HAL open science

Online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry-based untargeted volatile metabolomics for studying emerging complex biopesticides: A proof of concept

Hikmat Ghosson, Delphine Raviglione, Marie-Virginie Salvia, Cédric Bertrand

To cite this version:

Hikmat Ghosson, Delphine Raviglione, Marie-Virginie Salvia, Cédric Bertrand. Online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry-based untargeted volatile metabolomics for studying emerging complex biopesticides: A proof of concept. Analytica Chimica Acta, 2020, 1134, pp.58-74. 10.1016/j.aca.2020.08.016 . hal-02926025

HAL Id: hal-02926025

https://hal.science/hal-02926025

Submitted on 5 Sep 2022

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

Version of Record: https://www.sciencedirect.com/science/article/pii/S0003267020308473
Manuscript_382f2372ab390db5798d720af32021a2

Online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry-based untargeted volatile metabolomics for studying emerging complex biopesticides: a proof of concept

Hikmat Ghosson[1,2,*], Delphine Raviglione[1,2], Marie-Virginie Salvia[1,2,†], Cédric Bertrand[1,2,3,†]

1: PSL Université Paris: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

2: UFR Sciences Exactes et Expérimentales, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

3: S.A.S. AkiNaO, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

†: Equal Contribution

***: Corresponding author: Hikmat Ghosson**

E-mail addresses: hikmat.ghosson@univ-perp.fr / hikmatghosson@gmail.com

Postal address: USR3278 CRIOBE EPHE-CNRS-UPVD, Bâtiment T, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

Tel: +33.6.19.47.12.81

Keywords

Untargeted Metabolomics; Complex biopesticides; Solid Phase Microextraction; Headspace; Gas Chromatography-Mass Spectrometry

Highlights

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

- A novel online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry-based untargeted metabolomics approach was developed to study the environmental fate of an emerging complex bioherbicide: the *Myrica gale* extract.
- A green, non-destructive automated HS-SPME-GC-MS method was developed and applied for a comparative 38-day kinetics experiment to study the fate of herbicide's volatile xenometabolome after its application on soil samples.
- Untargeted metabolomics and multivariate statistical analyses explained the evolution of herbicide residues over time and allowed for the putative annotation of 96 compounds.
- The approach proved its reliability for high throughput analyses and only required a small number of samples that were not destroyed during the study.

Graphical Abstract

Placed at the end of the manuscript.

Abstract

This work introduces a novel online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry-based untargeted metabolomics approach, suggested as an alternative tool to study the environmental fate of volatile xenometabolites in emerging complex biopesticides, *e.g.* the *Myrica gale* methanolic extract herbicide containing several unknown metabolites. A "living" microcosm sample was designed for non-destructive analysis by a 35-minute HS-SPME automated extraction and a 36-minute GC-MS run. A 38-day kinetics study was then applied on two groups of soil samples: control and spiked. Statistical tools were used for the comparative kinetics study. The Principal Component Analysis revealed and explained the evolution and the dissipation of the herbicide volatile xenometabolome over time. The time-series Heatmap and Multivariate Empirical Bayes Analysis of Variance allowed the prioritization of 101 relevant compounds including 22

degradation by-products. Out of them, 96 xenometabolites were putatively identified. They included 63 compounds that are identified as herbicide components for the first time. The Orthogonal Projections to Latent Structures Discriminant Analysis and its Cross-Validation test were used to assess the total dissipation of the herbicide volatile residues. The reproducibility of the method was also assessed. The highest inter-samples *(n = 3)* Peak Area RSD was 7.75%. The highest inter-samples *(n = 3)* and inter-days *(n = 8)* Retention Time SD were 0.43 sec and 3.44 sec, respectively. The work presents a green, non-laborious and high-throughput approach. It required a small number of environmental samples (6 microcosms) that were analyzed 8 times and were not destroyed during the study.

1. Introduction

Pesticides research, development and production are constantly expanding since these chemicals and agents are essential for several anthropogenic and economic activities (*e.g.* agriculture, food production and protection, disease vectors control). Their development however, faces numerous problems due to their potential impact on human health [1–3] and ecology [4,5]. These issues reinforce the requirement and the importance of prior in-depth studies of their fate, impacts and risks on health and environment. Also, the development of new pesticides of natural origins, known as “biopesticides” or “biocontrol agents” (BA), is one of the suggested alternatives to chemical/synthetic pesticides, as they are presumed to be less harmful for human health and environment. Moreover, their dissipation is likely to be relatively fast [6].

Extracted from plants or different types of microorganisms, these emerging natural products are mostly based on one or several bioactive compounds which usually act in a synergic and/or pleiotropic mode of action. Their complex (bio)chemical nature containing several different and unknown molecules and/or macromolecules is requiring new conceptual and

69 analytical challenges for the assessment of their transformation and dissipation. The classic
70 concepts of fate assessment, such as the DT50 approach [7], are non-applicable for such types
71 of complex pesticides. These classic targeted approaches are limited to known compounds
72 and molecules. In addition, the DT50 approach does not consider the transformation products
73 (TPs) of the pesticide, in particular the unknown TPs. Additional protocols and approaches
74 are therefore needed in order to assess the pesticide transformation in the environment, and to
75 study the impact of its application on environmental biodiversity.

76 New analytical proxies were thus suggested as alternative approaches for the emerging
77 complex biopesticides, mainly based on untargeted metabolomics strategies [8,9]. A new
78 approach called Environmental Metabolic Footprinting (EMF) was recently introduced by
79 Patil *et al.* [10] and Salvia *et al.* [11]. This new approach presents the application of the
80 untargeted metabolomics as a universal tool for kinetics studies in order to assess both the fate
81 and impact of different types of complex pesticides. This aims to introduce an integrative
82 concept called "resilience time".

83 In the two previous mentioned works [10,11], kinetics studies were performed on an
84 important number of samples by applying destructive Solid-Liquid Extractions (SLE)
85 followed by Liquid Chromatography-Mass Spectrometry (LC-MS) analysis. They were
86 restricted to the solid phase of the environmental matrices (soil and sediments). However,
87 studying the volatile part of the xenometabolome, *i.e.* pesticide compounds and their TPs, is
88 essential for the risk assessment of these emerging biopesticides, and in particular for
89 products based on plant essential oils, which contain an important amount of volatile and
90 semi-volatile organic compounds. The OECD guidelines for the testing of chemicals and their
91 transformation recommend the consideration of the volatile part [12,13]. In fact, studying
92 pesticide volatile residues can provide complementary information to better understand its
93 environmental fate. In addition, pesticide volatile residues screening allows to assess the

exposure risk to pesticide compounds for farmers/workers, insects and plants, as well as the exposure to their TPs that might be more toxic. Therefore, the aim of the present work is to introduce the concept of a new untargeted metabolomics-based approach, dedicated to analyze and study the volatile residues of emerging complex biopesticides applied on environmental matrices, by using online Headspace-Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry (HS-SPME-GC-MS).

HS-SPME is an appropriate technique for volatile organic compounds analysis. It is based on extracting and isolating these analytes from the sample by adsorbing and concentrating them on the layer of a coated fiber. Thus, they can be eventually desorbed and introduced in the analytical instrument with or without the need of extraction solvents [14,15]. Since its introduction in 1989 by Belardi & Pawliszyn [16], the SPME is still being widely developed and extensively used for different types of targeted and untargeted analytical approaches as broadly described by Reyes-Garcés *et al.* [17]. For pesticides research, several works have been reported and were mainly focused on targeted screening and quantification of pesticide residues in different environmental and food matrices [18–20]. Untargeted screening to study the transformation of pesticides and to identify their by-products was also reported, but in a much fewer number of publications [18,21,22].

SPME presents several advantages as a green, non-destructive and cost-effective technique. Its automation provides additional advantages, particularly for metabolomics approaches, mainly by enhancing the robustness and the reproducibility of the applied extraction method. Moreover, reducing the laborious time-consuming manual work and sample preparation steps is essential for high throughput analyses and to minimize errors related to sample handling. Otherwise, as a green non-destructive method, the application of the HS-SPME reduces the number of environmental samples needed, by making it possible to analyze the same sample for several time points, particularly in case of kinetics tracking study. This can also enhance

119 the performance of the approach by reducing sample preparation and random biological
120 variations-related biases.

121 On the other hand, GC-MS analytical technique provides several advantages concerning
122 untargeted metabolomics. The GC is a suitable separation technique for volatile and semi-
123 volatile organic compounds. It is well known for its significant analytical robustness,
124 affording high chromatographic resolution and precise retention time repeatability [23]. GC
125 also provides a tool for compounds' identification by allowing the calculation of Kováts
126 Retention Index (RI) [24], which is an advantage for the identification of unknown
127 xenometabolites. Mass spectrometers are highly sensitive detectors capable of characterizing
128 and quantifying compounds. In this work, the chosen detector is a Single Quadrupole MS,
129 equipped with an Electron Impact (EI) ionization system. The main advantages of this
130 spectrometer are the large dynamic range of the Quadrupole mass analyzer, its high scan
131 frequency, and the ability of the EI to provide reproducible fragmentations for the analyzed
132 compounds [23]. This presents an essential tool for characterizing unknown compounds by
133 fast spectral library search and/or by structural elucidation.

134 All of these advantages were considered for the development of an online HS-SPME-GC-MS
135 method, which was dedicated for studying the environmental fate of an emerging bioherbicide
136 applied on soil: the *Myrica gale* methanolic extract.

137 Introduced by Popovici *et al.* [25,26], the herbicide composition was partially identified by
138 several studies [25–30]. Its bioactive compound is Myrigalone A, an allelochemical, mixed
139 with several other compounds: mainly triketones and terpenes. The herbicide mode of action
140 was described by Oracz *et al.* [31]. This research work revealed a potential synergic activity
141 between Myrigalone A and terpenes. This activity was recently confirmed and explained by
142 Khaled *et al.* [32]. Therefore, an optimal herbicide activity requires the application of the total

complex mixture of the plant extract. However, several components in this complex mixture are still unknown, and their transformation in nature is not deeply understood. Thus, the untargeted metabolomics approach is a potential solution for studying the environmental fate of this bioherbicide. Therefore, in order to prove the concept of the suggested HS-SPME-GC-MS-based untargeted metabolomics approach, the *Myrica gale* methanolic extract was selected as a typical complex bioherbicide in order to study the dissipation of its volatile residues after its application on soil, through a 38-days kinetics study. The study targets exclusively volatile residues that are spontaneously released to the gas phase above soil (the headspace) during imitated environmental conditions applied to microcosm samples.

2. Material and methods

2.1. Chemicals

Methanol HPLC grade was purchased from VWR International (Fontenay-sous-Bois, France). The dry methanolic extract of *Myrica gale* was prepared as described Popovici *et al.* [25]. The spiking herbicide solution for application on soil samples was prepared at a concentration of 72 mg mL^{-1} of dry extract dissolved in Methanol (containing 18 mg mL^{-1} of the bioactive compound Myrigalone A). C7-C30 Saturated Alkanes mix (1000 µg mL^{-1} of each component in Hexane) was purchased from Sigma-Aldrich (Saint-Quentin-Fallavier, France).

2.2. Soil material

Soil sample was collected from an arable field at the agricultural domain of the “Institut Universitaire de Technologie” (IUT) of Perpignan, France (42°40'55.1"N 2°53'51.2"E). The surface layer (15 cm) of soil was collected on 3 different points separated by 1.5 meter. After collection, the soil was homogenized and passed through a 2 mm sieve. Then, it was stocked in the dark at 4 °C until the experiment. The soil composition analysis and characterization were performed by Arterris Laboratory (Toulouges, France) accredited by the French

167 Accreditation Committee (Cofrac). Results were the following: 13.9 % of clay, 60.5 % of silt,
168 25.6 % of sand, 20 % of soil humidity, 1.7 % of organic matter, 0.98 % of organic Carbon,
169 15.5 meq 100 g^{-1} cation exchange capacity (CEC), 214 % Ca^{2+}/CEC and pH of 8.1 in water.
170 According to the Soil Textural Triangle of the United States Department of Agriculture [33],
171 this soil is classified as a silt loam soil. It was never been contaminated or exposed to
172 herbicides.

173 ***2.3. Soil samples set-up***

174 Samples consisted of 6 g of soil weighted in 20 mL HS-SPME vials (Thermo Fisher
175 Scientific, Courtabœuf, France). This weight was optimized in order to keep 2/3 of the vial
176 volume as headspace. After, vials were hermetically closed by a crimped septum, and two
177 18G×1 ½" (1.2 × 38 mm) Agani™ needles (Terumo®, Leuven, Belgium) were implanted on
178 the extremity sides of the septum (Figure A 1 – Appendix A). This is to assure aerobic
179 conditions by allowing air exchange between the internal headspace and the outside. The
180 prepared soil vials were incubated in a GC 401 growth chamber (Nüve, Saracalar, Turkey) for
181 24 hours before the spiking in order to reestablish the biological and microbial activity.
182 Incubation conditions were 24 hours day/night cycle with alternation of light/dark, 28 °C/18
183 °C of temperature, and 40 % RH/65 % RH of humidity (Figure A 2 – Appendix A). The soil
184 moisture was maintained at 20 % during the incubation and throughout the experiment,
185 following a standardized environmental protocol implemented and published in previous
186 works [10,34], aiming to assure conditions that are comparable to real environmental cases.
187 The aim of implementing this sample design was to assure a "living system". As mentioned
188 previously, samples will be used for several kinetic time points, so measures were taken to
189 ensure that they will not be destroyed during the study.

190 The preparation of herbicide-spiked soil samples was performed by applying the *Myrica gale*
191 methanolic extract with a dose equivalent to 300 µg of the active compound (Myrigalone A)
192 per gram of soil (1.2 mg of dry *Myrica gale* methanolic extract per gram of soil). This
193 corresponds to ten-times the agronomical field dose, following testing guidelines
194 recommendations [12,13] in order to assess their transformation and risks on health and
195 environment in an extreme pollution scenario [34].

196 ***2.4. Headspace-Solid Phase Microextraction development***

197 Automated Headspace-Solid Phase Microextraction (HS-SPME) was performed using a
198 TriPlus™ RSH™ autosampler (Thermo Fisher Scientific, Waltham, U.S.). The extraction
199 method was developed by optimizing the following conditions and parameters: the SPME
200 fiber coating, the incubation time, the extraction time, and the extraction temperature. Tests
201 were performed by analyzing herbicide-spiked soil samples (prepared following the protocol
202 described in Section 2.3.)

