Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of Saccharomyces cerevisiae

Olivier Perruchon, Isabelle Schmitz-Afonso, Cécile Grondin, Serge Casaregola, Carlos Afonso, Abdelhakim Elomri

Combination of UHPLC-MS/MS-molecular networking approach and FTICR-MS for the metabolic profiling of *Saccharomyces cerevisiae*

Olivier Perruchon[a], Isabelle Schmitz-Afonso[a], Cécile Grondin[b], Serge Casaregola[b], Carlos Afonso[a], Abdelhakim Elomri[a]*.
[a] Normandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA (UMR 6014), Rouen 76000, France
[b] Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, CIRM-Levures, 78350, Jouy-en-Josas, France

*Corresponding author: hakim.elomri@univ-rouen.fr

Abstract
Natural products are a reliable source of bioactive molecules and represent an industrial and pharmaceutical stake. Indeed, the model yeast species *Saccharomyces cerevisiae* is a well-known eukaryotic organism largely used as a biotechnological tool, but still a topical subject of study. In this work, the exploration of *Saccharomyces cerevisiae* is taken further through an untargeted metabolomics workflow. The aim is to enrich databases and bring new information about the standard *S. cerevisiae* strain in a given medium. Analytical methods and bioinformatics tools were combined in a high-throughput methodology useable to dereplicate many types of biological extracts and cartography secondary metabolites. Ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) analyses were carried out and spectral data were pre-processed to build molecular networks. Annotations were attributed to compounds through comparison with databases and manual investigation of networks. Ultra-high-resolution Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) brought additional information thanks to a higher dynamic range and enhanced UHPLC-MS/MS results by unveiling ambiguities and bringing accurate molecular formulae. Therefore, accurate and reliable annotated features resulted from the UHPLC-MS/MS data while FTICR-MS provided an overall cartography of metabolites thanks to van Krevelen diagrams. Various small molecules such as amino acids derivatives and indole alkaloids have been determined for the first time in this yeast. The complementarity of FTICR-MS and UHPLC-MS/MS for secondary metabolite annotation brought this new mapping of *S. cerevisiae*.

Keywords FTICR-MS; mass spectrometry; dereplication; molecular networking; *Saccharomyces cerevisiae*; metabolomics.

1. Introduction

Saccharomyces cerevisiae is one of the best-known model systems for eukaryotic cell biology. It belongs to the *Saccharomycotina* yeasts, a subphylum of Ascomycete unicellular fungi found on a wide variety of substrates such as insects, plants or soils [1]. The *S. cerevisiae* strain S288c is a widely used laboratory strain and was the first eukaryotic organism to have its genome fully sequenced through a systematic sequencing project. This reference sequence is stored in the Saccharomyces Genome Database (SGD) [2]. Yeasts are involved in numerous agri-food applications like winemaking, brewery, bakery [3] or drugs as probiotics, but *S. cerevisiae* is also an effective biotechnological tool for the production of biopharmaceuticals: hormones like insulin, or vaccines such as hepatitis B vaccine and human papillomavirus vaccine [4–6]. Fermentation and other biological processes make yeasts capable of producing interesting secondary metabolites that represent a large structural biodiversity. As natural products can show biological effects against many diseases, determining what *S. cerevisiae* naturally synthetize represents an industrial, pharmaceutical and scientific stake. Previous studies have characterized some of these small molecules by gas chromatography-mass spectrometry (GC-MS) analyses [7,8], liquid chromatography-mass spectrometry (LC-MS) analyses [9] or carried out metabolic profiling by nuclear magnetic resonance (NMR) techniques [10].

The last two decades have brought many technological advances in mass spectrometry, concerning ionization sources, analyzers or hyphenation with other separation techniques. Therefore, metabolomics research has evolved and among common techniques, high performance liquid chromatography (HPLC) is an efficient starting point to separate compounds according to their polarity. The coupling of liquid chromatography with high resolution mass spectrometry (HPLC-HRMS), with atmospheric pressure ionization such as electrospray ionization, allows a screening of a wide range of metabolites. In addition, tandem mass spectrometry (MS/MS), and in particular high resolution MS/MS spectra, affords structural information on metabolites [11,12].

High resolution mass spectrometry analyses in metabolomics generate an important amount of data that require automated data processing. Indeed, the complete analytical workflow in metabolomics studies implies data acquisition, data processing and finally data interpretation. With regard to data processing, computerized and automated tools are great time savers as manual MS-data deciphering is heavily time-consuming. The use of algorithms also avoids human error. As available applications are in development, there are a lot of parameters to optimize. For example, the open software MZMine2 processes HPLC-MS/MS data to convert spectral data in feature lists [13,14].

With regard to data interpretation the first descriptor used is the mass determination. Accurate mass-to-charge ratios (*m/z*) measurement allow to attribute molecular formulae that can be compared to databases to annotate the features (*e.g.* Yeast Metabolome Database, Human Metabolome Database, MassBank [15–18]). High resolution analyzers such as time-of-flight and orbitrap present a mass accuracy up to 5 ppm. With such accuracy there are however often several possible molecular formulas for a given signal even considering specific criteria such as the seven golden rules [19]. Fourier Transform Ion Cyclotron Resonance systems (FTICR-MS) have the required resolution (>1,000,000) and accuracy (<0.5 ppm) to solve this issue [20]. FTICR-MS is used in metabolomics for rapid metabolome fingerprinting or exploratory profiling with direct injection approaches [21,22]. The high quality of high field FTICR spectra allow to attribute unique molecular formulae to each ion with great confidence. The use of isotopic fine structure can be used to confirm the presence of some specific elements such as sulfur [23]. This annotated data can be used to generate molecular mapping using for instance van Krevelen diagrams [24,25].

The other descriptors used in data interpretation are the tandem mass spectra of metabolite ions. Fragmentation data are used on the Global Natural Products Social molecular networking platform to build molecular networks (https://gnps.ucsd.edu), as structurally related molecules share similar fragmentation patterns [26]. In addition, GNPS compares uploaded MS/MS spectra to constantly enrich databases and annotate compounds from the feature list. This recent bioinformatics tool is already a valuable instrument in the exploration of microorganisms and in drug discovery [27,28]. The combination of MZMine2 and molecular networking proves to be very worthwhile to accelerate the *dereplication* of biological samples and subsequently focus on potential unknown compounds.

In metabolomics studies, annotation is carried out based on the Metabolomics Standard Initiative annotation levels thanks to all descriptors obtained previously [29]. Ranging across four different levels, it respectively covers fully identified molecules, putatively annotated compounds, putatively annotated compound classes and unknown metabolites.

Thus, we present here an in-depth exploration of *S. cerevisiae* reference strain S288c extracts *via* the complementarity of UHPLC-MS/MS and Direct Injection FTICR-MS, completed by molecular networking data treatment. This approach accelerated significantly the deciphering of the biological extracts, and could be applied to other complex mixtures.

2. Materials and methods

2.1. Statement of Human and Animal Rights

This article does not contain any studies with human participants or animals.

2.2. Yeast culture

Saccharomyces cerevisiae S288c (CLIB 338) standard laboratory strain is preserved at the Biological Resource Centre CIRM-Levures (INRAE, Jouy-en-Josas, France; www.inra.fr/cirm_eng/Yeasts). This strain was routinely grown on YPD (1 % yeast extract, 1 % peptone, 1 % glucose) agar medium at 26.5 °C. For analysis, strains were streaked from an overnight culture on YNB (6.7 % Yeast Nitrogen Base, 1 % peptone, 1 % glucose) agar medium and incubated for 3 days at 26.5 °C. One culture have been extracted, and control samples, also called culture medium extracts were prepared in the same way, except that no yeast strain were added to the culture medium.

2.3. Metabolites extraction

Dichloromethane, ethyl acetate and methanol are HPLC gradient grade solvents. Intracellular and extracellular metabolites were extracted by addition of 15 mL of 1:1 *v/v* dichloromethane and ethyl acetate on the culture, helped by a 30 min sonication (Branson B-5200-54, 450 W). The extract was filtered through a sintered glass funnel (porosity from 16 µm to 40 µm). 10 mL of dichloromethane and ethyl acetate (1:1, v/v) were added to the medium, sonicated 10 min and filtered. The two gathered dichloromethane / ethyl acetate extracts were concentrated by solvent dry evaporation under vacuum at 40 °C to give a brown residue of 8.5 mg named "DE extract" and stored at – 20 °C. 10 mL of methanol were added to the same growing medium, the mixture was sonicated 5 min and filtered. The extracted metabolites were then concentrated by evaporation. A brown gel residue of 183.5 mg was obtained, named "M extract" and stored at – 20 °C. Culture medium without yeasts went through the same process (leading to a CDE extract and a CM extract), prepared and analyzed in the same way.

