Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed Taonia atomaria

Nathan Carriot, Benoît Paix, Stephane Greff, Bruno Viguier, Jean-François Briand, Gérald Culioli

▶ **To cite this version:**

Nathan Carriot, Benoît Paix, Stephane Greff, Bruno Viguier, Jean-François Briand, et al.. Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed Taonia atomaria. Talanta, 2021, 225, pp.121925. 10.1016/j.talanta.2020.121925 . hal-03108198

HAL Id: hal-03108198
https://amu.hal.science/hal-03108198

Submitted on 13 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

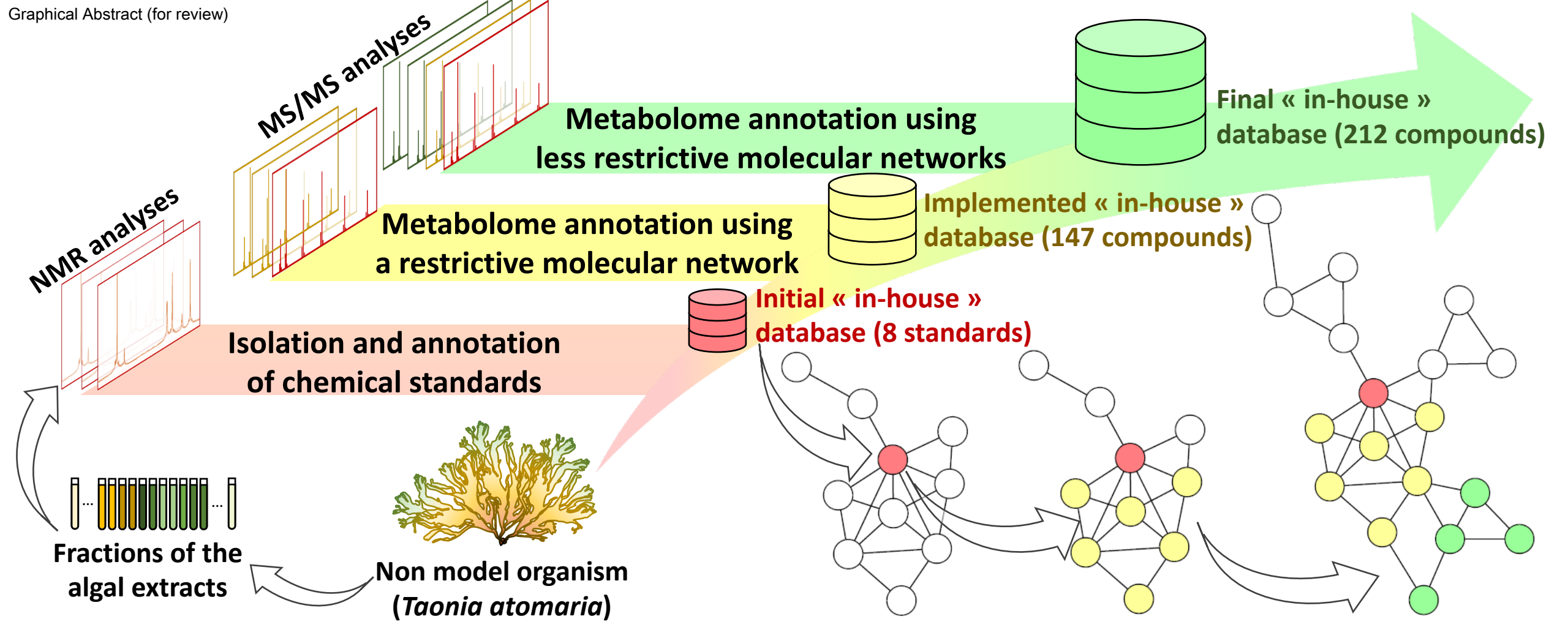
Highlights for the Ms:

Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed Taonia atomaria

by Carriot et al.

- An annotation method using MS/MS molecular networking was developed.
- A workflow coupling molecular networking and phytochemistry was described.
- The MS/MS fragmentation pattern of new families of algal lipids was deciphered.
- 212 metabolites were annotated from the metabolome of the brown alga *T. atomaria*.

Graphical Abstract (for review)



Revised Manuscript [Unmarked Text]

Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed *Taonia atomaria*

Nathan Carriot[a,1], *Benoît Paix*[a,1], *Stéphane Greff*[b], *Bruno Viguier*[a], *Jean-François Briand*[a], *Gérald Culioli*[a,*]

[a] Université de Toulon, MAPIEM, EA 4323, Toulon, France

[b] Aix Marseille Université, CNRS, IRD, Avignon Université, Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), Station marine d'Endoume, Marseille, France.

[1]These authors contributed equally to this work.

[*]Corresponding author.

E-mail addresses: culioli@univ-tln.fr and gerald.culioli@univ-avignon.fr (G. Culioli)

New affiliation: Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, UMR CNRS-IRD, Avignon Université, Aix-Marseille Université, Avignon, France.

Abstract

Untargeted LC-MS based metabolomics is a useful approach in many research areas such as medicine, systems biology, environmental sciences or even ecology. In such an approach, annotation of metabolomes of non-model organisms remains a significant challenge. In this study, an analytical workflow combining a classical phytochemical approach, using the isolation and the full characterization of the chemical structure of natural products, together with the use of MS/MS-based molecular networking with various levels of restrictiveness was developed. This protocol was applied to the marine brown seaweed *Taonia atomaria*, a cosmopolitan algal species, and allowed to annotate more than 200 metabolites. First, the algal organic crude extracts were fractionated by flash-chromatography and the chemical structure of eight of the main chemical constituents of this alga were fully characterized by means of spectroscopic methods (1D and 2D NMR, HRMS). These compounds were further used as chemical standards. In a second step, the main fractions of the algal extracts were analyzed by UHPLC-MS/MS and the resulting data were uploaded to the Global Natural Products Social Molecular Networking platform (GNPS) to create several molecular networks (MNs). A first MN (**MN-1**) was built with restrictive parameters and allowed the creation of clusters composed by nodes with highly similar MS/MS spectra. Then, using database hits and chemical standards as "seed" nodes and/or similarity between MS/MS fragmentation pattern, the main clusters were easily annotated as common glycerolipids and phospholipids, much rare lipids -such as acylglycerylhydroxymethyl-*N*,*N*,*N*-trimethyl-*ß*-alanines or fulvellic acid derivatives- but also new glycerolipids bearing a terpene moiety. Lastly, the use of less and less constrained MNs allowed to further increase the number of annotated metabolites.

Keywords: Molecular networking; Metabolomics; Lipidomics; UHPLC-MS/MS; Macroalga; *Taonia atomaria*.

Graphical abstract

1. Introduction

Marine chemical ecology is a field of research that aims to address the role of chemicals in marine ecosystems. Indeed, molecules expressed by marine organisms have preponderant ecological functions, e.g. foraging, defense or reproduction [1]. There is a growing interest in this field of study because it benefits from many recent technical advances including those of environmental metabolomics [2].

Metabolomics is the latest in the so-called "omics" sciences that brings together genomics, transcriptomics and proteomics [3]. Ultimately, in such a research area the aim is to decipher the metabolome, which represents the whole set of low-molecular-weight organic compounds present in an organism or more generally in a biological sample, and to evaluate and interpret its variations [4]. The metabolome is constituted by all quantifiable metabolites and appears to be highly complementary with other “omics” sciences for a global systems biology approach [5–7].

Untargeted LC-MS-based metabolomic studies are currently the most common ones but metabolite identification is generally the limiting factor when interpreting the resulting data [8]. Indeed, the diversity of the metabolism is vast in nature and metabolites implied in the discrimination between sample groups often remain unidentified [9]. With an estimation of less than 2% of mass spectra which can be annotated from untargeted studies, the characterization of the chemical structures is the most challenging step of the metabolomic workflow [10]. Knowing that each LC-MS analysis generates a large number of spectra [11,12], spectral identification of each compound, one by one, using classical dereplication tools with reference databases (DBs) such as NCBI, Metlin, NIST or MassBank, is a too time-consuming approach. For the past few years, MS/MS molecular networking (MN) has been proposed as a new emerging tool to optimize this dereplication process [13]. Based on the idea that MS/MS fragmentation patterns are similar for molecules sharing close chemical structures, the Global

Natural Products Social Molecular Networking (GNPS) Web-platform (http://gnps.ucsd.edu) allows to generate MNs where compounds are linked according to their molecular relatedness [14–16]. The resulting MNs show then diverse clusters composed by several *m/z* features (defined as "nodes") characterized by their MS/MS data. The clusters generated in a MN could then gather metabolites belonging to a same chemical family or sharing closely related chemical structures.

In this study, the use in a first step of a restricted MN will allow to gather compounds having very similar chemical structures in order to facilitate their annotation. Indeed, to unlock clusters and to putatively identify all nodes of a cluster, two situations are expected: (i) if a compound is identified, all nodes of a cluster could be annotated using mass differences from this "seed node" [17–19], (ii) if all compounds remain unidentified, the annotation can be done by determining the common structural characteristics of the nodes thanks to the comparison of their MS/MS spectra. The enlargement of the MN is then considered in order to continue the annotation of the metabolome by using less restrictive parameters for the connection between nodes.

The selected model in this study was *Taonia atomaria*, a brown alga commonly reported along the Mediterranean and North-Eastern Atlantic coasts. To date, several types of chemical compounds, mainly sesquiterpenes [20–24], meroditerpenes [25–29] and geranylgeranylglycerol (GGG) [24], have been characterized from the genus *Taonia* by natural products chemistry. Because of its morphology and distribution, this algal species represents a model of choice for studies in chemical ecology, and more particularly for those dealing with interactions between algae and their associated microbiota. For this reason and in order to carry out metabolomic analyses, the metabolome of this alga needs to be better characterized.

In this context, the aim of the present study was to develop an optimal workflow dedicated to the annotation of a largely unexplored metabolome of a non-model species, namely *T. atomaria*.

The proposed workflow was based on three successive steps: (i) a classical phytochemical approach permitting to isolate chemical standards which can then be used as seed nodes, initiating a "in-house" MS/MS DB ; (ii) a GNPS networking approach with high restrictiveness allowing only the most structurally-related metabolites to cluster together and to enrich the DB, and (iii) a GNPS networking approach with less restrictive parameters allowing to gather other related metabolites in order to complete the DB.

The present approach is expected to significantly extend our knowledge on algal metabolomes through the MS/MS fragmentation pattern of lipid families presumed to be new and at the same time to propose new solutions for the annotation of the metabolome of little-studied species.

2. Experimental

The metabolome of *T. atomaria* was investigated through an original workflow (Fig. 1) including the analysis of its main chemical constituents by natural products chemistry and the building of several less and less restricted MNs coupled to DB queries and MS/MS data analysis.

2.1. Chemicals

Methanol (MeOH) and dichloromethane (CH_2Cl_2) used for the extraction of the algal material were of analytical grade. MeOH, CH_2Cl_2 and acetonitrile (ACN) used for Flash-chromatography and semi-preparative HPLC experiments were of HPLC grade. ACN and MeOH used for UHPLC-MS experiments were of LC-MS grade. All these solvents were purchased from VWR (Fontenay-sous-Bois, France). Milli-Q water was generated by a Millipore ultrapure water system (Waters-Millipore Corporation, Milford, MA, USA). Formic acid of LC-MS grade (99%) used for UHPLC analysis and deuterated chloroform and methanol (99,8%; $CDCl_3$ and CD_3OD, respectively) used for NMR experiments were purchased from Sigma-Aldrich (St Quentin-Fallavier, France).

2.2. Algal material

Thalli of the brown alga *Taonia atomaria* (Woodward) J. Agardh (family Dictyotaceae, class Phaeophyceae, phylum Ochrophyta) were collected by hand (0.5m depth) on the south eastern French Mediterranean coast (Tamaris; N 43°5'35.56, E 5°54'31.81) in February 2017. Algal samples were placed in plastic bags filled with surrounding seawater and transported within an hour to the laboratory in a cool box maintaining the seawater temperature.

2.3. Isolation of chemical standards from *T. atomaria*

2.3.1. Extraction of the algal material and fractionation of the crude extracts by Flash-chromatography

The collected algal samples were air-dried for 48h in the dark at room temperature (22 ± 3°C), crushed and weighed. The dried algae (140g) were extracted using a mixture of MeOH/CH_2Cl_2 (1:1, v/v; 3 × 500 mL). The extracts were combined, filtered (filter paper), concentrated *in vacuo*, and weighted (25g; extraction yield: 17.9%, w/w).

A part of the resulting crude extracts (4.7 g) was submitted to fractionation by flash chromatography (Spot Flash system; Armen Instruments, Saint Ave, France) on a reversed-phase column (SuperVarioFlash D40-RP18 model, 40-63 μm, 84g; Merck, Darmstadt, Germany). For fractionation, the extracts were dissolved in 25 mL MeOH and mixed with 15g of reversed-phase silica gel (Sepra C18e, 50 μm; Phenomenex, Le Pecq, France), dried under reduced pressure and loaded on a guard column on the top of the column. The fractionation was monitored by UV detection at 205 nm. For the flow rate, a linear ramp was used from 2 to 32 mL/min during 2 min and then this final flow rate was maintained throughout the entire experiment. The separation process was carried out using a ternary eluent system: H_2O (solvent A), MeOH (solvent B) and CH_2Cl_2 (solvent C). After an initial isocratic step with 90% of A and 10% of B from 0 to 3 min, a linear gradient up to 100% of B from 3 to 13 min, an isocratic step with 100% of B from 13 to 18 min, and a linear gradient up to 100% of C from 18 to

29 min, an isocratic step with 100% of C was finally used from 29 to 40 min. This process led to 39 fractions which were dried *in vacuo*, weighted and stored at -20°C (more details are given in Supporting Information, Table S1).

2.3.2. Purification and structural characterization of chemical standards from T. atomaria

For sixteen of the 39 fractions obtained by flash chromatography, selected according to available quantities, an aliquot (5 to 10 mg) was solubilized in the appropriate deuterated solvent ($CDCl_3$ or CD_3OD) and analyzed by ^1H NMR (400 MHz Avance NMR; Bruker BioSpin, Wissembourg, France) at room temperature.