203 SPME fiber coating tests were performed by comparing 6 different types of coatings: 100 µm
204 Polydimethylsiloxane (100 µm PDMS, Fused Silica, 23 Ga, Autosampler), 7 µm
205 Polydimethylsiloxane (7 µm PDMS, Fused Silica, 24 Ga, Autosampler), 85 µm Polyacrylate
206 (85 µm PA, Fused Silica, 23 Ga, Autosampler), 65 µm Polydimethylsiloxane/Divinylbenzene
207 (65 µm PDMS/DVB, Stableflex, 23 Ga, Autosampler), 85 µm
208 Carboxen/Polydimethylsiloxane (85 µm CAR/PDMS, Stableflex, 23 Ga, Autosampler), and
209 50/30 µm Divinylbenzene/Carboxen/Polydimethylsiloxane (50/30 µm DVB/CAR/PDMS,
210 Stableflex, 23 Ga, Autosampler), all purchased from Supelco (Bellefonte, U.S.). Tests were
211 performed by applying the following HS-SPME conditions: 5 min of incubation time, 30 min
212 of extraction time, and 40 °C of extraction temperature.

213 Next, the duration of sample incubation before the SPME extraction (incubation time) was
214 assessed in order to choose the optimal condition. 3 different incubation times were tested
215 using the selected 50/30 µm DVB/CAR/PDMS fiber: 5 min, 15 min, and 30 min (extraction
216 time: 30 min, extraction temperature: 40 °C).

217 After, the exposure duration of the SPME fiber to the Headspace (extraction time or
218 adsorption time) was assessed. 7 different values were tested: 5 min, 10 min, 20 min, 30 min,
219 40 min, 50 min and 60 min (incubation time: 5 min, extraction temperature: 40 °C, fiber
220 coating: 50/30 µm DVB/CAR/PDMS).

221 Regarding extraction temperature, 3 values were tested in order to assess the impact of
222 increasing temperature on volatile metabolic profiles. Tested temperatures are the following:
223 40 °C, 60 °C, and 80 °C (incubation time: 5 min, extraction time: 30 min, fiber coating: 50/30
224 µm DVB/CAR/PDMS).

225 Finally, a dose response curve was applied after adapting optimal conditions. This in order to
226 examine fiber's over-saturation. 6 different herbicide doses were applied on 6 different
227 batches of soil samples (with 3 biological replicates for each dose batch), and then analyzed
228 and compared to control untreated soil samples (3 biological replicates) in order to assess
229 method's detection limit. The 6 applied doses corresponded to: 10^{-3}-time, 10^{-2}-time, 10^{-1}-time,
230 1-time, 10-times, and 20-times the agronomic field dose of the herbicide.

231 For all optimization tests and method's application, the incubated sample vial was shaken
232 vigorously throughout the incubation and the extraction procedures in order to enhance the
233 homogenization of sample temperature.

234 ***2.5. Gas Chromatography-Mass Spectrometry***

235 Gas Chromatography-Mass Spectrometry analyses were performed on a Focus GC system
236 coupled to an Electron Impact-Single Quadrupole DSQ II Mass Spectrometer (Thermo Fisher

237 Scientific, Waltham, U.S.; Bremen, Germany). An Agilent J&W DB-5MS GC column was
238 used (length: 30 m, inner diameter: 0.25 mm, film thickness: 0.25 µm, Agilent Technologies,
239 Santa Clara, U.S.). Desorption was performed in Splitless mode for a duration of 1 min at an
240 inlet temperature of 230 °C, followed by a 5 min post-injection fiber conditioning at 260 °C in
241 order to prevent fiber carryovers. The 36-min GC run was developed for an optimal
242 compounds separation. It consisted of a 1 mL min^{-1} constant flow method with Helium as
243 carrier gas. The oven temperature was programed as the following: an initial temperature of
244 60 °C was held for 1 min, and was then followed by a first ramp of 10 °C min^{-1} in order to
245 reach 100 °C. After, a second ramp of 3 °C min^{-1} was applied and held until a temperature of
246 182 °C was reached. Finally, the last ramp of 25 °C min^{-1} was applied until a temperature of
247 230 °C was reached. This end temperature was held for 2 min in order to prevent any
248 potential column carryover. GC-MS transfer line temperature was maintained on 240 °C
249 throughout the run.

250 The MS acquisition method was a Full MS scan for positive ions with an *m/z* range of 40-400.
251 The scan rate was 5 scans sec^{-1} (2027.11 amu sec^{-1}). The source temperature was set to 250
252 °C and the applied detector gain was 30000.

253 ***2.6. Software and data processing***

254 GC-MS piloting and data acquisition were performed using Xcalibur 3.0.63 (Thermo Fisher
255 Scientific, Waltham, U.S.). Data were acquired in RAW format and then converted to ANDI
256 format (NetCDF) in order to upload and process them using Galaxy Workflow4Metabolomics
257 platform [35–37]. The automated processing workflow used the metaMS package (Galaxy
258 Version 2.1.1) [38] dedicated for GC-MS data. All of its conditions and parameters were
259 published on the platform [39,40]. In brief, a "matchedFilter" algorithm was used for peak
260 piking, with a Full Width at Half Maximum (FWHM) of 5 (Gaussian model peak) [41]. In

261 addition, GC-MS peaks were considered for peak piking only if: i) their pseudo-spectra
262 contained a minimum of 5 *m/z* features, ii) if these peaks were present in at least 70 % of
263 samples belonging to a defined condition. Between the different injections/runs, the similarity
264 threshold between peaks pseudo-spectra was set to 0.7, and maximum peak Retention Time
265 (RT) variation was set to 15 sec in order to prevent any potential splitting of a metabolite
266 feature into two different features. After generating the data matrix, statistical analyses were
267 performed using the R-based MetaboAnalyst platform [42–44]. Xcalibur 4.1.31.9 (Thermo
268 Fisher Scientific, Waltham, U.S.) and AMDIS 2.72 (National Institute of Standards and
269 Technology, Gaithersburg, U.S.) were used for the deconvolution of MS spectra and the
270 manual data processing to cross-check the results obtained by the automated processing.
271 Compass DataAnalysis 4.3 (Bruker Daltonik GmbH, Bremen, Germany) was used for EIC
272 peak area integration and for counting molecular features' number. NIST 14 library search for
273 putative identification of compounds was performed using NIST MS Search 2.2 (National
274 Institute of Standards and Technology, Gaithersburg, U.S.). Welch Two Sample T-test for
275 independent means comparison was performed using the R Commander 2.4-2 "Rcmdr"
276 package [45] of R 3.3.3 software.

277 ***2.7. Application for a kinetics study***

278 After all analytical conditions were optimized and set-up, a 38-day kinetics tracking study
279 was conducted to prove the concept of the suggested approach. The studied environmental
280 samples consisted of two different groups of soil vials/microcosms (described in the Section
281 2.3.) with 3 replicates of each: an untreated control soil (UnTr), and an herbicide-spiked soil
282 (MyrN). After spiking, samples were incubated in the growth chamber with the day/night
283 cycle conditions mentioned in Section 2.3., in order to imitate natural conditions for herbicide
284 transformation in soil.

285 Next, 8 different kinetic time points were analyzed: day 1, day 2, day 3, day 4, day 8, day 17,
286 day 24, and day 38 after spiking. The same soil samples were analyzed by the HS-SPME-GC-
287 MS developed method for all the 8 time points. The order of injections of the different
288 samples was randomized in order to reduce the impact of potential analytical drifts. Blank
289 injections were performed during each time point analysis, by extracting and analyzing the
290 headspace of an empty 20 mL vial using the same HS-SPME-GC-MS method. For Kováts RI
291 calculation, 20 µL of the C7-C30 Alkanes mix solution were introduced to a 20 mL vial, then
292 it was analyzed by applying the same HS-SPME-GC-MS method.

293 After each analysis, soil microcosms were re-incubated in the growth chamber until the next
294 kinetics time point.

295 **3. Results and discussion**

296 ***3.1. Headspace-Solid Phase Microextraction optimization***

297 The HS-SPME method was optimized in order to establish a compromise between three
298 major criteria: i) assuring an optimal sensitivity for a wide-range detection of different types
299 of volatile compounds, ii) applying non-destructive conditions to soil samples, iii) preventing
300 an induced volatilization of compounds that are relatively less volatile in the imitated
301 environmental conditions, as the approach targets exclusively volatile residues that are
302 spontaneously released to the gas phase above soil.

303 For the selection of the SPME fiber coating, Results of tests are shown in Figure A 3 and
304 Table A 1 (Appendix A). PDMS/DVB, DVB/CAR/PDMS and CAR/PDMS showed better
305 results in term of total TIC area and number of molecular features when compared to the 2
306 PDMS and the PA coatings. In addition, CAR/PDMS fiber coating showed the highest total
307 TIC area and the highest number of molecular features, followed by the DVB/CAR/PDMS,
308 and then the PDMS/DVB.

309 Nonetheless, performances of PDMS/DVB, DVB/CAR/PDMS and CAR/PDMS coatings
310 were in-depth examined. A data matrix was generated by processing GC-MS raw data of fiber
311 tests (using the same processing method described in Section 2.6.), and then a Heatmap
312 analysis was applied on the dataset. Heatmap (Figure A 4 – Appendix A) shows that
313 PDMS/DVB and CAR/PDMS coatings differ by their specificity for different types of
314 herbicide compounds (as highlighted with yellow boxes in the Figure A 4). However,
315 DVB/CAR/PDMS coating is able to extract simultaneously a part of compounds that are
316 extracted with the PDMS/DVB exclusively, and another part of compounds that are extracted
317 with the CAR/PDMS exclusively (as outlined by the green boxes in the Figure A 4).
318 Therefore, for the current work, the use of the DVB/CAR/PDMS coating is considered as the
319 best compromise between the highest sensitivity and the widest molecular diversity.

320 Regarding the duration of sample incubation before the SPME extraction (incubation time),
321 Figure A 5 and Table A 2 (Appendix A) show that the increase of incubation time decreases
322 the sensitivity of the method (in term of total TIC area and number of molecular features).
323 This decrease of sensitivity can be hypothetically explained by the accumulation of a higher
324 ratio of water vapor in the headspace. This may prevent the optimal adsorption of some
325 compounds to the SPME fiber, such as L-α-bornyl acetate containing an Ester function, and
326 epi-γ-Eudesmol and α-Terpineol both containing a Hydroxyl function (Table A 3 – Appendix
327 A). Therefore, an incubation time of 5 min was chosen as an optimum for sensitivity.

328 Concerning the exposure duration of the SPME fiber to the Headspace (extraction time or
329 adsorption time), results in Figure A 6 and Table A 4 (Appendix A) show that a significant
330 difference (in term of total TIC area and number of molecular features) is observed when
331 comparing 5 min, 10 min and 20 min, *vs.* 30 min, 40 min, 50 min and 60 min. For those last 4
332 values of extraction time, total TIC areas and numbers of molecular features seem to be no

333 more evolving. Therefore, a 30 min extraction time was chosen as a compromise between
334 sensitivity and short-time analysis.

335 Regarding extraction temperature, this parameter is constrained by two problematics: i) the
336 application of relatively high temperatures risks to deteriorate the environmental samples.
337 These risks should be avoided as the current study aims to implement a non-destructive
338 method. ii) As mentioned previously, the scope of the approach is to target exclusively
339 volatile residues that are spontaneously released to the headspace during the imitated
340 environmental conditions. Applying relatively high temperature can provoke an induced
341 volatilization of compounds that are relatively less volatile in those conditions, which should
342 be avoided in order to prevent a conceptual bias. The provocation of this induced
343 volatilization was proved by testing 3 extraction temperatures: 40 °C, 60 °C, and 80 °C.
344 According to results in Figure A 7 and Table A 5 (Appendix A), the increase of extraction
345 temperature led to a decrease in signal for compounds eluted between 40 °C and 130 °C (0
346 min to 11 min of RT), meanwhile an increase in signal for compounds eluted between 130 °C
347 and 230 °C (11 min to 21 min) was observed. Therefore, beside its destructive aspect,
348 increasing extraction temperature seems to decrease method's sensitivity for the relatively
349 volatile compounds, meanwhile it increases the signal of compounds that are relatively less
350 volatile in environmental conditions.

351 On the other hand, temperatures below 40 °C were non-applicable in the current work due to
352 problems in stabilizing incubator temperature. This problem risks deteriorating the
353 reproducibility of the extraction. Thus, 40 °C is considered as the optimal compromise for
354 extraction temperature.

355 To sum up, the optimal HS-SPME conditions applied for the study are the following: 50/30
356 µm DVB/CAR/PDMS as fiber coating, 5 min of incubation time, 30 min of extraction time,

357 and 40 °C of extraction temperature. To assess the over-saturation of the fiber with these
358 conditions, a dose response curve was examined. Results in Figure A 8 and Figure A 9
359 (Appendix A) show that at 20-times the field dose, the Total TIC area and the number of
360 detected molecular features are still increasing. This means that at the optimized HS-SPME
361 conditions, the fiber is not yet over-saturated when analyzing 10-times the field dose (*i.e.* the
362 dose applied for the kinetics study), as the fiber is still able to adsorb higher number and
363 quantity of compounds.

364 It is worth mentioning that despite the important influence of moisture ratio on the detection
365 of several volatile metabolites, the variation of this parameter is constrained by the
366 complexity of the environmental context. In fact, the moisture ratio fixed at 20 % throughout
367 the current study aims to assure conditions that are comparable to real environmental cases
368 (following previously published protocols [10,34]). Setting a moisture ratio that does not
369 represent the standardized environmental/biochemical conditions question of the study risks
370 to change the abiotic and biotic transformation pathways of xenometabolites during the
371 kinetics study. In addition, the variation of moisture ratio can *de facto* provoke the
372 volatilization of metabolites that are relatively less volatile when the standardized
373 environmental conditions are in-place.

374 ***3.2. Herbicide residues detection and low matrix background***

375 After 1 day of spiking, a rich profile of extracted herbicide volatile residues was detected by
376 HS-SPME-GC-MS analysis, as shown in Figure 1. The detected analytes were eluted between
377 60 °C and 175 °C (1 min to 30 min of RT), presenting a complex volatile fingerprint with
378 several major and minor compounds.

379 In contrast to the spiked soil, the HS-SPME extract of the untreated control soil samples did
380 not contain an important number of detected compounds (Figure 1). Compared to the blank

381 GC-MS profiles, there was no significant difference. In both groups, all detected peaks mainly
382 consisted of silicon-derivate compounds. These compounds are probably issued from the
383 bleeding of septum, fiber and/or GC column.

384 ***Figure 1:*** *GC-MS chromatograms of HS-SPME extracts after one day of spiking for spiked samples "MyrN" (red), control*
385 *untreated "UnTr" samples (green) and blanks (grey).*
386 *For the two sample groups, figures consist of three overlaid chromatograms (TIC) of the three biological replicates. For*
387 *blanks, two chromatograms of the two analytical replicates are overlaid.*
388 *Chromatograms were performed with Compass DataAnalysis 4.3 software. The intensity scale is fixed to 3.50E8.*

389 The poor GC-MS profile of the untreated control soil HS-SPME extracts reveals the difficulty
390 in extracting and/or detecting endogenous metabolites originating from soil. Thus, this
391 method is not suitable for studying the impact of the applied herbicide on the soil biodiversity.
392 The advantage, however, is the selectivity of the HS-SPME-GC-MS method to the residues of
393 *Myrica gale* extract in the current study, leading to a low matrix background. This can
394 improve the study of the environmental fate of the herbicide, by enhancing the detection, the
395 quantification and the identification of volatile compounds issued from its xenometabolome,
396 and preventing matrix effects and interferences originating from the matrix.