2.4. Sample preparation for analysis

Methanol, water and acetonitrile are Fisher Chemical Optima LC-MS grade. DE and CDE extracts were resuspended in methanol, M and CM extracts were resuspended in water to a concentration of 50 mg/mL. Every sample was half-diluted with water / acetonitrile (1:1 v/v), which corresponds to the mobile phase, for analysis. All samples were filtered through PTFE 0.45 µm filters. Analytical blanks were prepared with the same solvents as the mobile phase and analyzed before and at the end of each analytical sequence to further clean-up the spectral data of background interference signals, like plasticizers or solvent impurities. Pure chemical standards (Table S1) were individually dissolved in water / acetonitrile (1:1 v/v) and mixed, then diluted to 0.01 mg/mL and analyzed by UHPLC-MS/MS.

2.5. UHPLC-ESI-QTOF-MS/MS analysis

Chromatographic separation was carried out using a UHPLC system (Dionex Ultimate 3000 UPLC+, Thermo Scientific, San Jose, CA, USA) equipped with a C18 silica-based Column (Acquity UPLC HSS T3 1.8 µm x 1.0 mm x 100 mm, Waters Corporation, Milford, MA, USA) with a prefilter of 0.2 µm. The column was kept at 40 °C during the analysis. An autosampler kept the samples at 4 °C. The injection volume was 1 µL. The solvents used for gradient separation are the following: water + 0.1 % formic acid (solvent A); acetonitrile + 0.1 % formic acid (solvent B); solvents are Fisher Chemical Optima LC-MS grade; formic acid is LiChropur (Merck) for LC-MS. The flow rate was 0.1 mL/min. Gradient consisted of 1 % B at 0 min; 1 % B at 1 min; going up linearly to 100 % B at 23 min then maintained 100 % B for 2 minutes; finally, 1 % B as initial conditions at 25.2 min. The run ended after 30 min including the stabilization step.

UHPLC was coupled to a hybrid quadrupole-time of flight analyzer (QTOF, SYNAPT G2 HDMS, Waters MS Technologies, Manchester, UK) equipped with an Electrospray Ionization source (ESI). The instrument was driven by MassLynx 4.1 software. Experiments were realized in positive and in negative ionization mode. *m/z* range was 50-2000. The ionization parameters and acquisition parameters for the data dependent scan method are indicated in Table S2. The instrument was externally calibrated with a sodium formate mixture (10 % formic acid / 0.1 M NaOH / acetonitrile in 1:1:8, v/v/v). The lock mass used during runs as internal calibrant was a Leucine-Enkephalin solution (2 ng/µL in water/isopropanol 1:1, v/v); the lock mass ions were *m/z* 556.2771 and *m/z* 554.2615 in positive and negative ionization mode respectively.

2.6. Spectral data processing

UHPLC-MS/MS data were firstly centroided on MassLynx 4.1 with the "Automatic peak detection" module. Raw data (Waters ".raw" format) were then imported within the MZMine2 software. Features are characterized by variables as *m/z*

ratio, retention time (RT in minutes), intensity, and are linked to corresponding MS/MS spectra. Retention time is considered so isomers can be discriminated. Peaks were detected with the "Exact mass" algorithm from the "Mass detection" module then detected masses were built into a chromatogram with "ADAP Chromatogram Builder" [30]. The chromatogram was deconvoluted with the Wavelet (ADAP) algorithm, removing background noise features then MS2 scans were paired to corresponding MS1 features. Isotopes were grouped and duplicate features were removed. Then, feature lists from samples and controls analyzed with the same ionization polarity were aligned. The "Formula prediction" module attributed molecular formulae following the seven golden rules [19]. Fragments and adducts were automatically searched with adequate modules. The detailed parameters for each module are shown in Table S3. Feature lists were then exported in a fitted format for GNPS [14].

2.7. Molecular networking on Global Natural Products Social molecular networking

We exported MZMine2 pre-processed feature lists and imported it on the GNPS platform. A molecular network was created with the feature based molecular networking workflow (https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/) on the GNPS website (http://gnps.ucsd.edu). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and a MS/MS fragment ion tolerance of 0.02 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 4 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 0, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. Molecular networks were visualized through Cytoscape.

2.8. FTICR-MS analysis

Analyses were carried out on a FTICR instrument (SolariX XR FTMS, Bruker Daltonics) equipped with a 12 Tesla superconducting magnet and a dynamically harmonized ICR cell. The instrument is equipped with an electrospray ionization source. The same samples analyzed by UHPLC-MS/MS were analyzed by direct injection FTICR-MS after dilution (1/10,000) with water / acetonitrile (1:1, v/v), at a flow of 120 µL/h. Each sample was analyzed both in positive and negative ionization mode. Source and analyzer parameters (shown in Table S4) were inspired by previous work and chosen in order to guarantee a robust and sensitive analysis [31]. FTICR-MS data were treated with Data Analysis 5.0 (Bruker). Molecular formulae were attributed considering M + H, Na and K adducts for positive mode, M − H and Cl for negative mode, for $C_{1\text{-}100}H_{0\text{-}200}N_{0\text{-}8}O_{0\text{-}20}S_{0\text{-}1}$ (then $P_{0\text{-}1}$) species. The tolerance was set to 0.3 ppm, RDBE from 0 to 80 and H/C ratio from 0 to 3. To exclude media components and artefacts, masses from analytical blanks and culture medium were subtracted from the sample MS scan for each extract in both ionization modes ("Xpose" mode with a ratio of 5).

3. Results and discussion

3.1. Feature determination and molecular networking

Dichloromethane / ethyl acetate (DE) and methanol (M) extracts from the strain *S. cerevisiae* S288c, together with culture medium control extracts, have been analyzed by UHPLC-MS/MS and Direct Injection FTICR-MS. Figure 1 summarizes the entire analysis and interpretation workflow. Figure S1 shows the base peak ion chromatogram (BPC for DE and M extracts in positive and negative ionization mode). UHPLC-MS/MS data were processed within MZMine2. During the automated data treatment, background noise features, blank features, in-source fragments, adducts and dimers were discarded. Yeast extract data and control sample data were aligned. The feature tables for each ionization mode were exported to build molecular networks on GNPS. Alignment and molecular networking allowed the removal of features exclusively found in culture medium extracts. The medium components present in networks are kept as they can help us to decipher yeast extract component that are linked to molecular networks, requiring the investigation of these additional networks. It is important to note that primary metabolites (e.g. amino-acids) are in the culture medium, and often cluster with structurally similar yeast metabolites. The feature tables and the molecular networks are the first form of results needed to begin data interpretation. Manual interpretation of result lists is a crucial step to confidently annotate the yeast extract compounds.

Annotations were attributed *via* database comparison and were supported by molecular network investigation. For instance, a feature was detected at *m/z* 231.1126 in positive mode, RT = 7.98 min. This node was included in a large molecular network, where indole alkaloids have already been annotated. Molecular networks built on GNPS displayed structurally related nodes. By comparing MS/MS spectra with tryptophol MS/MS spectrum and with MS/MS databases (e.g. HMDB), we could lead to an annotation that fitted with the molecular formula of the medium feature at *m/z* 217.0974: 1,2,3,4-

Tetrahydro-beta-carboline-3-carboxylic acid (Figure S2a). As this node was a crossroad of two other nodes, we used it to determine hypotheses for *m/z* 190.0864 and *m/z* 231.1126, attributed respectively 3-Indolepropionic acid and 1-Methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid. These annotations corroborate of an indole alkaloid network (Figure 2a). As compound hypotheses could be given, the annotation level is 2 for these compounds. Molecular networking permits the annotation spread and accelerates the whole annotation process. 25 level 2 annotations were provided thanks to progressing through molecular networks (21 in positive ionization mode and 4 in negative). Figure S3 shows all molecular networks for *S. cerevisiae* analysis including annotations after manual investigation. We observed that positive ionization mode led to bigger and more complex molecular networks. This is due to the structural families of the detected compounds, which led to more informative MS/MS spectra, therefore more networks. The largest network is related to alkaloids, as mentioned before, with an indole part and a quinoline part. Another small network corresponds to sulfated amino-acids. The last one clusters only medium and is also found in negative mode, alongside small networks of sulfated amino-acids, tryptophan derivatives and alkaloids.

To assess the confidence level of the annotations, we considered the metabolomics annotation level system [29]. The compounds may be given *putative annotations* on the basis of molecular formula, database search and molecular networking. *Level 1* are identifications allowed by retention time and MS/MS correlation between the studied feature and a pure standard. In our work, six pure standards were compared and six features were identified as level 1 (Table 1). Structural similarities associated to database match led to *level 2* annotations. When structural similarities were characterized between an unknown compound and a putatively annotated compound, molecular networking was incredibly useful to get *level 3* annotations. Finally, all compounds that cannot be associated to known molecules are considered *level 4* or unknown compounds. In parallel, additional data treatment and bibliographical search was done to increase the molecular network information. Figure 2b summarizes the number of annotated features for each level after the first manual data deciphering that led to annotations for 55 % of the 92 features in positive mode and for 44 % of the 55 features in negative, based on UHPLC-MS/MS analyses.