Six fractions were selected based on the available quantities (m > 150 mg) and the presence of NMR signals due to algal compounds other than common fatty acid derivatives. These fractions were then fractionated using a semi-preparative HPLC system (Prostar 210; Varian) equipped with a refractive index detector (Varian Prostar Model 350 RI) and a 50 μL injection loop. Purifications were done by iterative injections using various isocratic eluent conditions (mixtures of H_2O/ACN) at a flow rate of 3 mL/min on a reversed-phase C18 semi-preparative column (Purospher Star RP-18e, 5 μm, 10×250 mm; Merck) maintained at room temperature. This purification step yielded sub-fractions which were checked by ^1H NMR and allowed the obtention of eight pure compounds.

The chemical structure of these pure compounds was determined through extensive 1D and 2D-NMR analysis (^{13}C, DEPT experiments, ^1H-^1H COSY, ^1H-^1H NOESY, HSQC and HMBC) and by comparison with data previously reported in the literature. The chemical shifts of the different spectra were fixed with respect to the residual signals of the solvent used: δ_H 7.26 and δ_C 77.16 for $CDCl_3$ and δ_H 3.31 and δ_C 49.00 for CD_3OD. All the purified compounds were analyzed by UHPLC-ESI-QToF-MS/MS and these data [accurate mass, fragmentation pattern and retention time (RT)] were implemented in our in-house DB.

2.4. Annotation of the metabolome of *T. atomaria* using molecular networking (MN)

2.4.1. UHPLC-ESI-QToF-MS/MS analyses

For this analysis, the sixteen fractions selected in Part 2.3.2. were solubilized in MeOH (5 mg/mL) and injected on a Dionex Ultimate 3000 Rapid Separation chromatographic system (Thermo Fisher Scientific, Sunnyvale, CA, USA) coupled via an ESI interface to a QToF Impact II mass spectrometer (Bruker Daltonics, Champs-sur-Marne, France). Chromatographic separations were performed on an analytical core-shell reversed-phase column (150 × 2.1 mm, 1.7 μm, Kinetex Phenyl-Hexyl; Phenomenex) with a flow rate fixed at 0.5 mL/min. The column was maintained at 40°C, the injected sample volume was 5 μL and the autosampler temperature was kept at 4°C. An elution gradient involving a binary programming of acidified H_2O and ACN (0.1% of formic acid, v/v), was applied. The elution gradient started with 5% ACN and was kept for 2 min, then reached 100% ACN (linear ramp) in 8 min and was kept for 4 min; then came back to 5% ACN over 0.01 min and was maintained 1.99 min, for a total run time of 16 min (See Supporting information Fig. S1 for the chromatograms of these sixteen fractions). The MS acquisition was performed in positive ionization and full scan range (*m/z* 50 to 1200 at a frequency of 2 Hz) modes. The following MS conditions were used: nebulizing gas (N_2) pressure: 0.4 bars, drying gas (N_2) flow: 4 L/min, drying temperature: 180°C, capillary voltage: 4.5 kV. The mass spectrometer was calibrated with a solution of formate/acetate forming clusters at the beginning of the sequence run and the same solution was automatically injected before each sample for internal mass calibration. MS/MS acquisition experiments were conducted by collision induced dissociation (CID) on each sample. Three major precursor ions were used and, after the acquisition of three MS/MS spectra per precursor ion, the corresponding *m/z* value was excluded from the precursor ion selection for 1 min.

2.4.2. MS/MS molecular networking using GNPS

Raw UHPLC-MS/MS data were converted into ".mzXML" files using DataAnalysis software (version 4.3; Bruker Daltonics). The files corresponding to the sixteen fractions previously selected (See Part 2.3.2.) were grouped into six folders corresponding to six polarity ranges. The Global Natural Products Social Molecular Networking online workflow (GNPS; http://gnps.ucsd.edu) was used to create MNs of related compounds based on the similarity of their MS/MS fragmentation patterns [14]. Three MNs (MN-1 to MN-3, from the most to the least restrictied MN) were then generated using various values of "*Minimum cosine score* (CS)" (a similarity score between two MS/MS spectra allowing the creation of a link between them), "*Minimum matched fragment ions* (MF)" (the minimum number of common fragment ions that are shared by two MS/MS spectra in order to be connected by an edge in the MN) and "*Network TopK*" (the maximum number of neighbor nodes for a single node) (Table 1). Data from GNPS were imported into Cytoscape (version 3.7.0) using nodes to represent each spectrum (each *m/z* feature) and edges to represent the similarity of MS/MS fragmentation between two connected nodes [14]. A specific color was attributed to each node according to the chemical family of the annotated metabolite or to the corresponding RT. Width of edges connecting each node was proportional to the spectral similarity using CS as a default parameter.

2.4.3. Metabolite annotation in the MNs

The strategy was to focus on clusters with the highest number of nodes. In such clusters, each node could be tentatively annotated by one or several of the following possibilities: (1) during the computational process dedicated to the construction of MNs on the GNPS platform, fragment similarity searches were performed in several public DBs (e.g. MassBank, HMDB, MoNA, Metlin…) for each *m/z* feature and the results were directly available into the MNs subsequently generated by Cytoscape, (2) MS data (accurate mass and fragmentation pattern) of a specific metabolite can also be compared with those found in other DBs (PubChem, ChemSpider…) using various tools [30], (3) MS data and RT of the detected *m/z* features were

compared with those of pure compounds isolated from *T. atomaria* (See Part 2.3.2.), and (4) in the case of absence of a hit in a DB, the MS data of all the nodes in a same cluster were carefully analyzed to highlight specific common fragments or mass losses which could allow the identification of a specific chemical family in relation to the data available in the literature or through mass fragmentation pattern analysis.

In all these cases, once one *m/z* feature was annotated in a cluster, the main part of the corresponding cluster could be annotated from this seed node according to similarity of MS/MS spectra and to accurate masses differences between nodes.

3. Results and discussion

3.1. Isolation of chemical standards from *T. atomaria*: a "phytochemical approach"

Previous studies related to compounds isolated from extracts of the brown alga *T. atomaria* allowed the isolation and the structural characterization of GGG [20,21,24,31] and other terpenes (mainly germacrane and cadinane sesquiterpenoids) [20–24,31,32].

In this study, after the fractionation by flash-chromatography of the crude extracts of *T. atomaria* and the purification by semi-preparative HPLC of the main chemical components from a selection of fractions, eight metabolites were characterized. Seven of these compounds were already described from *T. atomaria*: cadina-4(14),5-diene, cubebol, 4-*epi*-cubebol, (+)-gleenol, germacra-4(15),5,10(14)-trien-9-ol, (1*S*, 5*E*, 7*S*) 1-acetoxygermacra-4(15),5,10(14)-triene and geranylgeranylglycerol [20,23,24] (Fig. S2). An additional compound was identified as dictyol A, a diterpene previously reported from the Phaeophyceae *Dictyota dichotoma* [33] but described here for the first time from *T. atomaria* (Fig. S3-S5). As *Dictyota* spp. are morphologically and taxonomically close to *T. atomaria* and in order to avoid a possible contamination of the crude extracts studied here, the presence of this compound was confirmed by LC-MS/MS on specific crude extracts obtained within a single thallus of *T. atomaria*.

The MS data and RT of these eight compounds allowed to initiate the elaboration of an in-house DB dedicated to this algal model.

3.2. Annotation of the metabolome of *T. atomaria* using a restricted MN (MN-1)

The building of a MN (named **MN-1**) with very restrictive parameters allowed to obtain a simplified MN keeping only strong similarities between nodes. Three parameters have been specifically optimized to have a more restricted MN: (i) the minimum CS which was fixed at 0.85 (default value: 0.7) for a higher similarity between the MS/MS spectra of two connected nodes, (ii) the minimum matched fragment ions (MF) was set at 10 (default value: 6) and (iii) the Network TopK value was 7 (default value: 10). Many studies using MNs for annotation of the metabolome of microorganisms use low values of CS and MF [34,35] in order to increase the size of the clusters by inducing the clustering of less similar MS/MS spectra. This strategy allows to obtain a lot of links between nodes and less unbound nodes.

Our proposed strategy for annotation is completely opposite, the creation at first of a constrained network will give small clusters with highly correlated nodes which would allow an easier and faster determination of their characteristic MS/MS fragmentation patterns. The enlargement of the network, by using less restrictive parameters values, would connect some unbound nodes in clusters already annotated and should allow their quick annotation.

The building of a highly restricted MN (**MN-1**) showed the occurrence of ten main clusters named **1A** to **1J** which were carefully analyzed (Fig. 2).

3.2.1. Annotation of known lipid classes: Clusters 1A to 1E

During the building process of **MN-1**, the only *m/z* feature matching within a DB (MoNA) belonged to cluster **1A** and was identified as a triacylglycerol (TG). This node at *m/z* 794.7233 ($[C_{49}H_{96}NO_6]^+$, calc. *m/z* 794.7232, Δ -0.1 ppm; $[M + NH_4]^+$) was annotated as TG (14:0/16:0/16:1) (See Supporting information for MS/MS data and fragmentation pattern, Fig.

S6) [36]. Just using mass differences from this seed node (Fig. 2), thirteen other TGs were readily detected in the neighboring nodes and their annotation was confirmed after examination of their MS/MS data. RT constitutes an essential parameter when analyzing lipids. In reversed phase chromatography, retention increases with the number of carbon atoms and decreases with the number of unsaturations, in a homologous series. In that respect, a RT-based representation of **MN-1** (See supporting information, Fig. S7) confirmed that these nodes belong to the same lipid family. Interestingly, several nodes connected to TGs in the same cluster showed a lower retention: a similar MS/MS fragmentation pattern and lower molecular masses were then used to easily annotated these nodes as diacylglycerols (DGs). It's noteworthy that a TG which was connected to other TGs or DGs always bore with these nodes two common acyl chains [e.g. TG (14:0/16:0/16:1) and TG (15:0/16:0/16:1), TG (16:0/16:0/18:1) and DG (16:0/16:0) or TG (16:0/18:1/18:2) and DG (16:0/18:1)]. In cluster **1A**, a detailed analysis of the MS data of the remaining non-annotated nodes allowed to gather them in two subgroups. Some of these nodes showed a characteristic neutral loss of 261 Da which is typical of the loss of the polar sulfoquinovosyl group from the ammoniated adduct of sulfoquinovosyldiacylglycerols (SQDGs) [37]. Seven of these lipids were then identified in cluster **1A**, including the node at *m/z* 834.5396 ($[C_{43}H_{80}NO_{12}S]^+$, calc. *m/z* 834.5396, Δ 0.0 ppm; $[M + NH_4]^+$) annotated as SQDG (16:0/18:3) (See Supporting information for MS/MS data and fragmentation pattern, Fig. S6). MS data of nodes of the second subgroup were characterized by a neutral loss of *m/z* 197 which allowed to annotate them as monogalactosyldiacylglycerols (MGDGs) [37,38]. For example, the node at *m/z* 746.5770 ($[C_{41}H_{80}NO_{10}]^+$, calc. *m/z* 746.5777, Δ 0.9 ppm; $[M + NH_4]^+$) was annotated as MGDG (14:0/18:1) (See Supporting information for MS/MS data and fragmentation pattern, Fig. S6). Ultimately, all the nodes of cluster **1A** were annotated. The corresponding metabolites belonged to four lipid classes but the acyl chains of all these glycerolipids were mainly C16:0, C16:1 and C18:1. It can be noticed that MGDG (16:0/18:2)

and MGDG (16:1/18:1) corresponded to a same single node on the network because of their identical *m/z* value and their similar RT and MS fragmentation pattern.

Interestingly, seven nodes with characteristic RT of TGs were observed in cluster **1B** (See supporting information, Fig. S8) and the detailed analysis of their MS data confirmed their annotation. As reported for cluster **1A**, the same approach allowed to annotate the other nodes as DGs and MGDGs. However, the MS data of the node at *m/z* 976.5977 ($[C_{53}H_{86}NO_{15}]^+$, calc. *m/z* 976.5992, Δ 1.6 ppm; $[M + NH_4]^+$) matched with none of the chemical families previously defined in this MN. This *m/z* feature was then annotated as a digalactosyldiacylglycerol, DGDG (18:4/20:5) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S8) [37,38]. It could be pointed out that glycerolipids of cluster **1B** were composed by longer FAs (eg. C18:4, C20:4 and C20:5) than those found in metabolites of cluster **1A**.

A third cluster (**1C**) consisted of nodes, which showed a characteristic MS/MS fragment ion at *m/z* 236.1493 corresponding to the raw formula $[C_{10}H_{22}NO_5]^+$. A comparison of these MS data with those from the literature indicated that this fragment ion is typical of mono- and diacylglycerylhydroxymethyl-*N*,*N*,*N*-trimethyl-*ß*-alanines (MGTAs and DGTAs) [12]. This class of betaine lipids has already been observed in *T. atomaria* [39,40] and such glycerolipids can be differentiated from the very similar mono- and diacylglyceryl-*N*,*N*,*N*-trimethylhomoserines (MGTSs and DGTSs) via the absence of a characteristic mass loss of *m/z* 87 [41]. Fifteen MGTAs were identified in cluster **1C** using the MGTA (14:0) at *m/z* 446.3481 ($[C_{24}H_{48}NO_6]^+$, calc. *m/z* 446.3476, Δ 2.0 ppm) as a seed node. (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S9). These last fragment ions observed in all the MS/MS spectra of the nodes found in cluster **1C** were thus characteristic of MGTAs (but also of DGTAs) and allowed to group these lipids within a single cluster.

A fourth cluster (**1D**) gathered nodes among which some showed a characteristic neutral loss of 261 Da on their MS/MS spectra. Based on the annotation previously carried out for the

SQDGs and considering their MS and MS/MS data, the compounds of this cluster were annotated as sulfoquinovosylmonoacylglycerols (SQMGs). For example, the node at *m/z* 600.3411 ($[C_{27}H_{54}NO_{11}S]^+$, calc. *m/z* 600.3412, Δ 0.1 ppm; $[M + NH_4]^+$) was thus annotated as SQMG (18:1) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S10). Six other SQMGs were identified in cluster **1D** together with several structurally related lipids: three monogalactosylmonoacylglycerols (MGMGs) and a monoacylglycerol (MG). It can be noticed that in this restricted MN the SQMGs appeared in a separate cluster than the SQDGs (only found in cluster **1A**) showing that the fragment ions due to the acyl chains were in this context more important for the gathering than those of the sulfoquinovosyl part.

