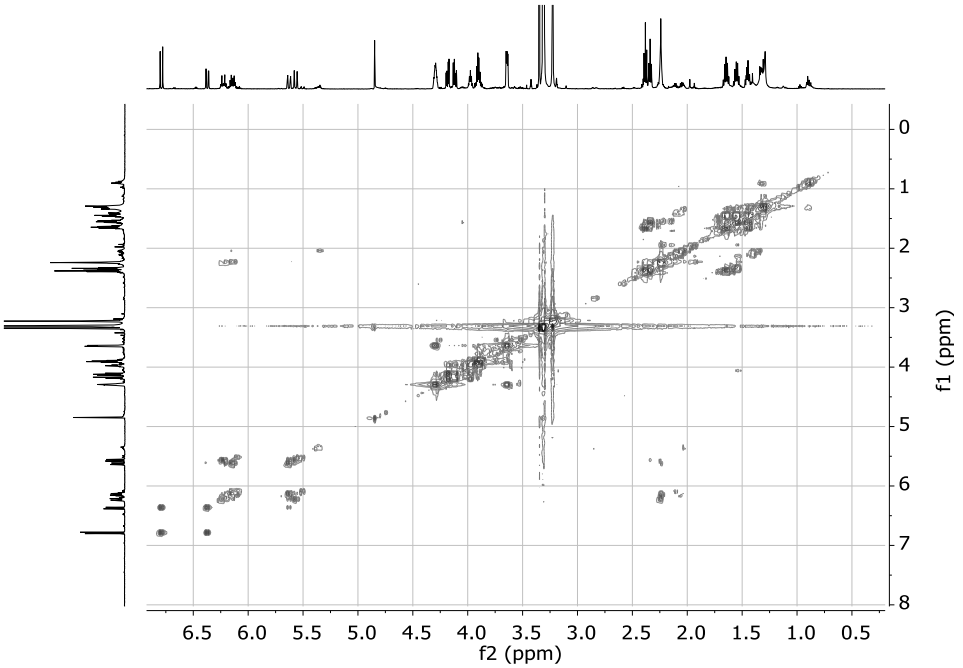

Figure S3: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **2**.

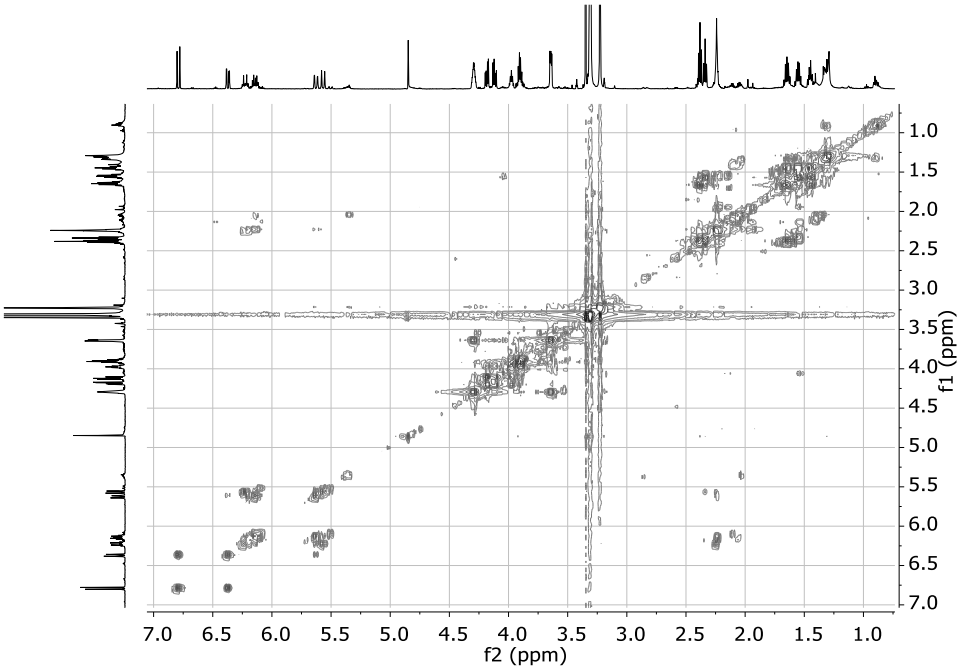

Figure S4: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **3**.