3.2. Contribution of FTICR-MS

The direct injection ESI-FTICR analyses, yielded an average mass accuracy of 0.08 ppm. After analysis subtraction, 1676 ions have been detected from positive ion spectra of DE extract, and 794 in negative ion spectra, with signal to noise of 9. M extract yielded 1604 ions in positive ion spectra, and 1471 in negative. From which unique molecular formulae have been attributed considering parameters told in Material and methods section (603 for DE POS, 403 for DE NEG, 404 for M POS, 304 for M NEG). Isotopic fine structure was used to confirm the presence of specific elements. In a few cases in which 2 possible molecular formulae are proposed, manual control was used to discard wrong attribution (keeping the formulae with the best accuracy in ppm, or discarding aberrant formulae with O/C ratio > 1 or formulae without enough H or too much N for the number of C).
The data from the two techniques were manually linked by searching corresponding *m/z* between UHPLC-MS/MS data and FTICR-MS data. Some molecular formulae were not attributed in UHPLC-MS/MS due to the lower time-of-flight precision. The highest FTICR mass measure precision was able to enhance molecular formulae determination. Therefore, more than 91 % of the features already detected by UHPLC-MS/MS could be linked to an accurate molecular formula in positive mode, and 96 % in negative mode. The compounds whose formulae were not retrieved tend to *level 4* annotations and remain unknown. These values prove the efficacy of analyzing yeast extracts with UHPLC-MS/MS on a side and FTICR-MS on the other. It should be pointed out that compared to the UHPLC-MS/MS analysis the DI FTICR analysis yielded a large amount of sodium and potassium adduct due to the lack of chromatographic separation.

It can be shown that the manual linking of FTICR-MS molecular formulae to UHPLC-MS/MS features enhanced significantly the quality of the obtained metabolomics results. The major interest is revealed when the chromatographic workflow did not lead to a precise annotation. The lower precision of TOF analysis, with errors around 2 to 5 ppm, could lead to several possible molecular formulae, from two to more than seven for high *m/z* ratio compounds, and then prevent the feature annotation. FTICR helped to discriminate the right formula. *Level 2* annotations were confirmed if the given formula corresponds to the compound hypothesis. Some *level 3* compounds had a chemical class attributed (*e.g.* alkaloid) thanks to molecular networking and common fragments, and FTICR-MS provided the missing molecular formula that allowed, combined with MS/MS data, a *level 2* annotation. *Level 4* unknown compounds at least have a formula. An empiric example of a hypothesis ambiguity was a feature detected at RT 12.76 min (Figure S4). The FTICR-MS data allowed us to determine the molecular formula among the three possible ones, confirming the Paspalic acid ($C_{16}H_{16}N_2O_3$) hypothesis. The enhancement of the deeper manual interpretation after FTICR-MS is shown in Figure 2c. Thus, 89 % of annotations is reached in positive mode and 87 % in negative mode. FTICR-MS enhanced previous results as a fast and complementary technique. In fact, the separation based on LC-MS methods has the advantage on isomer identification and matrix ion suppression effects reduction whereas the direct FTICR injection has the advantage on mass measure accuracy, leading to unique molecular formula determination. The critical annotation step is therefore improved. FTICR-MS also have the

advantage of the sensitivity and dynamic range that allows the observation of features that could not be detected in UHPLC-MS/MS.

Based on FTICR-MS data, global metabolic profiling was established for *S. cerevisiae* extracts. Molecular formulae were automatically attributed during data treatment, then O/C ratios were plotted against H/C ratios to create van Krevelen diagrams. These diagrams are bidimensional graphical representations of the global elemental composition of each sample. The van Krevelen diagrams presented in Figure 3a depicts a qualitative description of the four extracts, that is to say a snapshot of what *S. cerevisiae* can produce when cultivated on the studied medium. The dot color testifies the chemical composition in C, H, O, N and S of the determined formulae, so the compositional differences between extracts may be examined. In Figure S5, blue dots represent the common masses detected both by FTICR-MS and LC-MS/MS. Thus, more features are observed in FTICR-MS analyses, which enhanced the metabolite coverage of the study by its sensitivity and dynamic range. Namely, 603 and 404 features were detected in positive ion mode (respectively in DE extract and in M extract) and 404 and 304 features in negative ion mode. The accuracy of this MS technology reveals the richness of metabolite classes within the yeast extracts. By comparing our data with previously described LC-MS/MS data (Figure S6), alkaloids, lipids and amino-acids were placed. Van Krevelen diagrams from the literature [21,32] allowed to confirm lipid region and amino-acid region, and to locate small acids. Compared to these papers, a very few compounds of the "oxy-aromatics" region were detected.

The *S. cerevisiae* extracts appear to have a highly diverse content of alkaloids of diverse polarities, as it was seen according to LC-MS/MS results. Many were detected in the DE extract. Alkaloids are often found in non-polar extracts, but in this case, there is a lot of CHON and CHONS molecules in the methanol extract too. These alkaloids present in the M extract are more polar than average and may contain one or several hydroxyl groups. A lot of CHON signals are found in the M POS van Krevelen diagram, in an area classically dedicated to lipids and non-polar compounds. These could be saturated alkaloids or amino-acids derivatives with hydroxyls for example. Van Krevelen diagrams were built to discriminate nitrogen-rich features from others (Fig. S7): the bluer the dot, the more nitrogen-rich is the formula. We can see compounds with a lot of nitrogen, confirming an overlay between the alkaloid region and lipid region (indicated on Figure 3a). The methanol extracted a lot of alkaloids with a high H/C ratio. Van Krevelen diagrams are a useful tool for the comparison and the differentiation of the different yeast extracts, here from the same strain, but it could be used to compare extracts from different strains or species.

The histograms shown in Figure 3b represent the distribution of masses among chemical compositions (CHO, CHOS, CHON and CHONS). The most represented composition is CHON, which corresponds to the numerous alkaloids and amino-acid derivatives found in the samples, followed by CHO, CHOS and CHONS, and far ahead from CH, CHNS and phosphorus type molecules. The fermentation ability of yeast explains the prominence of CHON and CHONS chemical space. Data were compared to build Venn diagrams (Figure 3c), comparing for each polarity the presence of detected masses between both DE and M extracts. As in LC-MS/MS data, DE extract analyses provided more features than M extract analyses, which can be attributable to the sequence of extraction (DE comes first and extracts more metabolites). Positive ionization mode also led to more detected masses. Ions that are common between DE and M extract are found in the intersection of the two circles. The amount of common masses is significantly smaller than each extract exclusive species. This is explained by the polarity difference of the used solvent mixtures, and by the order of the extractions (the culture is extracted by dichloromethane / ethyl acetate, then by methanol). The methanol extraction picked some metabolites that dichloromethane / ethyl acetate did not. It confirms the value of using two different extractions to expand the metabolite coverage.

3.3. *Saccharomyces cerevisiae* secondary metabolites

The annotated compounds are listed in Table 1, together with their annotation level, the origin of the extract and the ionization polarity. We also summarized in Figure 2d the proportions of the different substance types. All the detected lipids may be part of the membrane system remaining after the solvent-based extraction. Except for glycero-phosphoryl choline, they are all detected in the DE extract, the most non-polar extract. Amino-acids may come from strain primary metabolism, but are also derived compounds such as N-Acetyl-methyltryptophan or Kynurenic acid. "Other" compile diverse families, such as sugars. Alkaloids are a major family part of the small molecules of this strain, in the form of indole alkaloids or quinoline alkaloids. All these compounds are presumably secondary metabolites produced by the yeast, such as Paspalic acid (Peak ID 69), which is an indolo-isoprenic alkaloid identified in *Claviceps purpurea* [33] and used in the hemisynthesis of therapeutic alkaloids. The (2S,5R)-5-(1H-Indol-3-yl)-4-oxo-2,3,5,6-tetrahydro-1H-azepino[4,5-b]indole-2-carboxylic acid (Peak ID 75, also called Malassezindole A) is an indole alkaloid involved in the tryptophan metabolism [34]. Some other features have been sparsely described like Ethyl 3-methyl-9H-carbazole-9-carboxylate (Peak ID 62) and 5-(3-Methyl-2-butenyl)-1H-indole-3-carboxylic acid (Peak ID 84). To the best of our knowledge, this is the first time that these compounds have been described in *S. cerevisiae*. It is important to note that focused structural studies, such as NMR

experiments or MS experiments with standards, are necessary to bring structural evidences and convert putative annotations into identifications. The possibility of several isomers and isobars still exists. However, we obtained evidence of the presence of many unexpected compounds so far, such as the indole and quinoline alkaloids described here for the first time in this species.

Previous studies have implemented gas chromatography to analyze *S. cerevisiae* secondary metabolites [7,8]. Compared to this work, the *S.cerevisiae* strain was cultivated in a solid medium, the extraction method started with a dichloromethane / ethyl acetate mixture to obtain a moderately polar fraction followed by a second extraction with methanol. It was expected to find more diversified chemical classes, the method was rapid and presented a good level of annotation confidence. The GC-MS analysis by Al-Jassani et al. [7] led to non-polar and volatile compound determination such as terpenoids, whereas we characterized more polar or averagely polar compounds. The two studies appear to be complementary, the GC-MS workflow determined the volatile molecules synthetized during fermentation, while we analyzed intracellular and excreted secondary metabolites. The differences may also be explained by the difference between the strains, the difference of culture media or the sample derivatization. Palomino-Schätzlein et al. [10] have profiled *S. cerevisiae* global metabolism by NMR. Apart from the analytical technique, the differences in the used strains, the culture media or the extraction method are the reason why we determined different secondary metabolites and less molecules from primary metabolism, such as nucleotides and all the amino-acids. The same observation is done by comparing the work by Farrés et al. [9] who studied *S. cerevisiae* metabolism by LC-MS; the primary amino-acids L-Tryptophan or L-Methionine are also found in our work but not displayed as they were found in the culture medium too.