A further cluster (**1E**) was composed by five compounds belonging to a known lipid family. All these compounds showed in their MS/MS spectrum a characteristic fragment ion at *m/z* 184.0733 which corresponded to the head polar group of phosphatidylcholines (PCs) [42]. As an example, the node with *m/z* 542.3237 ($[C_{28}H_{49}NO_7P]^+$, calc. *m/z* 542.3241, Δ 0.8 ppm; $[M + H]^+$) was annotated as *lyso*-PC (20:5) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S11). Thus, using further fragment ions, such as acylium ions, it was possible to identify the five other nodes as four *lyso*-PCs and a PC.

3.2.2. Putative annotation of new lipid classes: Clusters 1F to 1J

*a) Geranylgeranylglycerol derivatives (Clusters **1F** and **1G**)*

Geranylgeranylglycerol (GGG), previously isolated using the "phytochemical approach", was identified in cluster **1F** using the experimental data (RT, MS and MS/MS data) obtained from the chemical standard. In this situation, this known metabolite served as a seed node to annotate the rest of this cluster as no other metabolites matched with a hit. Thanks to the MS fragmentation of the precursor adduct ion of GGG at *m/z* 382.3322 $[C_{23}H_{44}NO_3]^+$, calc. *m/z* 382.3316, Δ -1.6 ppm; $[M + NH_4]^+$), many of the resulting fragment ions observed on the

MS/MS spectrum of GGG were also found in the MS/MS spectra of the other compounds of this cluster (Fig. 3). More particularly, the characteristic fragment ion at *m/z* 273.2577 $[C_{20}H_{33}]^+$ corresponding to the geranylgeranyl (GG) moiety was observed for all nodes of the cluster **1F**. After a detailed analysis of their MS/MS data, most of the nodes of this cluster have been putatively annotated as belonging to a new class of glycerolipids, monoradylgeranylgeranylglycerols (MGGGs), which could be divided in two main groups, namely monoacyl and monoalkyl derivatives of GGG. Eight monoacyl derivatives were then identified by first using the high similarity (CS = 0.90) between the MS/MS data of GGG and those of the node at *m/z* 640.5293 ($[C_{41}H_{70}NO_4]^+$, calc. *m/z* 640.5299, Δ 1.0 ppm; $[M + NH_4]^+$) (Fig. 3). For this particular node, in addition to the characteristic ion fragment of the GG moiety at *m/z* 273.2577 ($[C_{20}H_{33}]^+$, calc. *m/z* 273.2577, Δ -0.2 ppm), a fragment ion at *m/z* 351.2534 ($[C_{21}H_{35}O_4]^+$ calc. *m/z* 351.2530, Δ -1.1 ppm; $[M - GG + H]^+$) resulting from the loss of the GG part and an acylium ion at *m/z* 259.2056 ($[C_{18}H_{27}O]^+$, calc. *m/z* 259.2056, Δ 0.0 ppm) together with its dehydrated counterpart at *m/z* 241.1950 ($[C_{18}H_{25}]^+$, calc. *m/z* 241.1951, Δ 0.4 ppm) were also observed allowing its annotation as MGGG (18:4). Seven other nodes of the cluster with acyl chains varying from C16 to C20 with various unsaturations levels were then easily identified in the same way. In the same cluster, six other nodes showed molecular formulae lacking one oxygen atom which were inconsistent with a substitution of GGG by an acyl chain but in accordance with the occurrence of an ether-linked chain. For example, in addition to characteristic fragment ions of the GG chain, MGGG (O-17:2) was characterized by an adduct ion at *m/z* 616.5652 ($[C_{40}H_{74}NO_3]^+$, calc. *m/z* 616.5663, Δ 1.9 ppm; $[M + NH_4]^+$), a fragment ion at m/z 327.2892 ($[C_{20}H_{39}O_3]^+$, calc. *m/z* 327.2894, Δ 0.4 ppm; $[M - GG + H]^+$) resulting of the loss of the GG moiety, and two fragment ions specific to the alkyl chain at *m/z* 252.2680 ($[C_{17}H_{34}N]^+$, calc. *m/z* 252.2686, Δ 2.3 ppm; $[C_{17}H_{30} + NH_4]^+$) and *m/z* 235.2418 ($[C_{17}H_{31}]^+$,

calc. *m/z* 235.2420, Δ 1.2 ppm) (Fig. 3). Five other nodes with a similar fragmentation pattern were putatively annotated as monoalkyl derivatives of GGG in cluster **1F**.

Another cluster (**1G**) was composed of *m/z* features which showed a similar fragmentation pattern than those of cluster **1F**, the main differences in their MS/MS data being a greater number of oxygen atoms in their molecular formulae and a characteristic ion fragment at *m/z* 271.2419 ($[C_{20}H_{31}]^+$, calc. *m/z* 271.2419, Δ 0.6 ppm). An in-depth analysis of the MS/MS spectra of these nodes allowed to putatively annotate them as GGG derivatives bearing an oxidized GG side chain. More particularly, the node at *m/z* 398.3273 ($[C_{23}H_{44}NO_4]^+$, calc. *m/z* 398.3265, Δ -2.1 ppm; $[M + NH_4]^+$) was characterized using the fragment ions at *m/z* 289.2529 ($[C_{20}H_{33}O]^+$, calc. *m/z* 289.2529, Δ -1.1 ppm; $[M – glycerol + H]^+$) and *m/z* 271.2427 ($[C_{20}H_{31}]^+$, calc. *m/z* 271.2420, Δ -2.5 ppm; $[M – glycerol - H_2O + H]^+$) as an analog of GGG with one hydroxyl group on the GG chain (named here GGGOH). Through their characteristic MS/MS fragmentation pattern, six other nodes of cluster **1G** were easily identified as monoacyl derivatives of GGGOH. For example, among these derivatives MGGGOH (18:4) was characterized through the adduct ion at *m/z* 656.5245 ($[C_{41}H_{70}NO_5]^+$ calc. *m/z* 656.5249, Δ 0.5 ppm; $[M + NH_4]^+$), the characteristic fragment ions at *m/z* 351.2527 ($[C_{21}H_{35}O_4]^+$, calc. *m/z* 351.2530, Δ 0.8 ppm; $[M – GGGOH + H]^+$) and *m/z* 289.2526 $[C_{20}H_{33}O]^+$, calc. *m/z* 289.2526, Δ -0.0 ppm; $[M - Glycerol – C_{18:4} + H]^+$) and their respective dehydrated counterparts at *m/z* 333.2426 ($[C_{21}H_{33}O_3]^+$ calc. *m/z* 333.2424, Δ -0.4 ppm) and *m/z* 271.2419 ($[C_{20}H_{31}]^+$, calc. *m/z* 271.2424, Δ 0.6 ppm), and the acylium ion at *m/z* 259.2057 ($[C_{18}H_{27}O]^+$, calc. *m/z* 259.2056, Δ 0.4 ppm) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S12).

*b) Farnesylglycerol derivatives (Cluster **1H**)*

Cluster (**1H**) was composed by nodes showing a characteristic ion fragment at *m/z* 205.1951 $[C_{15}H_{25}]^+$ which matched with the occurrence of a farnesyl (i.e. C_{15}-terpenoid) side chain. This

putative assumption was strengthened by the strong similarity of the MS/MS fragmentation pattern of this ion with that of farnesol found in online DBs (e.g. Metlin).

A detailed analysis of the MS/MS data of the nodes of cluster **1H** allowed to putatively annotate them as mono- and diacyl derivatives of farnesylglycerol (MFGs and DFGs, respectively). For example, the node at *m/z* 578.5146 ($[C_{36}H_{68}NO_4]^+$, calc. *m/z* 578.5143, Δ -0.5 ppm; $(M + NH_4]^+$) was characterized as MFG (18:1) through: (i) a fragment ion at *m/z* 357.3010 ($[C_{21}H_{41}O_4]^+$, calc. *m/z* 357.2999, Δ -2.8 ppm, $[M - FG + H]^+$) resulting from the loss of the farnesyl part and the corresponding dehydrated ion at *m/z* 339.2891 ($[C_{21}H_{39}O_3]^+$, calc. *m/z* 339.2894, Δ 0.9 ppm; $[M - FG – H_2O + H]^+$), (ii) the characteristic ion fragment of the farnesyl chain at *m/z* 205.1950 ($[C_{15}H_{25}]^+$, calc. *m/z* 205.1951, Δ 0.6 ppm; $[M - Glycerol - C_{18:1} + H]^+$) and (iii) an acylium ion at *m/z* 283.2632 ($[C_{18}H_{35}O_2]^+$, calc. *m/z* 283.2632, Δ 0.0 ppm) and its dehydrated counterpart at *m/z* 265.2529 ($[C_{18}H_{33}O]^+$, calc. *m/z* 265.2526, Δ -1.0 ppm) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S13). A similar MS fragmentation was also observed for five other nodes while two other nodes showed the occurrence of oxygenated acyl chains.

In the same cluster, 14 other nodes showing higher molecular masses and retention times were annotated as DFGs. For example, DFG (18:1/16:1) was putatively identified on the basis of the adduct ion at *m/z* 814.7280 ($[C_{52}H_{96}NO_5]^+$, calc. *m/z* 814.7283, Δ 0.3 ppm; $[M + NH_4]^+$) and the characteristic fragment ions at *m/z* 575.5034 ($[C_{37}H_{67}O_4]^+$, calc. *m/z* 575.5034, Δ 0.0 ppm; $[M - FG – H_2O + H]^+$), *m/z* 339.2889 ($[C_{21}H_{39}O_3]^+$, calc. *m/z* 339.2894, Δ 1.4 ppm; $[M - FG - H_2O - C_{16:1} +H]^+$) and *m/z* 311.2579 ($[C_{19}H_{35}O_3]^+$, calc. *m/z* 311.2581, Δ 0.5 ppm; $[M - FG - H_2O - C_{18:1} + H]^+$) due to the respective losses of the farnesyl chain, a molecule of water and one acyl chain.

*c) Fulvellic acid derivatives (Clusters **1I** and **1J**)*

Two other clusters (**1I** and **1J**) were composed by nodes which showed molecular formulae and MS/MS fragmentation pattern, more particularly the characteristic fragment ion at *m/z* 85.0284 $[C_4H_5O_2]^+$, corresponding to those of mono- and diacylglycerols substituted by a $C_4H_5O_2$ additional chemical group. These findings were in accordance with the chemical structure of fulvellic acid derivatives (i.e. acylglycerols bearing a methacrylic acid moiety) which have been previously described from Japanese samples of the brown alga *Sargassum fulvellum* [43]. Thus, we proposed to putatively annotate these compounds as fulvellic acid derivatives even if the lack of MS/MS data in the literature did not allow their unambiguous characterization.

In more detail, the seven nodes of cluster **1I** were described as monoacyl fulvellic acid derivatives (MFuAs) based on their MS/MS fragmentation. For example, MFuA (20:4) was characterized through the adduct ion at *m/z* 480.3329 ($[C_{27}H_{46}NO_6]^+$, calc. *m/z* 480.3320, Δ -1.8 ppm; $[M + NH_4]^+$) and the following ion fragments found on its MS/MS spectrum such as: (i) the ion at *m/z* 445.2963 ($[C_{27}H_{41}O_5]^+$, calc. *m/z* 445.2949, Δ -3.2 ppm, $[M - H_2O + H]^+$) resulting of a loss of water, (ii) the ion at *m/z* 361.2746 ($[C_{23}H_{37}O_3]^+$, calc. *m/z* 361.2737, Δ -2.4 ppm; $[M - H_2O - C_4H_5O_2 + H]^+$) due to the subsequent loss of the methacrylic acid part, (iii) the acylium ion at *m/z* 287.2376 ($[C_{20}H_{31}O]^+$, calc. *m/z* 287.2369, Δ -2.4 ppm) and (iv) the characteristic fragment ion of the methacrylic acid moiety at *m/z* 85.0288 ($[C_4H_5O_2]^+$, calc. *m/z* 85.0284, Δ -4.6 ppm) (Fig. 4).

Concerning the six nodes of cluster **1J**, their RT, molecular formulae and MS/MS data were in accordance with their putative annotation as diacyl fulvellic acid derivatives (DFuAs). As an example, DFuA (20:4/20:5) was characterized on the basis of the adduct ion at *m/z* 764.5455 ($[C_{47}H_{74}NO_7]^+$, calc. *m/z* 764.5460, Δ 0.7 ppm; $[M + NH_4]^+$), the fragment ions characteristic of the presence of the fulvellic acid part at *m/z* 645.4867 ($[C_{43}H_{65}O_4]^+$, calc. *m/z* 645.4877, Δ 1.5 ppm, $[M - H_2O - C_4H_5O_2 + H]^+$) and at *m/z* 85.0282 ($[C_4H_5O_2]^+$, calc. *m/z* 85.0284,

Δ 2.6 ppm), the fragment ions due to the loss of an acyl chain at *m/z* 445.2943 ($[C_{27}H_{41}O_5]^+$, calc. *m/z* 445.2949, Δ 1.1 ppm, $[M - H_2O - C_{20:5} + H]^+$) and *m/z* 443.2788 ($[C_{27}H_{39}O_5]^+$, calc. *m/z* 443.2792, Δ 0.9 ppm; $[M - H_2O - C_{20:4} + H]^+$), and the acylium ions at *m/z* 287.2368 ($[C_{20}H_{31}O]^+$, calc. *m/z* 287.2369, Δ 0.5 ppm) and *m/z* 285.2214 ($[C_{20}H_{29}O]^+$, calc. *m/z* 285.2213, Δ -0.2 ppm) (See Supporting information for MS/MS data and proposed fragmentation pattern, Fig. S14).

*3.2.3. Assessment of the annotation carried out using the restricted **MN-1***

This MN approach allowed to quickly identify chemically related compounds gathered in a same cluster because these metabolites showed common MS fragment ions and/or similar MS/MS fragmentation patterns. Indeed, as soon as a metabolite was annotated (seed node), the chances of annotating the whole cluster were very strong. In the case of **MN-1**, nearly 90% of the metabolites belonging to a cluster where a seed node has been identified were annotated. Nevertheless, it should be pointed out that to isobaric molecules [e.g. DG (16:1/18:1) & DG (16:0/18:2)] constituted one limitation of this type of approach as they are commonly detected in a same node.

Finally, the careful analysis of **MN-1** led to the annotation of 147 metabolites distributed in eighteen chemical classes (Fig. 2). Even if the nodes belonging to the major clusters were annotated, certain compounds and/or clusters are still to be determined. Then, the use of less restricted MNs would allow to include them within a cluster where most of the nodes were already annotated, which could facilitate their identification.

