

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

**Global Tracking of Transformation Products of  
Environmental Contaminants by  ${}^2\text{H}$ -labeled Stable Isotope-  
Assisted Metabolomics**

Journal:	<i>Analytical Chemistry</i>
Manuscript ID	ac-2022-00500v.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Chen, Ke; South-Central University for Nationalities, College of Resources and Environmental Science Xiang, Yuhui; South-Central University for Nationalities Yan, Xiaoyu; Renmin University of China, Department of Chemistry Li, Zhenghui; South-Central University for Nationalities Qin, Rui; South-Central University for Nationalities Sun, Jie; South-Central University for Nationalities, Key Laboratory of Catalysis and Materials Science of the State Ethnic Affairs Commission & Ministry of Education

SCHOLARONE™  
Manuscripts

1  
2      1   **Global Tracking of Transformation Products of Environmental Contaminants**  
3                2   **by  $^2\text{H}$ -labeled Stable Isotope-Assisted Metabolomics**

4                3   Ke Chen<sup>1,\*</sup>, Yuhui Xiang<sup>1</sup>, Xiaoyu Yan<sup>2</sup>, Zhenghui Li<sup>3</sup>, Rui Qin<sup>4</sup> and Jie Sun<sup>1</sup>

5                4   <sup>1</sup>. College of Resources and Environmental Science, Key Laboratory of Resources  
6                5   Conversion and Pollution Control of the State Ethnic Affairs Commission, South-  
7                6   Central University for Nationalities, Wuhan, 430068, P.R. China;

8                7   <sup>2</sup>. Department of Chemistry, Renmin University of China, Beijing, 100872, P.R. China;

9                8   <sup>3</sup>. School of Pharmaceutical Sciences, South-Central University for Nationalities,  
10                9   Wuhan, Hubei 430074, P.R. China;

11                10   <sup>4</sup>. College of Life Sciences, South-Central University for Nationalities, Wuhan, 430068,  
12                11   P.R. China.

13                13   **Abbreviation List:**

14                14   TPs, transformation products;

15                15   GC, Gas Chromatography; LC, Liquid Chromatography;

16                16   UPLC, Ultra Performance Liquid Chromatography;

17                17   MS, Mass Spectrometry; HRMS, High Resolution Mass Spectrometry;

18                18   EI, Electron Ionization; ESI, Electrospray Ionization;

19                19   m/z, mass-to-charge ratio;

20                20   RT, retention time; EIC, Extracted Ion Chromatogram;

21                21   NMR, nuclear magnetic resonance spectrometer;

22                22   NTA, non-targeted analysis;

23                23   SIAM, Stable Isotope Assisted Metabolomics;

24                24   M, natural compounds; M', isotopically labeled M;

25

26

1  
2  
3     27 **Abstract**  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Stable Isotope Assisted Metabolomics (SIAM) enables global tracking of isotopic labels in non-targeted metabolomics in living organisms. However, its application in tracking transformation products (TPs, as metabolites of contaminants) of environmental contaminants is still a challenge due to limits in methodology, unmatured algorithms, and the high cost of  $^{13}\text{C}$ -labeled contaminants. Therefore, we developed a  $^2\text{H}$ -SIAM pipeline coupled with a highly flexible algorithm  $^2\text{H}$ -SIAM(V1.0) (<https://github.com/kechen1984/2H-SIAM>), facilitating tracking TPs of contaminants in the environment matrix. A detailed discussion illustrates the theory, behavior, and prospect of  $^2\text{H}$ -SIAM. We demonstrate that the proposed  $^2\text{H}$ -SIAM pipeline has unique advantages over  $^{13}\text{C}$ -SIAM, for example, cost-effective  $^2\text{H}$ -labeled contaminants, easy synthesis of  $^2\text{H}$ -labeled emerging contaminants, and providing more structure information. A pyrene soil degradation study further confirmed its high performance. It efficiently discarded 99% noise signals and extracted 52 features from the non-targeted High-Resolution Mass Spectrometry (HRMS) data. Among them, 13 features were annotated as TPs of pyrene with identification confidence from Level 2a to Level 5, and 5 TPs were reported for the first time. In conclusion, the proposed  $^2\text{H}$ -SIAM pipeline is powerful in tracking potential TPs of environmental contaminants with unique advantages.

**Keywords:** Deuterium, Organic Contaminants, Transformation Products, Stable Isotope, Non-targeted Analysis, High-Resolution Mass Spectrometry.  
**Synopsis statement:** We provided a  $^2\text{H}$ -labeled Stable Isotope-Assisted Metabolomics pipeline as a powerful tool for tracking transformation products of environmental contaminants.

## 52      Introduction

53      Anthropogenic organics in the earth generate countless transformation products  
54 (TPs) by biological and non-biological factors. Their fates in the environment is of  
55 increasing concern, as they may constitute higher toxicity than their parent  
56 contaminants<sup>1,2</sup>. Over the past decade, the growth, evolution, and accessibility of High  
57 Resolution Mass Spectrometry (HRMS) has witnessed the progress of non-targeted  
58 analysis (NTA), also referred to as "non-target screening", "untargeted metabolomics",  
59 and "untargeted screening", among several other related terms<sup>3-5</sup>. Additionally, Stable  
60 Isotope Assisted Metabolomics (SIAM) enables global tracking of isotopic labels from  
61 parent compounds in non-targeted metabolomics in living organisms<sup>6-9</sup>. Numbers of  
62 algorithms and software contribute to their raw data process, statistical analysis, and  
63 annotation<sup>10-13</sup>.

64      However, the extremely high cost of <sup>13</sup>C-labeled contaminants (e.g., <sup>13</sup>C-labeled  
65 persistent organic pollutants) limits the application of <sup>13</sup>C-SIAM in the environmental  
66 studies. As listed in Cambridge Isotope Laboratories ([www.isotope.com](http://www.isotope.com)), 0.12 mg of  
67 pyrene-<sup>13</sup>C3 costs \$955, which is about 15000 times more expensive than pyrene-*d*10.  
68 Then, it is only available to carry out <sup>13</sup>C-SIAM in a very small microcosm, such as a  
69 few grams of soil. The <sup>13</sup>C-SIAM study, in a planted soil system or a 500-liter bioreactor,  
70 is unacceptable due to the high cost of <sup>13</sup>C-labeled contaminants.

71      An alternative solution is the use of <sup>2</sup>H-labeled contaminants, that is <sup>2</sup>H-SIAM,  
72 costing much less than <sup>13</sup>C-SIAM. Additionally, direct H/D exchange reactions enable  
73 to add <sup>2</sup>H-labels to emerging contaminants with mild reaction conditions<sup>14</sup>, broadening  
74 the source of isotope-labeled contaminants. However, misunderstanding of <sup>2</sup>H-labeled  
75 contaminants, unmatured methodology and algorithms for SIAM limit the application  
76 of <sup>2</sup>H-SIAM in the environment studies.