The methodology described here, summarized in Figure 1, is very versatile and could also be applied to many types of biological extracts. Globally, more features were described by positive ionization mode analysis, and more masses were detected in the first extract (DE extract), whether by visualizing FTICR-MS results or UHPLC-MS/MS molecular networks. The solvent mixture permitted to extract a wide range of polar, non-polar and averagely polar metabolites. The extraction step is crucial and has to be optimized for every type of analyzed sample. Treatment steps are also essential; the optimization of process parameters is a crucial point to care about for each analysis batch. Finally, decisive steps in this kind of study are manual investigation and data interpretation after the automated processes.

4. Conclusion

Throughout this study, *Saccharomyces cerevisiae* S288c strain cultures were extracted then dereplicated *via* a combination of mass spectrometry techniques. UHPLC-MS/MS-feature-based molecular networking provides accurate and reliable annotations, with the advantage of ionization suppression reduction and discrimination of isomers. In addition, FTICR-MS represents a valuable technology to complement our results by throwing away doubts within compound annotation, by means of accurate molecular formula determination. FTICR-MS data also brought another dimension to the metabolite mapping through a global molecular representation.
The most important challenge in most of metabolomics studies is the spectral data interpretation. Thanks to the large polarity extraction range, the data processing steps and the multiple available databases, 128 features *S. cerevisiae* exclusive features were detected, with 55 % of them annotated in positive ion mode, and 44 % in negative (*level 1, 2* and *3* annotations). The annotation propagation was permitted by the molecular networks that were built, with bigger networks in positive ion mode. A rich molecular network was obtained for the indole and quinoline-type alkaloids.
The direct injection FTICR-MS analysis detected many more features, due to its great sensitivity. It also largely enhanced the annotation levels of the UHPLC-MS/MS analysis, going up to 89 % annotations in positive ion mode, and to 87 % in negative ion mode. This analysis brought van Krevelen diagrams as global metabolic profiling representations and demonstrating the richness of the yeast extracts in term of metabolites.
Among the annotated features, 58 % of alkaloids were found. Our work revealed many alkaloids produced by this strain that could expand fungal databases. The overall representation shows the many alkaloids and amino-acid derived molecules that *S. cerevisiae* can produce when cultivated on the studied medium.
As for *S. cerevisiae* strain S288c, our work provided an effective mapping of compounds by complementing previous studies. This methodology represents the laid foundations for a fast exploration of the metabolites from other yeast strains or species, and can also be applied to various biological samples, as plants or soils.

Acknowledgments
Access to a CNRS FTICR research infrastructure (FR3624) is gratefully acknowledged. Authors would like to thank the CIRM-Levures (Micalis Institute, INRAE, Université Paris-Saclay) for the yeast strain cultures and scientific partnership; the COBRA team for all the support.

Funding

This work was partially supported by Normandie Université (NU), the Centre National de la Recherche Scientifique (CNRS), Université de Rouen Normandie (URN), INSA Rouen Normandie; the European Regional Development Fund (ERDF) N° HN0001343, the European Union's Horizon 2020 Research Infrastructures program (Grant Agreement 731077), the Région Normandie, and the Laboratoire d'Excellence (LabEx) SynOrg (ANR-11-LABX-0029).

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Availability of data and material

The metabolomics and metadata reported in this paper are available *via* MetaboLights [35] on www.ebi.ac.uk/metabolights/MTBLS1701 study identifier MTBLS1701.

Author Contributions

Olivier Perruchon: Methodology, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Isabelle Schmitz-Afonso: Conceptualization, Methodology, Investigation, Ressources, Writing - Review & Editing. Cécile Grondin: Conceptualization, Ressources, Writing - Review & Editing. Serge Casaregola: Conceptualization, Ressources, Writing - Review & Editing. Carlos Afonso: Conceptualization, Investigation, Ressources, Writing - Review & Editing, Funding acquisition. Abdelhakim Elomri: Conceptualization, Methodology, Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

References

[1] C.P. Kurtzman, J.W. Fell, T. Boekhout, The Yeasts: A Taxonomic Study 5th Edition, 2011.

[2] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, Life with 6000 Genes, Science (80-.). 274 (1996) 546–567. https://doi.org/10.1126/science.274.5287.546.

[3] J.L. Legras, D. Merdinoglu, J.M. Cornuet, F. Karst, Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history, Mol. Ecol. 16 (2007) 2091–2102. https://doi.org/10.1111/j.1365-294X.2007.03266.x.

[4] J. Nielsen, Production of biopharmaceutical proteins by yeast, Bioengineered. 4 (2013) 207–211. https://doi.org/10.4161/bioe.22856.

[5] W. Jilg, M. Schmidt, G. Zoulek, B. Lorbeer, B. Wilske, F. Deinhardt, Clinical Evaluation of a Recombinant Hepatitis B Vaccine, Lancet. 324 (1984) 1174–1175. https://doi.org/10.1016/S0140-6736(84)92740-5.

[6] K.J. Hofmann, J.C. Cook, J.G. Joyce, D.R. Brown, L.D. Schultz, H.A. George, M. Rosolowsky, K.H. Fife, K.U. Jansen, Sequence Determination of Human Papillomavirus Type 6a and Assembly of Virus-like Particles in Saccharomyces cerevisiae, Virology. 209 (1995) 506–518. https://doi.org/10.1006/viro.1995.1283.

[7] M.J. Al-Jassani, G.J. Mohammed, I.H. Hameed, Secondary metabolites analysis of Saccharomyces cerevisiae and evaluation of antibacterial activity, Int. J. Pharm. Clin. Res. 8 (2016) 303–314.

[8] A. Tejero Rioseras, D. Garcia Gomez, B.E. Ebert, L.M. Blank, A.J. Ibáñez, P.M.L. Sinues, Comprehensive Real-Time Analysis of the Yeast Volatilome, Sci. Rep. 7 (2017) 14236. https://doi.org/10.1038/s41598-017-14554-y.

[9] M. Farrés, B. Piña, R. Tauler, Chemometric evaluation of Saccharomyces cerevisiae metabolic profiles using LC–MS, Metabolomics. 11 (2014) 210–224. https://doi.org/10.1007/s11306-014-0689-z.

[10] M. Palomino-Schätzlein, M.M. Molina-Navarro, M. Tormos-Pérez, S. Rodríguez-Navarro, A. Pineda-Lucena, Optimised protocols for the metabolic profiling of S. cerevisiae by 1H-NMR and HRMAS spectroscopy, Anal. Bioanal. Chem. 405 (2013) 8431–8441. https://doi.org/10.1007/s00216-013-7271-9.

[11] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z. Lai, S.S. Mehta, G. Wohlgemuth, D.K. Barupal, M.R. Showalter, M. Arita, O. Fiehn, Identification of small molecules using accurate mass MS/MS search, Mass Spectrom. Rev. 37 (2018) 513–532. https://doi.org/10.1002/mas.21535.

[12] J.L. Wolfender, J.M. Nuzillard, J.J.J. Van Der Hooft, J.H. Renault, S. Bertrand, Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics, Anal. Chem. 91 (2019) 704–742. https://doi.org/10.1021/acs.analchem.8b05112.

[13] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing,

and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics. 11 (2010) 395. https://doi.org/10.1186/1471-2105-11-395.

[14] F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MZmine 2 Data-Preprocessing to Enhance Molecular Networking Reliability, Anal. Chem. 89 (2017) 7836–7840. https://doi.org/10.1021/acs.analchem.7b01563.

[15] T. Jewison, C. Knox, V. Neveu, Y. Djoumbou, A.C. Guo, J. Lee, P. Liu, R. Mandal, R. Krishnamurthy, I. Sinelnikov, M. Wilson, D.S. Wishart, YMDB: The yeast metabolome database, Nucleic Acids Res. 40 (2012) 815–820. https://doi.org/10.1093/nar/gkr916.

[16] M. Ramirez-Gaona, A. Marcu, A. Pon, A.C. Guo, T. Sajed, N.A. Wishart, N. Karu, Y.D. Feunang, D. Arndt, D.S. Wishart, YMDB 2.0: A significantly expanded version of the yeast metabolome database, Nucleic Acids Res. 45 (2017) D440–D445. https://doi.org/10.1093/nar/gkw1058.

[17] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: The human metabolome database, Nucleic Acids Res. 35 (2007) 521–526. https://doi.org/10.1093/nar/gkl923.

[18] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, MassBank: A public repository for sharing mass spectral data for life sciences, J. Mass Spectrom. 45 (2010) 703–714. https://doi.org/10.1002/jms.1777.