3.3. Further annotation of the metabolome of *T. atomaria* using less restricted MNs (MN-2 and MN-3)

The previous MN (**MN-1**) was based on restrictive parameters that only permitted to highlight strong similarities between the MS/MS spectra of metabolites within a same cluster. In order to

further enrich the chemical DB of *T. atomaria*, the metabolites previously identified by the phytochemical approach and the analysis of **MN-1** were incorporated into a less restricted MN (MN-2; See Supporting information , Fig. S15), and then an even less restricted (**MN-3**) network (Fig. 5). Reducing this restriction allowed certain unbound metabolites (self-loop nodes) to be gathered in clusters containing known compounds. Their annotation became easier by highlighting fragment ions and MS/MS fragmentation pathways which were common with previously identified nodes.

Moreover, the use of less restricted MNs allowed the gathering of nodes belonging to close chemical classes in a same cluster. Thus, MN enlargement with **MN-2** and **MN-3** brought together two clusters (**1A** and **1B**), which were separated in **MN-1**, in a single cluster (**2A**-**B** and **3A**-**B,** respectively) gathering several classes of structurally related lipids (DGs, TGs, MGDGs and SQDGs). In the same way, the two clusters **1F** and **1G** comprising various GGG derivatives grouped together (**3F-G**). Moreover, several terpenes, some of which have been identified in the phytochemical approach, namely dictyol A, (1*S*, 5*E*, 7*S*) 1-acetoxygermacra-4(15),5,10(14)-triene and germacra-4(15),5,10(14)-trien-9-ol, which appeared previously in **MN-1** as self-loop nodes, were found clustered with the MGGGs in **MN-2** (cluster **2G**) and in **MN-3** (cluster **3F-G**). These clusters also allowed to confirm the annotation of other terpenes, such as retinol, retinal, and dictyolene identified through online DBs and *trans*-calamenene and 1,4-peroxymuurol-5-ene already isolated by our group from this algal species and referenced in our in-house DB, which appeared as self-loop nodes in **MN-1**. This strategy made it possible to annotate 61 additional metabolites very easily and to further enrich the DB. Finally, a total of 210 metabolites (Table 2) were putatively identified within the metabolome of *T. atomaria* using the workflow described in this study.

3.4. Taxonomical and ecological significance of the annotated metabolites of *T. atomaria*

In line with previous observations on *T. atomaria* from the French Mediterranean coasts [20,21,39,40], no meroditerpenes were observed here either with the phytochemical approach or using MNs. In these previous studies and in the present one, the main isolated terpenes were GGG and sesquiterpenes with cadinane, germacrane or spiroaxane carbon skeletons. The same trend was also observed for specimens of *T. atomaria* collected in the Adriatic Sea (Croatia) [22,23], or for samples of *Taonia pseudociliata* from the Gulf of Catania (Sicily, Italy) (previously identified as *Dilophus fasciola*, [24,31,32]) and no meroditerpenes were reported either in these studies.

However, several meroditerpenes were identified in previous works focusing on algal species reported as *Taonia atomaria* from Canary Islands (e.g. taondiol and atomaric acid) [27,28], or from the Aegean Sea (e.g. taondiol, atomaric acid, sargaquinone, and atomarianones A and B) [26,29,44]. In these studies, which did not report any sesquiterpenes or GGG, such a chemical composition appeared to be strikingly close to those of *Stypopodium* species widely reported for their meroditerpenes production [45–49]. For instance, atomaric acid and taondiol have been previously isolated from *Stypopodium zonale* [50] while *epi*-taondiol and sargaquinone were described from *Stypopodium flabelliforme* [51]. *Stypopodium* and *Taonia* are morphologically and phylogenetically close genera among the Dictyotaceae family [52] and, as already suggested [50], the hypothesis of a botanical misidentification of *Stypopodium* spp. samples as *T. atomaria* could be a plausible scenario to consider. However, more thorough chemotaxonomic studies are necessary to confirm this hypothesis and an integrative approach coupling morphological, phylogenetic and metabolomics analyses could provide the information needed to clarify this situation.

In an ecological point of view, sesquiterpenes and GGG produced by *T. atomaria* have demonstrated anti-adhesion activities against several epiphytes including diverse bacterial

strains [20,21] and barnacle larvae [21], with gleenol and GGG showing the strongest activities. Through a specific extraction procedure developed for this seaweed [20], these two compounds were also characterized as main components of the surface metabolome of *T. atomaria* [20,21,39,40]. These studies brought evidence that these chemical compounds may be implied in the regulation of the epibiosis at the algal surface. Sesquiterpenes have been noticed mainly in apical parts, where the meristem -which is composed of cells involved in the algal growth- is located. In line with the optimal defense theory which assumes that chemical resources are allocated within the organism to maximize the plant fitness [53], these molecules, known as defense compounds, may allow to protect the merismatic cells against fouling organisms [40]. The biosynthetic pathway and mechanisms allowing the expression of these compounds at the algal surface remain uncertain and not documented, implying the need to improve the annotation of its metabolome.

In this study, we also annotated two new lipid classes, GGG and FG derivatives, which are likely to play an important role in the biogenesis and storage of GGG and sesquiterpenes, respectively, since glycerides are generally known for their fundamental storage role. More precisely, GGG derivatives could act as a lipid form of storage for GGG; while mono- and di-FG, harboring a farnesyl chain, could be used by the alga as storage compounds for biosynthetic precursors of sesquiterpenes.

Concerning the fulvellic acid derivatives, their description remains uncertain because they have only been described from the Japanese alga *Sargassum fulvellum* [43], but their presence, in addition to those of all the other lipid families described here, shows that *T. atomaria,* and more broadly marine macroalgae, constitute an immeasurable source of still untapped lipid derivatives.

3.5. Integration of phytochemical approaches and the use of less restrictive parameters in the building of MNs can improve the annotation of relevant metabolites

Several studies have already successfully combined phytochemical approaches and the building of MNs (see [18] for review) but the main part of them were implemented with purposes different than the one driving the present work. Thus, a first set of such studies have employed MNs as a dereplication tool, using in-house or online DBs fed with experimental or *in silico* MS/MS data, to focus isolation efforts on putatively new compounds (e.g. [54,55]). A second set of studies using MNs have been implemented in order to target specific families of chemical compounds (e.g. [56,57]). Finally, a large third set of studies have described the use of MNs for the identification or the targeted isolation of bioactive compounds (e.g. [58,59]). In our case, starting from a non-model species, the aim was to isolate the main chemical components of the extracts of *T. atomaria* and to use them, in combination with MNs, to annotate as much as possible the metabolome of this alga without any condition of novelty or bioactivity of the characterized metabolites.

The first particularity of the present study was to include a fractionation step of the crude extracts of T. atomaria and the subsequent MS/MS analysis of the resulting fractions in order to maximize the annotation of the metabolome of this alga. By way of comparison, it can be noted that most of the previous studies dealing with MNs have only analyzed crude extracts that can hamper the detection of numerous compounds, probably due to lower detection limits and competition between molecules during the ionization process.

The second particularity of the present analytical workflow consisted in the use of less and less restrictive parameters for the construction of more and more complex MNs. As most studies to date use the default set of values proposed on the GNPS platform (CS: 0.70; Minimum matched fragments ions: 6; Network TopK: 10), only a limited number of them have discussed the impact of such parameters on the structure of the MNs and on the annotation process. However,

these parameters need to be appropriate for each analysis according to the expected results [16]. Consistently, in several previous studies the use of default parameters gives complex MNs with large clusters composed by mostly unknown metabolites. In such cases, the use of less stringent conditions for the building of MNs could have allowed an easiest determination of common MS/MS fragmentation patterns between nodes which composed a same cluster and thus could have led to a better annotation.

4. Conclusion

An efficient method dedicated to the annotation of the metabolome of a non-model organism, the brown alga *T. atomaria*, has been developed by coupling two types of approaches, one based on a classical study of the organic extracts of this alga by natural products chemistry and the other, more recent, highlighting the power of annotation of MNs built with fractions of these extracts. Following the elaboration of a first very restricted MN (i.e. implying a very high similarity between the MS/MS data of the metabolites within a same cluster), the annotation was carried out using either seed nodes (i.e. metabolite identified through the interrogation of online DBs or isolated standards during the phytochemical analysis of the alga) or by analyzing the similarities and/or interpreting the MS/MS data of the nodes belonging to a same cluster. This first step allowed to annotate 147 compounds in the ten main clusters of the restricted MN (**MN-1**) and to putatively characterize two new lipid classes including a geranylgeranyl or a farnesyl part together with a rare class of acylglycerols bearing a methacrylic acid moiety (fulvellic acid derivatives). The implementation, in a second phase, of less constrained MNs allowed to enlarge the size of the clusters and to annotate 65 additional metabolites.

The use of such a method, which is simple to implement, could allow to broaden our knowledge of the metabolome of species, particularly marine ones, which are still little studied to date. This method could thus be an essential tool for the study of such organisms by environmental metabolomics.

Notes

The authors declare no competing financial interest.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This work was financially supported by the French "Sud Provence-Alpes-Côte d'Azur (Sud PACA)" regional council (PhD grant of Benoît Paix) and the MAPIEM laboratory of the University of Toulon (Master internship grant of Nathan Carriot). LC-MS experiments were conducted on the regional platform MALLABAR funded by the Institute of Ecology and Environment (INEE) of the French National Centre for Scientific Research (CNRS), the Total Foundation and the French Sud PACA regional council. The authors are grateful to François Audrezet and Damien Devilly (MAPIEM laboratory, University of Toulon, France) for their help during the isolation and the structural characterization of the chemical standards.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://.

Data availability

Raw data for LC-ESI-(+)-MS/MS experiments were deposited and are publicly available in the MassIVE platform under the IDs MSV000082235.

References

[1] M.E. Hay, Marine chemical ecology: chemical signals and cues structure marine populations, communities, and ecosystems, Ann. Rev. Mar. Sci. 1 (2009) 193-212. https://doi.org/10.1146/annurev.marine.010908.163708.

[2] C. Kuhlisch, G. Pohnert, Metabolomics in chemical ecology, Nat. Prod. Rep. 32 (2015) 937-955. https://doi.org/10.1039/C5NP00003C.

[3] R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G.J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, Nat. Biotechnol. 30 (2012) 826-828. https://doi.org/10.1038/nbt.2348.

[4] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast genome, Trends Biotechnol. 16 (1998) 373-378. https://doi.org/10.1016/S0167-7799(98)01214-1.

[5] S.M. Dittami, D. Eveillard, T. Tonon, A metabolic approach to study algal-bacterial interactions in changing environments, Mol. Ecol. 23 (2014) 1656-1660. https://doi.org/10.1111/mec.12670.

[6] A.A. Monte, C. Brocker, D.W. Nebert, F.J. Gonzalez, D.C. Thompson, V. Vasiliou, Improved drug therapy: triangulating phenomics with genomics and metabolomics, Hum. Genomics. 8 (2014) 16. https://doi.org/10.1186/s40246-014-0016-9.

[7] G.J. Patti, R. Tautenhahn, G. Siuzdak, Meta-analysis of untargeted metabolomic data from multiple profiling experiments, Nat. Protoc. 7 (2012) 508-516. https://doi.org/10.1038/nprot.2011.454.

[8] R. Chaleckis, I. Meister, P. Zhang, C.E. Wheelock, Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics, Curr. Opin. Biotech. 55 (2019) 44-50. https://doi.org/10.1016/j.copbio.2018.07.010.

[9] A. Vaniya, O. Fiehn, Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics, TRAC-Trends Anal. Chem. 69 (2015) 52-61. https://doi.org/10.1016/j.trac.2015.04.002.

[10] R.R. da Silva, P.C. Dorrestein, R.A. Quinn, Illuminating the dark matter in metabolomics, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 12549-12550. https://doi.org/10.1073/pnas.1516878112.

[11] K. Scheubert, F. Hufsky, D. Petras, M. Wang, L.-F. Nothias, K. Dührkop, N. Bandeira, P.C. Dorrestein, S. Böcker, Significance estimation for large scale metabolomics annotations by spectral matching, Nat. Commun. 8 (2017) 1-10 . https://doi.org/10.1038/s41467-017-01318-5.

[12] T. Kind, J.K. Meissen, D. Yang, F. Nocito, A. Vaniya, Y.-S. Cheng, J.S. VanderGheynst, O. Fiehn, Qualitative analysis of algal secretions with multiple mass spectrometric platforms, J. Chromatogr. A. 1244 (2012) 139-147. https://doi.org/10.1016/j.chroma.2012.04.074.

[13] J.Y. Yang, L.M. Sanchez, C.M. Rath, X. Liu, P.D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. de Felicio, A. Fenner, W. Ruh Wong, R.G. Linington, L. Zhang, H.M. Debonsi, W.H. Gerwick, P.C. Dorrestein, Molecular networking as a dereplication strategy, J. Nat. Prod. 76 (2013) 1686-1699. https://doi.org/10.1021/np400413s.

[14] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.-T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.-C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A. Boya P, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.Ø. Palsson, K. Pogliano, R.G. Linington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social molecular networking, Nat. Biotechnol. 34 (2016) 828-837. https://doi.org/10.1038/nbt.3597.

[15] L.-F. Nothias, M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A. Sarvepalli, P. Leyssen, D. Touboul, J. Costa, J. Paolini, T. Alexandrov, M. Litaudon, P.C. Dorrestein, Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation, J. Nat. Prod. 81 (2018) 758-767. https://doi.org/10.1021/acs.jnatprod.7b00737.

[16] A.T. Aron, E.C. Gentry, K.L. McPhail, L.-F. Nothias, M. Nothias-Esposito, A. Bouslimani, D. Petras, J.M. Gauglitz, N. Sikora, F. Vargas, J.J.J. van der Hooft, M. Ernst, K.B. Kang, C.M. Aceves, A.M. Caraballo-Rodríguez, I. Koester, K.C. Weldon, S. Bertrand, C. Roullier, K. Sun, R.M. Tehan, C.A.B. P, M.H. Christian, M. Gutiérrez, A.M. Ulloa, J.A. Tejeda Mora, R. Mojica-Flores, J. Lakey-Beitia, V. Vásquez-Chaves, Y. Zhang, A.I. Calderón, N. Tayler, R.A. Keyzers, F. Tugizimana, N. Ndlovu, A.A. Aksenov, A.K. Jarmusch, R. Schmid, A.W. Truman, N. Bandeira, M. Wang, P.C. Dorrestein, Reproducible molecular networking of untargeted mass spectrometry data using GNPS, Nat. Protoc. 15 (2020) 1954-1991. https://doi.org/10.1038/s41596-020-0317-5.