397 ***3.3. Untargeted metabolomics analyses***

398 To prove the concept of the suggested untargeted approach, the 38-day kinetics tracking was
399 performed by applying the HS-SPME-GC-MS analysis on the two groups of samples; the
400 control untreated soil and the soil spiked with the *Myrica gale* extract herbicide (as described
401 in Section 2.7.). After the end of the kinetics tracking and the acquisition of all data, RAW
402 files were converted to ANDI format (NetCDF) and then uploaded on the Galaxy
403 Workflow4Metabolomics platform for data preprocessing (Section 2.6.). The generated data
404 matrix consisted of 64 analyzed samples (24 untreated control samples, 24 spiked samples,
405 and 16 blank injections), and 376 variables. Each of these variables represents a "picked"
406 pseudo-spectrum after it was defined by retention time-based clustering of its *m/z* fragment
407 ion signals using CAMERA package [38,46]. *In fine*, depending on the applied parameters of

408 the preprocessing [39], each variable should represent a relevant detected compound (without
409 neglecting the high possibility of considering noise and artefacts). This acquired data matrix
410 was used for the statistical analyses.

411 *3.3.1. Principal Component Analysis*

412 ***Figure 2:*** *Principal Component Analysis (PCA).*
413 *Plot generated using MetaboAnalyst.*

414 First, the Principal Component Analysis (PCA) was applied. All kinetics time points of both
415 untreated control (UnTr) and spiked (MyrN) samples were integrated. The PCA played an
416 important role for understanding the results that were acquired with this approach. It shows
417 that over time, the volatile metabolic profiles of the spiked samples tend to converge with
418 those of the untreated control samples (Figure 2). According to the first principal component
419 axis (PC1), the later kinetics time points, *i.e.* days 17, 24 and 38 after spiking, were more
420 similar to the control profiles in comparison with the earlier kinetics time points. This means
421 that after 17 days of herbicide application, an important dissipation of its xenometabolome
422 had occurred. In fact, the PC1 that explains 81.4 % of variations, consists of the regression of
423 the main features issued from the xenometabolome. This was confirmed by exploring the
424 loadings of the PC1, revealed by the loading plot of the PCA and the Biplot (Figure 3). The 6
425 most significant features of the PC1 were only present in the extracts of spiked soils as shown
426 in Figure 4. They were more abundant particularly in the earlier kinetics time points, *i.e.* days
427 1, 2, 3, 4 and 8 after spiking.

428 ***Figure 3:*** *Loading plot and Bioplot showing correlations between samples and features of the PC1 and the PC2.*
429 *Plots generated using MetaboAnalyst.*

430 Another important result regarding the degradation of the herbicide was revealed by the PCA.
431 In fact, a progressive evolution of the volatile profiles of the earlier kinetics time points, *i.e.*
432 from day 1 (T01) to day 8 (T08), was significantly observed on the second principal
433 component axis (PC2). The explanation of this result is that the PC2, which accounts for 7 %

434 of variations, consists of two main types of volatile xenometabolites: the major herbicide
435 volatile compounds contained in the *Myrica gale* extract, and the volatile degradation by-
436 products issued from herbicide compounds. These two "families" of xenometabolites
437 constituted the two opposed sides of the PC2 as shown in Figure 2. This explanation was
438 confirmed by the loadings of the PC2 (Figure 3). The 6 most significant features of the upper
439 part of the PC2 axis were the compounds of the herbicide. Their highest abundance was at
440 day 1 (T01), and then started to decay over time (Figure 5A). For the 6 features with the
441 highest contributions in variation on the lower part of the PC2 axis, their abundance increased
442 over time, before starting to decay in the later kinetics time points (Figure 5B). Thus, these
443 features represent the by-products issued from the degradation of the herbicide mixture.

444 ***Figure 4:*** *Boxplots of features with the highest contributions in variation of PC1. The abundance in the two groups of soil*
445 *samples and their evolution over time are represented.*
446 *Boxplots show the null abundance of these features in all control samples (UnTr). The abundance decay over time in spiked*
447 *samples is also shown (MyrNT01 to MyrNT38, respectively).*
448 *Features plots are sorted according to the descending order of their PC1 scores (in absolute value), from the left to the right,*
449 *and then from the top to the bottom.*
450 *Plots generated using MetaboAnalyst.*

451 ***Figure 5:*** *Boxplots of features with the highest contributions in variation on the two opposite sides of the PC2.*
452 *A: Most significant features on the upper part of PC2 axis. Their highest abundance is at day 1 (MyrNT01), and then it starts*
453 *to decay through time.*
454 *B: Most significant features on the lower part of PC2 axis. Their abundance increases through time until starting to decay by*
455 *days 8 and/or 17 (MyrNT08 and/or MyrNT17).*
456 *All these features show a null intensity in the control samples (UnTr).*
457 *Features plots are sorted according to the descending order of their PC2 scores (in absolute value), from the left to the right.*
458 *Plots generated using MetaboAnalyst.*

459 It is worth mentioning that according to PCA, there was no significant difference between
460 volatile profiles of the untreated control samples over time. This proves another advantage of
461 reducing the matrix background, by eliminating soil biochemical evolution factor from
462 analyses. Therefore, tracking and understanding herbicide's environmental fate are enhanced
463 from a chemical-analytical point of view.

464 Ultimately, PCA provided a general understanding of the evolution of xenometabolome
465 through the time. In-depth analyses were then conducted to explain this evolution by filtering
466 and tracking xenometabolites over time, in order to identify their nature and to annotate them.

467 *3.3.2. Xenometabolome features prioritization by time-series Heatmap*

468 As mentioned previously, PCA represented a good tool for an overview understanding of
469 xenometabolome evolution through the time. However, only major molecular traces were
470 revealed by this model. In-depth xenometabolome discovery required different statistical tools
471 dedicated to prioritize and filtrate molecular features of the detected volatile xenometabolites.
472 Therefore, a second different statistical analysis was exploited in this work: the time series-
473 based Heatmap.

474 The time series-based Heatmap considers all features present in the data matrix, in order to
475 visualize the evolution of their abundances over time. Moreover, it performs a clustering in
476 order to group the correlated variables (molecular features), basing on their evolution profiles
477 over the time, and their abundances in the different sample conditions. Thus, it facilitates the
478 "fishing" of features according to their chemical/biochemical nature.

479 Results of the applied time series-based Heatmap are shown in Figure 6. The Heatmap was
480 applied on all of the kinetics time points (day 1 to day 38) for both untreated control soil and
481 herbicide-spiked soil groups. In this Heatmap, samples were not clustered but sorted
482 according to treatment condition as the first factor, and then according to time evolution as the
483 second. Features however, were clustered without *a priori*, according to the correlation of
484 their abundances between the different samples.

485 ***Figure 6:*** *Time series-based Heatmap. Clustering algorithm: Ward, distance measure: Euclidean.*
486 *Plots generated using MetaboAnalyst.*

487 The clustering of features led to the identification of 4 main zones of interest. Zone A, divided
488 into two sub-zones, A_1 and A_2, consisted of compounds that were only present in the spiked
489 samples. The majority of these features were at their highest level of abundance at day 1 and
490 then started to decay over time. Thus, they are considered as components of the *Myrica gale*
491 extract herbicide. The difference between the two A sub-zones was that in comparison to the

492 A_2 sub-zone features, A_1 sub-zone features presented a higher intensity at the beginning of the
493 kinetics tracking, and their decay over time was faster. A_2 sub-zone features however, had
494 lower intensities at the beginning of the tracking compared to A_1 sub-zone features. They
495 decayed more slowly and some of these features were still present on day 38.

496 Zone B consisted of features that appeared at the middle of the kinetics tracking. Therefore,
497 those features were considered as degradation by-products. Zone C features were also
498 considered as degradation by-products. They appeared at the end of the kinetics tracking,
499 however.

500 Zone D, also divided into two sub-zones, D_1 and D_2, represented features that were abundant
501 in both control and spiked samples. Most of these features were considered as noise and
502 artefacts as they showed a random dispersion of abundances between replicates. For sub-zone
503 D_1, features were identified as random artefacts and noise issued from the complex
504 xenometabolome profile. As this complexity is relatively higher in spiked samples at day 1
505 and day 2, this can explain the reason why this noise is higher in those samples, and less
506 intense in the control samples. For sub-zone D_2 the most relevant of these features were
507 examined by a fast putative annotation using NIST 14 library. All of these features were
508 silicon-derivate compounds. Thus, they were considered as products of septum, fiber and
509 column bleeding. These features were also found in blank injections, which confirmed this
510 hypothesis. It is worth mentioning that the significant features of the D_2 sub-zone were more
511 abundant in the untreated control soil samples. This can be explained by the possible fact that
512 in the spiked samples, the adsorption sites of the SPME fiber were less “available” due to the
513 presence of a rich volatile xenometabolome, bleeding compounds originating from the vial
514 septum were thus less able to fixate on the fiber.

515 Afterwards, as important numbers of features were prioritized by the Heatmap, a verification
516 procedure was performed in order to filter and remove the eventual hidden artefacts. This
517 procedure was performed using the Multivariate Empirical Bayes Analysis of Variance
518 (MEBA) approach for time series, based on the timecourse method [47], and designed for the
519 comparison of temporal profiles across different conditions or groups of treatments.

520 *3.3.3. Xenometabolome kinetics tracking and putative compounds identification*

521 All of the prioritized significant features, revealed by the Heatmap and verified by the MEBA,
522 were manually tracked over time by integrating their GC-MS pseudo-spectra peak areas over
523 all the RAW files. This was done in order to draw their time evolution curves according to the
524 38-day kinetics tracking. In addition, this manual tracking is recommended in order to
525 crosscheck the automatically generated results and to avoid any false positives that may occur
526 due to the potential errors of the automated data preprocessing.

527 The manual tracking finally led to consider 101 features as relevant, including 22 features that
528 were considered as degradation by-products according to the kinetics profiles/curves
529 evolution over time. All of these kinetics profiles are shown in Appendix B. Two orthogonal
530 tools were used for putative identification of compounds: the EI-MS fragmentation patterns
531 search on NIST 14 spectral library, and the calculation of Kováts RI that were compared to RI
532 values reported in the NIST library. Kováts RI calculation was performed following the
533 method of Lucero *et al.* [48]. Out of the 101 relevant features, 96 compounds including 20
534 degradation by-products, representing 99.83 % of the total TIC area after blank subtraction
535 were putatively identified on the levels "2" and "3" of identification confidence according to
536 Sumner *et al.* [49]. The most abundant compounds and all identified degradation by-products
537 are shown in Table 1. Detailed annotations of all the 101 prioritized features are summarized
538 in Table A 6 (Appendix A). Furthermore, out of the 96 annotated compounds, 33 were

539 reported in the literature as *Myrica gale* essential oil components [28–30]. All of these 33
540 compounds found in the literature were abundant at day 1 after spiking, representing 67.82 %
541 of the total TIC area after blank subtraction. Meanwhile 63 compounds (47 herbicide
542 components and 16 degradation by-products) are identified for the first time as compounds
543 originating from the *Myrica gale* extract. Figure 7 shows kinetic profiles of the 6 major
544 compounds identified: Eucalyptol, L-terpinen-4-ol, α-Terpineol, α-Terpineol acetate, 3,7(11)-
545 Selinadiene and Germacrone. The rest of the xenometabolites were predominantly terpenes,
546 aromatic and aliphatic esters, alcohol and ketones (Table A 6 – Appendix A).

547 ***Figure 7:*** *Kinetic profiles of the 38-day degradation tracking of the 6 major compounds detected.*
548 *The Peak Areas represent the sum of EICs of the major EI-fragments/ions.*

549 Otherwise, several of the identified degradation by-products could be hypothetically related to
550 the detected *Myrica gale* extract compounds. For instance, Figure 8 shows the kinetics
551 profiles of 2,3-Dehydro-1,8-cineole, Camphor, and Camphene hydrate, that are hypothetically
552 the by-products of Eucalyptol and Borneol after oxidation, and Camphene after hydration,
553 respectively.

554 ***Figure 8:*** *Kinetics profiles of the 38-day degradation tracking of the 3 volatile degradation by-products: 2,3-Dehydro-1,8-*
555 *cineole, Camphor, and Camphene hydrate.*
556 *The Peak Areas represent the sum of EICs of the major EI-fragments/ions.*

557 It is worth to mention that several degradation by-products (14 out of 22, representing 5.49 %
558 of the total TIC area after blank subtraction), were detected at the day 1 after spiking. This can
559 be explained by three different hypotheses: i) the degradation of their parents was very fast so
560 they started to appear after 1 day of the application of the herbicide, ii) they were already
561 present in the applied herbicide mixture due to a slight degradation of their parents during the
562 extraction and/or the stock of the *Myrica gale* extract, iii) these compounds are not only
563 degradation by-products but also essential components of the *Myrica gale* extract. This last
564 case can be considered for the Camphor and the Camphene hydrate that were reported in the

565 literature as components of the *Myrica gale* essential oil, as well as their hypothetic parents,
566 *i.e.* Borneol and Camphene, respectively.

567 Thereby, this fast putative identification, based on fast library search for EI-MS fragmentation
568 spectra and the Kováts RI calculations, presents one of the main advantages of this approach.
569 Indeed, it allowed studying the complex mixture of the emerging natural herbicide, where
570 several of its unknown components and TPs could be putatively identified. The EI
571 fragmentation patterns and their reproducibility allowed the fast annotation of several of these
572 unknown xenometabolites by a simple spectral library search, despite the low resolution of
573 Quadrupole mass analyzer in measuring ions' *m/z* values. Kováts RI calculations assured
574 higher identification confidence by providing an additional and orthogonal tool for
575 metabolites characterization.