77      Credentialing features and ALLocator are designed solely for <sup>13</sup>C-labeled  
78 isotopologue (Table S1)<sup>15, 16</sup>. MetExtract II, HiTIME and ALLocator, collect data  
79 directly from scans of the mass spectrum when isotopologue pairs are coeluted<sup>16-18</sup>.  
80 <sup>13</sup>C-labels always occur in the backbone of chemicals, and <sup>2</sup>H-labels are located in the  
81 surface layer of molecule architecture, which may result in an apparent difference of  
82 retention time (RT) between <sup>2</sup>H-labeled isotopologue pairs<sup>19</sup>. In this scenario, the scan-  
83 based algorithm could not delicately deal with <sup>2</sup>H-SIAM.

84      Additionally, some published algorithms for SIAM, for instance, X13CMS,  
85 geoRge, ALLocator, and MzMatch-ISO, were designed with XCMS<sup>16, 20-22</sup>, or with

1  
2  
3     86 embedded data processing algorithms ([Table S1](#))<sup>17, 18, 23</sup>. The application of them needs  
4     87 fundamental knowledge about those bioinformatics tools or limits the choice of data  
5     88 processing algorithms to obtain a features list (features with m/z, RT, and intensities).  
6  
7     89 Other widely used data processing algorithms, for instance, MZmine2 and MS-DIAL,  
8     90 were excluded. They are user-friendly, flexible, and easily extendable, covering the  
9     91 entire LC-MS data analysis workflow, which could be a powerful tool in SIAM.  
10  
11

12  
13     92 Then, we explored a <sup>2</sup>H-SIAM pipeline to address the issue. A features-based  
14     93 algorithm, namely <sup>2</sup>H-SIAM(V1.0) (For Win10, by Visual Basic .NET), was developed  
15     94 to extract <sup>2</sup>H-labeled isotopologue pairs. The accuracy of the pipeline was evaluated by  
16     95 non-targeted recovery of 7 isotopologue pairs mixed in soil extract. The possible  
17     96 application of H/D exchange in <sup>2</sup>H-SIAM was evaluated by the synthesis and structure  
18     97 analysis of 6PPD-d9 (6PPD, N-(1,3-dimethylbutyl)-N'-phenyl-p-phenylenediamine).  
19     98 The overall performance of the pipeline was finally evaluated by a soil pyrene  
20     99 degradation study.  
21  
22

## 23     100 Methods

### 24     101 Algorithm <sup>2</sup>H-SIAM(1.0)

25  
26     102 When a features list (.csv format) is imported into <sup>2</sup>H-SIAM(V1.0), it starts with  
27     103 the calculation of means of replicates. Then, the features list is inspected within an  
28     104 indicated duration of RT for pairs of two features as potential isotopologue pairs; the  
29     105 potential natural compounds are denoted as M, and the potential isotope-labeled  
30     106 compounds are denoted as M'. The mean intensities of the features M and M' from  
31     107 samples Mix<sub>1:3</sub> and Mix<sub>3:1</sub> are denoted as M<sub>Mix1:3</sub>, M<sub>Mix3:1</sub>, M'<sub>Mix1:3</sub> and M'<sub>Mix3:1</sub>  
32     108 respectively. They are used for the calculation of the following 3 actual ratios (f<sub>n</sub>):  
33  
34

$$43 \quad 109 f_1 = \frac{M_{Mix1:3}}{M'_{Mix1:3}}, f_2 = \frac{M_{Mix3:1}}{M'_{Mix3:1}}, f_3 = \frac{M'_{Mix1:3}}{M'_{Mix3:1}}$$

44  
45     110 users defined parameters R<sub>n</sub> and Tol.<sub>n</sub> are used to set triple filters (F<sub>1</sub>, F<sub>2</sub> and F<sub>3</sub>)  
46  
47     111 for TPs as following:  
48

$$49 \quad 112 \text{Filter n, } F_n: R_n \times \text{Tol.}_n < f_n < R_n \times \text{Tol.}_n^{-1}$$

50  
51     113 where, R<sub>n</sub> is theoretic ratios for F<sub>n</sub>, and Tol.<sub>n</sub> is tolerances for R<sub>n</sub>, 0 < Tol.<sub>n</sub> < 1.

52  
53     114 When f<sub>1</sub> follows the requirement of F<sub>1</sub>, a pair of two features (isotopologue pair)  
54     115 is tracked by the algorithm and further filters F<sub>2</sub> and F<sub>3</sub> will be sequentially evaluated.  
55  
56     116 The features passing triple filters contain features as potential TPs, deserving further  
57     117 annotation and identification.  
58

### 59     118 Recovery of 7 Isotopologue Pairs from Soil Extract by <sup>2</sup>H-SIAM

1  
2  
3       119 Crude soil was air-dried and sieved (< 2 mm) to remove debris and ready for use  
4       120 as blank soil as our previous report<sup>24</sup>. Per gram of blank soil was extracted by 10 mL  
5       121 of acetone (ACE) and hexane (HEX) (1:3, v/v) by microwave extraction (Anton Paar  
6       122 GmbH, Multiwave PRO, Austria). After that, ACE and HEX were replaced by  
7       123 acetonitrile (MeCN) by solvent exchange, and anhydrous Na<sub>2</sub>SO<sub>4</sub> (baked at 450 °C for  
8       124 4 h prior to use) was used to remove residual water. Extracts were concentrated to 2 mL  
9       125 under nitrogen flow and divided into three aliquots, and one of them was used as blank.  
10  
11  
12  
13  
14

15       126 10 ppm of 7 natural contaminants (dichlorvos, atrazine, sulfamethoxazole,  
16       127 naphthalene, pyrene, fluorene, anthracene) and 30 ppm of their <sup>2</sup>H-labeled  
17       128 isotopologues (dichlorvos-*d*6, atrazine-*d*5, sulfamethoxazole-*d*4, naphthalene-*d*8,  
18       129 pyrene-*d*10, fluorene-*d*10, anthracene-*d*10) were added into one aliquot to obtain Mix<sub>1:3</sub>  
19       130 sample. Meanwhile, 30 ppm of natural contaminants and 10 ppm of <sup>2</sup>H-labeled  
20       131 isotopologues were added to another aliquot to obtain a Mix<sub>3:1</sub> sample. 20 ppm of 7  
21       132 isotope pairs were prepared as standard. Fluorene-<sup>13</sup>C<sub>6</sub> was added to all the  
22       133 extracts as an internal standard.  
23  
24  
25  
26  
27  
28

29       134 Standard, blanks and mixed samples were analyzed by UPLC-ESI-HRMS,  
30       135 Ultimate 3000 (Dionex) coupled with a Q Exactive Orbitrap mass spectrometers  
31       136 (Thermo) and heated electrospray ionization (ESI) source. More details were provided  
32       137 in [Supporting Information 1](#).