[17] K.R. Duncan, M. Crüsemann, A. Lechner, A. Sarkar, J. Li, N. Ziemert, M. Wang, N. Bandeira, B.S. Moore, P.C. Dorrestein, P.R. Jensen, Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species, Chem. Biol. 22 (2015) 460-471. https://doi.org/10.1016/j.chembiol.2015.03.010.

[18] A.E.F. Ramos, L. Evanno, E. Poupon, P. Champy, M.A. Beniddir, Natural products targeting strategies involving molecular networking: different manners, one goal, Nat. Prod. Rep. 36 (2019) 960-980. https://doi.org/10.1039/C9NP00006B.

[19] C.M. Grim, G.T. Luu, L.M. Sanchez, Staring into the void: demystifying microbial metabolomics, FEMS Microbiol. Lett. 366 (2019) fnz135. https://doi.org/10.1093/femsle/fnz135.

[20] A. Othmani, J.-F. Briand, M. Ayé, M. Molmeret, G. Culioli, Surface metabolites of the brown alga *Taonia atomaria* have the ability to regulate epibiosis, Biofouling. 32 (2016) 801-813. https://doi.org/10.1080/08927014.2016.1198954.

[21] A. Othmani, R. Bunet, J.-L. Bonnefont, J.-F. Briand, G. Culioli, Settlement inhibition of marine biofilm bacteria and barnacle larvae by compounds isolated from the Mediterranean brown alga *Taonia atomaria*, J. Appl. Phycol. 28 (2016) 1975-1986. https://doi.org/10.1007/s10811-015-0668-4.

[22] I. Jerković, M. Kranjac, Z. Marijanović, M. Roje, S. Jokić, Chemical diversity of headspace and volatile oil composition of two brown algae (*Taonia atomaria* and *Padina pavonica*) from the Adriatic Sea, Molecules. 24 (2019) 495. https://doi.org/10.3390/molecules24030495.

[23] S. De Rosa, A. De Giulio, C. Iodice, N. Zavodink, Sesquiterpenes from the brown alga *Taonia atomaria*, Phytochemistry. 37 (1994) 1327-1330. https://doi.org/10.1016/S0031-9422(00)90407-2.

[24] C. Tringali, M. Piattelli, C. Spatafora, Sesquiterpenes and geranylgeranylglycerol from the brown algae *Taonia lacheana* and *Taonia atomaria* f. *ciliata*: their chemotaxonomic significance, Phytochemistry. 40 (1995) 827–831. https://doi.org/10.1016/0031-9422(95)00357-D.

[25] M. Pellegrini, R. Valls, L. Pellegrini, Chimiotaxonomie et marqueurs chimiques dans les algues brunes, Lagascalia 19 (1997) 145-164.

[26] L.-A. Tziveleka, D. Abatis, K. Paulus, R. Bauer, C. Vagias, V. Roussis, Marine polyprenylated hydroquinones, quinones, and chromenols with inhibitory effects on leukotriene formation, Chem. Biodivers. 2 (2005) 901-909. https://doi.org/10.1002/cbdv.200590066.

[27] A.G. González, J. Darias, J.D. Martin, C. Pascual, Marine natural products of the atlantic zone-V: The structure and chemistry of taondiol and related compounds, Tetrahedron. 29 (1973) 1605-1609. https://doi.org/10.1016/S0040-4020(01)83404-9.

[28] A.G. González, J. Darias, J.D. Martín, M. Norte, Atomaric acid, a new component from *Taonia atomaria*, Tetrahedron Letters. 15 (1974) 3951-3954. https://doi.org/10.1016/S0040-4039(01)92054-4.

[29] D. Abatis, C. Vagias, D. Galanakis, J.N. Norris, D. Moreau, C. Roussakis, V. Roussis, Atomarianones A and B: two cytotoxic meroditerpenes from the brown alga *Taonia atomaria*, Tetrahedron Letters. 46 (2005) 8525-8529. https://doi.org/10.1016/j.tetlet.2005.10.007.

[30] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: A web-based platform to process untargeted metabolomic data, Anal. Chem. 84 (2012) 5035-5039. https://doi.org/10.1021/ac300698c.

[31] V. Amico, G. Oriente, M. Piattelli, C. Tringali, E. Fattorusso, S. Magno, L. Mayol, (-)-(*R*)-1-*O*-Geranylgeranylglycerol from the brown alga *Dilophus fasciola*, Experientia. 33 (1977) 989-990. https://doi.org/10.1007/BF01945922.

[32] V. Amico, G. Oriente, M. Piattelli, C. Tringali, E. Fattorusso, S. Magno, L. Mayol, Sesquiterpenes based on the cadalane skeleton from the brown alga *Dilophus fasciola*, Experientia. 35 (1979) 450-451. https://doi.org/10.1007/BF01922698.

[33] E. Fattorusso, S. Magno, L. Mayol, C. Santacroce, D. Sica, V. Amico, G. Oriente, M. Piattelli, C. Tringali, Dictyol A and B, two novel diterpene alcohols from the brown alga *Dictyota dichotoma*, J. Chem. Soc., Chem. Commun. (1976) 575-576. https://doi.org/10.1039/C39760000575.

[34] M. Crüsemann, E.C. O'Neill, C.B. Larson, A.V. Melnik, D.J. Floros, R.R. da Silva, P.R. Jensen, P.C. Dorrestein, B.S. Moore, Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols, J. Nat. Prod. 80 (2017) 588-597. https://doi.org/10.1021/acs.jnatprod.6b00722.

[35] K. Saurav, M. Macho, A. Kust, K. Delawská, J. Hájek, P. Hrouzek, Antimicrobial activity and bioactive profiling of heterocytous cyanobacterial strains using MS/MS-based molecular networking, Folia Microbiol. 64 (2019) 645-654. https://doi.org/10.1007/s12223-019-00737-9.

[36] V. Samburova, M.S. Lemos, S. Hiibel, S. Kent Hoekman, J.C. Cushman, B. Zielinska, Analysis of triacylglycerols and free fatty acids in algae using ultra-performance liquid chromatography mass spectrometry, J. Am. Oil. Chem. Soc. 90 (2013) 53-64. https://doi.org/10.1007/s11746-012-2138-3.

[37] K.J. Popendorf, H.F. Fredricks, B.A.S.V. Mooy, Molecular ion-independent quantification of polar glycerolipid classes in marine plankton using triple quadrupole MS, Lipids. 48 (2013) 185-195. https://doi.org/10.1007/s11745-012-3748-0.

[38] G. Isaac, R. Jeannotte, S.W. Esch, R. Welti, New mass-spectrometry-based strategies for lipids, in: J.K. Setlow (Eds), Genetic Engineering, vol 28, Springer, Boston, 2007, pp. 129-157. https://doi.org/10.1007/978-0-387-34504-8_8.

[39] B. Paix, A. Othmani, D. Debroas, G. Culioli, J.-F. Briand, Temporal covariation of epibacterial community and surface metabolome in the Mediterranean seaweed holobiont *Taonia atomaria*, Environ. Microbiol. 21 (2019) 3346-3363. https://doi.org/10.1111/1462-2920.14617.

[40] B. Paix, N. Carriot, R. Barry-Martinet, S. Greff, B. Misson, J.-F. Briand, G. Culioli, A multi-omics analysis suggests links between the differentiated surface metabolome and epiphytic microbiota along the thallus of a Mediterranean seaweed holobiont, Front. Microbiol. 11 (2020) 494. https://doi.org/10.3389/fmicb.2020.00494.

[41] S.A. Roche, J.D. Leblond, Betaine lipids in chlorarachniophytes, Phycol. Res. 58 (2010) 298-305. https://doi.org/10.1111/j.1440-1835.2010.00590.x.

[42] S. Suárez-García, L. Arola, A. Pascual-Serrano, A. Arola-Arnal, G. Aragonès, C. Bladé, M. Suárez, Development and validation of a UHPLC-ESI-MS/MS method for the simultaneous quantification of mammal lysophosphatidylcholines and lysophosphatidylethanolamines in serum, J. Chromatogr. B. 1055-1056 (2017) 86-97. https://doi.org/10.1016/j.jchromb.2017.04.028.

[43] T. Kusumi, M. Ishitsuka, T. Iwashita, H. Naoki, T. Konno, H. Kakisawa, A novel type of glycerides bearing a methacrylic acid moiety from the brown alga, *Sargassum fulvellum*, Chem. Lett. 10 (1981) 1393-1396. https://doi.org/10.1246/cl.1981.1393.

[44] R. Nahas, D. Abatis, M.A. Anagnostopoulou, P. Kefalas, C. Vagias, V. Roussis, Radical-scavenging activity of Aegean Sea marine algae, Food Chem. 102 (2007) 577-581. https://doi.org/10.1016/j.foodchem.2006.05.036.

[45] W.H. Gerwick, W. Fenical, Ichthyotoxic and cytotoxic metabolites of the tropical brown alga *Stypopodium zonale* (Lamouroux) Papenfuss, J. Org. Chem. 46 (1981) 22-27. https://doi.org/10.1021/jo00314a005.

[46] J. Rovirosa, M. Sepulveda, E. Quezada, A. San-Martin, Isoepitaondiol, a diterpenoid of *Stypopodium flabelliforme* and the insecticidal activity of stypotriol, epitaondiol and derivatives, Phytochemistry. 31 (1992) 2679-2681. https://doi.org/10.1016/0031-9422(92)83610-B.

[47] M. Wessels, G.M. König, A.D. Wright, A new tyrosine kinase inhibitor from the marine brown alga *Stypopodium zonale*, J. Nat. Prod. 62 (1999) 927-930. https://doi.org/10.1021/np990010h.

[48] B. Gil, M.L. Ferrándiz, M.J. Sanz, M.C. Terencio, A. Ubeda, J. Rovirosa, A. San-Martin, M.J. Alcaraz, M. Payá, Inhibition of inflammatory responses by epitaondiol and other marine natural products, Life Sci. 57 (1995) 25-30. https://doi.org/10.1016/0024-3205(95)00260-d.

[49] O.M.M. Sabry, S. Andrews, K.L. McPhail, D.E. Goeger, A. Yokochi, K.T. LePage, T.F. Murray, W.H. Gerwick, Neurotoxic meroditerpenoids from the tropical marine brown alga *Stypopodium flabelliforme*, J. Nat. Prod. 68 (2005) 1022-1030. https://doi.org/10.1021/np050051f.

[50] A.R. Soares, V.L. Teixeira, R. Pereira, R.C. Villaça, Variation on diterpene production by the Brazilian alga *Stypopodium zonale* (Dictyotales, Phaeophyta), Biochem. Syst. Ecol. 31 (2003) 1347-1350. https://doi.org/10.1016/S0305-1978(03)00116-9.

[51] C. Areche, A. San-Martín, J. Rovirosa, M.A. Muñoz, A. Hernández-Barragán, M.A. Bucio, P. Joseph-Nathan, Stereostructure reassignment and absolute configuration of isoepitaondiol, a meroditerpenoid from *Stypopodium flabelliforme*, J. Nat. Prod. 73 (2010) 79-82. https://doi.org/10.1021/np900553p.

[52] J.C.D. Paula, M.A. Vallim, V.L. Teixeira, What are and where are the bioactive terpenoids metabolites from Dictyotaceae (Phaeophyceae), Rev. Bras. Farmacogn. 21 (2011) 216-228. https://doi.org/10.1590/S0102-695X2011005000079.

[53] D. McKey, The distribution of secondary compounds within plants, in: G.A. Rosenthal, D.H. Janzen, S.W.Applebaum (Eds.), Herbivores-their interaction with secondary plant metabolites, Academic Press, New York, 1979, pp. 55-134. https://ci.nii.ac.jp/naid/10012096248/.

[54] C.F. Alcover, G. Bernadat, F.A. Kabran, P. Le Pogam, K. Leblanc, A.E. Fox Ramos, J.-F. Gallard, E. Mouray, P. Grellier, E. Poupon, M.A. Beniddir, Molecular networking reveals serpentinine-related bisindole alkaloids from *Picralima nitida*, a previously well-investigated species, J. Nat. Prod. 83 (2020) 1207−1216.

[55] Y.-M. Ren, S.-Z. Zhou, T. Zhang, M. Qian, R. Zhang, S. Yao, H. Zhu, C. Tang, L. Lin, Y. Ye, Targeted isolation of two disesquiterpenoid macrocephadiolides A and B from *Ainsliaea macrocephala* using a molecular networking-based dereplication strategy. Org. Chem. Front. 7 (2020) 1481-1489. https://doi.org/10.1039/d0qo00030b.

[56] C.Y.G. Ding, L.M. Pang, Z.-X. Liang, K.K.K. Goh, E. Glukhov, W.H. Gerwick, L.T. Tan, MS/MS-based molecular networking approach for the detection of aplysiatoxin-related compounds in environmental marine cyanobacteria, Mar Drugs. 16 (2018) 505. https://doi.org/10.3390/md16120505.

[57] Y. Li, N. He, M. Luo, B. Hong, Y. Xie, Application of untargeted tandem mass spectrometry with molecular networking for detection of enniatins and beauvericins from complex samples, Journal of Chromatography A. 1634 (2020) 461626. https://doi.org/10.1016/j.chroma.2020.461626.

[58] L. Buedenbender, F.A. Astone, D. Tasdemir, Bioactive molecular networking for mapping the antimicrobial constituents of the baltic brown alga *Fucus vesiculosus*, Marine Drugs. 18 (2020) 311. https://doi.org/10.3390/md18060311.

[59] F. Olivon, P.-M. Allard, A. Koval, D. Righi, G. Genta-Jouve, J. Neyts, C. Apel, C. Pannecouque, L.-F. Nothias, X. Cachet, L. Marcourt, F. Roussi, V.L. Katanaev, D. Touboul, J.-L. Wolfender, M. Litaudon, Bioactive natural products prioritization using massive multi-informational molecular networks, ACS Chem. Biol. 12 (2017) 2644–2651. https://doi.org/10.1021/acschembio.7b00413.