576 ***Table 1:*** *Summary of putative identifications of the most relevant features (herbicide xenometabolites with EICs area/total*
577 *TIC area ≥ 1 %, and identified degradation by-products).*
578 *†: The given-code represents the retention time of the compound (in minutes) preceded by the Retention Time "RT"*
579 *abbreviation.*
580 *‡: If MF ≥ 700, and Δ between experimental and NIST RI ≤ 10, the considered level of identification confidence is "2". If MF*
581 *< 700, or Δ between experimental and NIST RI > 10, the considered level of identification confidence is "3" (levels defined*
582 *by Sumner et al. 2007 [49]).*
583 *⸸: The percentage of the "sum of major fragments EICs area/total TIC area" ratio, calculated at day 1 after spiking.*
584 *N/A: Not Available. N/C: Not Calculated. The relative intensity was not calculated for degradation by-products that were not*
585 *detected at day 1.*

586 *3.3.4. Dissipation assessment by Orthogonal Projections to Latent Structures Discriminant*
587 *Analysis (OPLS-DA)*

588 Regarding limitations of classic concepts for environmental fate assessment of complex
589 biopesticides, the targeted tracking is not applicable for the present study as described
590 previously. Thus, the comparison of volatile metabolic profiles of both spiked samples and
591 untreated control samples can be considered as an alternative concept to determine the
592 dissipation of volatile compounds of the studied bioherbicide. The total dissipation is
593 considered when the difference between the volatile metabolic profiles of the compared
594 groups is no more significant. Therefore, the choice of the comparative statistical approach

595 should lay on its ability to reveal the minor differences between the compared profiles. In
596 addition, the significance of those minor differences should be also assessed to avoid the
597 misleading conclusions or the loss of information.

598 The PCA model is a suitable tool for a holistic overview of the acquired data, as for revealing
599 the major differences. However, minor differences will be hidden and it is difficult to
600 determine them with this descriptive multivariate analysis. Thus, a discriminant analysis is
601 needed for this aim. In this work, Orthogonal Projections to Latent Structures Discriminant
602 Analysis (OPLS-DA) [50,51], and its Cross-Validation (CV) test, were considered to quarry
603 and validate the significance of minor differences that are still present after 38 days of kinetics
604 tracking between spiked samples and control samples.

605 PCA, OPLS-DA and the CV test of OPLS-DA were applied to compare the volatile metabolic
606 profiles of both spiked soil and untreated control soil samples at day 38 after spiking. This in
607 order to check if the total dissipation of the herbicide has occurred. First, the PCA showed a
608 discrimination between the two conditions according to both PC1 and PC2, explaining 84.9 %
609 and 13.2 % of variations, respectively (Figure A 10 – Appendix A). However, PCA loadings
610 showed that the significance of the two major discriminant features of the PC1 was unreliable,
611 as an important intra-group variation had been noticed (Figure A 11 – Appendix A).
612 Contrariwise, the three major features of the PC2 showed a significant difference between
613 groups (Figure A 11 – Appendix A). Two of those features were considered as persistent
614 xenometabolites as they were not detected in the untreated control samples (RT5.501: o-
615 Cymene and RT1.766: Methyl benzyl sulfoxide). The third feature showing a higher
616 abundance in the untreated control samples was identified as a silicon derivate compound
617 issued from bleeding. It was also detected in the blank injections.

618 Results of PCA led to conclude that in this descriptive unsupervised multivariate analysis,
619 minor significant discriminations between groups risk to be hidden by the random
620 contaminations and artefacts. Therefore, the OPLS-DA and its CV test were applied as
621 explicative supervised multivariate analyses, in order to reveal significant differences related
622 to the two pre-defined groups (untreated control samples *vs.* spiked samples). These
623 differences in variables (molecular features) will be revealed by the predictive component (p)
624 of the OPLS-DA. Moreover, the significance of these features will be assessed by introducing
625 the confidence dimension represented by the orthogonal (o) component of the OPLS-DA.

626 As described in Figure A 12 (Appendix A), the T score shows that the predictive (p)
627 component explains 55.3 % of variations between the volatile profiles of spiked soil and the
628 control soil. The orthogonal component, that explains intra-group variations, represents 16.3
629 % of variations (Orthogonal T score). Thus, the CV test was performed to assess the
630 significance of "between-groups" and "intra-group" discriminations. The "between-groups"
631 discrimination is assessed by calculating the correlation of samples (R2Y) according to the
632 regressed features of the p component, and by the prediction/significance estimated by the Q2
633 value. The "intra-group" variations are also assessed by calculating the R2Y and Q2 applied
634 to the orthogonal component.

635 The CV test results shown in Appendix A (Figure A 13) were the following: for the p
636 component, R2Y and Q2 were 98.7 % and 92.3 %, respectively. For the o component, R2Y
637 and Q2 were 1.25 % and 1.94 %, respectively. These results show both R2Y and Q2 above 90
638 %, with R2Y higher than R2X and Q2, meaning that the OPLS-DA model is valid. Thus, the
639 discrimination between the two defined groups of samples is significant. In addition, there is a
640 high confidence in the significance of discrimination as the "intra-group" variations were not
641 significant (R2Y and Q2 below 50 %, with R2Y lower than R2X and Q2).

642 Hence, this result means that at day 38 after spiking, the total dissipation of the volatile
643 xenometabolome was not reached. Therefore, to reveal the persistent xenometabolites, the
644 OPLS-DA score plot (S-Plot) can be used as shown in Figure 9. The S-Plot showed several
645 persistent herbicide compounds that were still significantly abundant in the volatile profiles of
646 spiked soil at day 38, *e.g.* RT1.766: Methyl benzyl sulfoxide, RT5.501: o-Cymene, RT25.500:
647 Germacrone, RT20.371: 3,5,11-Eudesmatriene, RT4.333: Methyl 2-methylhexanoate,
648 RT19.447: Calamenene, RT8.073: Camphene hydrate, RT22.148: cis-β-Elemenone. The last
649 6 mentioned features were difficult to reveal using the PCA loading plot. In addition, the
650 kinetics curves proved coherent results with the S-Plot by showing the persistence of these
651 features after 38 days of herbicide application (Appendix B).

652 Another advantage of the OPLS-DA S-Plot was its ability to explain the high
653 risk/insignificance of artefacts and contamination features (RT2.671 and RT4.641) previously
654 revealed by the PCA loadings, despite their high contribution in discrimination between
655 groups. This is thanks to the confidence/reliability dimension represented by the p(corr)[1]
656 axis, as explained in Figure 9.

657 It is worth mentioning that the determination of the total dissipation time of the *Myrica gale*
658 extract herbicide necessitated a longer kinetics study. This however was not in the scope of
659 the present work.

660 ***Figure 9:*** *OPLS-DA Score Plot showing the markers of discrimination between the two defined groups of soil samples.*
661 *The further the feature from the 0 of the p[1] axis, the higher the magnitude of its variation between the two groups.*
662 *The further the feature from the 0 of the p(corr)[1] axis, the lower its intra-group variation, thus the higher the confidence of*
663 *its variation significance [52–54].*
664 *Plot generated using MetaboAnalyst.*

665 ### *3.4. Repeatability and detection limit assessment*

666 #### *3.4.1. Repeatability*

667 The analytical repeatability of the HS-SPME-GC-MS method was assessed by selecting
668 several important compounds to calculate their Retention Time (RT) and Peak Area (PA)

669 deviations. The 5 chosen compounds were distributed on the chromatogram RT range (Table
670 2). "Inter-samples" Peak Area Relative Standard Deviation (RSD) and Retention Time
671 Standard Deviation (SD) were calculated using the 3 biological replicates at the same day
672 (day 1). The highest PA RSD was 7.75 % for the 3,7(11)-Selinadiene, and the highest RT SD
673 was 0.43 sec for Germacrone (Table 2). This proved that the method was highly repeatable.
674 "Inter-days" Retention Time Standard Deviation (SD) was also calculated for the selected
675 compounds using the same sample that was injected 8 times with the following time gaps: 1,
676 2, 3, 4, 8, 17, 24, 38 days. The highest SD was for the Germacrone with 3.44 sec of deviation
677 (Table 2). "Inter-days" PA variation, however, was not assessed due to the difficulty of the
678 application of Internal Standards (IS) with this type of approach. In fact, as the sample is a
679 living system analyzed for several time points for a period of 38 days, a degradation of IS
680 may occur during the experiment.

681 ***Table 2:*** *"Inter-samples" and "inter-days" variations of Peak Area (PA) and Retention Time (RT) over the experiment.*
682 *"Inter-samples" PA RSD and "inter-samples" RT SD were calculated using the 3 biological replicates.*
683 *"Inters-days" RT SD was calculated after the same sample was injected 8 times with the following time gaps: 1, 2, 3, 4, 8,*
684 *17, 24, 38 days.*

685 *3.4.2. Detection limit*

686 As the current study is suggesting an untargeted metabolomics-based approach, classic
687 protocols for targeted method validation are not reasonable (*e.g.* absolute quantification of
688 targeted compounds using reference standards and calibration curves). Thus, a different
689 concept was applied to assess the detection limit at day 1 of the kinetics study, based on a
690 comparative approximation related to herbicide's field dose. Untreated control samples were
691 compared to each dose level of the spiked soil samples (described in Section 2.4.).
692 Comparisons were performed using 3 indicators: the Total TIC area integration, the number
693 of the detected molecular features, and by applying OPLS-DA Cross-Validation tests using a
694 data matrix generated after raw data of dose response curve were processed (using the same
695 processing method described in Section 2.6.). Results in Table 3 show that the method is able

696 to discriminate between the 2 conditions (spiked *vs.* untreated) from 20-times and until 10^{-1}-
697 time the field dose, as significant differences between the compared conditions are observed
698 for Total TIC areas and for numbers of the detected molecular features (Welch Two Sample
699 T-test p-Values < 5 %), and as a reliable predictivity of the OPLS-DA model is observed for
700 these dose levels (Q2 of the p1 > 50 %). Therefore, a detection limit relative to herbicide dose
701 is estimated between 10^{-1}-time and 10^{-2}-time the field dose at day 1 of the kinetics study.

702 ***Table 3:*** *Summary of OPLS-DA CV test and Welch Two Sample T-test results.*

703 ### *3.5. Sample design: a living system after 8 extraction operations*

704 As previously mentioned, the sample design described in Section 2.3. was optimized in order
705 to create a "living system", such that the same prepared samples (soil vials) could then be
706 tracked by several kinetics time point analyses, as the HS-SPME extraction is a non-
707 destructive method.

708 After the application of 8 extractions on each vial/sample during the 38-day kinetics study,
709 green plants were observed on top of the soil layer of untreated control samples. The
710 development of this plant layer was progressive during the kinetics study and was even
711 observed 44 days after the end of the kinetics tracking as shown in Figure 10. This indicates
712 that the implemented sample design and the optimized HS-SPME extraction was successful in
713 providing appropriate conditions to sustain a living micro-ecosystem.

714 ***Figure 10:*** *The evolution of the untreated control soil vials/microcosms during the experiment.*
715 *Photo A was taken before the kinetics tracking began. Photo B was taken at the end of the kinetics tracking (at day 38, after*
716 *analyses). Photos C was taken 44 days after the last time point was analyzed (i.e. after 82 days of the beginning of the*
717 *kinetics tracking).*

718 ## 4. Conclusion

719 The present work aimed to introduce a novel HS-SPME-GC-MS-based untargeted
720 metabolomics approach dedicated to study the environmental fate of complex biopesticides.
721 The approach was developed and applied to study the volatile residues of the *Myrica gale*

722 methanolic extract; an emerging natural herbicide applied to soil, consisting of a complex
723 mixture of identified and non-identified compounds. The developed analytical method was
724 proved reliable in the detection of a rich volatile profile originating from the herbicide, with a
725 low matrix background and a significant robustness. This allowed the fast putative
726 identification of 96 xenometabolites including 33 compounds reported in the literature, 47
727 compounds identified for the first time as *Myrica gale* extract components, and 16 new
728 degradation by-products, by a 38-day kinetics tracking experiment. A comparison of
729 herbicide-spiked and untreated control soil samples over time demonstrated the advantages of
730 applying the untargeted metabolomics and its statistical tools as an alternative concept for
731 complex pesticides study. The evolution of the herbicide volatile xenometabolome over time
732 can be visualized and explained by using the PCA. The time-series Heatmap method is a
733 suitable tool for prioritizing the relevant xenometabolites and sorting them according to their
734 temporal evolution in the different groups of samples. This is done in order to characterize
735 and identify new xenometabolites and TPs, which can then help to better understand the
736 environmental fate of the herbicide, as well for assessing its potential risk and toxicity on the
737 health and the environment. The OPLS-DA and its CV test provided a sensitive and confident
738 determination of minor discriminations between the different groups of samples in order to
739 assess the total dissipation of the herbicide volatile xenometabolome.

740 The developed approach successfully revealed all of the significant results and conclusions
741 through an analysis of only 6 environmental samples that were not destroyed throughout the
742 course of the study. Thus, this non-destructive green automated method has now been shown
743 to be capable of cost-effective high throughput analyses. Nevertheless, further analytical and
744 technical developments should be performed to improve the current approach, in order to
745 expand its potential application in environmental fate studies and emerging pesticides
746 research.

Acknowledgments

Authors would like to acknowledge Jeanine Almany (École Pratique des Hautes Études) for providing English language editing (as well as constructing comments) which improved the manuscript. Acknowledgments for Dr. Nicolas Le Yondre (Université de Rennes 1), Dr. Chandrashekhar Patil (Université de Perpignan Via Domitia), and Mathieu Lazarus, M.Sc. (Université de Perpignan Via Domitia) for their scientific and technical advices. Authors also acknowledge two anonymous reviewers for their constructive reviews and critical comments that helped improving the current work.

This work was supported by the French National Research Agency (ANR) under TRICETOX project (ANR-13-CESA-0002), and the European Regional Development Fund (ERDF) under the Interreg POCTEFA PALVIP project (POCTEFA 2014-2020). The funding institutions had no role in the experimental design, the data processing, or in writing and reviewing the manuscript.

Ph.D. fellowship grant was awarded to HG by the French Ministry of Higher Education, Research and Innovation (MESRI), *via* the Doctoral School ED 305 "Energie et Environnement" (Université de Perpignan Via Domitia).

The HS-SPME-GC-MS method developments and analyses had been performed using the Biodiversité et Biotechnologies Marines (Bio2Mar) facilities – Métabolites Secondaires Xénobiotiques Métabolomique Environnementale (MSXM) platform at the Université de Perpignan Via Domitia (http://bio2mar.obs-banyuls.fr/).

Authors' contributions

CB, MVS and HG designed the study and edited the draft of the manuscript. HG developed the HS-SPME-GC-MS method, carried out the experimental work and data analyses, and wrote the first and the revised drafts of the manuscript. DR contributed in method's

771 development and instruments' verification and piloting. All authors read and approved the
772 final manuscript.