### 33       138 **Synthesis and Structure of 6PPD-*d*9 and its Application in <sup>2</sup>H-SIAM**

34  
35       139 500 mg 6PPD, 100 mg 5% Pt/C, 50 mg 10% Pd/C and 15 mL D<sub>2</sub>O were mixed in  
36       140 a 100 mL glass reactor. Then, the air in the reactor was replaced by H<sub>2</sub> (99.999%), and  
37       141 the reaction was stirred at 80 °C under an H<sub>2</sub> atmosphere for 24 h<sup>25,26</sup>. Then the mixture  
38       142 was extracted with ethoxyethane (ETH) and dried in a centrifuge evaporator (RVC 2-  
39       143 25 CD Plus, Christ, Germany) to obtain brown color products (1<sup>st</sup> round products). The  
40       144 reaction was repeated twice to obtain highly <sup>2</sup>H-labeled 6PPD-*d*9 (3<sup>rd</sup> round products,  
41       145 about 50 mg). 6PPD, 1<sup>st</sup>, and 3<sup>rd</sup> rounds of H/D exchange products were characterized  
42       146 by GC-MS (Thermo, Trace 1300, TSQ9000EVO triple quadrupole). The 3<sup>rd</sup> round  
43       147 product was further characterized by a 600 MHz nuclear magnetic resonance  
44       148 spectrometer (NMR, Ascend 600, Bruker). <sup>1</sup>H NMR spectra were obtained by using  
45       149 DMSO-*d*6 as solvent. The deuterium incorporation levels were determined by the <sup>1</sup>H  
46       150 NMR spectrum, which showed hydrogen residue signals.

47  
48  
49  
50  
51       151 6PPD and synthesized 6PPD-*d*9 were added into soil extract to prepare Mix<sub>1:3</sub> and  
52       152 Mix<sub>3:1</sub> samples independently as introduced above. Sulfadiazine-<sup>13</sup>C<sub>6</sub> was used as an  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 153 internal standard, and other details were provided in [Supporting Information 1](#).  
4  
5

#### 6 154 **Non-targeted Analysis of TPs of Pyrene from Soil by $^2\text{H}$ -SIAM Pipeline**

7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

155 To evaluate the performance of the  $^2\text{H}$ -SIAM pipeline, a soil pyrene degradation  
156 study was carried out. Pyrene and pyrene-*d*10 were dissolved in ACE and mixed with  
157 one-quarter blank soil. After solvent evaporation, they were mixed respectively with  
158 the rest bulk soils to obtain 100 ppm pyrene and 100 ppm pyrene-*d*10 contaminated  
159 soils. Then, blank and contaminated soils were placed at room temperature for 60 days  
160 and watered twice per month to keep soil moisture. One aliquot of contaminated soils  
161 was stored in -80 °C to verify the stability of  $^2\text{H}$ -labels.

162 After 60 days incubation, soils were air-dried at 60 °C until to constant weight. Per  
163 gram of soils were extracted by 10 mL ACE and HEX (1:3, v/v) by microwave  
164 extraction with the addition of ortho-terphenyl (OTP) as extraction surrogate, and  
165 anhydrous Na<sub>2</sub>SO<sub>4</sub> were used to remove residual water. The solvent was replaced by  
166 acetonitrile (MeCN), and extracts were concentrated to 1 mL under nitrogen flow.

167 Extracts from pyrene and pyrene-*d*10 treatments were mixed with the ratio of 1:3  
168 and 3:1 to obtain Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples. Anthracene-*d*10 was added into samples  
169 as an internal standard. Then, samples were subjected to UPLC-ESI-HRMS analysis  
170 and GC-EI-HRMS analysis, and details were provided in [Supporting Information 1](#).

171 Mass spectrum raw data were formatted to .mzXML and MZmine2 was used to  
172 obtain features lists<sup>27,28</sup>. Features of isotopologue pairs were tracked by  $^2\text{H}$ -SIAM(1.0),  
173 and it tracked 162 features from GC-EI-HRMS data (21955 features) and 52 features  
174 from UPLC-ESI-HRMS data (4376 features).

175 The tracked features from HRMS were recorded as MS<sup>1</sup> precursors, and they were  
176 annotated against KEGG and PubChem database (by MZmine2), or manually. Then,  
177 the MS<sup>1</sup> precursors from UPLC-ESI-HRMS analysis were further fragmented in  
178 quadrupole mode with a 0.4 m/z isolation window and nominal collision energy of 40,  
179 and MS<sup>2</sup> fragmentations were scanned with orbitrap at a resolution of 140000. The MS<sup>2</sup>  
180 fragments from UPLC-ESI-HRMS and fragments from GC-EI-HRMS were matched  
181 against the mass spectrum predicted by CFM-ID 3.0 for further identification<sup>29</sup>.  
182 Detailed information was provided in [Supporting Information 1](#).

#### 183 **Identification Confidence**

184 Compared to common NTA, the tracked features by  $^2\text{H}$ -SIAM coeluted with their  
185 isotopologues, confirming its origination from substrates and providing basic structure  
186 information. Additionally, the number of  $^2\text{H}$ -labels provides potential structure

1  
2  
3 187 information. Thus, we annotated TPs with identification confidence levels proposed by  
4 188 Schymanski et al. but with little modification<sup>30</sup>.  
5  
6

7 189 Level 1 was achieved by matching the RT and MS<sup>2</sup> with reference standards; Level  
8 190 2a was achieved by matching the number of possible <sup>2</sup>H-labels and at least two major  
9 191 fragment ions with MS<sup>2</sup> libraries (from *in silico* or references spectrum), or achieved  
10 192 by matching at least three major fragment ions with MS<sup>2</sup> libraries; Level 2b was  
11 193 achieved by matching the number of possible <sup>2</sup>H-labels and one major fragment ions  
12 194 with MS<sup>2</sup> libraries, or achieved by matching at least two major fragment ions with MS<sup>2</sup>  
13 195 libraries; Level 3 was achieved by matching the number of possible <sup>2</sup>H-labels and  
14 196 natural isotope pattern, or achieved by matching one major fragment ions with MS<sup>2</sup>  
15 197 libraries; Level 4 was achieved by matching the number of possible <sup>2</sup>H-labels, or natural  
16 198 isotope pattern; Level 5 was achieved by matching the exact mass.  
17  
18

#### 199 **Quality Control and Data Availability**

200 For qualitative analysis, quality control (QC) samples were prepared by a pool of  
201 extracts, and it was used to stabilize the chromatographic system and to verify the  
202 stability of the measurements. Indicated internal standard was added into samples to  
203 calibrate the fluctuation of the instrument. Quantitative of pyrene and pyrene-*d*10 by  
204 GC-MS were analyzed using one-way analysis of variance (ANOVA) followed by the  
205 Duncan test, where different letters indicate significant differences at *p* < 0.05.