Tables and Figures

Table 1. Selected building parameters and results for the MNs constructed with the MS/MS data of fractions obtained from the crude extracts of the brown alga *Taonia atomaria*

Table 2: Summary of the annotation of the metabolome of the brown alga *Taonia atomaria*

Figure 1. Analytical workflow used in this study

Figure 2. Restrained molecular network (**MN-1**) built with MS/MS data of fractions obtained from the crude extracts of the brown alga *Taonia atomaria*

Node colors were chosen according to the chemical classes (see color code below the network). Nodes with a diamond shape correspond to fragments ions while those with a triangle shape represent compounds with a distinct adduct from the rest of the cluster. Thickness of an edge between two nodes is proportional to the cosine score (CS, from 0.851 to 0.995). Only clusters with at least three nodes are represented. The whole network also includes 39 clusters of two nodes and 512 unbound nodes.

Figure 3. Annotation of geranylgeranylglycerol (GGG) derivatives in cluster **1F** and proposed MS/MS fragmentation pattern for GGG, MGGG (18:4) and MGGG (O-17:2) (**Abbreviations:** MGGG: Monoradylgeranylgeranylglycerol) (Circles: metabolites; Diamonds: fragments)

Figure 4. Annotation of cluster **1I** and proposed fragmentation pattern for the compound MFuA (20:4) (**Abbreviation:** MFuA: Monoacylfulvellic acid) (Circles: metabolites; Triangles: adducts)

Figure 5. Less restricted MN (**MN-3**) built with MS/MS data of fractions obtained from the crude extracts of the brown alga *Taonia atomaria*

Node colors were chosen according to their chemical classes (see color code below the network). Nodes with a diamond shape correspond to fragments ions, while those with a triangle shape represent compounds with a distinct adduct from the rest of the cluster. Nodes with a thick border correspond to newly annotated compounds via the MN enlargement. Thickness of an edge between two nodes is proportional to the cosine score (CS, from 0.701 to 0.995). Only clusters with at least three nodes were represented. The whole network also includes 37 clusters of two nodes and 326 unbound nodes.

Table 1. Selected building parameters and results for the MNs constructed with the MS/MS data of fractions obtained from the crude extracts of the brown alga *T. atomaria*

Molecular network	Minimum cosine score	Minimum Matched fragment ions	Network TopK	Number of annotated metabolites	Number of clusters with 5 or more nodes
MN-1	0.85	10	7	147	15
MN-2	0.80	8	7	198	19
MN-3	0.70	6	10	212	21

Table 2

Chemical family	Abbreviation	Color code	Characteristic ion fragment (Bold: most intense fragment ions)	Characteristic neutral mass loss	Other MS/MS fragmentation pattern	Annotated nodes (MN-1)	Annotated nodes (MN-1 to MN-3)
Monoacylglycerol	MG		-	-	$[R\text{-}C{\equiv}O]^+$	1	3
Diacylglycerol	DG		-	-	$[M - FA]^+$, $[R\text{-}C{\equiv}O]^+$	13	20
Triacylglycerol	TG		-	-	$[M - FA]^+$, $[M - 2FA + OH]^+$, $[R\text{-}C{\equiv}O]^+$	21	38
Monogalactosylmonoacylglycerol	MGMG		-	*m/z* 197.0900 [-Gal ; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	3	6
Monogalactosyldiacylglycerol	MGDG		-		$[M - Gal - FA]^+$, $[R\text{-}C{\equiv}O]^+$	17	25
Digalactosylmonoacylglycerol	DGMG		-	*m/z* 359.1428 [-2Gal ; $-H_2O$; $-NH_3$] *m/z* 341.1322 [-2Gal ; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	0	4
Digalactosyldiacylglycerol	DGDG		-		$[M - 2\ Gal - FA]^+$, $[M - 2\ Gal - H_2O - FA]^+$, $[R\text{-}C{\equiv}O]^+$	1	2
Monoacylglycerylhydroxymethyl-*N*,*N*,*N*-trimethyl-*ß*-alanine	MGTA		***m/z* 236.1493 [$C_{10}H_{22}NO_5$]$^+$** *m/z* 218.1387 [$C_{10}H_{20}NO_4$]$^+$ *m/z* 162.1125 [$C_7H_{16}NO_3$]$^+$ *m/z* 144.1019 [$C_7H_{14}NO_2$]$^+$ *m/z* 100.1121 [$C_6H_{14}N$]$^+$	m/z 161.1052 [-$C_7H_{12}O_3$; $-NH_3$]	-	14	19
Sulfoquinovosylmonoacyglycerol	SQMG		-	*m/z* 261.0519 [-$C_6H_{12}O_8S$; $-NH_3$] *m/z* 243.0413 [-$C_6H_{10}O_7S$]	$[R\text{-}C{\equiv}O]^+$	7	7
Sulfoquinovosyldiacylglycerol	SQDG		-		$[M - C_6H_{12}O_8S - FA]^+$, $[R\text{-}C{\equiv}O]^+$	7	11
Lyso-Phosphatidylcholine	*Lyso*-PC		*m/z* 258.1101 [$C_8H_{21}NO_6P$]$^+$ ***m/z* 184.0733 [$C_5H_{14}NO_4P$]$^+$** *m/z* 104.1070 [$C_4H_{14}NO$]$^+$ *m/z* 86.0964 [$C_5H_{12}N$]$^+$	*m/z* 183.0660 [-$C_5H_{14}NO_4P$]	-	4	4
Phosphatidylcholine	PC				-	1	1
Terpene (and GGG)	-		-	-	-	1	9
Monoacylgeranylgeranylglycerol	MGGG		***m/z* 273.2577 [$C_{20}H_{33}$]$^+$**	*m/z* 289.2770 [-GGG ; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	8	8
Monoalkylgeranylgeranylglycerol	MGGG (O)				-	6	7
Monoacylhydroxy-geranylgeranylglycerol	MGGG-OH		*m/z* 289.2526 [$C_{20}H_{33}O$]$^+$ ***m/z* 271.2420 [$C_{20}H_{31}$]$^+$**	*m/z* 323.2824 [-$C_{20}H_{34}O_2$; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	6	6
Monoacylfarnesylglycerol	MFG		***m/z* 205.1951 [$C_{15}H_{25}$]$^+$**	*m/z* 239.2249 [-FG ; $-H_2O$; $-NH_3$] *m/z* 221.2144 [-FG ; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	8	8
Diacylfarnesylglycerol	DFG				$[M - FG - FA]^+$, $[R\text{-}C{\equiv}O]^+$	15	15
Monoacylglycerylfulvellic acid	MFuA		*m/z* 141.0546 [$C_7H_9O_3$]$^+$ ***m/z* 85.0284 [$C_4H_5O_2$]$^+$**	m/z 119.0583 [-FuA ; $-H_2O$; $-NH_3$]	$[R\text{-}C{\equiv}O]^+$	7	8
Diacylglycerylfulvellic acid	DFuA				$[M - FA]^+$, $[R\text{-}C{\equiv}O]^+$	7	11
					Total:	**147**	**212**

Figure 1

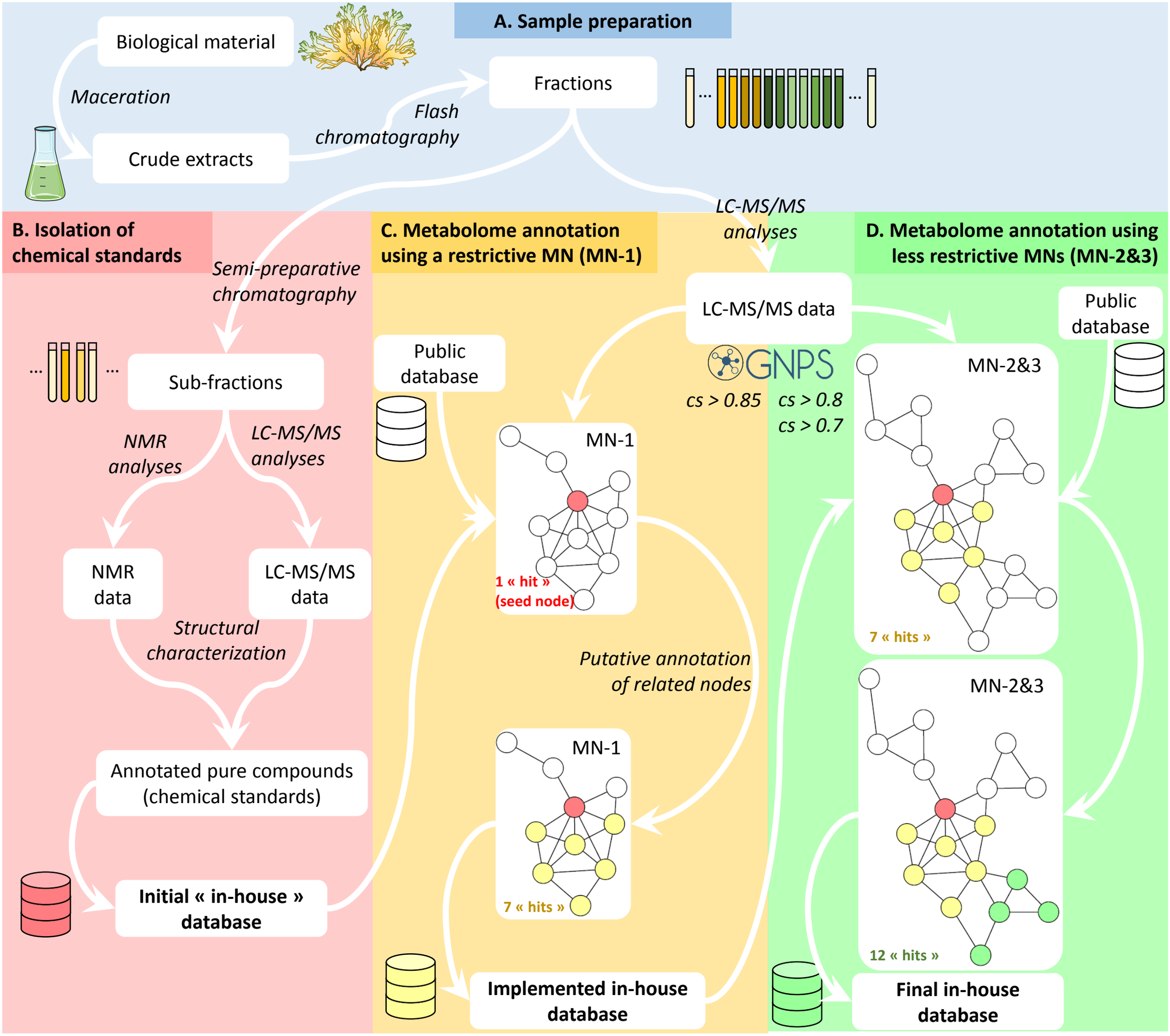

A. Sample preparation
Biological material
Maceration
Crude extracts
Flash chromatography
Fractions
LC-MS/MS analyses
B. Isolation of chemical standards
Semi-preparative chromatography
Sub-fractions
NMR analyses
LC-MS/MS analyses
NMR data
LC-MS/MS data
Structural characterization
Annotated pure compounds (chemical standards)
Initial « in-house » database
C. Metabolome annotation using a restrictive MN (MN-1)
Public database
LC-MS/MS data
MN-1
1 « hit » (seed node)
Putative annotation of related nodes
MN-1
7 « hits »
Implemented in-house database
D. Metabolome annotation using less restrictive MNs (MN-2&3)
GNPS
cs > 0.85
cs > 0.8
cs > 0.7
Public database
MN-2&3
7 « hits »
MN-2&3
12 « hits »
Final in-house database