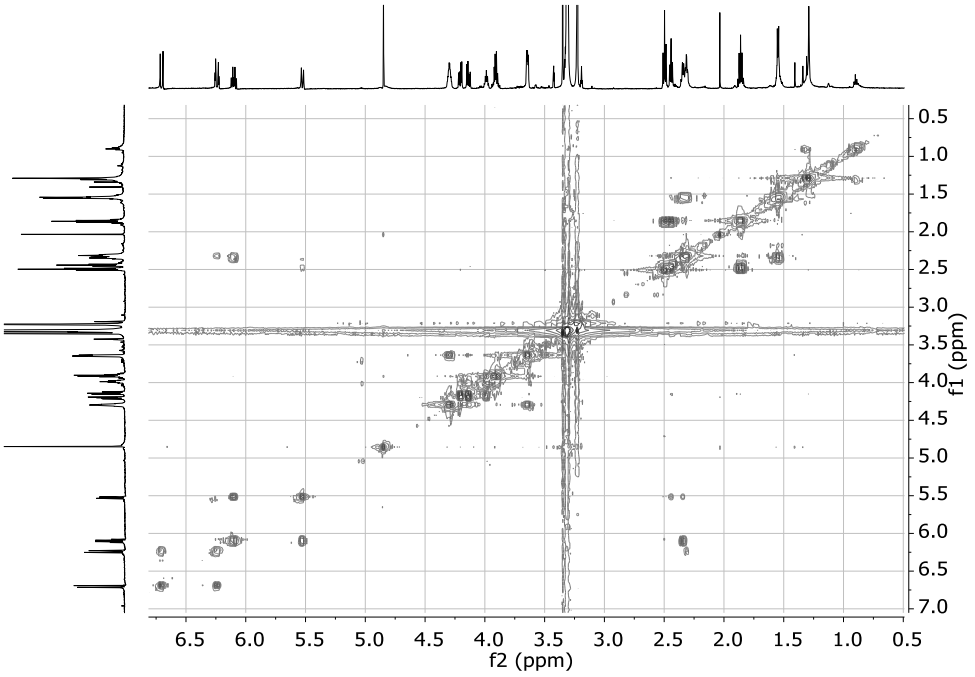

Figure S5: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **4**.

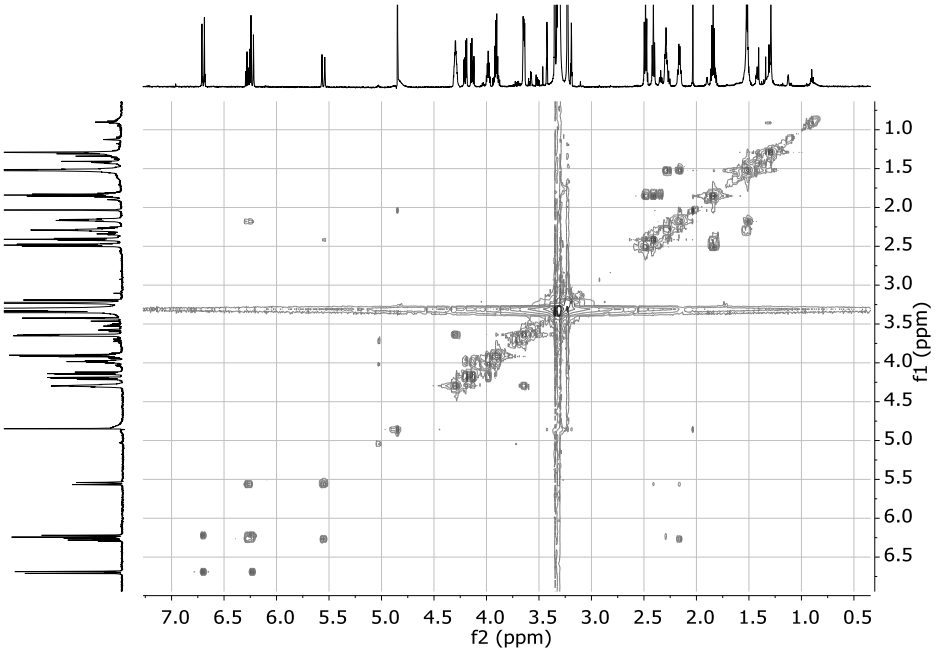

Figure S6: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **5**.