206 The code of <sup>2</sup>H-SIAM (V1.0) was published in  
207 <https://github.com/kechen1984/2H-SIAM>, which provide a link from Mendeley Data  
208 with example data of the soil pyrene degradation study, software, and details for use of  
209 <sup>2</sup>H-SIAM(1.0).  
210

## 43 **Results and Discussion**

### 44 211 **Methodology of <sup>2</sup>H-SIAM**

45 212 The application of <sup>2</sup>H-SIAM in the environmental studies still faces challenges.  
46 213 People may worry about the stability of <sup>2</sup>H-labeled contaminants due to the so-called  
47 214 deuterium loss or H/D back exchange, resulting in the loss of <sup>2</sup>H-labels during the study.  
48 215 The stability of <sup>2</sup>H-labels in contaminants depends on their location. H/D back  
49 216 exchange generally occurs in the labile <sup>2</sup>H atom, e.g., <sup>2</sup>H in carboxyl and amino groups,  
50 217 which may result in the loss of 20-30% <sup>2</sup>H atoms, and it does not occur in the sites of  
51 218 C-D, which is the most popular <sup>2</sup>H-labeled site in contaminants<sup>31, 32</sup>. Our study also  
52 219 proved that 60 days incubation of pyrene and pyrene-*d*10 in soil under -80 °C or ambient  
53 220 temperature (biotransformed by microorganisms) did not result in a significant  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 221 difference between signals from pyrene and pyrene-*d*10 ([Fig. S1](#)), confirming the  
4 222 stability of <sup>2</sup>H-labeled contaminants over the experiment.  
5  
6

7 223 Additionally, unmatured methodology and algorithms limit the use of <sup>2</sup>H-SIAM in  
8 224 the environment studies ([Table S1](#)). We developed a features-based algorithm <sup>2</sup>H-  
9 225 SIAM(V1.0) ([Fig S2](#)) to carry out SIAM, and the proposed pipeline is shown in [Fig. 1](#).  
10 226 It is a fully GUI-based pipeline, lowering the skill barriers for its application. Parent  
11 227 isotopologue pairs of contaminants are added into the environmental matrix  
12 228 respectively and incubated for the indicated duration. They are extracted and mixed  
13 229 with indicated ratios (e.g., 1:1 and 1:2, 1:3 and 3:1, 1:9 and 9:1). We propose a 1:3 and  
14 230 3:1 ratio (Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples), keeping an appropriate equilibrium between  
15 231 accuracy and sensitivity for tracking isotopologue pairs. That is because a high mix  
16 232 ratio will dilute extracts, leading to the loss of positive signals.  
17  
18

19 233 Mixed samples are then determined by MS platforms (e.g., UPLC-ESI-HRMS,  
20 234 GC-EI-HRMS). MS raw data from Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples are transformed  
21 235 to .mzXML format and imported into MZmine2 to obtain a features list ([Fig. S3](#)). Any  
22 236 other MS data processing algorithm, e.g., XCMS and MS-DIAL, are compatible when  
23 237 they could output a features list. The key for efficiently tracking TPs is to acquire a  
24 238 high-quality features list, and the optimized parameters for the selected algorithm are  
25 239 critically important.  
26  
27

28 240 When the features list (.csv format) is imported into <sup>2</sup>H-SIAM(V1.0), it starts with  
29 241 qualitatively tracking of isotopologue pairs with the indicated isotope labels, m/z, and  
30 242 RT tolerance ([Fig. S4](#)). Then, 3 quantitative filters F<sub>1</sub>, F<sub>2</sub>, and F<sub>3</sub> are used to track  
31 243 features of TPs with isotopologues. After that, <sup>2</sup>H-SIAM(V1.0) will output a new  
32 244 features list, and only a small portion of features (1% - 2%) will be tracked by triple  
33 245 filters and be marked with numbers of isotope labels ([Table 1, Supporting Information](#)  
34 246 [2](#) and [3](#)). This step will exclude many noise features, and potential TPs could be  
35 247 annotated among the tracked features, contributing to further identification. Although  
36 248 the <sup>2</sup>H-SIAM pipeline is designed for the tracking of <sup>2</sup>H-labeled TPs by UPLC-ESI-  
37 249 HRMS, data from any other types of LC-MS or GC-MS platform or any other types of  
38 250 isotope-labeled atoms, e.g., <sup>13</sup>C, <sup>15</sup>N, and <sup>18</sup>O, are compatible ([Table S2](#)).  
39  
40

41 251 Additionally, even though many MS<sup>2</sup>-based algorithms have been recently  
42 252 developed for the discovery of unknown compounds <sup>33-35</sup>, it is still critically important  
43 253 to obtain TPs information by isotope labels and MS<sup>1</sup> features. That is because isotope  
44 254 labels provide direct evidence for the origination of TPs. Meanwhile, the signals from  
45  
46

1  
2  
3 255 TPs of substrates are always weak, and their signals may be lost in data dependent or  
4 256 independent acquisition of MS<sup>2</sup>.  
5  
6

### 257 Accuracy of the <sup>2</sup>H-SIAM Pipeline

7  
8 258 To verify the accuracy of the proposed pipeline, 7 isotopologue pairs of typical  
9 environmental contaminants were added into soil extract to examine whether the  
10 designed pipeline could pick them up from soil extracts. Same amounts of contaminants  
11 (10 ppm) were ionized by ESI with different efficiency, and signals varied from  $3 \times 10^5$   
12 to  $7 \times 10^7$  (height, [Fig. 2a](#) and [Fig. S5](#)). Signals from polycyclic aromatic hydrocarbons  
13 (PAHs, anthracene, pyrene, fluorene, and naphthalene) were lower than pesticides,  
14 herbicides, and antibiotics due to relatively lower ionization efficiency.  
15  
16

17 265 The features-based algorithm <sup>2</sup>H-SIAM(V1.0) compared features by their  
18 maximum intensities (height/area) with indicated RT tolerance, avoiding the loss of  
19 positive signals. All of the 7 isotopologue pairs were correctly tracked from thousands  
20 of noises ([Table 1](#), [Supporting Information 2](#)), and two examples with triple filters were  
21 shown in [Fig. 2b](#) and [2c](#). Filters F<sub>1</sub> and F<sub>2</sub> compare intensities of features from natural  
22 and <sup>2</sup>H-labeled compounds from Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples, and the ratios theoretically  
23 should be 1/3 and 3, respectively. Filter F<sub>3</sub> compared intensities of <sup>2</sup>H-labeled  
24 compounds from Mix<sub>1:3</sub> and Mix<sub>3:1</sub>, and theoretically, it should be 3. Then combination  
25 of the theoretic ratio and appropriate settled tolerance constitute triple filters embedded  
26 in the algorithm. Additionally, since signals from the same amount of natural and <sup>2</sup>H-  
27 labeled contaminants are not always equivalent, F<sub>1</sub> and F<sub>2</sub> constitute relaxed filters with  
28 a lower Tol. of 0.3, and F<sub>3</sub> constitute a stricter filter with a higher Tol. of 0.5.  
29  
30

31 277 The peak times of isotopologue pairs of PAHs were different, which may affect the  
32 performance of SIAM analyzed by scan-based algorithms ([Table S1](#)). They compare  
33 natural and isotope-labeled signals from the same scan, relying on the coelute of  
34 isotopologue pairs. [Fig. 2d](#) and [Fig. 2e](#) depicted mass spectrums (MS scan #1691 and  
35 MS scan #1726) from peaks of EIC (Extracted Ion Chromatogram) of pyrene and  
36 pyrene-d10 of Mix<sub>1:3</sub> sample at 14.75 and 15.05 minutes. As shown, when isotopologue  
37 pairs elute at different times, there is an uncrossable gap between the theoretic (1:3,  
38 equal to 0.333) and observed (0 and 53) ratio of height between isotopologue pairs.  
39 That may result in the loss of positive signals for scan-based algorithms.  
40  
41