[19] T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics. 8 (2007) 105. https://doi.org/10.1186/1471-2105-8-105.

[20] M. Ghaste, R. Mistrik, V. Shulaev, Applications of fourier transform ion cyclotron resonance (FT-ICR) and orbitrap based high resolution mass spectrometry in metabolomics and lipidomics, Int. J. Mol. Sci. 17 (2016) 816. https://doi.org/10.3390/ijms17060816.

[21] C. Roullier-Gall, J. Signoret, D. Hemmler, M.A. Witting, B. Kanawati, B. Schäfer, R.D. Gougeon, P. Schmitt-Kopplin, Usage of FT-ICR-MS Metabolomics for Characterizing the Chemical Signatures of Barrel-Aged Whisky, Front. Chem. 6 (2018) 29. https://doi.org/10.3389/fchem.2018.00029.

[22] E. Chekmeneva, G. Dos Santos Correia, M. Gómez-Romero, J. Stamler, Q. Chan, P. Elliott, J.K. Nicholson, E. Holmes, Ultra-Performance Liquid Chromatography-High-Resolution Mass Spectrometry and Direct Infusion-High-Resolution Mass Spectrometry for Combined Exploratory and Targeted Metabolic Profiling of Human Urine, J. Proteome Res. 17 (2018) 3492–3502. https://doi.org/10.1021/acs.jproteome.8b00413.

[23] T. Nagao, D. Yukihira, Y. Fujimura, K. Saito, K. Takahashi, D. Miura, H. Wariishi, Power of isotopic fine structure for unambiguous determination of metabolite elemental compositions: In silico evaluation and metabolomic application, Anal. Chim. Acta. 813 (2014) 70–76. https://doi.org/10.1016/j.aca.2014.01.032.

[24] D.W. van Krevelen, Graphical-statistical method for the study of structure and reaction processes of coal, Fuel. 29 (1950) 269–284.

[25] S. Kim, R.W. Kramer, P.G. Hatcher, Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, Anal. Chem. 75 (2003) 5336–5344. https://doi.org/10.1021/ac034415p.

[26] 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.P. Boya, 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. 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.

[27] J. Watrous, P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, P.C. Dorrestein, Mass spectral molecular networking of living microbial colonies, Proc. Natl. Acad. Sci. 109 (2012) E1743–E1752. https://doi.org/10.1073/pnas.1203689109.

[28] R.A. Quinn, L.F. Nothias, O. Vining, M. Meehan, E. Esquenazi, P.C. Dorrestein, Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy, Trends Pharmacol. Sci. 38 (2017) 143–154. https://doi.org/10.1016/j.tips.2016.10.011.

[29] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics. 3 (2007) 211–221. https://doi.org/10.1007/s11306-007-0082-2.

[30] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks, Anal. Chem. 89 (2017) 8696–8703. https://doi.org/10.1021/acs.analchem.7b00947.

[31] J. Maillard, J. Ferey, C.P. Rüger, I. Schmitz-Afonso, S. Bekri, T. Gautier, N. Carrasco, C. Afonso, A. Tebani, Optimization of ion trajectories in a dynamically harmonized Fourier-transform ion cyclotron resonance cell using a design of experiments strategy, Rapid Commun. Mass Spectrom. 34 (2020). https://doi.org/10.1002/rcm.8659.

[32] R.D. Gougeon, M. Lucio, M. Frommberger, D. Peyron, D. Chassagne, H. Alexandre, F. Feuillat, A. Voilley, P. Cayot, I. Gebefügi, N. Hertkorn, P. Schmitt-Kopplin, The chemodiversity of wines can reveal a metabologeography expression of cooperage oak wood, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 9174–9179. https://doi.org/10.1073/pnas.0901100106.

[33] T. Correia, N. Grammel, I. Ortel, U. Keller, P. Tudzynski, Molecular Cloning and Analysis of the Ergopeptine Assembly System in the Ergot Fungus Claviceps purpurea, Chem. Biol. 10 (2003) 1281–1292. https://doi.org/10.1016/j.chembiol.2003.11.013.

[34] B. Irlinger, A. Bartsch, H.J. Krämer, P. Mayser, W. Steglich, New tryptophan metabolites from cultures of the lipophilic yeast Malassezia furfur, Helv. Chim. Acta. 88 (2005) 1472–1485. https://doi.org/10.1002/hlca.200590118.

[35] K. Haug, K. Cochrane, V.C. Nainala, M. Williams, J. Chang, K.V. Jayaseelan, C. O'Donovan, MetaboLights: a resource evolving in response to the needs of its scientific community, Nucleic Acids Res. 48 (2020) D440–D444. https://doi.org/10.1093/nar/gkz1019.

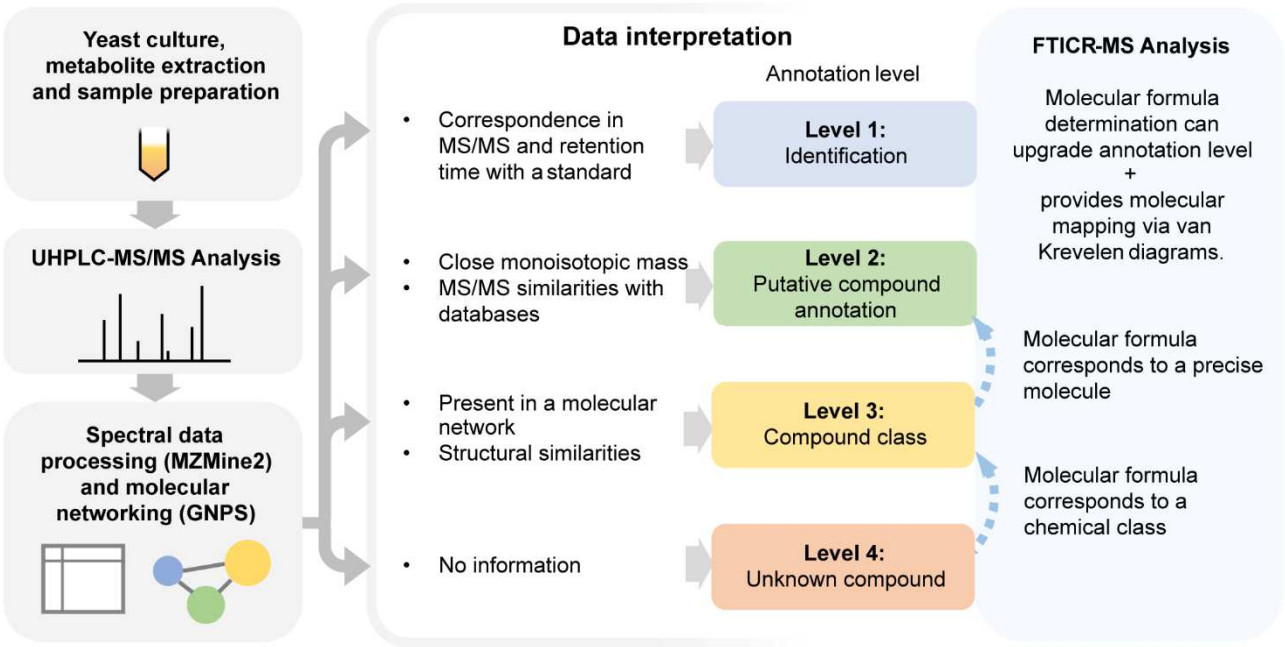


Figure 1: Summary diagram of the yeast extract analysis workflow with the different annotation levels. Metabolites are extracted from yeast cultures to prepare the samples that are analyzed through UHPLC-MS/MS. Spectral data are processed and used to build molecular networks. Annotation are attributed and can be upgraded thanks to FTICR-MS analysis.