773 **List of Abbreviations**

774 BA: Biocontrol Agent

775 TP: Transformation Product

776 EMF: Environmental Metabolic Footprinting

777 SLE: Solid-Liquid Extraction

778 LC: Liquid Chromatography

779 MS: Mass Spectrometry

780 OECD: Organization for Economic Co-operation and Development

781 HS: Headspace

782 SPME: Solid Phase Microextraction

783 GC: Gas Chromatography

784 RI: Kováts Retention Index

785 EI: Electron Impact

786 HPLC: High Performance Liquid Chromatography

787 CEC: Cation Exchange Capacity

788 RH: Relative Humidity

789 PDMS: Polydimethylsiloxane

790 PA: Polyacrylate

791 DVB: Divinylbenzene

792 CAR: Carboxen

793 ANDI: Analytical Data Interchange

794 NetCDF: Network Common Data Form

795 FWHM: Full Width at Half Maximum

796 RT: Retention Time

797 EIC: Extracted Ion Chromatogram

798 TIC: Total Ion Chromatogram

799 PCA: Principal Component Analysis

800 PC1: First Principal Component

801 PC2: Second Principal Component

802 MEBA: Multivariate Empirical Bayes Analysis of Variance

803 MF: Matching Factor

804 OPLS-DA: Orthogonal Projections to Latent Structures Discriminant Analysis

805 CV: Cross-Validation

806 S-Plot: Score Plot

807 PA: Peak Area

808 SD: Standard Deviation

809 RSD: Relative Standard Deviation

810 IS: Internal Standard

811 CAS: Chemical Abstracts Service

812 **References**

813 [1] R. McKinlay, J.A. Plant, J.N.B. Bell, N. Voulvoulis, Endocrine disrupting pesticides:
814 Implications for risk assessment, Environment International. 34 (2008) 168–183.
815 https://doi.org/10.1016/j.envint.2007.07.013.

816 [2] J. Dich, S.H. Zahm, A. Hanberg, H.-O. Adami, Pesticides and cancer, Cancer Causes
817 & Control. 8 (1997) 420–443. https://doi.org/10.1023/A:1018413522959.

818 [3] B.N. Ames, L.S. Gold, Pesticides, Risk, and Applesauce, Science. 244 (1989) 755–
819 757. https://www.jstor.org/stable/1703501.

820 [4] M. Gavrilescu, Fate of Pesticides in the Environment and its Bioremediation, Eng.
821 Life Sci. 5 (2005) 497–526. https://doi.org/10.1002/elsc.200520098.

822 [5] H.M.G. van der Werf, Assessing the impact of pesticides on the environment,
823 Agriculture, Ecosystems & Environment. 60 (1996) 81–96. https://doi.org/10.1016/S0167-
824 8809(96)01096-1.

825 [6] C. Bertrand, C. Prigent-Combaret, A. Gonzales-Coloma, Chemistry, activity, and
826 impact of plant biocontrol products, Environ Sci Pollut Res. 25 (2018) 29773–29774.
827 https://doi.org/10.1007/s11356-018-3209-2.

828 [7] European Food Safety Authority, EFSA Guidance Document for evaluating laboratory
829 and field dissipation studies to obtain DegT50 values of active substances of plant protection
830 products and transformation products of these active substances in soil, EFSA Journal. 12
831 (2014) 3662. https://doi.org/10.2903/j.efsa.2014.3662.

832 [8] K.A. Aliferis, M. Chrysayi-Tokousbalides, Metabolomics in pesticide research and
833 development: review and future perspectives, Metabolomics. 7 (2011) 35–53.
834 https://doi.org/10.1007/s11306-010-0231-x.

835 [9] M.R. Viant, U. Sommer, Mass spectrometry based environmental metabolomics: a
836 primer and review, Metabolomics. 9 (2013) 144–158. https://doi.org/10.1007/s11306-012-
837 0412-x.

838 [10] C. Patil, C. Calvayrac, Y. Zhou, S. Romdhane, M.-V. Salvia, J.-F. Cooper, F.E.
839 Dayan, C. Bertrand, Environmental Metabolic Footprinting: A novel application to study the
840 impact of a natural and a synthetic β-triketone herbicide in soil, Science of The Total
841 Environment. 566–567 (2016) 552–558. https://doi.org/10.1016/j.scitotenv.2016.05.071.

842 [11] M.-V. Salvia, A. Ben Jrad, D. Raviglione, Y. Zhou, C. Bertrand, Environmental
843 Metabolic Footprinting (EMF) vs. half-life: a new and integrative proxy for the discrimination
844 between control and pesticides exposed sediments in order to further characterise pesticides'
845 environmental impact, Environ Sci Pollut Res. 25 (2018) 29841–29847.
846 https://doi.org/10.1007/s11356-017-9600-6.

847 [12] OECD, Test No. 307: Aerobic and Anaerobic Transformation in Soil, 2002.
848 https://doi.org/10.1787/9789264070509-en.

849 [13] OECD, Test No. 308: Aerobic and Anaerobic Transformation in Aquatic Sediment
850 Systems, 2002. https://doi.org/10.1787/9789264070523-en.

851 [14] J. Pawliszyn, Handbook of Solid Phase Microextraction, First Edition, Elsevier, 2012.
852 https://doi.org/10.1016/C2011-0-04297-7.

853 [15] J. Pawliszyn, Theory of Solid-Phase Microextraction, Journal of Chromatographic
854 Science. 38 (2000) 270–278. https://doi.org/10.1093/chromsci/38.7.270.

855 [16] R.P. Belardi, J.B. Pawliszyn, The Application of Chemically Modified Fused Silica
856 Fibers in the Extraction of Organics from Water Matrix Samples and their Rapid Transfer to
857 Capillary Columns, Water Quality Research Journal. 24 (1989) 179–191.
858 https://doi.org/10.2166/wqrj.1989.010.

859 [17] N. Reyes-Garcés, E. Gionfriddo, G.A. Gómez-Ríos, Md.N. Alam, E. Boyacı, B.
860 Bojko, V. Singh, J. Grandy, J. Pawliszyn, Advances in Solid Phase Microextraction and
861 Perspective on Future Directions, Analytical Chemistry. 90 (2018) 302–360.
862 https://doi.org/10.1021/acs.analchem.7b04502.

863 [18] M. Llompart, M. Celeiro, C. García-Jares, T. Dagnac, Environmental applications of
864 solid-phase microextraction, TrAC Trends in Analytical Chemistry. 112 (2019) 1–12.
865 https://doi.org/10.1016/j.trac.2018.12.020.

866 [19] D. Liang, W. Liu, R. Raza, Y. Bai, H. Liu, Applications of solid-phase micro-
867 extraction with mass spectrometry in pesticide analysis, Journal of Separation Science. 42
868 (2019) 330–341. https://doi.org/10.1002/jssc.201800804.

869 [20] J. Beltran, F.J. López, F. Hernández, Solid-phase microextraction in pesticide residue
870 analysis, Journal of Chromatography A. 885 (2000) 389–404. https://doi.org/10.1016/S0021-
871 9673(00)00142-4.

872 [21] J.L. Martínez Vidal, P. Plaza-Bolaños, R. Romero-González, A. Garrido Frenich,
873 Determination of pesticide transformation products: A review of extraction and detection
874 methods, Journal of Chromatography A. 1216 (2009) 6767–6788.
875 https://doi.org/10.1016/j.chroma.2009.08.013.

876 [22] V. Andreu, Y. Picó, Determination of pesticides and their degradation products in soil:
877 critical review and comparison of methods, TrAC Trends in Analytical Chemistry. 23 (2004)
878 772–789. https://doi.org/10.1016/j.trac.2004.07.008.

879 [23] M. Bedair, L.W. Sumner, Current and emerging mass-spectrometry technologies for
880 metabolomics, TrAC Trends in Analytical Chemistry. 27 (2008) 238–250.
881 https://doi.org/10.1016/j.trac.2008.01.006.

882 [24] L.S. Ettre, The Kováts Retention Index System, Anal. Chem. 36 (1964) 31A-41A.
883 https://doi.org/10.1021/ac60214a727.

884 [25] J. Popovici, C. Bertrand, D. Jacquemoud, F. Bellvert, M.P. Fernandez, G. Comte, F.
885 Piola, An Allelochemical from Myrica gale with Strong Phytotoxic Activity against Highly
886 Invasive Fallopia x bohemica Taxa, Molecules. 16 (2011) 2323–2333.
887 https://doi.org/10.3390/molecules16032323.

888 [26] J. Popovici, C. Bertrand, G. Comte, Use of a Myrica gale plant for producing a
889 herbicide agent, US008734858B2, 2014. https://patents.google.com/patent/US8734858B2/en.

890 [27] J. Popovici, G. Comte, E. Bagnarol, N. Alloisio, P. Fournier, F. Bellvert, C. Bertrand,
891 M.P. Fernandez, Differential Effects of Rare Specific Flavonoids on Compatible and
892 Incompatible Strains in the Myrica gale-Frankia Actinorhizal Symbiosis, Applied and
893 Environmental Microbiology. 76 (2010) 2451–2460. https://doi.org/10.1128/AEM.02667-09.

894 [28] J. Popovici, C. Bertrand, E. Bagnarol, M.P. Fernandez, G. Comte, Chemical
895 composition of essential oil and headspace-solid microextracts from fruits of *Myrica gale* L.
896 and antifungal activity, Natural Product Research. 22 (2008) 1024–1032.
897 https://doi.org/10.1080/14786410802055568.

898 [29] K.P. Svoboda, A. Inglis, J. Hampson, B. Galambosi, Y. Asakawa, Biomass
899 production, essential oil yield and composition of Myrica gale L. harvested from wild
900 populations in Scotland and Finland, Flavour and Fragrance Journal. 13 (1998) 6.
901 https://doi.org/10.1002/(SICI)1099-1026(199811/12)13:6<367::AID-FFJ724>3.0.CO;2-M.

902 [30] R.R. Carlton, P.G. Waterman, A.I. Gray, Variation of leaf gland volatile oil within a
903 population of sweet gale (Myrica gale) (Myricaceae), Chemoecology. 3 (1992) 45–54.
904 https://doi.org/10.1007/BF01261456.

905 [31] K. Oracz, A. Voegele, D. Tarkowská, D. Jacquemoud, V. Turečková, T. Urbanová, M.
906 Strnad, E. Sliwinska, G. Leubner-Metzger, Myrigalone A Inhibits Lepidium sativum Seed
907 Germination by Interference with Gibberellin Metabolism and Apoplastic Superoxide
908 Production Required for Embryo Extension Growth and Endosperm Rupture, Plant and Cell
909 Physiology. 53 (2012) 81–95. https://doi.org/10.1093/pcp/pcr124.

910 [32] A. Khaled, M. Sleiman, E. Darras, A. Trivella, C. Bertrand, N. Inguimbert, P. Goupil,
911 C. Richard, Photodegradation of Myrigalone A, an Allelochemical from *Myrica gale*:
912 Photoproducts and Effect of Terpenes, J. Agric. Food Chem. 67 (2019) 7258–7265.
913 https://doi.org/10.1021/acs.jafc.9b01722.

914 [33] United States Natural Resources Conservation Service - Soil Science Division, Soil
915 Survey Manual, revised, United States Department of Agriculture, 2017.
916 https://books.google.fr/books?id=dieUtAEACAAJ.

917 [34] S. Romdhane, M. Devers-Lamrani, J. Beguet, C. Bertrand, C. Calvayrac, M.-V.
918 Salvia, A.B. Jrad, F.E. Dayan, A. Spor, L. Barthelmebs, F. Martin-Laurent, Assessment of the
919 ecotoxicological impact of natural and synthetic β-triketone herbicides on the diversity and
920 activity of the soil bacterial community using omic approaches, Science of The Total
921 Environment. 651 (2019) 241–249. https://doi.org/10.1016/j.scitotenv.2018.09.159.

922 [35] Galaxy Workflow4Metabolomics, Galaxy Workflow4Metabolomics. (n.d.).
923 https://galaxy.workflow4metabolomics.org/ (accessed February 28, 2020).

924 [36] F. Giacomoni, G. Le Corguille, M. Monsoor, M. Landi, P. Pericard, M. Petera, C.
925 Duperier, M. Tremblay-Franco, J.-F. Martin, D. Jacob, S. Goulitquer, E.A. Thevenot, C.

926 Caron, Workflow4Metabolomics: a collaborative research infrastructure for computational
927 metabolomics, Bioinformatics. 31 (2015) 1493–1495.
928 https://doi.org/10.1093/bioinformatics/btu813.

929 [37] Y. Guitton, M. Tremblay-Franco, G. Le Corguillé, J.-F. Martin, M. Pétéra, P. Roger-
930 Mele, A. Delabrière, S. Goulitquer, M. Monsoor, C. Duperier, C. Canlet, R. Servien, P.
931 Tardivel, C. Caron, F. Giacomoni, E.A. Thévenot, Create, run, share, publish, and reference
932 your LC–MS, FIA–MS, GC–MS, and NMR data analysis workflows with the
933 Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics, The
934 International Journal of Biochemistry & Cell Biology. 93 (2017) 89–101.
935 https://doi.org/10.1016/j.biocel.2017.07.002.

936 [38] R. Wehrens, G. Weingart, F. Mattivi, metaMS: An open-source pipeline for GC–MS-
937 based untargeted metabolomics, Journal of Chromatography B. 966 (2014) 109–116.
938 https://doi.org/10.1016/j.jchromb.2014.02.051.

939 [39] H. Ghosson, D. Raviglione, M.-V. Salvia, C. Bertrand, Online HS-SPME-GC-MS-
940 based untargeted volatile metabolomics for studying emerging complex biopesticides: a proof
941 of concept - W4M workflow and parameters, Galaxy Workflow4Metabolomics. (2020).
942 https://galaxy.workflow4metabolomics.org/u/hghosson/w/online-hs-spme-gc-ms-based-
943 untargeted-metabolomics-for-emerging-complex-biopesticides-study-can-we-footprint-the-
944 volatilome (accessed February 28, 2020).

945 [40] H. Ghosson, D. Raviglione, M.-V. Salvia, C. Bertrand, Online HS-SPME-GC-MS-
946 based untargeted volatile metabolomics for studying emerging complex biopesticides: a proof
947 of concept - W4M data, Galaxy Workflow4Metabolomics. (2020).
948 https://galaxy.workflow4metabolomics.org/u/hghosson/h/online-hs-spme-gc-ms-based-

949 untargeted-metabolomics-for-emerging-complex-biopesticides-study-can-we-footprint-the-
950 volatilome (accessed February 28, 2020).

951 [41] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing
952 Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment,
953 Matching, and Identification, Anal. Chem. 78 (2006) 779–787.
954 https://doi.org/10.1021/ac051437y.

955 [42] MetaboAnalyst, MetaboAnalyst. (n.d.). https://www.metaboanalyst.ca/ (accessed
956 February 28, 2020).

957 [43] J. Chong, D.S. Wishart, J. Xia, Using MetaboAnalyst 4.0 for Comprehensive and
958 Integrative Metabolomics Data Analysis, Current Protocols in Bioinformatics. 68 (2019) e86.
959 https://doi.org/10.1002/cpbi.86.

960 [44] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D.S. Wishart, J. Xia,
961 MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis, Nucleic
962 Acids Research. 46 (2018) W486–W494. https://doi.org/10.1093/nar/gky310.