### 286 H/D Exchange and its Application in <sup>2</sup>H-SIAM

42 287 Direct H/D exchange reaction enables unique or site-specific H/D-exchange,  
43 providing valuable and cost-effective isotope-labeled contaminants <sup>14</sup>, <sup>36</sup>. 6PPD-  
44  
45

1  
2  
3      289 quinone is the TPs of 6PPD, and the toxicity of 6PPD-quinone to *Oncorhynchus kisutch*  
4      290 is confirmed by *Science* in 2021<sup>37</sup>. Here we show how <sup>2</sup>H-labels be selectively added  
5      291 into this emerging contaminant when a <sup>13</sup>C-labeled compound is unreachable. The H/D-  
6      292 exchange reacted in a glass tube with catalysts, H<sub>2</sub>, and D<sub>2</sub>O as deuterium source at 80  
7      293 °C for 24 hours ([Fig. 3a](#)) . After the 1<sup>st</sup> round reaction, we found 6PPD-*d*6, 6PPD-*d*7,  
8      294 6PPD-*d*8, and 6PPD-*d*9 in the mass spectrum ([Fig. 3d](#)), and 6PPD-*d*9 signal accounts  
9      295 for approximately 21% of all <sup>2</sup>H-labeled 6PPD. After the 3<sup>rd</sup> round reaction, the 6PPD-  
10     296 *d*9 (~50mg) accounts for approximately 51% of all <sup>2</sup>H-labeled 6PPD, and 80% of <sup>2</sup>H-  
11     297 labels were added onto aromatic rings ([Fig. 3be](#)), which could be used to track 6PPD  
12     298 by <sup>2</sup>H-SIAM(1.0) ([Fig. 3fg](#)).

13  
14  
15     299 As 100 grams of D<sub>2</sub>O (99.9%, Sigma-Aldrich) cost only ~\$300, and catalysts could  
16     300 be reused, the possible H/D-exchange reaction offers various affordable <sup>2</sup>H-labeled  
17     301 contaminants. Thus, it enables us to carry out <sup>2</sup>H-SIAM in a complex environmental  
18     302 matrix, for instance, an ecotron with several cubic meters of reconstituted aquatic  
19     303 ecosystem.

20  
21     304 **Performances of <sup>2</sup>H-SIAM in Tracking TPs of Pyrene in Soil**

22  
23  
24     305 The performance of <sup>2</sup>H-SIAM in tracking possible TPs of contaminants is evaluated  
25     306 by a soil PAHs degradation study. Pyrene was selected as the parent contaminant, as  
26     307 TPs of pyrene have been widely studied and recorded in KEGG (<https://www.kegg.jp>),  
27     308 enabling us to evaluate the accuracy of the <sup>2</sup>H-SIAM pipeline. Additionally, the peak  
28     309 times of its isotopologue pairs are different, and its ionization efficiency is relatively  
29     310 low, providing a challenging situation for the proposed <sup>2</sup>H-SIAM pipeline.

30  
31     311 Overview of the performance of the <sup>2</sup>H-SIAM pipeline carried out by UPLC-ESI-  
32     312 HRMS and GC-EI-HRMS were provided in [Table 1](#), and the pipeline efficiently  
33     313 discarded ~99% features, and only ~1% features had been tracked by triple filters (162  
34     314 and 52). It is, arbitrarily, more efficient than a Credentialing Features study, <sup>13</sup>C-SIAM  
35     315 of glucose, which discarded 93% features and ~1800 features left (~7%)<sup>15</sup>. It also  
36     316 confirmed the importance of triple filters. If only one quantitative filter was used (F<sub>1</sub>),  
37     317 1122 and 318 features were left ([Table 1](#)), which is six times more than that from triple  
38     318 filters (162 and 52 features). Thus, the algorithms with triple filters are better than  
39     319 algorithms with only one filter ([Table S1](#)).

40  
41     320 For data from UPLC-ESI-HRMS, the algorithm tracked 52 potential isotopologue  
42     321 pairs, and 13 isotopologue pairs were annotated as potential TPs ([Fig. 4](#), [Table S3](#), [Fig.](#)  
43     322 [S6-S10](#)), which accounts for highly to 25% of the total extracted features. Cheng

1  
2  
3 323 successfully annotated 7 isotopologue pairs as potential TPs of mono-*n*-butyl phthalate  
4 324 using  $^2\text{H}$ -labeled contaminants and MetExtract II <sup>38</sup>. Tian annotated 6 isotopologue  
5 325 pairs as potential TPs of pyrene by  $^{13}\text{C}$ -labeled pyrene and X13CMS <sup>39</sup>. By our  
6 326 proposed pipeline, 13 TPs of pyrene were annotated proving its higher performance  
7 327 with cost-effective properties.  
8  
9

10 328 4 isotopologue pairs detected from UPLC-ESI-HRMS data (e.g., M<sub>221</sub>, [M+H]<sup>+</sup>)  
11 329 were further tracked by GC-EI-HRMS (e.g., M<sub>220</sub>, [M]<sup>+</sup>, [Table S4](#)). Their fragment  
12 330 from natural and  $^2\text{H}$ -labeled contaminants fit well to the fragment predicted by CFM-  
13 331 ID ([Fig. 4e](#), [Fig S11-S13](#)) <sup>29</sup>, contributing to improve the identification confidence  
14 332 levels of TPs obtained by UPLC-ESI-HRMS data ([Table S3](#)).  
15  
16

17 333 Additionally, isotope labels are important for tracking possible TPs in the  
18 334 environment because the environment matrix may contain TPs naturally. As shown in  
19 335 [Fig. S14](#), we detected a feature annotated as phthalic acid, TPs of pyrene listed in KEGG.  
20 336 However, we could not find its  $^2\text{H}$ -labeled isotopologue, indicating that the detected  
21 337 phthalic acid is natural existed rather than TPs of pyrene in our experiment regime.  
22  
23

24 338 M<sub>233</sub> at 12.8 min and 13.6 min from UPLC-ESI-HRMS data were annotated as  
25 339 pyrenequinone ([Fig. 4b](#), [Table S3](#)), and their isotopologue pairs M<sub>233-d8</sub> were correctly  
26 340 paired. According to the difference of LogK<sub>ow</sub> of pyrenequinone estimated by EPISuite  
27 341 (<https://www.epa.gov>) and RT of detected features, we could tentatively annotate M<sub>233</sub>  
28 342 at 12.8 min as 1,6- pyrenequinone and annotate M<sub>233</sub> at 13.6 min as 4,5-pyrenequinone  
29 343 respectively. MS<sup>2</sup> of M<sub>233</sub> and M'<sub>233-d8</sub> at 13.6 min were plotted in [Fig. 4d](#). The detected  
30 344 MS<sup>2</sup> spectrum and isotope pattern of M<sub>233</sub> and M<sub>233-d8</sub> matched the MS<sup>2</sup> spectrum and  
31 345 isotope pattern predicted by algorithms within 5 ppm mass error, indicating  
32 346 identification confidence at Level 2a.  
33  
34