Figure 2

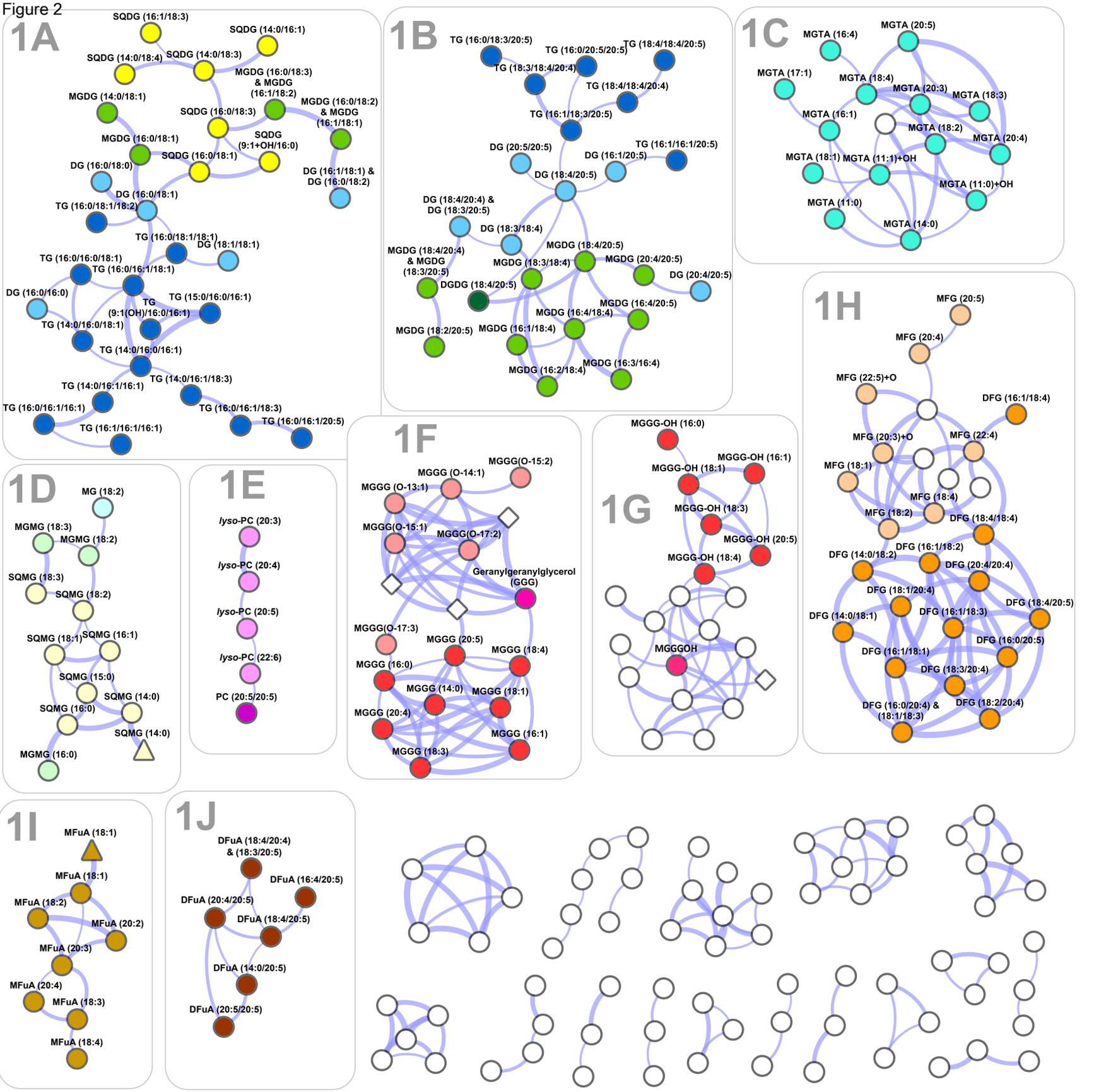

1A
SQDG (16:1/18:3)
SQDG (14:0/16:1)
SQDG (14:0/18:4)
SQDG (14:0/18:3)
MGDG (16:0/18:3) & MGDG (16:1/18:2)
MGDG (14:0/18:1)
SQDG (16:0/18:3)
MGDG (16:0/18:2) & MGDG (16:1/18:1)
MGDG (16:0/18:1)
SQDG (9:1+OH/16:0)
SQDG (16:0/18:1)
DG (16:0/18:0)
DG (16:1/18:1) & DG (16:0/18:2)
DG (16:0/18:1)
TG (16:0/18:1/18:2)
TG (16:0/18:1/18:1)
DG (18:1/18:1)
TG (16:0/16:0/18:1)
TG (16:0/16:1/18:1)
DG (16:0/16:0)
TG (15:0/16:0/16:1)
TG (9:1(OH)/16:0/16:1)
TG (14:0/16:0/18:1)
TG (14:0/16:0/16:1)
TG (14:0/16:1/16:1)
TG (14:0/16:1/18:3)
TG (16:0/16:1/16:1)
TG (16:0/16:1/18:3)
TG (16:0/16:1/20:5)
TG (16:1/16:1/16:1)
1B
TG (16:0/18:3/20:5)
TG (16:0/20:5/20:5)
TG (18:4/18:4/20:5)
TG (18:3/18:4/20:4)
TG (18:4/18:4/20:4)
TG (16:1/18:3/20:5)
DG (20:5/20:5)
TG (16:1/16:1/20:5)
DG (16:1/20:5)
DG (18:4/20:5)
DG (18:4/20:4) & DG (18:3/20:5)
DG (18:3/18:4)
MGDG (18:4/20:4) & MGDG (18:3/20:5)
MGDG (18:4/20:5)
MGDG (18:3/18:4)
MGDG (20:4/20:5)
DG (20:4/20:5)
DGDG (18:4/20:5)
MGDG (16:4/20:5)
MGDG (16:4/18:4)
MGDG (18:2/20:5)
MGDG (16:1/18:4)
MGDG (16:3/16:4)
MGDG (16:2/18:4)
1C
MGTA (20:5)
MGTA (16:4)
MGTA (17:1)
MGTA (18:4)
MGTA (20:3)
MGTA (18:3)
MGTA (16:1)
MGTA (18:2)
MGTA (20:4)
MGTA (18:1)
MGTA (11:1)+OH
MGTA (11:0)+OH
MGTA (11:0)
MGTA (14:0)
1D
MG (18:2)
MGMG (18:3)
MGMG (18:2)
SQMG (18:3)
SQMG (18:2)
SQMG (18:1)
SQMG (16:1)
SQMG (15:0)
SQMG (14:0)
SQMG (16:0)
SQMG (14:0)
MGMG (16:0)
1E
lyso-PC (20:3)
lyso-PC (20:4)
lyso-PC (20:5)
lyso-PC (22:6)
PC (20:5/20:5)
1F
MGGG(O-15:2)
MGGG (O-14:1)
MGGG (O-13:1)
MGGG(O-15:1)
MGGG(O-17:2)
Geranylgeranylglycerol (GGG)
MGGG(O-17:3)
MGGG (20:5)
MGGG (18:4)
MGGG (16:0)
MGGG (14:0)
MGGG (18:1)
MGGG (20:4)
MGGG (16:1)
MGGG (18:3)
1G
MGGG-OH (16:0)
MGGG-OH (16:1)
MGGG-OH (18:1)
MGGG-OH (18:3)
MGGG-OH (20:5)
MGGG-OH (18:4)
MGGGOH
1H
MFG (20:5)
MFG (20:4)
MFG (22:5)+O
DFG (16:1/18:4)
MFG (20:3)+O
MFG (22:4)
MFG (18:1)
MFG (18:4)
MFG (18:2)
DFG (18:4/18:4)
DFG (16:1/18:2)
DFG (14:0/18:2)
DFG (20:4/20:4)
DFG (18:1/20:4)
DFG (18:4/20:5)
DFG (14:0/18:1)
DFG (16:1/18:3)
DFG (16:0/20:5)
DFG (16:1/18:1)
DFG (18:3/20:4)
DFG (16:0/20:4) & (18:1/18:3)
DFG (18:2/20:4)
1I
MFuA (18:1)
MFuA (18:1)
MFuA (18:2)
MFuA (20:2)
MFuA (20:3)
MFuA (20:4)
MFuA (18:3)
MFuA (18:4)
1J
DFuA (18:4/20:4) & (18:3/20:5)
DFuA (16:4/20:5)
DFuA (20:4/20:5)
DFuA (18:4/20:5)
DFuA (14:0/20:5)
DFuA (20:5/20:5)
Monoacylglycerols (MG)
Diacylglycerols (DG)
Triacylglycerols (TG)
Monogalactosylmonoacylglycerols (MGMG)
Monogalactosyldiacylglycerols (MGDG)
Digalactosyldiacylglycerols (DGDG)
Monoacylglyceryl-N,N,N-trimethyl-ß-alanines (MGTA)
Sulfoquinovosylmonoacylglycerols (SQMG)
Sulfoquinovosyldiacylglycerols (SQDG)
Lyso-phosphatidylcholines (Lyso-PC)
Phosphatidylcholines (PC)
Geranylgeranylglycerol (GGG)
Monoacylgeranylgeranylglycerols (MGGG)
Monoalkylgeranylgeranylglycerols (MGGG-O)
Monoacylfarnesylglycerols (MFG)
Diacylfarnesylglycerols (DFG)
Monoacylfulvellic acids (MFuA)
Diacylfulvellic acids (DFuA)

Figure 3

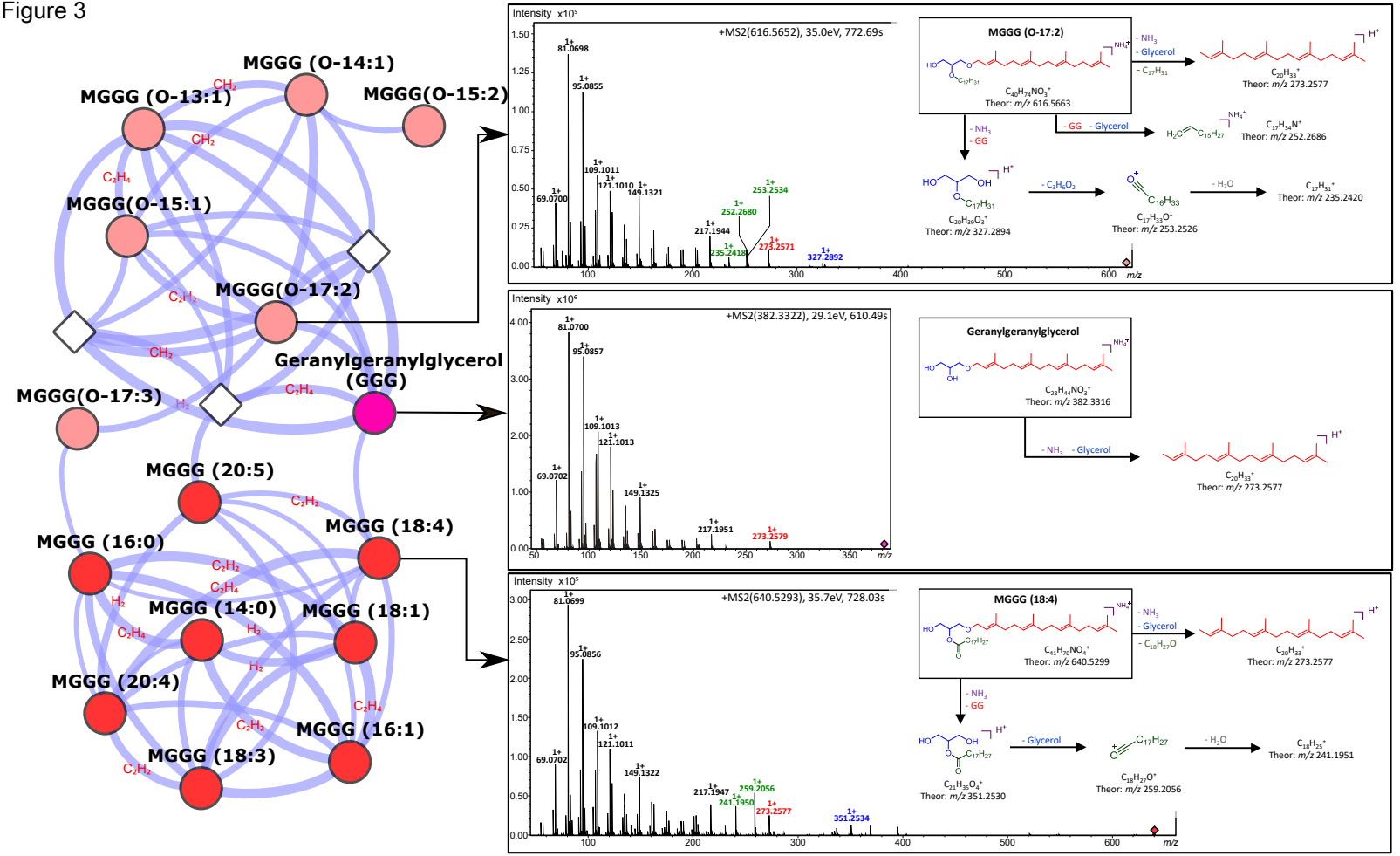

MGGG (O-14:1)
MGGG (O-13:1)
MGGG(O-15:2)
MGGG(O-15:1)
MGGG(O-17:2)
Geranylgeranylglycerol (GGG)
MGGG(O-17:3)
MGGG (20:5)
MGGG (18:4)
MGGG (16:0)
MGGG (14:0)
MGGG (18:1)
MGGG (20:4)
MGGG (16:1)
MGGG (18:3)
CH2
C2H4
H2
C2H2
Intensity x10^5
+MS2(616.5652), 35.0eV, 772.69s
MGGG (O-17:2)
C40H72NO3+
Theor: m/z 616.5663
- NH3
- Glycerol
- C17H31
C20H33+
Theor: m/z 273.2577
- GG
C17H34N+
Theor: m/z 252.2686
C20H39O3+
Theor: m/z 327.2894
- C3H6O2
C17H33O+
Theor: m/z 253.2526
- H2O
C17H31+
Theor: m/z 235.2420
+MS2(382.3322), 29.1eV, 610.49s
Geranylgeranylglycerol
C23H44NO2+
Theor: m/z 382.3316
C20H33+
Theor: m/z 273.2577
+MS2(640.5293), 35.7eV, 728.03s
MGGG (18:4)
C41H70NO4+
Theor: m/z 640.5299
C20H33+
Theor: m/z 273.2577
C21H35O4+
Theor: m/z 351.2530
- Glycerol
C18H27O
C18H25+
Theor: m/z 241.1951
C18H27O+
Theor: m/z 259.2056
m/z