963 [45] J. Fox, The R Commander: A Basic-Statistics Graphical User Interface to R, Journal
964 of Statistical Software. 14 (2005) 1–42. https://doi.org/10.18637/jss.v014.i09.

965 [46] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: An
966 Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid
967 Chromatography/Mass Spectrometry Data Sets, Anal. Chem. 84 (2012) 283–289.
968 https://doi.org/10.1021/ac202450g.

969 [47] Y.C. Tai, T.P. Speed, A multivariate empirical Bayes statistic for replicated
970 microarray time course data, Ann. Statist. 34 (2006) 2387–2412.
971 https://doi.org/10.1214/009053606000000759.

[48] M. Lucero, R. Estell, M. Tellez, E. Fredrickson, A retention index calculator simplifies identification of plant volatile organic compounds, Phytochemical Analysis. 20 (2009) 378–384. https://doi.org/10.1002/pca.1137.

[49] 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.

[50] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), J. Chemometrics. 16 (2002) 119–128. https://doi.org/10.1002/cem.695.

[51] H. Wold, Estimation of principal components and related models by iterative least squares, in: Multivariate Analysis, Krishnaiah, P. R., Academic Press, 1966: pp. 391–420. https://ci.nii.ac.jp/naid/20001378860/en/.

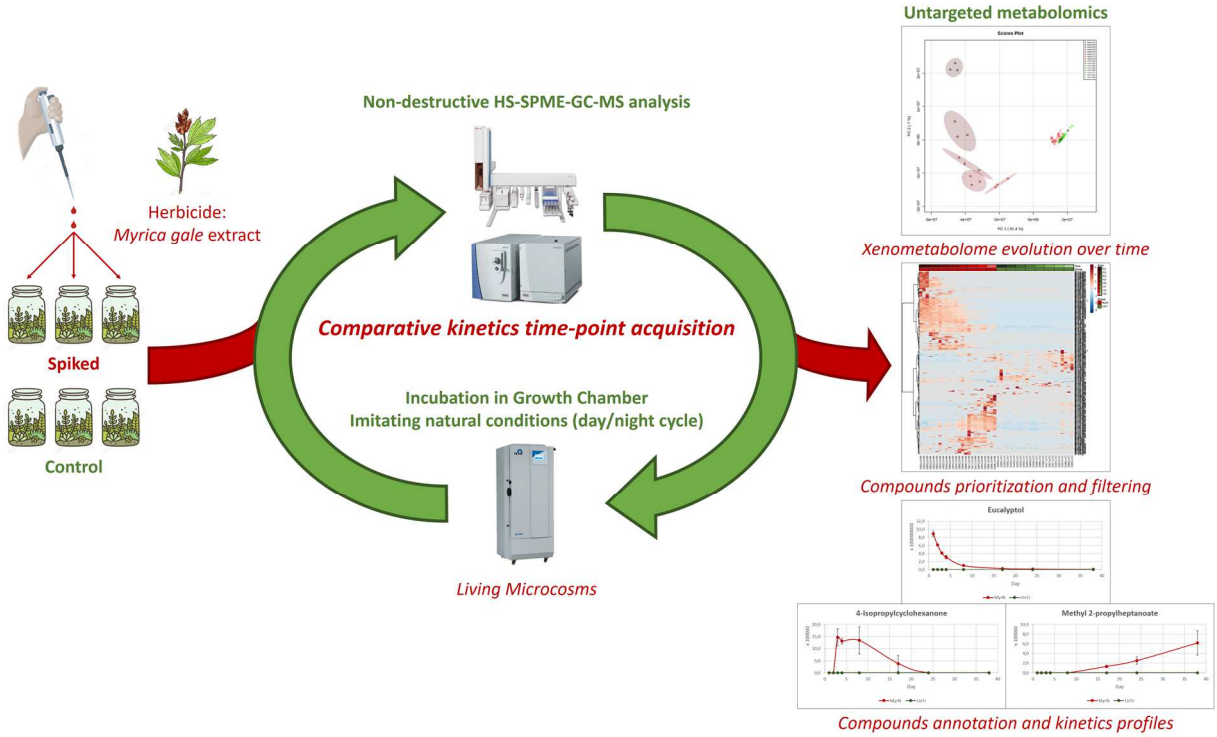
[52] J. Cohen, Things I have learned (so far)., American Psychologist. 45 (1990) 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304.

[53] A. Roux, Analysis of human urinary metabolome by liquid chromatography coupled to high resolution mass spectrometry, Theses, Université Pierre et Marie Curie - Paris VI, 2011. https://tel.archives-ouvertes.fr/tel-00641529.

[54] S. Wiklund, Multivariate data analysis for Omics, (2008). https://metabolomics.se/Courses/MVA/MVA%20in%20Omics_Handouts_Exercises_Solutions_Thu-Fri.pdf.

995

996



997

998 Graphical Abstract

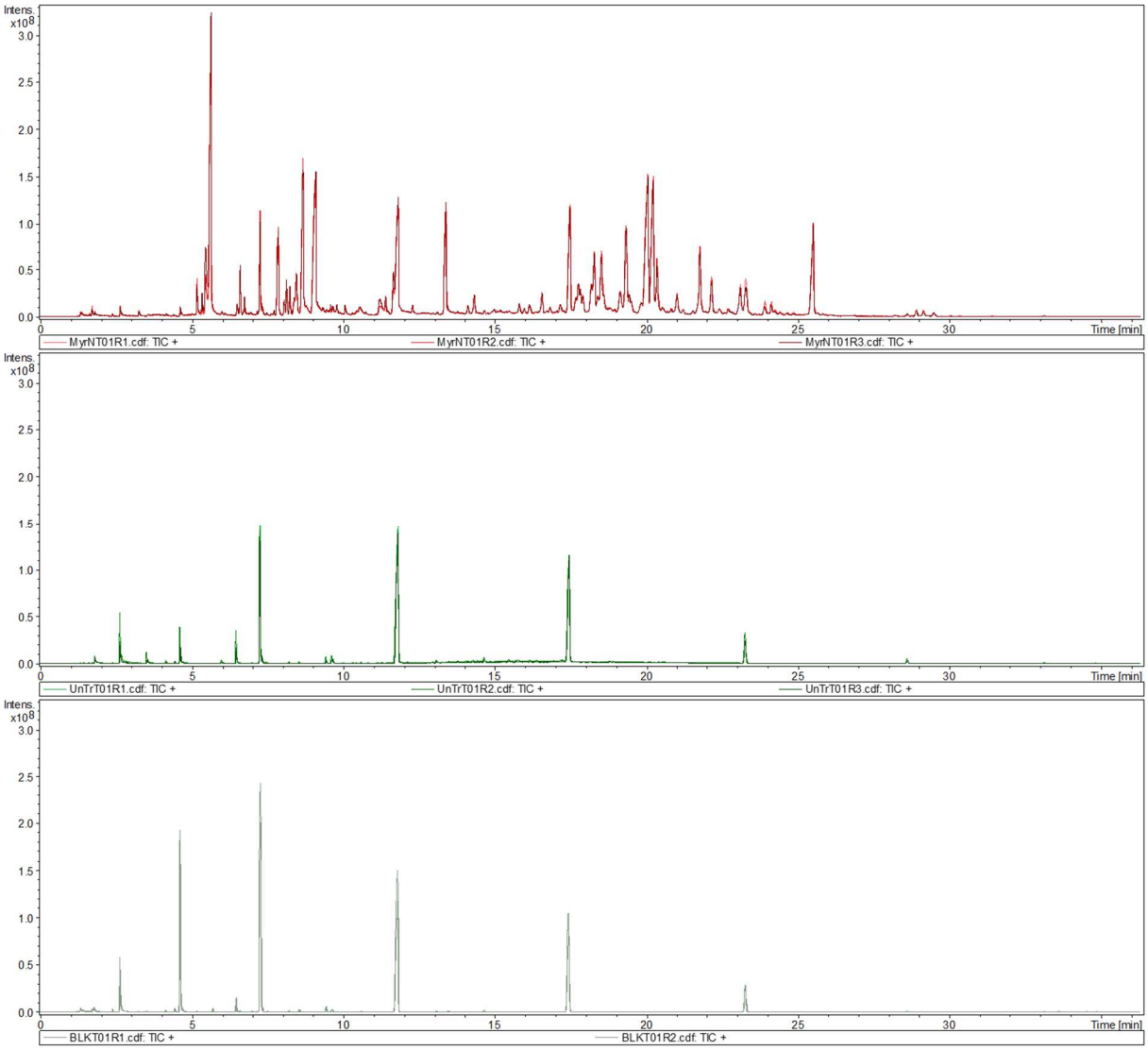
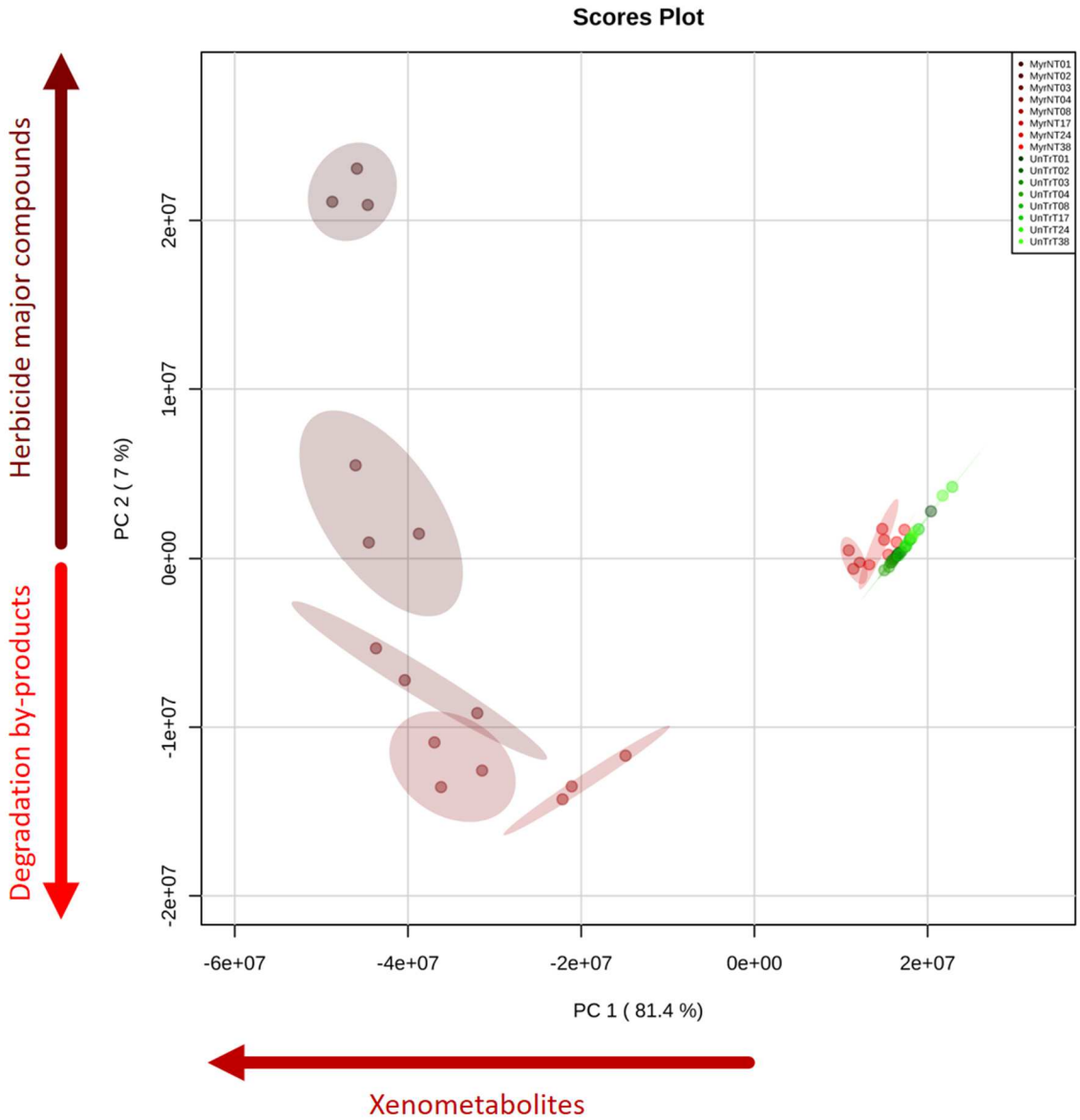
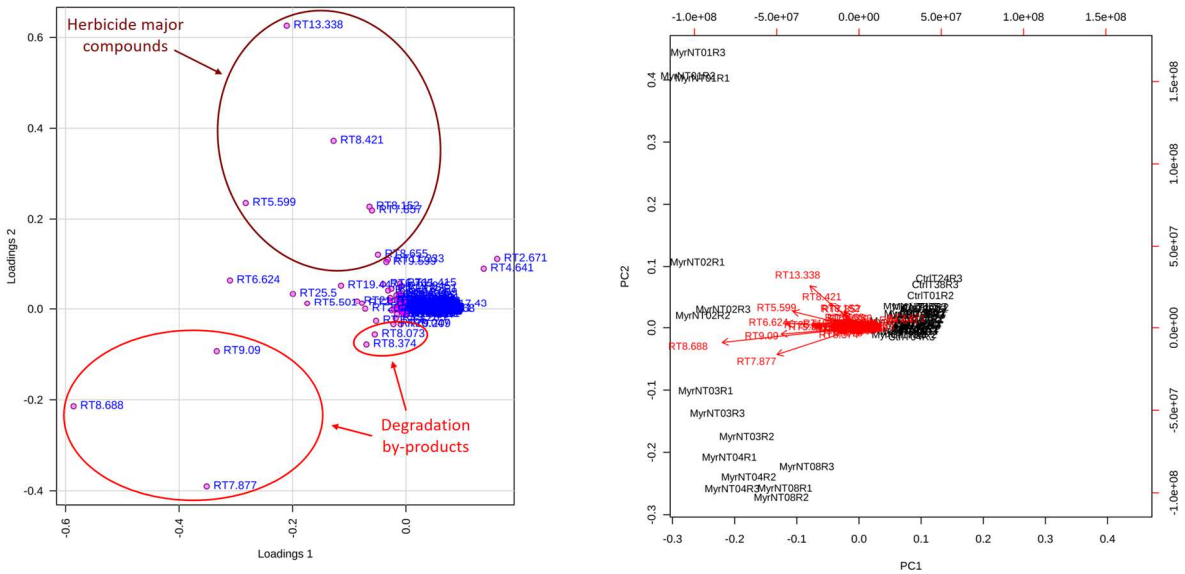