35 347 Except for reported TPs of pyrene, we also detected 5 unreported potential TPs of  
36 348 pyrene, M<sub>235</sub>, M<sub>225</sub>, M<sub>244</sub>, M<sub>247</sub>, and M<sub>231</sub>. M<sub>235</sub>, corresponding to [C<sub>16</sub>H<sub>10</sub>O<sub>2</sub>+H]<sup>+</sup> with  
37 349 isotope patent score of 97% (MZmine2), was annotated by KEGG as dihydropyrene or  
38 350 1-hydroxypyrene-6,7-oxide and 1-hydroxypyrene-7,8-oxide. However, the  
39 351 isotopologue pair of M<sub>235</sub> is M'<sub>235-d10</sub>, labeled with 10 of  $^2\text{H}$  atoms. The formation of  
40 352 pyreneoxide is accomplished by cytochrome P450 superfamily, an unspecific  
41 353 monooxygenase, which theoretically will not result in the loss of  $^2\text{H}$ -labels. Thus, M<sub>235</sub>  
42 354 was annotated as pyreno[4,5-b:9,10-b']bisoxirene, 3b,4a,7b,8a-tetrahydro- (PubChem  
43 355 CID 616438) according to 10 discovered  $^2\text{H}$  atoms in M'<sub>235-d10</sub>, which has not been  
44 356 included in KEGG. This speculation exhibited the advantage of  $^2\text{H}$ -labels over  $^{13}\text{C}$ -  
45  
46

1  
2  
3 357 labels, as the numbers of  $^2\text{H}$ -labels contribute to deducing possible structure. Thus, as  
4 358 introduced in the Methods, we slightly promote the identification confidence levels of  
5 359 potential TPs with correct  $^2\text{H}$ -labels.  
6  
7

8 360 Speculation of the structure of  $\text{M}_{231}$  and  $\text{M}_{247}$  were ambiguous due to the  
9 361 inconsistent numbers of  $^2\text{H}$ -labels. However,  $\text{M}_{231}$ ,  $\text{M}_{247}$ , and  $\text{M}_{244}$  point to the  
10 362 possibility that carbon atom is directly added to aromatic rings, which has not been  
11 363 reported before in the environmental studies. Due to the focus of the study, their  
12 364 identification will be further confirmed in the future by separation, purification, and  
13 365 NMR or  $\text{MS}^n$  structure analysis. Additionally, it should be noted that NTA by mass  
14 366 spectrometry is blind to stereochemistry and often regiochemistry<sup>40</sup>. Isomers of the TPs  
15 367 cannot be excluded, and exact annotations should be done by comparing with references  
16 368<sup>3, 41</sup>.

17 369 **Conclusion**

18 370 In summary, our study confirms the performance of the  $^2\text{H}$ -SIAM pipeline in  
19 371 tracking and discovering TPs of contaminants in the environment. The proposed fully  
20 372 GUI-based pipeline is friendly to users. Additionally, it has shown unique advantages  
21 373 over  $^{13}\text{C}$ -SIAM, for instance, affordable and highly customized  $^2\text{H}$ -labeled  
22 374 contaminants, and providing more structure information. Thus,  $^2\text{H}$ -SIAM is  
23 375 recommended as a powerful tool in understanding the fate of environmental  
24 376 contaminants.

25 377 **Acknowledgments**

26 378 This work was financially supported by the National Natural Science Foundation  
27 379 of China (41503067 and 31870513). This paper is dedicated to the memory of S.Y. Ke.

28 380 **Author Contributions**

29 381 K.C. designed the study; K.C., Y.H.X, X.Y.Y., and Z.H.L. collected and analyzed  
30 382 the data; K.C. wrote the manuscript; R.Q. and J.S. financially supported the study; all  
31 383 authors contributed substantially to the revision.

32 384 **Competing interests declares:**

33 385 The authors declare no competing financial interest.

34 386 **Supporting Information:**

35 387 [Supporting Information 1](#): supplemental methods, figures, and tables. [Supporting](#)  
36 388 [Information 2](#): output files of  $^2\text{H}$ -SIAM(1.0) for the recovery study of 7 contaminants  
37 389 (UPLC-ESI-HRMS). [Supporting Information 3](#): output files of  $^2\text{H}$ -SIAM(1.0) for the  
38 390 soil pyrene degradation study (GC-EI-HRMS).