Figure 4

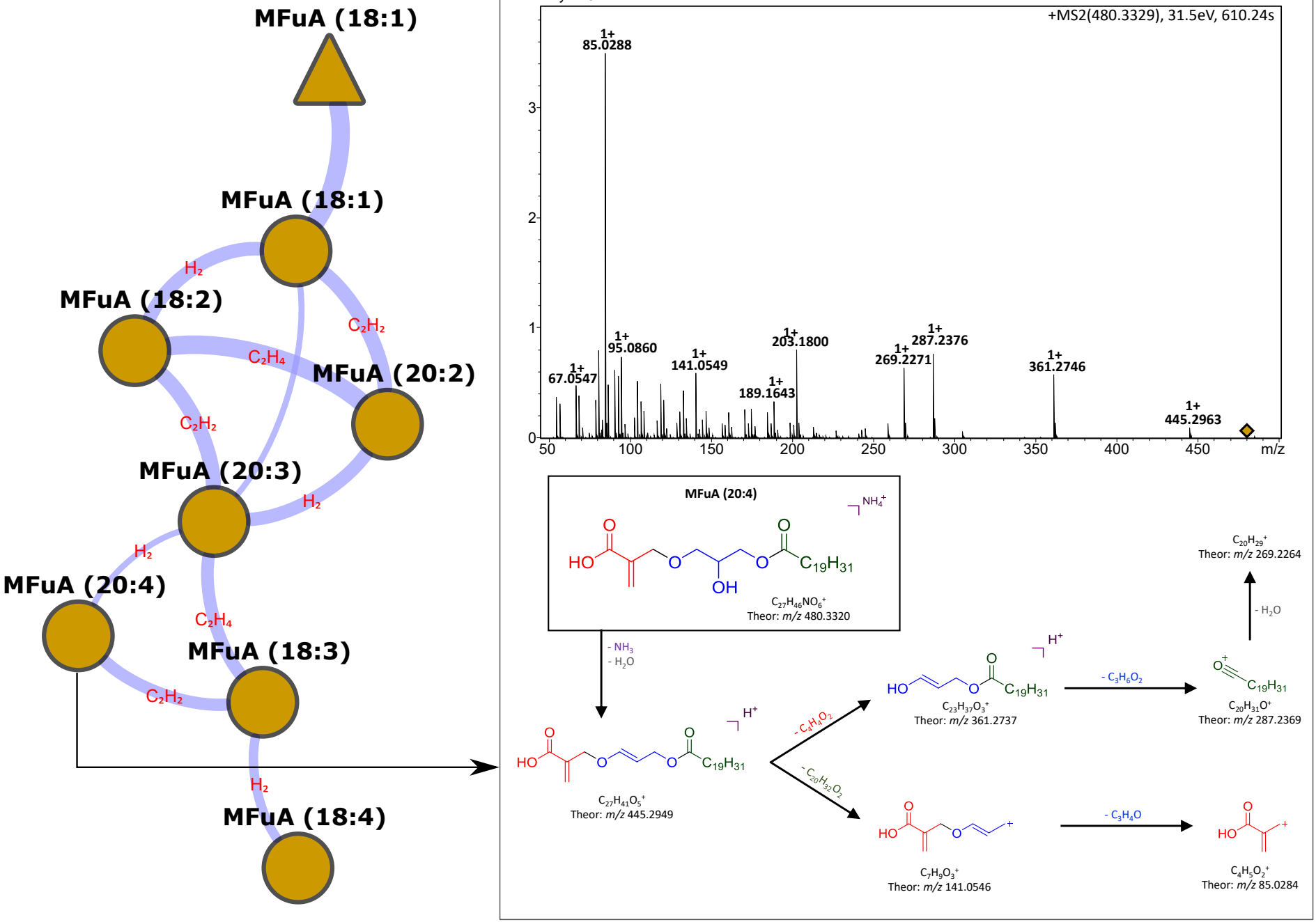
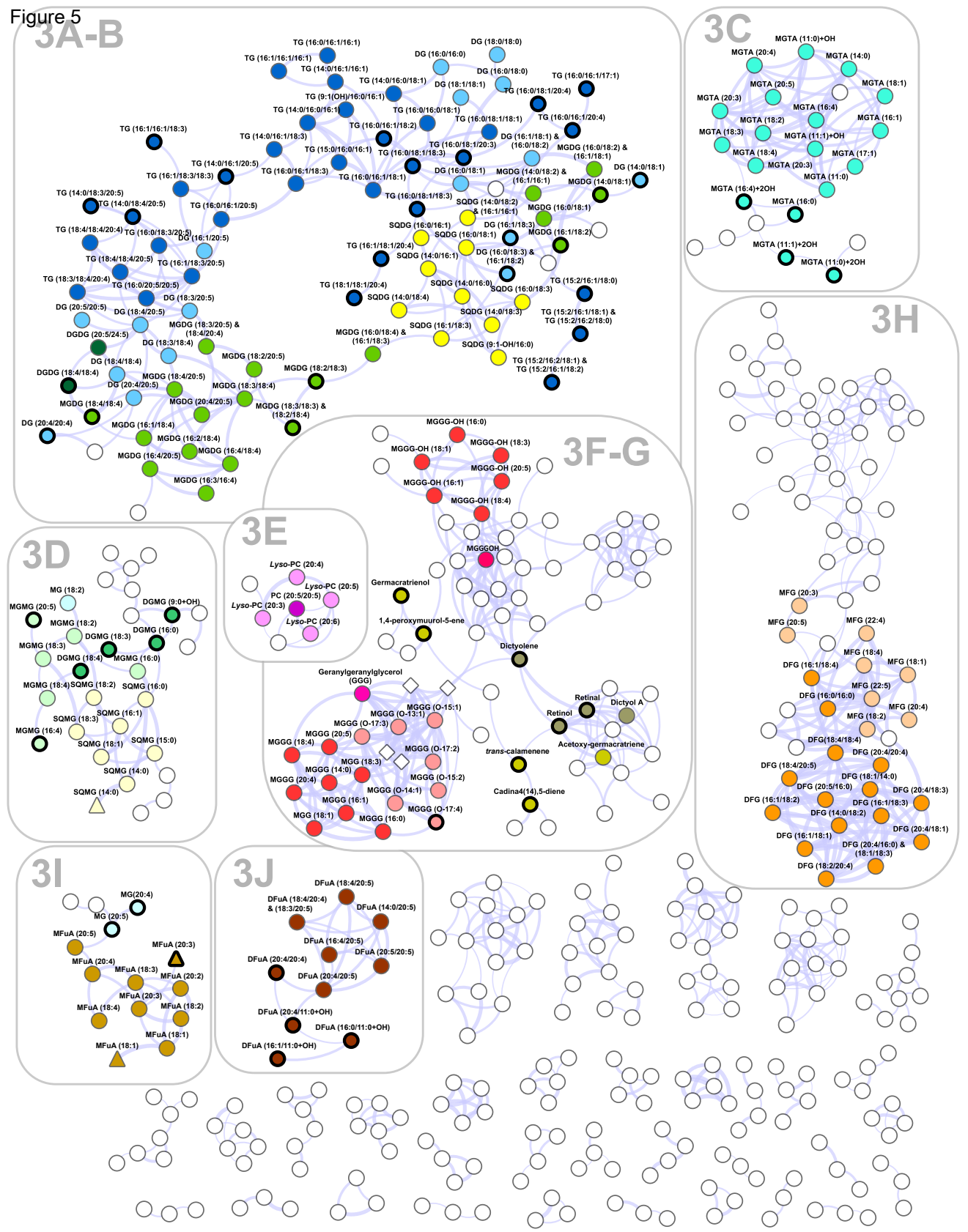

MFuA (18:1)
MFuA (18:1)
MFuA (18:2)
MFuA (20:2)
MFuA (20:3)
MFuA (20:4)
MFuA (18:3)
MFuA (18:4)
H2
C2H2
C2H4
Intensity x10⁵
+MS2(480.3329), 31.5eV, 610.24s
1+ 85.0288
1+ 67.0547
1+ 95.0860
1+ 141.0549
1+ 189.1643
1+ 203.1800
1+ 269.2271
1+ 287.2376
1+ 361.2746
1+ 445.2963
m/z
MFuA (20:4)
NH4+
C27H46NO6+
Theor: m/z 480.3320
- NH3
- H2O
H+
C27H41O5+
Theor: m/z 445.2949
- C4H4O2
- C20H32O2
C23H37O3+
Theor: m/z 361.2737
- C3H6O2
C20H31O+
Theor: m/z 287.2369
- H2O
C20H29+
Theor: m/z 269.2264
C7H9O3+
Theor: m/z 141.0546
- C3H4O
C4H5O2+
Theor: m/z 85.0284

Figure 5


3A-B
TG (16:0/16:1/16:1)
TG (16:1/16:1/16:1)
TG (14:0/16:1/16:1)
TG (14:0/16:0/18:1)
TG (9:1(OH)/16:0/16:1)
TG (14:0/16:0/16:1)
TG (16:0/16:0/18:1)
TG (16:0/16:1/18:2)
TG (14:0/16:1/18:3)
TG (15:0/16:0/16:1)
TG (16:0/16:1/18:3)
TG (16:0/16:1/18:1)
TG (16:1/16:1/18:3)
TG (14:0/16:1/20:5)
TG (16:1/18:3/18:3)
TG (16:0/16:1/20:5)
TG (14:0/18:3/20:5)
TG (14:0/18:4/20:5)
TG (18:4/18:4/20:4)
TG (16:0/18:3/20:5)
DG (16:1/20:5)
TG (18:4/18:4/20:5)
TG (16:1/18:3/20:5)
TG (18:3/18:4/20:4)
TG (16:0/20:5/20:5)
DG (18:3/20:5)
DG (20:5/20:5)
DG (18:4/20:5)
DGDG (20:5/24:5)
MGDG (18:3/20:5) & (18:4/20:4)
DG (18:3/18:4)
DG (18:4/18:4)
MGDG (18:2/20:5)
DGDG (18:4/18:4)
MGDG (18:4/20:5)
MGDG (18:2/18:3)
DG (20:4/20:5)
MGDG (18:3/18:4)
MGDG (18:4/18:4)
MGDG (20:4/20:5)
MGDG (18:3/18:3) & (18:2/18:4)
DG (20:4/20:4)
MGDG (16:1/18:4)
MGDG (16:2/18:4)
MGDG (16:4/20:5)
MGDG (16:4/18:4)
MGDG (16:3/16:4)
DG (18:0/18:0)
DG (16:0/16:0)
DG (16:0/18:0)
DG (18:1/18:1)
TG (16:0/16:1/17:1)
TG (16:0/18:1/20:4)
TG (16:0/18:1/18:1)
TG (16:0/16:1/20:4)
DG (16:1/18:1) & (16:0/18:2)
TG (16:0/18:1/20:3)
MGDG (16:0/18:2) & (16:1/18:1)
TG (16:0/18:1/18:3)
DG (16:0/18:1)
MGDG (14:0/18:2) & (16:1/16:1)
DG (14:0/18:1)
MGDG (14:0/18:1)
TG (16:0/18:1/18:3)
SQDG (14:0/18:2) & (16:1/16:1)
MGDG (16:0/18:1)
SQDG (16:0/16:1)
DG (16:1/18:3)
SQDG (16:0/18:1)
MGDG (16:1/18:2)
TG (16:1/18:1/20:4)
SQDG (14:0/16:1)
DG (16:0/18:3) & (16:1/18:2)
SQDG (14:0/16:0)
SQDG (16:0/18:3)
TG (15:2/16:1/18:0)
TG (18:1/18:1/20:4)
SQDG (14:0/18:4)
SQDG (14:0/18:3)
TG (15:2/16:1/18:1) & TG (15:2/16:2/18:0)
MGDG (16:0/18:4) & (16:1/18:3)
SQDG (16:1/18:3)
SQDG (9:1-OH/16:0)
TG (15:2/16:2/18:1) & TG (15:2/16:1/18:2)
3C
MGTA (11:0)+OH
MGTA (20:4)
MGTA (14:0)
MGTA (20:5)
MGTA (18:1)
MGTA (20:3)
MGTA (16:4)
MGTA (18:2)
MGTA (16:1)
MGTA (18:3)
MGTA (11:1)+OH
MGTA (18:4)
MGTA (17:1)
MGTA (20:3)
MGTA (11:0)
MGTA (16:4)+2OH
MGTA (16:0)
MGTA (11:1)+2OH
MGTA (11:0)+2OH
3H
MFG (20:3)
MFG (22:4)
MFG (20:5)
MFG (18:4)
MFG (18:1)
DFG (16:1/18:4)
MFG (22:5)
DFG (16:0/16:0)
MFG (20:4)
MFG (18:2)
DFG(18:4/18:4)
DFG (20:4/20:4)
DFG (18:4/20:5)
DFG (18:1/14:0)
DFG (20:5/16:0)
DFG (20:4/18:3)
DFG (16:1/18:2)
DFG (16:1/18:3)
DFG (14:0/18:2)
DFG (20:4/18:1)
DFG (16:1/18:1)
DFG (20:4/16:0) & (18:1/18:3)
DFG (18:2/20:4)
3F-G
MGGG-OH (16:0)
MGGG-OH (18:1)
MGGG-OH (18:3)
MGGG-OH (20:5)
MGGG-OH (16:1)
MGGG-OH (18:4)
MGGGOH
Germacratrienol
1,4-peroxymuurol-5-ene
Dictyolene
Geranylgeranylglycerol (GGG)
MGGG (O-15:1)
MGGG (O-13:1)
MGGG (O-17:3)
MGGG (20:5)
MGGG (18:4)
MGGG (O-17:2)
MGG (18:3)
MGGG (14:0)
MGGG (20:4)
MGGG (O-15:2)
MGGG (O-14:1)
MGGG (16:1)
MGG (18:1)
MGGG (16:0)
MGGG (O-17:4)
Retinal
Dictyol A
Retinol
trans-calamenene
Acetoxy-germacratriene
Cadina4(14),5-diene
3E
Lyso-PC (20:4)
Lyso-PC (20:5)
PC (20:5/20:5)
Lyso-PC (20:3)
Lyso-PC (20:6)
3D
MG (18:2)
MGMG (20:5)
DGMG (9:0+OH)
MGMG (18:2)
DGMG (18:3)
DGMG (16:0)
MGMG (18:3)
DGMG (18:4)
MGMG (16:0)
MGMG (18:4)
SQMG (18:2)
SQMG (16:0)
SQMG (16:1)
SQMG (18:3)
MGMG (16:4)
SQMG (18:1)
SQMG (15:0)
SQMG (14:0)
SQMG (14:0)
3I
MG(20:4)
MG (20:5)
MFuA (20:5)
MFuA (20:3)
MFuA (20:4)
MFuA (18:3)
MFuA (20:2)
MFuA (20:3)
MFuA (18:4)
MFuA (18:2)
MFuA (18:1)
MFuA (18:1)
3J
DFuA (18:4/20:5)
DFuA (18:4/20:4) & (18:3/20:5)
DFuA (14:0/20:5)
DFuA (16:4/20:5)
DFuA (20:4/20:4)
DFuA (20:5/20:5)
DFuA (20:4/20:5)
DFuA (20:4/11:0+OH)
DFuA (16:0/11:0+OH)
DFuA (16:1/11:0+OH)
Monoacylglycerols (MG)
Diacylglycerols (DG)
Triacylglycerols (TG)
Monogalactosylmonoacylglycerols (MGMG)
Monogalactosyldiacylglycerols (MGDG)
Digalactosylmonoacylglycerols (DGMG)
Digalactosyldiacylglycerols (DGDG)
Monoacylglyceryl-N,N,N-trimethyl-ß-alanines (MGTA)
Sulfoquinovosylmonoacylglycerols (SQMG)
Sulfoquinovosyldiacylglycerols (SQDG)
Lyso-phosphatidylcholines (Lyso-PC)
Phosphatidylcholines (PC)
Sesquiterpenes
Diterpenes
Geranylgeranylglycerol (GGG)
Monoacylgeranylgeranylglycerols (MGGG)
Monoalkylgeranylgeranylglycerols (MGGG-O)
Monoacylfarnesylglycerols (MFG)
Diacylfarnesylglycerols (DFG)
Monoacylulvellic acids (MFuA)
Diacylfulvellic acids (DFuA)

Credit Author Statement

Nathan Carriot: Data curation, Methodology, Visualization, Writing - original draft. **Benoît Paix:** Conceptualization, Data curation, Methodology, Visualization, Writing - original draft. **Stéphane Greff:** Data curation, Methodology, Resources. **Bruno Viguier:** Methodology, Resources. **Jean-François Briand:** Conceptualization, Funding acquisition, Supervision. **Gérald Culioli:** Conceptualization, Data Curation, Funding acquisition, Methodology, Supervision, Validation, Writing - review & editing.