Intens.
x10^8
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
25
30
Time [min]
MyrNT01R1.cdf: TIC +
MyrNT01R2.cdf: TIC +
MyrNT01R3.cdf: TIC +
UnTrT01R1.cdf: TIC +
UnTrT01R2.cdf: TIC +
UnTrT01R3.cdf: TIC +
BLKT01R1.cdf: TIC +
BLKT01R2.cdf: TIC +

Figure 1



1001

1002 Figure 2



a: PC1 × PC2 Loading plot

b: PC1 × PC2 Biplot

Figure 3

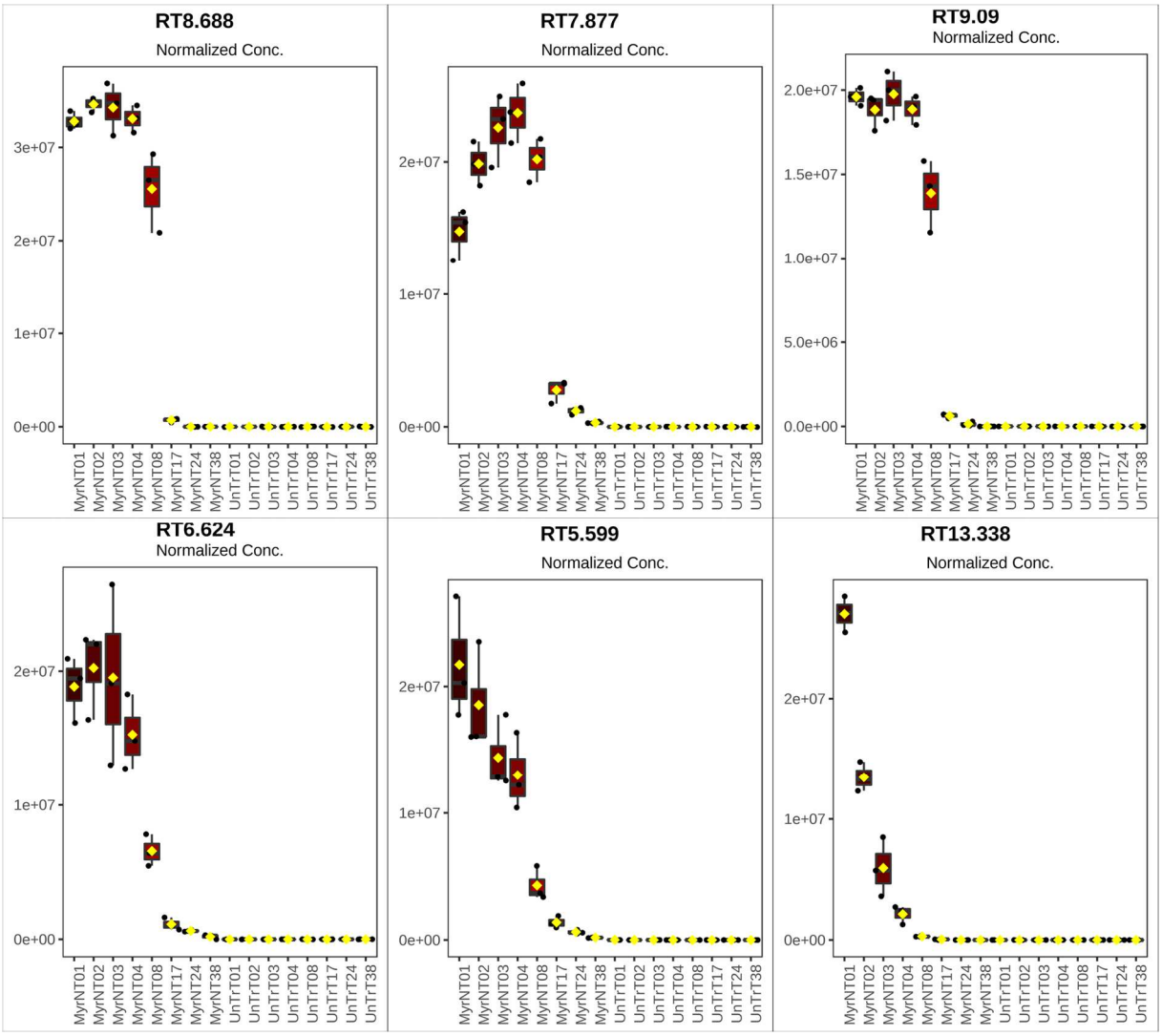


Figure 4

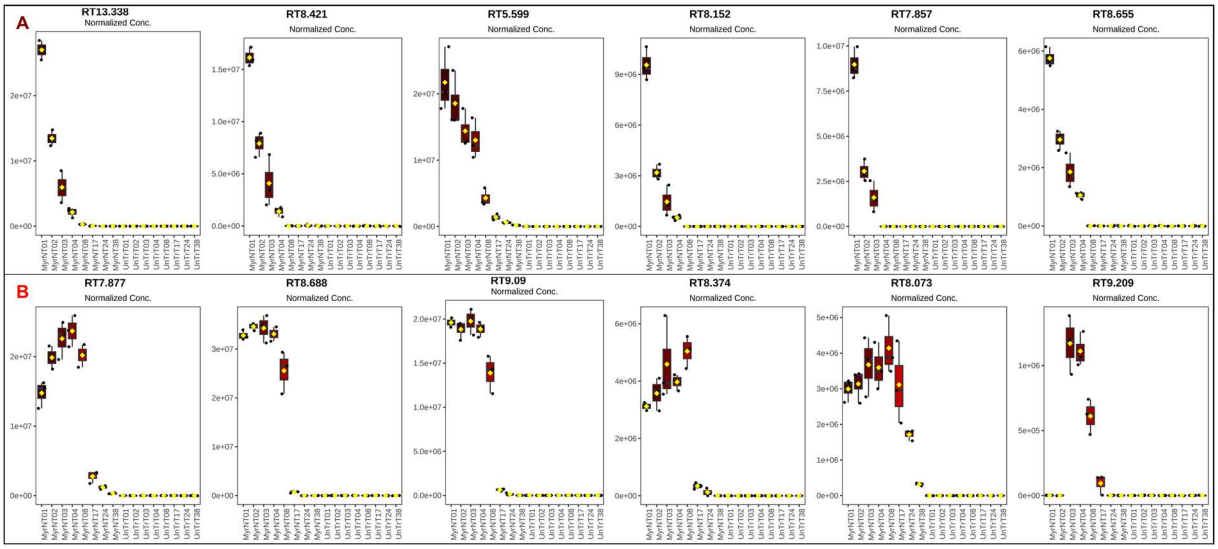


Figure 5

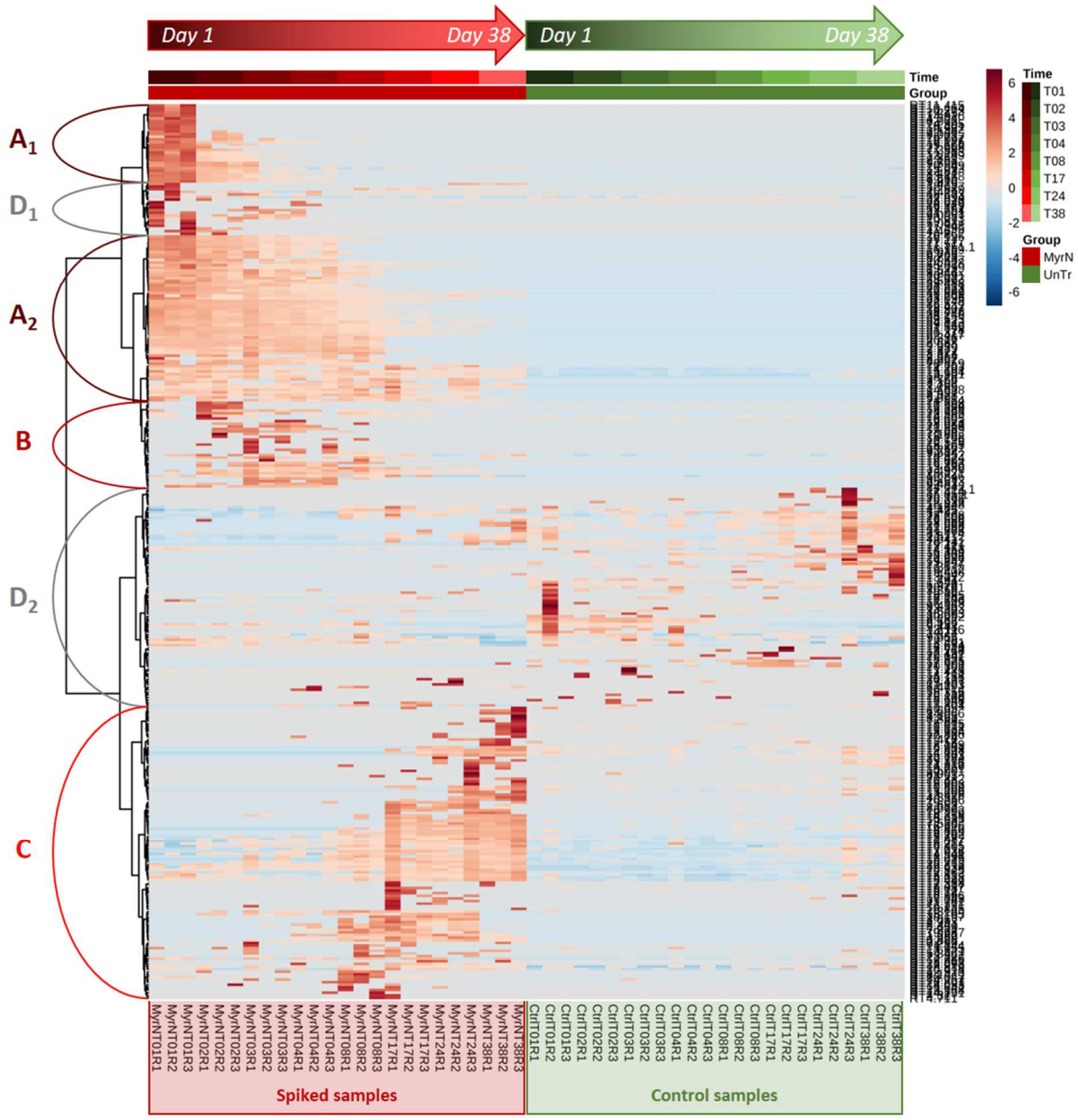
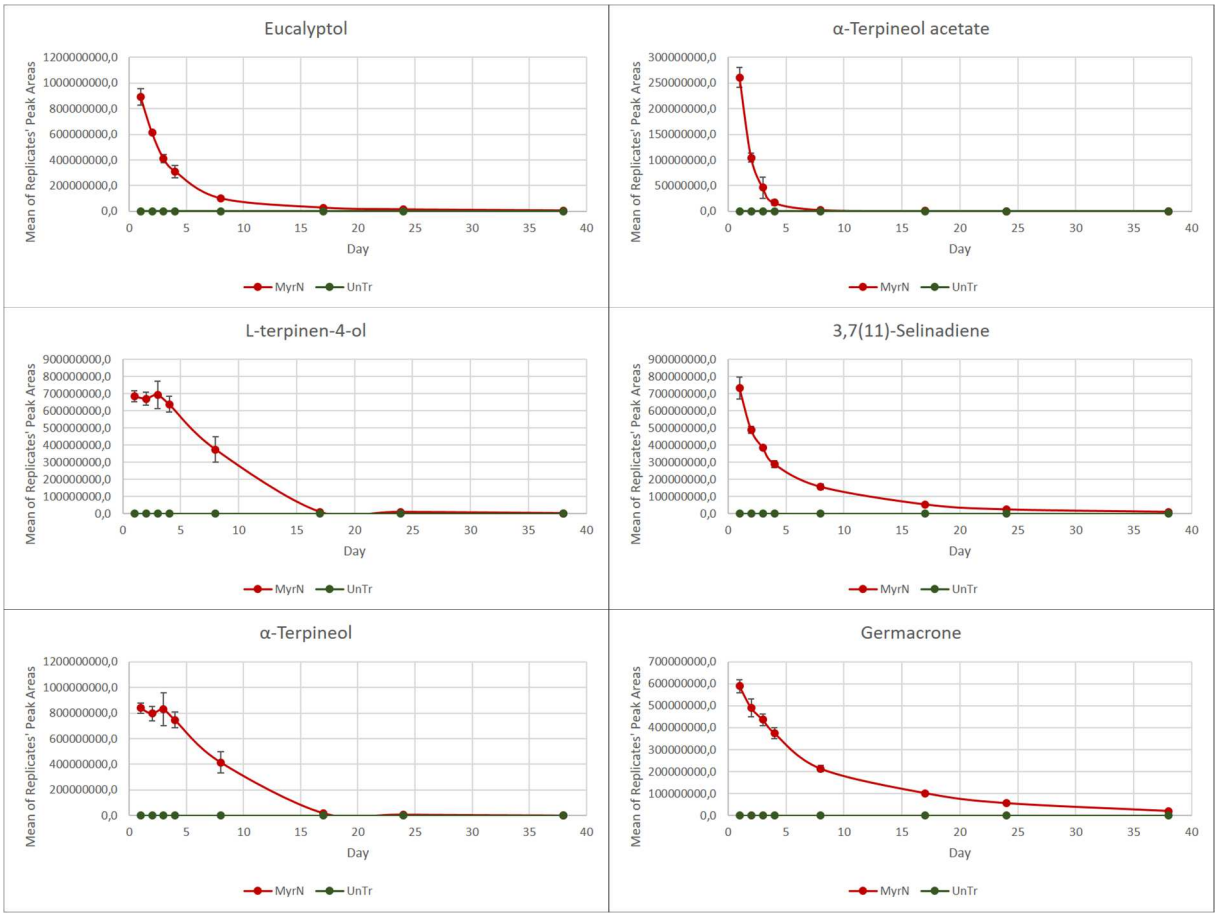

Day 1
Day 38
Day 1
Day 38
Time
Group
A1
D1
A2
B
D2
C
Time
T01
T02
T03
T04
T08
T17
T24
T38
Group
MyrN
UnTr
6
4
2
0
-2
-4
-6
MyrNT01R1
MyrNT01R2
MyrNT01R3
MyrNT02R1
MyrNT02R2
MyrNT02R3
MyrNT03R1
MyrNT03R2
MyrNT03R3
MyrNT04R1
MyrNT04R2
MyrNT04R3
MyrNT08R1
MyrNT08R2
MyrNT08R3
MyrNT17R1
MyrNT17R2
MyrNT17R3
MyrNT24R1
MyrNT24R2
MyrNT24R3
MyrNT38R1
MyrNT38R2
MyrNT38R3
CtrlT01R1
CtrlT01R2
CtrlT01R3
CtrlT02R1
CtrlT02R2
CtrlT02R3
CtrlT03R1
CtrlT03R2
CtrlT03R3
CtrlT04R1
CtrlT04R2
CtrlT04R3
CtrlT08R1
CtrlT08R2
CtrlT08R3
CtrlT17R1
CtrlT17R2
CtrlT17R3
CtrlT24R1
CtrlT24R2
CtrlT24R3
CtrlT38R1
CtrlT38R2
CtrlT38R3
Spiked samples
Control samples