**References**

1. Brussaard, C. P. D.; Peperzak, L.; Beggah, S.; Wick, L. Y.; Wuerz, B.; Weber, J.; Arey, J. S.; van  
2. der Burg, B.; Jonas, A.; Huisman, J.; van der Meer, J. R., Immediate ecotoxicological effects of  
3. short-lived oil spills on marine biota. *Nature Communications* **2016**, *7*, 11206.
4. Gonzalez-Gaya, B.; Martinez-Varela, A.; Vila-Costa, M.; Casal, P.; Cerro-Galvez, E.; Berrojalbiz,  
5. N.; Lundin, D.; Vidal, M.; Mompean, C.; Bode, A.; Jimenez, B.; Dachs, J., Biodegradation as an  
6. important sink of aromatic hydrocarbons in the oceans. *Nat Geosci* **2019**, *12*, (2), 119–125.
7. Sindelar, M.; Patti, G. J., Chemical Discovery in the Era of Metabolomics. *J Am Chem Soc* **2020**,  
8. *142*, (20), 9097-9105.
9. Perez de Souza, L.; Alseekh, S.; Scossa, F.; Fernie, A. R., Ultra-high-performance liquid  
10. chromatography high-resolution mass spectrometry variants for metabolomics research. *Nat  
11. Methods* **2021**.
12. Place, B. J.; Ulrich, E. M.; Challis, J. K.; Chao, A.; Du, B.; Favela, K.; Feng, Y.-L.; Fisher, C. M.;  
13. Gardinali, P.; Hood, A.; Knolhoff, A. M.; McEachran, A. D.; Nason, S. L.; Newton, S. R.; Ng, B.;  
14. Nuñez, J.; Peter, K. T.; Phillips, A. L.; Quinete, N.; Renslow, R.; Sobus, J. R.; Sussman, E. M.;  
15. Warth, B.; Wickramasekara, S.; Williams, A. J., An Introduction to the Benchmarking and  
16. Publications for Non-Targeted Analysis Working Group. *Anal Chem* **2021**, *93*, (49), 16289-16296.
17. Hootman, K. C.; Trezzi, J.-P.; Kraemer, L.; Burwell, L. S.; Dong, X.; Guertin, K. A.; Jaeger, C.;  
18. Stover, P. J.; Hiller, K.; Cassano, P. A., Erythritol is a pentose-phosphate pathway metabolite and  
19. associated with adiposity gain in young adults. *P Natl Acad Sci Usa* **2017**, *114*, (21), E4233-E4240.
20. Chen, Y. J.; Mahieu, N. G.; Huang, X. J.; Singh, M.; Crawford, P. A.; Johnson, S. L.; Gross, R. W.;  
21. Schaefer, J.; Patti, G. J., Lactate metabolism is associated with mammalian mitochondria. *Nat Chem  
22. Biol* **2016**, *12*, (11), 937-943.
23. Mueller, D.; Heinze, E., Stable isotope-assisted metabolomics to detect metabolic flux changes in  
24. mammalian cell cultures. *Curr Opin Biotech* **2013**, *24*, (1), 54-59.
25. Creek, D. J.; Chokkathukalam, A.; Jankevics, A.; Burgess, K. E. V.; Breitling, R.; Barrett, M. P.,  
26. Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal Chem*  
27. **2012**, *84*, (20), 8442-8447.
28. Llufrio, E. M.; Cho, K.; Patti, G. J., Systems-level analysis of isotopic labeling in untargeted  
29. metabolomic data by X-13 CMS. *Nat Protoc* **2019**, *14*, (7), 1970-1990.
30. Forsberg, E. M.; Huan, T.; Rinehart, D.; Benton, H. P.; Warth, B.; Hilmers, B.; Siuzdak, G., Data  
31. processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online.  
32. *Nat Protoc* **2018**, *13*, (4), 633-651.
33. Huan, T.; Forsberg, E. M.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M. L.;  
34. Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgeresen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M.  
35. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.;  
36. Goodacre, R.; Lairson, L. L.; Siuzdak, G., Systems biology guided by XCMS Online metabolomics.  
37. *Nat Methods* **2017**, *14*, (5), 461-462.
38. Zhu, Z. J.; Schultz, A. W.; Wang, J. H.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G.,  
39. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of  
40. metabolites in rat urine. *Anal Chem* **2013**, *85*, (18), 8630-8637.
41. Patti, G. J.; Huan, T.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M. L.;  
42. Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgeresen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M.  
43. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.;  
44. Goodacre, R.; Lairson, L. L.; Siuzdak, G., Systems biology guided by XCMS Online metabolomics.  
45. *Nat Methods* **2017**, *14*, (5), 461-462.
46. Zhu, Z. J.; Schultz, A. W.; Wang, J. H.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G.,  
47. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of  
48. metabolites in rat urine. *Anal Chem* **2013**, *85*, (18), 8630-8637.
49. Patti, G. J.; Huan, T.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M. L.;  
50. Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgeresen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M.  
51. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.;  
52. Goodacre, R.; Lairson, L. L.; Siuzdak, G., Systems biology guided by XCMS Online metabolomics.  
53. *Nat Methods* **2017**, *14*, (5), 461-462.
54. Zhu, Z. J.; Schultz, A. W.; Wang, J. H.; Johnson, C. H.; Yannone, S. M.; Patti, G. J.; Siuzdak, G.,  
55. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of  
56. metabolites in rat urine. *Anal Chem* **2013**, *85*, (18), 8630-8637.
57. Patti, G. J.; Huan, T.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M. L.;  
58. Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgeresen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M.  
59. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.;  
60. Goodacre, R.; Lairson, L. L.; Siuzdak, G., Systems biology guided by XCMS Online metabolomics.  
61. *Nat Methods* **2017**, *14*, (5), 461-462.

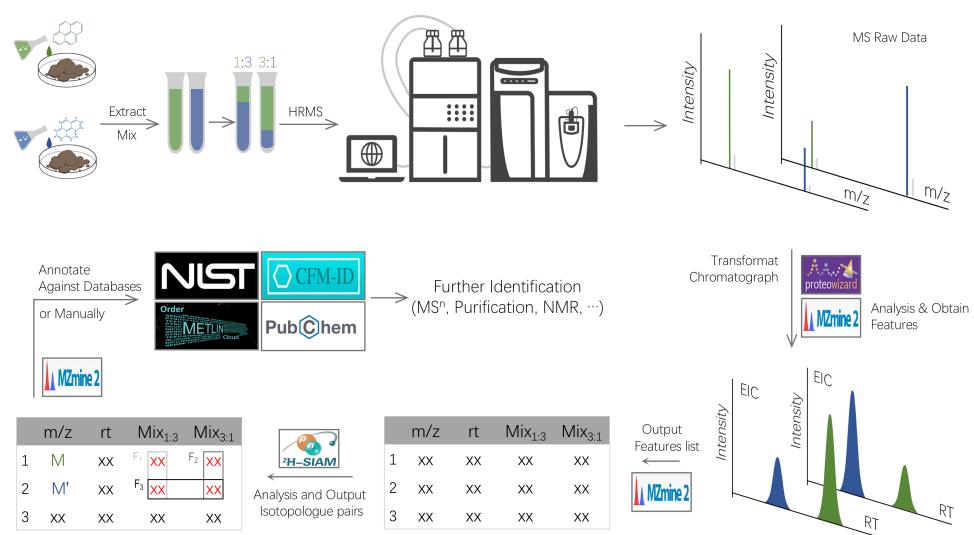
- 1  
2  
3       431       metabolites guided by the METLIN database. *Nat Protoc* **2013**, *8*, (3), 451-460.  
4       432       14. Atzrodt, J.; Derdau, V.; Kerr, W. J.; Reid, M., C-H functionalisation for hydrogen isotope exchange.  
5       433       *Angew Chem Int Edit* **2018**, *57*, (12), 3022-3047.  
6       434       15. Mahieu, N. G.; Huang, X.; Chen, Y., Jr.; Patti, G. J., Credentialing Features: A Platform to  
7       435       Benchmark and Optimize Untargeted Metabolomic Methods. *Anal Chem* **2014**, *86*, (19), 9583-9589.  
8       436       16. Kessler, N.; Walter, F.; Persicke, M.; Albaum, S. P.; Kalinowski, J.; Goesmann, A.; Niehaus, K.;  
9       437       Nattkemper, T. W., ALLOCATOR: An Interactive Web Platform for the Analysis of Metabolomic LC-  
10      438       ESI-MS Datasets, Enabling Semi-Automated, User-Revised Compound Annotation and Mass  
11      439       Isotopomer Ratio Analysis. *Plos One* **2014**, *9*, (11).  
12      440       17. Bueschl, C.; Kluger, B.; Neumann, N. K. N.; Doppler, M.; Maschietto, V.; Thallinger, G. G.; Meng-  
13      441       Reiterer, J.; Krska, R.; Schuhmacher, R., MetExtract II: A Software Suite for Stable Isotope-Assisted  
14      442       Untargeted Metabolomics. *Anal Chem* **2017**, *89*, (17), 9518-9526.  
15      443       18. Leeming, M. G.; Isaac, A. P.; Pope, B. J.; Cranswick, N.; Wright, C. E.; Ziogas, J.; O'Hair, R. A. J.;  
16      444       Donald, W. A., High-Resolution Twin-Ion Metabolite Extraction (HiTIME) Mass Spectrometry:  
17      445       Nontargeted Detection of Unknown Drug Metabolites by Isotope Labeling, Liquid Chromatography  
18      446       Mass Spectrometry, and Automated High-Performance Computing. *Anal Chem* **2015**, *87*, (8), 4104-  
19      447       4109.  
20      448       19. Itoh, N.; Numata, M.; Aoyagi, Y.; Yarita, T., Comparison of the behavior of <sup>13</sup>C- and deuterium-  
21      449       labeled polycyclic aromatic hydrocarbons in analyses by isotope dilution mass spectrometry in  
22      450       combination with pressurized liquid extraction. *Journal of chromatography. A* **2007**, *1138*, (1-2),  
23      451       26-31.  
24      452       20. Chokkathukalam, A.; Jankevics, A.; Creek, D. J.; Achcar, F.; Barrett, M. P.; Breitling, R., mzMatch-  
25      453       ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry  
26      454       data. *Bioinformatics* **2013**, *29*, (2), 281-283.  
27      455       21. Huang, X. J.; Chen, Y. J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J., X13CMS: Global  
28      456       Tracking of Isotopic Labels in Untargeted Metabolomics. *Anal Chem* **2014**, *86*, (3), 1632-1639.  
29      457       22. Capellades, J.; Navarro, M.; Samino, S.; Garcia-Ramirez, M.; Hernandez, C.; Simo, R.; Vinaixa,  
30      458       M.; Yanes, O., geoRge: A Computational Tool To Detect the Presence of Stable Isotope Labeling in  
31      459       LC/MS-Based Untargeted Metabolomics. *Anal Chem* **2016**, *88*, (1), 621-628.  
32      460       23. Hiller, K.; Wegner, A.; Weindl, D.; Cordes, T.; Metallo, C. M.; Kelleher, J. K.; Stephanopoulos, G.,  
33      461       NTFD-a stand-alone application for the non-targeted detection of stable isotope-labeled compounds  
34      462       in GC/MS data. *Bioinformatics* **2013**, *29*, (9), 1226-1228.  
35      463       24. Zhu, H.; Chen, L.; Xing, W.; Ran, S.; Wei, Z.; Amee, M.; Wassie, M.; Niu, H.; Tang, D.; Sun, J.;  
36      464       Du, D.; Yao, J.; Hou, H.; Chen, K.; Sun, J., Phytohormones-induced senescence efficiently promotes  
37      465       the transport of cadmium from roots into shoots of plants: A novel strategy for strengthening of  
38      466       phytoremediation. *J Hazard Mater* **2020**, *388*, 122080.  
39      467       25. Ito, N.; Esaki, H.; Maesawa, T.; Imamiya, E.; Maegawa, T.; Sajiki, H., Efficient and selective Pt/C-  
40      468       catalyzed H-D exchange reaction of aromatic rings. *B Chem Soc Jpn* **2008**, *81*, (2), 278-286.  
41      469       26. Sawama, Y.; Nakano, A.; Matsuda, T.; Kawajiri, T.; Yamada, T.; Sajiki, H., H-D Exchange  
42      470       Deuteration of Arenes at Room Temperature. *Org Process Res Dev* **2019**, *23*, (4), 648-653.