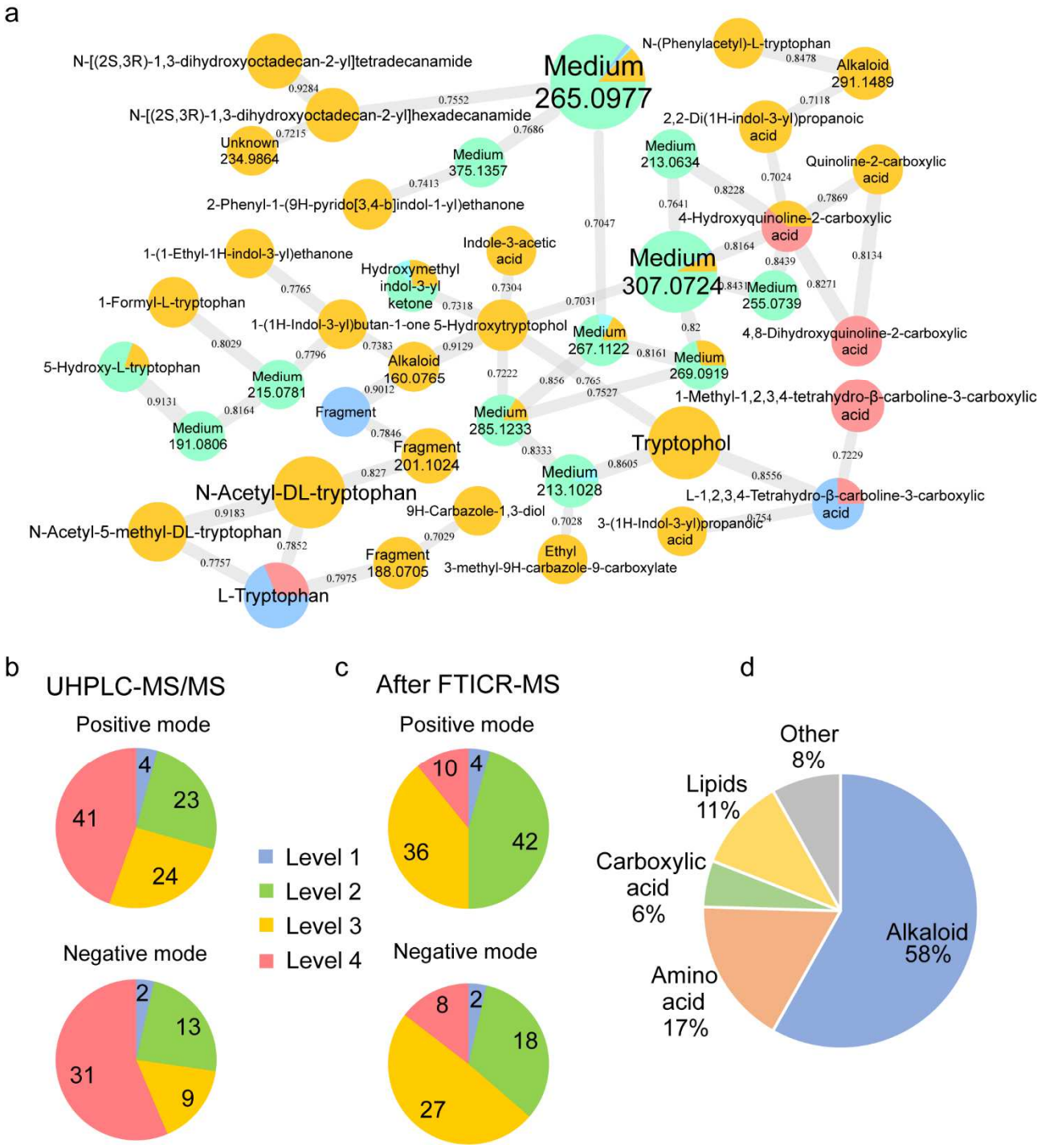


Figure 2: Results from UHPLC-MS/MS-molecular networking workflow on *Saccharomyces cerevisiae* extracts. a: example of alkaloid network built in GNPS (positive mode). Color is the sample origin (Orange: DE; Red: M; Green and Blue: culture medium). Node labels display the proposed metabolite, or the compound class with the *m/z* ratio. Node size is proportional to intensity, edge size to cosine score. b: features from the analysis by UHPLC-MS/MS. The number of features is displayed after a comparison with databases and GNPS annotations. c: feature number after FTICR-MS analysis and deeper investigation step. d: compound classes pie chart based on UHPLC-MS/MS analysis enhanced by FTICR-MS. Compounds from DE extract and M extract analyzed in positive and negative mode are present.

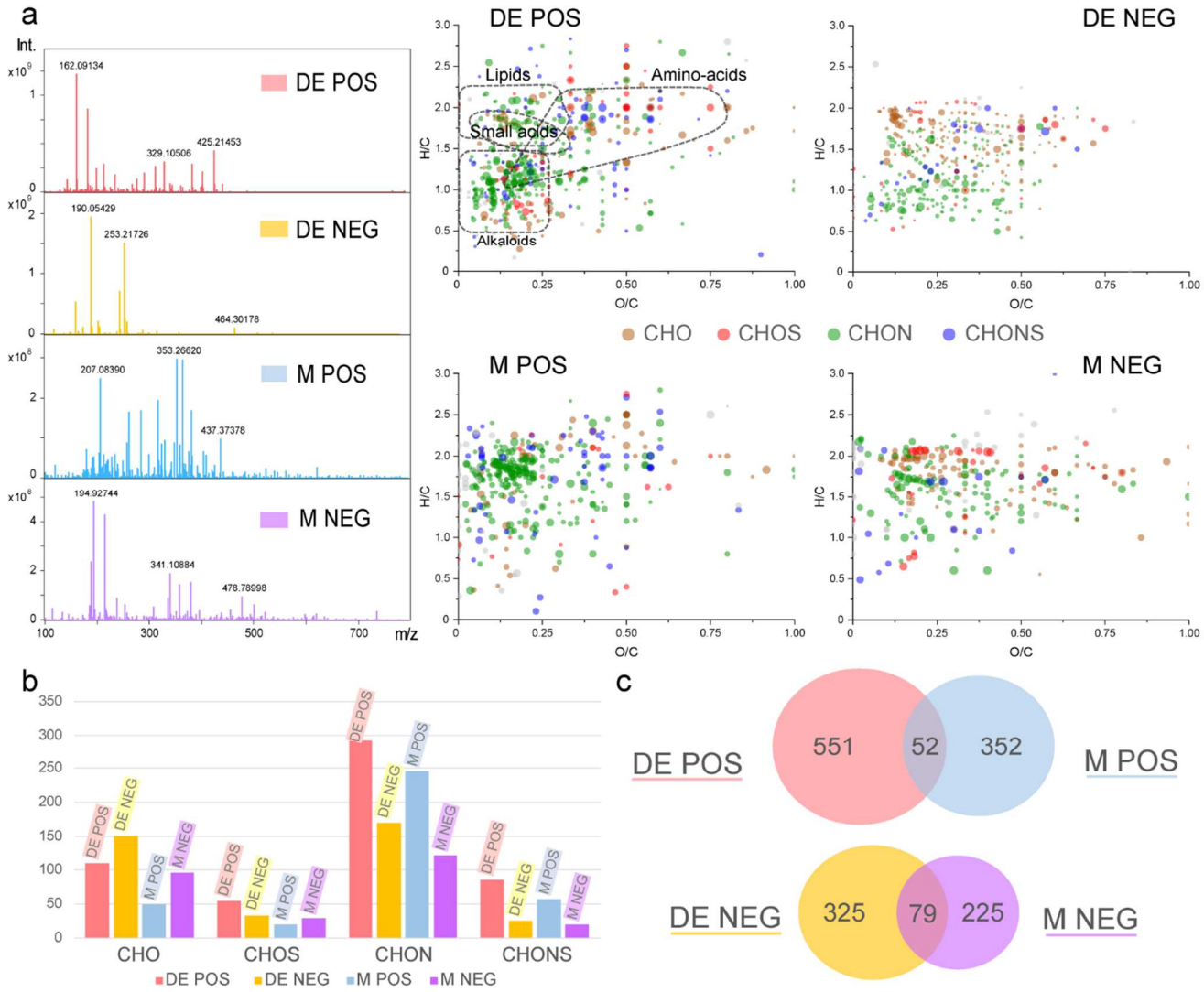


Figure 3: Representations of FTICR-MS data from *Saccharomyces cerevisiae* extract analysis after chemical formula attributions. a: FTICR-MS spectra and corresponding O/C vs. H/C van Krevelen diagrams. Brown dots are CHO features; red dots are CHOS; green dots are CHON; blue dots are CHONS. b: histograms of FTICR-MS data distribution among chemical composition types (four most numerous classes) in number of features. Red bars are for DE extract in POS ionization mode; yellow bars are for DE NEG; blue bars are for M POS; purple bars are for M NEG. c: Venn diagrams comparing attribution number between DE extracts and M extracts, in positive and negative ionization modes.

Table.1: Proposed metabolites from *Saccharomyces cerevisiae* analysis. 128 compounds from DE extract and M extract analyzed in positive and negative mode are present. Concerning features detected in both ionisation modes, the observed *m/z* value comes from the UHPLC-MS/MS analysis in positive (POS) and/or negative mode (NEG).