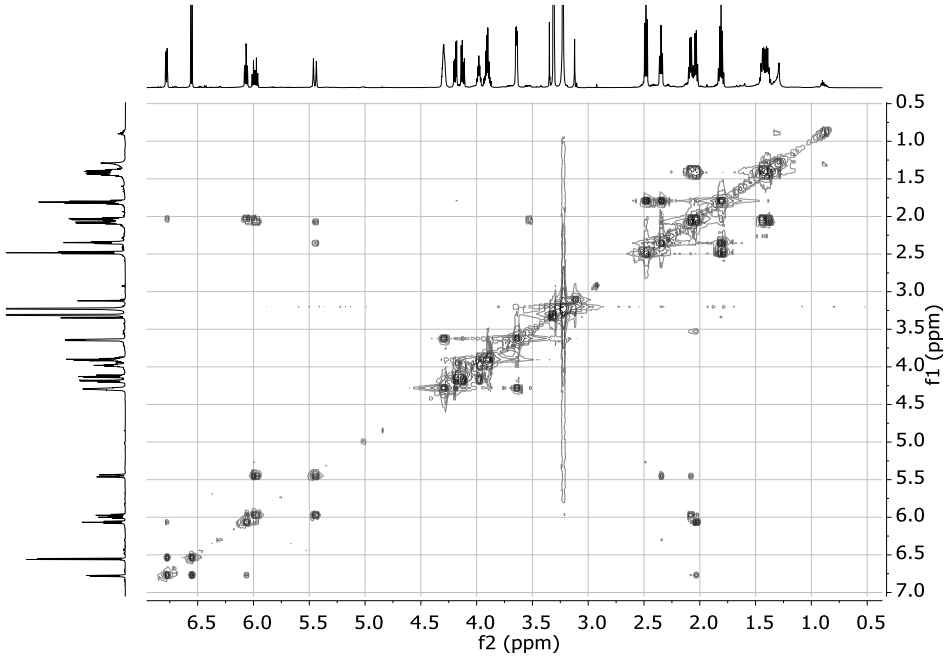

Figure S7: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **6**.