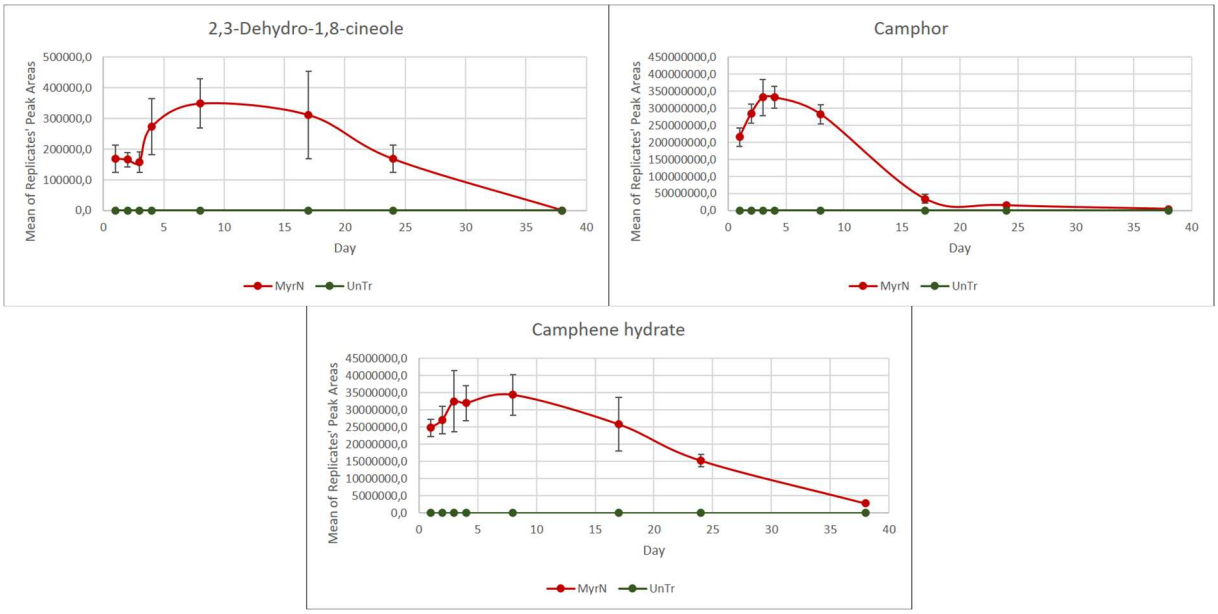
Figure 6

1011

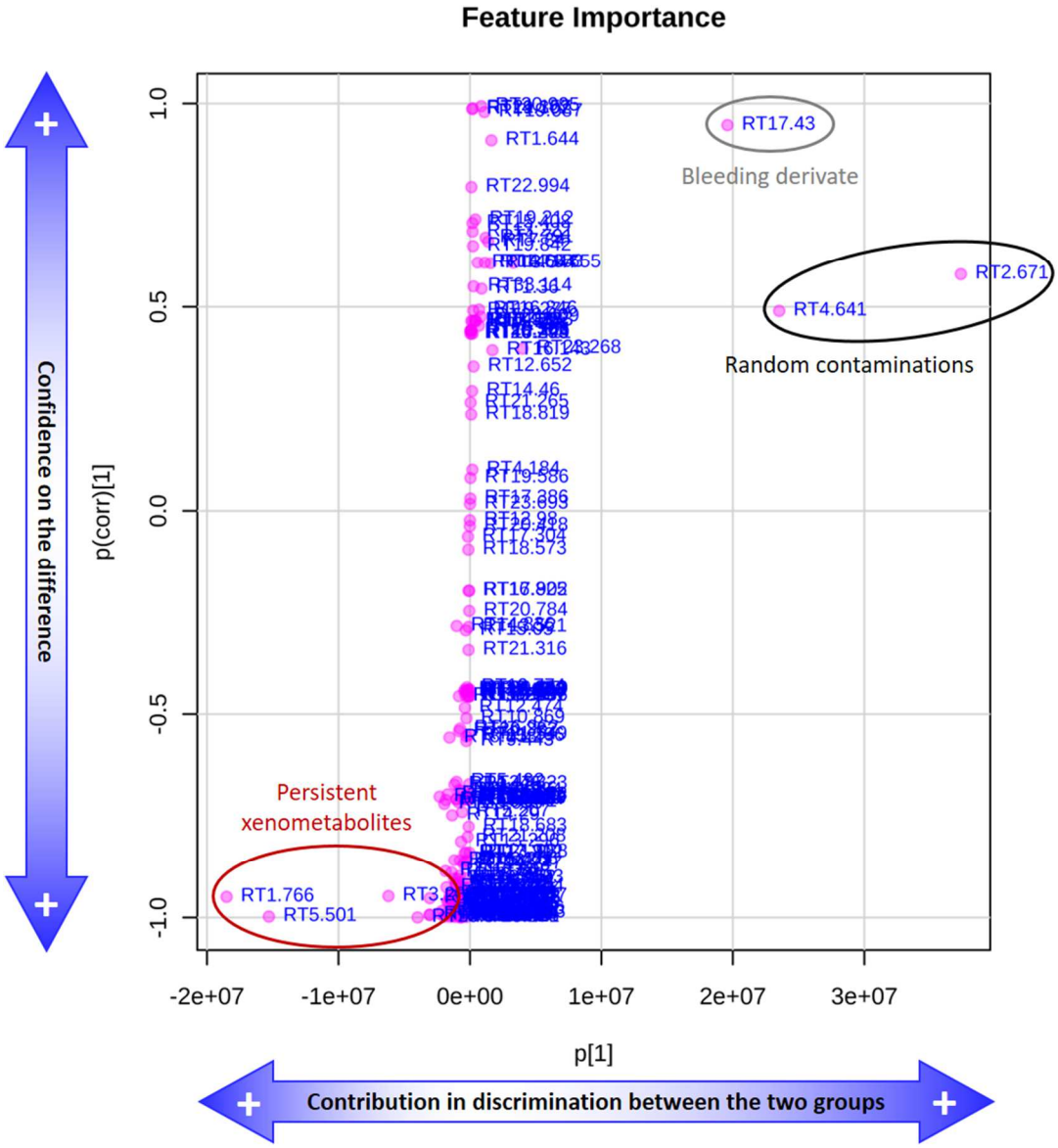


1012 Figure 7

1013



1014 Figure 8



1015

1016 Figure 9

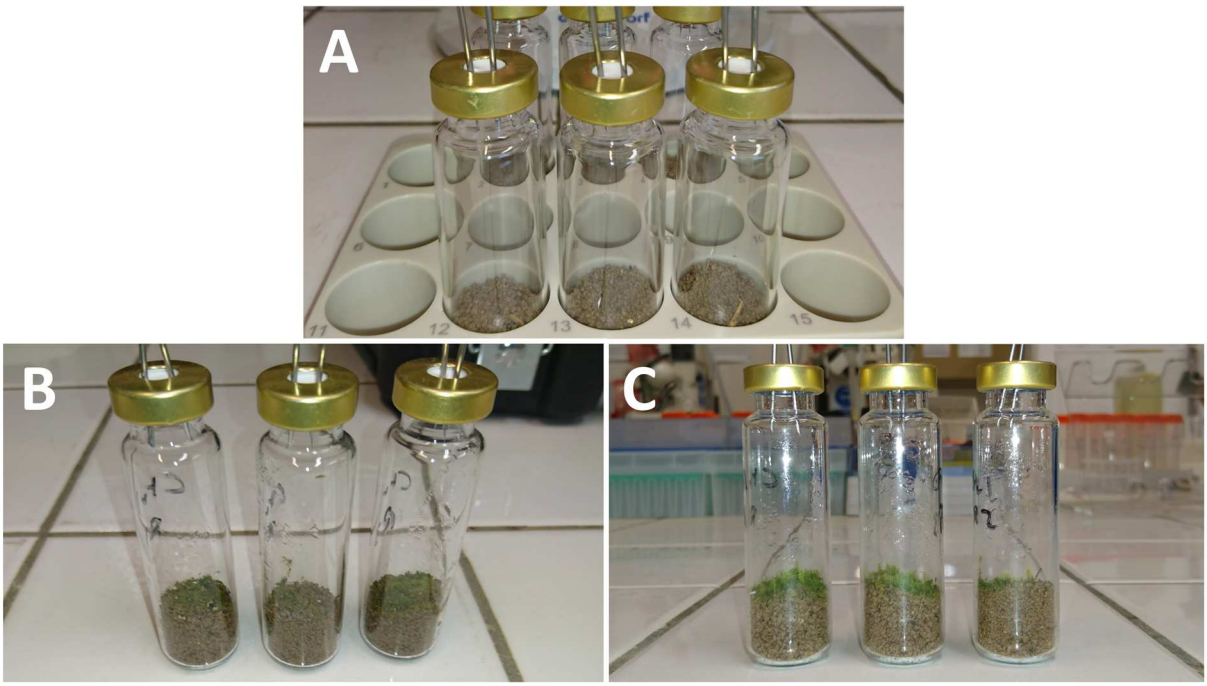


Figure 10

Compound given-code†	Putative identity‡ (level 2 or 3 of identification confidence)	MF	RI (Experimental)	RI (NIST)	Relative intensity (%)ǂ	Reference
Myrica gale methanolic extract components						
RT5.492	p-Cymene	881	1026	1025 ± 2	1.56	[28–30]
RT5.501	o-Cymene	902	1029	1022 ± 2	1.46	N/A
RT5.599	Eucalyptol	910	1036	1032 ± 2	11.09	[28–30]
RT8.421	Borneol	903	1172	1166 ± 7	1.25	[30]
RT8.688	L-terpinen-4-ol	934	1179	1182 ± 0	7.09	[29,30]
RT9.090	α-Terpineol	913	1193	1189 ± 2	7.94	[29,30]
RT11.233	Methyl hydrocinnamate	928	1267	1279 ± 2	1.28	N/A
RT11.751	2-Undecanone	922	1285	1294 ± 2	1.60	[28]
RT13.338	α-Terpineol acetate	934	1342	1350 ± 3	4.75	[30]
RT17.570	Aromadendrene, dehydro-	784	1466	1464 ± 1	5.74	N/A
RT18.448	(+)-β-Selinene	853	1488	1486 ± 3	2.49	[30]
RT18.573	α-Selinene	904	1492	1494 ± 3	1.70	[30]
RT19.299	β-Cadinene	850	1516	1518 ± 10	3.90	[29]
RT20.047	γ-Selinene	907	1541	1544 ± N/A	8.30	N/A
RT20.170	3,7(11)-Selinadiene	913	1547	1542 ± 3	5.96	[28]
RT21.778	Aristolene epoxide	817	1585	N/A	2.84	N/A
RT22.148	cis-β-Elemenone	905	1595	1593 ± 3	1.42	[29,30]
RT23.095	1,4-Benzenedipropanol, α,α',γ,γ,γ',γ'-hexamethyl-	710	1622	N/A	1.81	N/A
RT25.500	Germacrone	930	1691	1693 ± 3	5.81	[28–30]
Degradation by-products						
RT2.425	Methyl isovalerate	831	777	773 ± 5	0.04	N/A
RT3.050	Tyranton	834	843	838 ± 8	<0.01	N/A
RT4.259	Butanoic acid, 2,2-dimethyl-3-oxo-, methyl ester	800	946	936	N/C	N/A
RT4.333	Methyl 2-methylhexanoate	770	952	953 ± 2	N/C	N/A
RT4.441	Camphene	932	957	952 ± 2	0.02	[29,30]
RT4.459	β-Pinene	678	960	979 ± 2	<0.01	[29,30]
RT4.953	2,3-Dehydro-1,8-cineole	819	991	991 ± 2	<0.01	N/A
RT6.690	Methyl 2-propylheptanoate	720	1096	1155 ± N/A	N/C	N/A
RT6.885	3-Acetyl-2,5-dimethylfuran	590	1101	1099 ± 4	0.01	N/A
RT7.536	Methyl octanoate	584	1133	1126 ± 2	N/C	N/A
RT7.877	(+)-Camphor	932	1148	1143 ± 9	2.80	[28]
RT8.073	Camphene hydrate	870	1155	1148 ± 2	0.31	[28]
RT8.140	3-Isopropyl-2-methylcyclopentanone	715	1157	1174 ± N/A	N/C	N/A
RT8.267_2	cis-p-Menthan-3-one	809	1164	1164 ± 6	0.35	N/A
RT8.677	2(3H)-Benzofuranone, hexahydro-3a,7a-dimethyl-, cis-	729	1182	N/A	N/C	N/A
RT9.209	Tetrahydrocarvone	856	1200	1208 ± N/A	0.01	N/A
RT11.495	8,9-Dehydrothymol methyl ether	733	1281	1247 ± N/A	0.02	N/A
RT12.486	5-Methoxy-4,4,6-trimethyl-7-oxabicyclo[4.1.0]heptan-2-one	645	1314	N/A	N/C	N/A
RT16.983	Selinan	621	1450	1476 ± 12	0.01	N/A
RT20.371	3,5,11-Eudesmatriene	859	1547	1495 ± N/A	1.92	N/A

Table 1

Compound	Retention Time (min)	PA RSD (%) "Inter-samples" *(n = 3)*	RT SD (sec) "Inter-samples" *(n = 3)*	RT SD (sec) "Inter-days" *(n = 8)*
Eucalyptol	5.60	6.79	0.25	1.61
L-terpinen-4-ol	8.69	4.20	0.18	1.00
α-Gurjunene	17.17	5.34	0.36	0.92
3,7(11)-Selinadiene	20.17	7.75	0.18	2.53
Germacrone	25.50	2.86	0.43	3.44

Table 2

Applied dose (the field dose)	**p1**			**o1**			**T-test (-Log$_{10}$[p-Value])**	
	R2X	***R2Y***	***Q2***	***R2X***	***R2Y***	***Q2***	***Total TIC area***	***Number of Molecular Features***
20-times	70.30	99.90	98.30	07.68	00.09	00.28	*** 3.67	*** 5.48
10-times	69.40	100.00	98.30	07.64	00.02	00.45	*** 3.93	*** 6.54
1-time	56.60	99.80	94.80	12.70	00.21	00.97	** 3.15	*** 3.49
10^{-1}-time	44.60	99.40	88.50	16.30	00.55	03.76	*** 3.51	** 2.57
10^{-2}-time	26.20	84.80	48.30	24.30	14.00	11.10	0.70	1.19
10^{-3}-time	26.40	75.00	43.60	24.50	24.30	00.43	0.60	0.42

1032 Table 3

1033