Peak ID	RT (min)	Extract	Proposed metabolite	Type of ion	Observed *m/z*	Molecular formula	Annotation level
1	1.07	M	Unknown	Positive	381.0779	Unknown	4
2	1.07	M	Choline glycerophosphate	M + H	258.1101	$C_8H_{20}NO_6P$	2
3	1.10	DE	Alkylamine	M + H	102.1283	$C_6H_{15}N$	3
4	1.13	M	Unknown	M + H	365.1074	$C_{18}H_{20}O_6S$	4
5	1.13	M	β-D-Glucopyranose-3-amino-3-deoxy-β-D-glucopyranose (1:1)	M + H	360.1511	$C_{12}H_{25}NO_{11}$	2
6	1.13	DE	Amino acid	M + H	176.0911	$C_7H_{13}NO_4$	3
7	1.24	DE	N-Acetylmethionine sulfoxide	M + H M – H	208.0637 206.0485	$C_7H_{13}NO_4S$	2
8	1.28	M	Unknown	Positive	308.0913	Unknown	4
9	1.75	DE; M	Amino-acid	M + H	208.0636	$C_7H_{13}NO_4S$	3
10	3.45	DE	Lipids	M – H	145.0503	$C_6H_{10}O_4$	3
11	5.48	DE	Unknown	Negative	327.0581	Unknown	4
12	5.48	DE	2-Amino-3-hydroxybenzoic acid	M + H	154.0504	$C_7H_7NO_3$	2
13	5.58	DE	Indole-3-acetaldehyde	M + H	160.0760	$C_{10}H_9NO$	2
14	6.07	DE	Carboxylic acid	M – H	117.0550	$C_5H_{10}O_3$	3
15	6.66	DE	Hydroxylated sulfated ester	M – H	193.0532	$C_7H_{14}O_4S$	3
16	6.70	DE	Carboxylic acid	M – H	161.0814	$C_7H_{14}O_4$	3
17	6.86	DE	Unknown	Positive	293.9688	Unknown	4
18	6.99	DE; M	N-Acetyl-L-methionine	M + H M – H	192.0693 190.0539	$C_7H_{13}NO_3S$	2
19	7.04	M	4,8-Dihydroxyquinoline-2-carboxylic acid	M + H M – H	206.0446 204,0298	$C_{10}H_7NO_4$	2
20	7.37	DE	Carboxylic acid	M – H	194.0445	$C_9H_9NO_4$	3
21	7.40	DE M	4-Hydroxyquinoline-2-carboxylic acid	M + H M – H	190.0503 188,0356	$C_{10}H_7NO_3$	1
22	7.44	DE M	Carboxylic acid	M – H	175.0609	$C_7H_{12}O_5$	3
23	7.98	M	1-Methyl-1,2,3,4-tetrahydro-β-carboline-3-carboxylic acid	M + H M – H	231.1126 229.0974	$C_{13}H_{14}N_2O_2$	2
24	8.12	DE	2-Benzamidoacetic acid	M – H	178.0500	$C_9H_9NO_3$	2
25	8.21	DE	2-Isopropylmaleic acid	M – H	157.0502	$C_7H_{10}O_4$	2
26	8.24	DE	5-Hydroxytryptophol	M + H	178.0864	$C_{10}H_{11}NO_2$	2

27	8.24	DE M	N-Propionylmethionine	M + H M − H	206.0848 204.0693	$C_8H_{15}NO_3S$	2
28	8.40	DE	Carboxylic acid	M − H	131.0709	$C_6H_{12}O_3$	3
29	8.71	DE	Alkaloid	M + H	160.0765	$C_{10}H_9NO$	3
30	8.84	DE	N-Acetyl-DL-leucine	M − H	172.0973	$C_8H_{15}NO_3$	2
31	8.86	DE	(1H-Indol-3-yl)methanol	M + H	148.0761	C_9H_9NO	2
32	8.91	DE	Alkaloid	M + H	233.0925	$C_{12}H_{12}N_2O_3$	3
33	8.91	DE	Alkaloid	M + H M − H	251.1023 249.0865	$C_{12}H_{14}N_2O_4$	3
34	9.11	DE	Alkylphenylketone	M − H	209.0826	$C_{11}H_{14}O_4$	3
35	9.23	DE	3-Phenyllactic acid	M − H	165.0554	$C_9H_{10}O_3$	1
36	9.39	DE	Amino-acid	M + H	166.0867	$C_9H_{11}NO_2$	3
37	9.45	DE	N-Acetyl-L-phenylalanine	M − H	206.0808	$C_{11}H_{13}NO_3$	1
38	9.69	DE	1-Formyl-L-tryptophan	M + H	233.0922	$C_{12}H_{12}N_2O_3$	2
39	9.72	DE	2-Hydroxy-3-(1H-indol-3-yl)propanoic acid	M − H	204.0665	$C_{11}H_{11}NO_3$	2
40	9.81	DE	N-Acetyl-DL-tryptophan	M + H M − H	247.1080 245.0921	$C_{13}H_{14}N_2O_3$	2
41	9.86	DE	Unknown	M − H	277.0816	$C_{13}H_{14}N_2O_5$	4
42	9.95	DE	1H-Indazole-3-carbaldehyde	M + H	147.0555	$C_8H_6N_2O$	2
43	10.51	DE	Alkaloid	M − H	220.0965	$C_{12}H_{15}NO_3$	3
44	10.57	DE	Alkaloid	M + H	291.1488	$C_{19}H_{18}N_2O$	3
45	10.57	DE	Indole-3-acetic acid	M + H M − H	176.0705 174.0557	$C_{10}H_9NO_2$	1
46	10.57	DE	Tryptophol	M + H	162.0920	$C_{10}H_{11}NO$	1
47	10.57	DE	Unknown	M + H	319.1443	$C_{20}H_{18}N_2O_2$	4
48	10.63	DE	Alkaloid	M + H	335.1385	$C_{20}H_{18}N_2O_3$	3
49	10.65	DE	Sugar	M − H	327.0957	$C_{10}H_{21}N_2O_8P$	3
50	10.71	DE	N-Acetyl-5-methyl-DL-tryptophan	M + H M − H	261.1233 259.1078	$C_{14}H_{16}N_2O_3$	2
51	10.89	DE	Unknown	M − H	337.1240	$C_{16}H_{22}N_2O_4S$	4
52	10.98	DE	3-Acetylindole	M + H	160.0764	$C_{10}H_9NO$	1
53	11.12	DE	N-(5-Methyl-3-oxohexyl)alanine	M − H	200.1283	$C_{10}H_{19}NO_3$	2
54	11.18	DE	Amino-acid	M + H M − H	234.1155 232.1000	$C_{10}H_{19}NO_3S$	3
55	11.19	DE	Alkaloid	M + H	188.1091	$C_{12}H_{13}NO$	3
56	11.32	DE	3-(1H-Indol-3-yl)propanoic acid	M + H	190.0864	$C_{11}H_{11}NO_2$	2
57	11.57	DE	Unknown	M − H	323.0666	$C_{17}H_{12}N_2O_5$	4

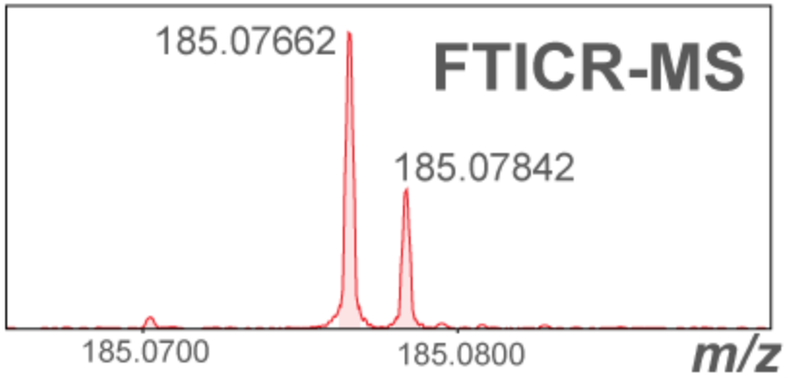
58	11.63	DE	Alkaloid	M − H	234.1125	$C_{13}H_{17}NO_3$	3
59	11.66	DE	Amino-acid	M + H M − H	268.1006 266.0849	$C_{13}H_{17}NO_3S$	3
60	11.74	DE	Alkaloid	M + H	379.1102	$C_{21}H_{18}N_2O_3S$	3
61	11.80	DE	6-Methylquinoline	M + H	144.0807	$C_{10}H_9N$	2
62	11.86	DE	Ethyl 3-methyl-9H-carbazole-9-carboxylate	M + H	254.1176	$C_{16}H_{15}NO_2$	2
63	12.21	DE	1,2,3,5-Tetrahydroxy-10H-acridin-9-one	M + H	260.0558	$C_{13}H_9NO_5$	2
64	12.27	DE	Unknown	Positive	234.9864	Unknown	4
65	12.41	DE	Alkaloid	M + H M − H	285.1229 283.1077	$C_{16}H_{16}N_2O_3$	3
66	12.48	DE	Alkaloid	M + H	254.1176	$C_{16}H_{15}NO_2$	3
67	12.56	DE	2,4-Dihydroxydodecanoic acid	M − H	231.1583	$C_{12}H_{24}O_4$	2
68	12.76	DE	6-(3-Methylbut-2-enyl)-1H-indole-3-carboxylic acid	M + H	230.1176	$C_{14}H_{15}NO_2$	2
69	12.76	DE	Paspalic acid	M + H	269.1290	$C_{16}H_{16}N_2O_2$	2
70	12.95	DE	Alkaloid	M − H	374.1151	$C_{21}H_{17}N_3O_4$	3
71	12.97	DE	Methanone, (6-hydroxy-9H-pyrido[3,4-b]indol-1-yl)(5-methoxy-1H-indol-3-yl)-	M + H	358.1188	$C_{21}H_{15}N_3O_3$	2
72	13.03	DE	Methyl 2-(1H-indol-3-yl)acetate	M + H	190.0868	$C_{11}H_{11}NO_2$	2
73	13.14	DE	Alkaloid	M + H	188.1075	$C_{12}H_{13}NO$	3
74	13.22	DE	Alkaloid	M + H	243.0761	$C_{13}H_{10}N_2O_3$	3
75	13.22	DE	(2S,5R)-5-(1H-Indol-3-yl)-4-oxo-2,3,5,6-tetrahydro-1H-azepino[4,5-b]indole-2-carboxylic acid	M + H M − H	360.1339 358.1188	$C_{21}H_{17}N_3O_3$	2
76	13.36	DE	Alkaloid	M + H	218.0812	$C_{12}H_{11}NO_3$	3
77	13.36	DE	Alkaloid	M + H	335.1385	$C_{20}H_{18}N_2O_3$	3
78	13.39	DE	Alkaloid	M − H	289.1345	$C_{19}H_{18}N_2O$	3
79	13.42	DE	1-Acetylindole	M + H	160.0767	$C_{10}H_9NO$	2
80	13.42	DE	Quinoline-2-carboxylic acid	M + H	174.0557	$C_{10}H_7NO_2$	2
81	13.45	DE	Alkaloid	M − H	303.1130	$C_{19}H_{16}N_2O_2$	3
82	13.49	DE	2,2-Di(1H-indol-3-yl)propanoic acid	M + H	305.1292	$C_{19}H_{16}N_2O_2$	2
83	13.49	DE	Alkaloid	M + H	407.1971	$C_{24}H_{26}N_2O_4$	3
84	13.57	DE	5-(3-Methyl-2-butenyl)-1H-indole-3-carboxylic acid	M + H	230.1180	$C_{14}H_{15}NO_2$	2