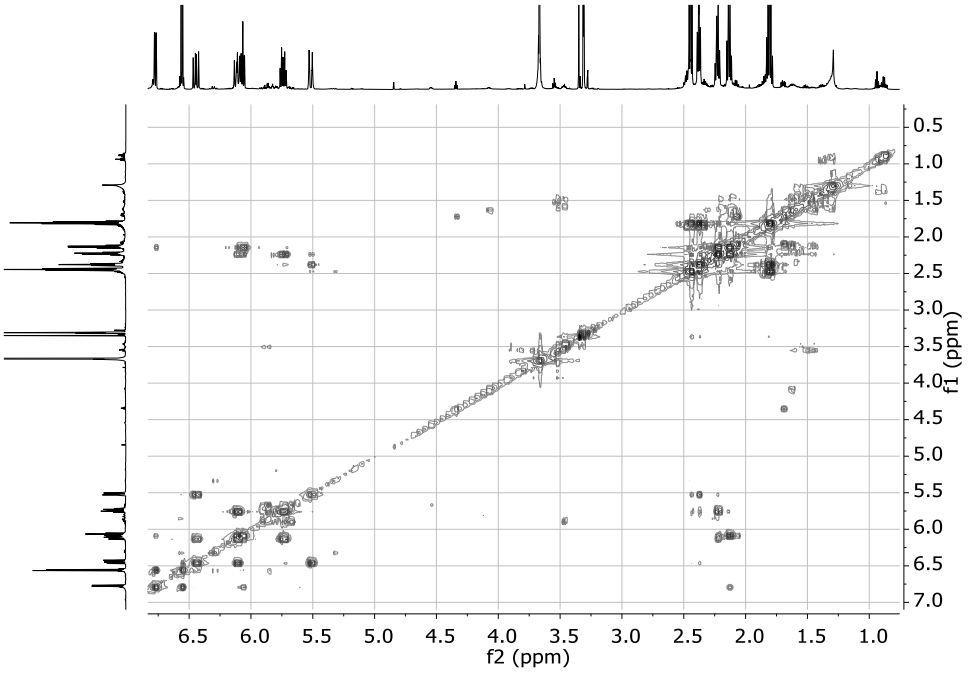

Figure S8: ^{1}H-^{1}H Correlation spectroscopy (COSY) spectrum (CH_3OH-d_4, 600 MHz) of Compound **7**.

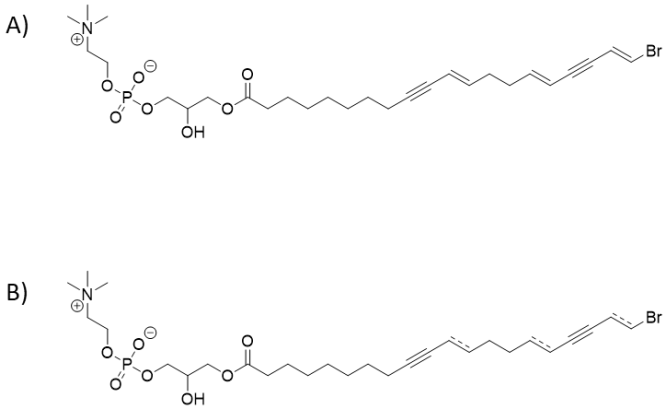

Figure S9: Putative structures characteristic for Group 3 A: m/z 618.20 and B: m/z 616.19

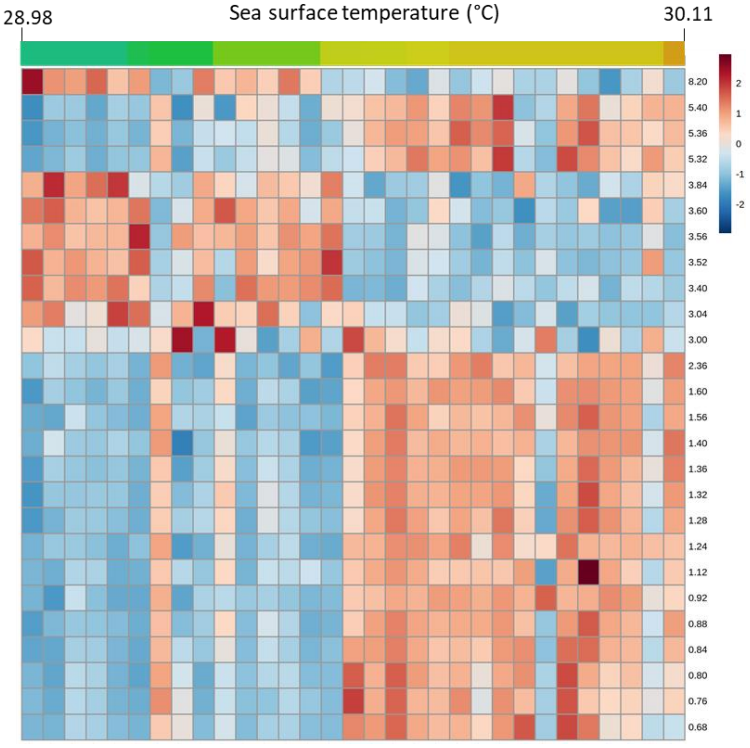

Figure S10: Heat map of characteristic signals from ^{1}H-NMR data obtained from the variable importance for the projection (VIP) plot of partial least-squares (PLS) of samples belonging to genetic group 3 using as Y variable the temperature.