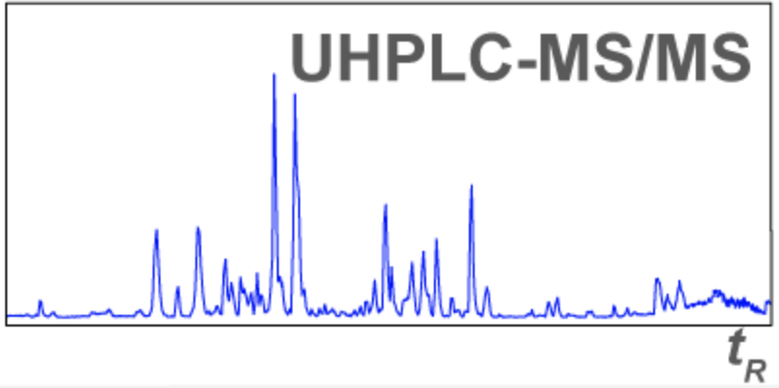
85	13.63	DE	Alkaloid	M + H	333.1236	$C_{20}H_{16}N_2O_3$	3
86	13.69	DE	Alkaloid	M + H	289.1332	$C_{19}H_{16}N_2O$	3
87	13.83	DE	Alkaloid	M + H	243.0765	$C_{13}H_{10}N_2O_3$	3
88	13.83	DE	Alkaloid	M + H M − H	360.1348 358.1189	$C_{21}H_{17}N_3O_3$	3
89	13.89	DE	Unknown	Negative	564.2285	Unknown	4
90	13.89	DE	Unknown	M + H	520.2361	$C_{36}H_{29}N_3O$	4
91	14.04	DE	Alkaloid	M + H	190.0871	$C_{11}H_{11}NO_2$	3
92	14.31	DE	3,4-Bis(1H-indole-3-yl)-2,5-dihydrofuran-2-one	M + H	315.1129	$C_{20}H_{14}N_2O_2$	2
93	14.39	DE	Alkaloid	M + H	188.1071	$C_{12}H_{13}NO$	3
94	14.41	M	Tauroursodeoxycholic acid	M + H M − H	500.3040 498.2892	$C_{26}H_{45}NO_6S$	2
95	14.43	DE	Unknown	M − H	347.1758	$C_{22}H_{24}N_2O_2$	4
96	14.51	DE	3-[[4-(3-Carboxyanilino)-3,6-dioxocyclohexa-1,4-dien-1-yl]amino]benzoic acid	M + H	379.0925	$C_{20}H_{14}N_2O_6$	2
97	14.68	DE	Alkaloid	M − H	287.1180	$C_{19}H_{16}N_2O$	3
98	14.72	DE	2-Phenyl-1-(9H-pyrido[3,4-b]indol-1-yl)ethanone	M + H	287.1178	$C_{19}H_{14}N_2O$	2
99	14.72	DE	Alkaloid	M + H M − H	333.1232 331.1078	$C_{20}H_{16}N_2O_3$	3
100	14.86	DE	Alkaloid	M + H M − H	291.1489 289.1337	$C_{19}H_{18}N_2O$	3
101	14.86	DE	Alkaloid	M + H	305.1288	$C_{19}H_{16}N_2O_2$	3
102	14.86	DE	N-(Phenylacetyl)-L-tryptophan	M + H	323.1396	$C_{19}H_{18}N_2O_3$	2
103	14.86	DE	Unknown	Positive	310.1227	Unknown	4
104	14.88	DE	Lipids	M − H	259.1902	$C_{14}H_{28}O_4$	3
105	15.01	DE	1-(1H-Indol-3-yl)butan-1-one	M + H	188.1067	$C_{12}H_{13}NO$	2
106	15.21	DE	Alkaloid	M + H	315.1133	$C_{20}H_{14}N_2O_2$	3
107	15.79	DE	2-Quinolinemethanethiol acetate	M − H	216.0481	$C_{12}H_{11}NOS$	2
108	15.82	DE	Unknown	Positive	616.1765	Unknown	4
109	16.07	DE	Unknown	M − H	315.1131	$C_{20}H_{16}N_2O_2$	4
110	16.09	DE	9H-Carbazole-1,3-diol	M + H	200.0691	$C_{12}H_9NO_2$	2
111	16.28	DE	Alkaloid	M − H	359.1389	$C_{22}H_{20}N_2O_3$	3
112	16.34	DE	2-Methylquinoline	M + H	144.0812	$C_{10}H_9N$	2
113	16.49	DE	(2S,3R)-2-aminooctadecane-1,3-diol	M + H	302.3047	$C_{18}H_{39}NO_2$	2
114	16.57	DE	Alkaloid	M + H	289.1330	$C_{19}H_{16}N_2O$	3

115	16.79	DE	Alkaloid	M + H	188.1070	$C_{12}H_{13}NO$	3
116	17.17	DE	Polycyclic alcohol	M − H	253.1801	$C_{15}H_{26}O_3$	3
117	17.52	DE	1-(1-Ethyl-1H-indol-3-yl)-ethanone	M + H	188.1073	$C_{12}H_{13}NO$	2
118	18.04	DE	Lipids	M + H	330.3366	$C_{20}H_{43}NO_2$	3
119	20.21	DE	Unknown	M − H	342.2635	$C_{19}H_{37}NO_4$	4
120	20.58	DE	Alkene	M + H	205.1944	$C_{15}H_{24}$	3
121	21.09	DE	Amino-acid	M − H	357.1886	$C_{14}H_{26}N_6O_5$	3
122	21.27	DE	Unknown	Positive	343.2023	Unknown	4
123	21.37	DE	Lipids	M + H M − H	255.2309 253.2163	$C_{16}H_{30}O_2$	3
124	21.44	DE	Lipids	M + H	319.2843	$C_{18}H_{38}O_4$	3
125	23.00	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]dodecanamide	M + H	484.4716	$C_{30}H_{61}NO_3$	2
126	23.85	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]tetradecanamide	M + H	512.5035	$C_{32}H_{65}NO_3$	2
127	24.63	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]hexadecanamide	M + H	540.5356	$C_{34}H_{69}NO_3$	2
128	25.51	DE	N-[(2S,3R)-1,3-dihydroxyoctadecan-2-yl]octadecanamide	M + H	568.5648	$C_{36}H_{73}NO_3$	2

Saccharomyces cerevisiae

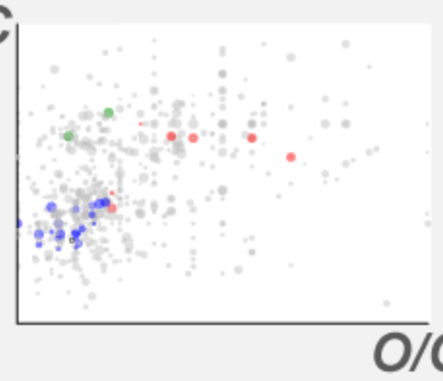
Mass spectrometry analyses

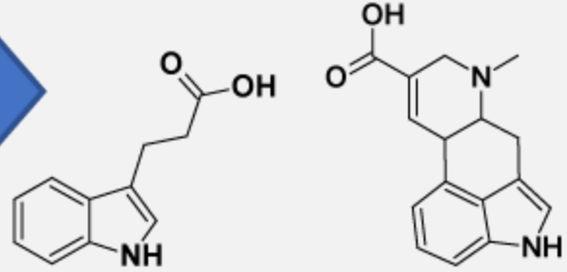




Data interpretation

Metabolic profiling





Metabolite annotation

Molecular networks