- 1  
2  
3 471 27. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: Modular framework for  
4 472 processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *Bmc*  
5 473 *Bioinformatics* **2010**, *11*, 11.  
6  
7 474 28. Holman, J. D.; Tabb, D. L.; Mallick, P., Employing ProteoWizard to Convert Raw Mass  
8 475 Spectrometry Data. *Curr Protoc Bioinformatics* **2014**, *46*, 13.24.1-9.  
9  
10 476 29. Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J. M.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.;  
11 477 Wishart, D. S., CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound  
12 478 Identification. *Metabolites* **2019**, *9*, (4), 23.  
13  
14 479 30. Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J., Identifying  
15 480 small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci*  
16 481 *Technol* **2014**, *48*, (4), 2097-2098.  
17  
18 482 31. Yee, A. W.; Blakeley, M. P.; Moulin, M.; Haertlein, M.; Mitchell, E.; Forsyth, V. T., Back-exchange  
19 483 of deuterium in neutron crystallography: characterization by IR spectroscopy. *J Appl Crystallogr*  
20 484 **2017**, *50*, 660-664.  
21  
22 485 32. Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W., Minimizing Back Exchange in the  
23 486 Hydrogen Exchange-Mass Spectrometry Experiment. *J Am Soc Mass Spectr* **2012**, *23*, (12), 2132-  
24 487 2139.  
25  
26 488 33. Hoffmann, M. A.; Nothias, L.-F.; Ludwig, M.; Fleischauer, M.; Gentry, E. C.; Witting, M.;  
27 489 Dorrestein, P. C.; Dührkop, K.; Böcker, S., High-confidence structural annotation of metabolites  
28 490 absent from spectral libraries. *Nat Biotechnol* **2021**.  
29  
30 491 34. Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.;  
31 492 Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.;  
32 493 Rabinowitz, J. D., Metabolite discovery through global annotation of untargeted metabolomics data.  
33 494 *Nat Methods* **2021**, *18*, (11), 1377-1385.  
34  
35 495 35. Nothias, L. F.; Petras, D.; Schmid, R.; Duhrkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst,  
36 496 M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P. M.; Barsch, A.;  
37 497 Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.;  
38 498 Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kamenik, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf,  
39 499 A.; Le Gouellec, A.; Ludwig, M.; Martin, H. C.; McCall, L. I.; McSayles, J.; Meyer, S. W.;  
40 500 Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweiger, H.; Nguyen, N. H.; Nothias-  
41 501 Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi,  
42 502 A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H. J.; Zhang, Z.; Zubeil, F.;  
43 503 Kohlbacher, O.; Bocker, S.; Alexandrov, T.; Bandeira, N.; Wang, M. X.; Dorrestein, P. C., Feature-  
44 504 based molecular networking in the GNPS analysis environment. *Nat Methods* **2020**, *17*, (9), 905-+.  
45  
46 505 36. Huang, L. W.; Liu, W.; Zhao, L. L.; Zhang, Z. Y.; Yan, X. Y., Base-Catalyzed H/D Exchange  
47 506 Reaction of Difluoromethylarenes. *J Org Chem* **2021**, *86*, (5), 3981-3988.  
48  
49 507 37. Tian, Z. Y.; Zhao, H. Q.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X. M.; Prat, J.; Mudrock,  
50 508 E.; Hettinger, R.; Cortina, A. E.; Biswas, R. G.; Kock, F. V. C.; Soong, R.; Jenne, A.; Du, B. W.;  
51 509 Hou, F.; He, H.; Lundein, R.; Gilbreath, A.; Sutton, R.; Scholz, N. L.; Davis, J. W.; Dodd, M. C.;  
52 510 Simpson, A.; McIntyre, J. K.; Kolodziej, E. P., A ubiquitous tire rubber-derived chemical induces

- 1  
2  
3 511 acute mortality in coho salmon. *Science* **2021**, *371*, (6525), 185-189.  
4 512 38. Cheng, Z. P.; Sun, H. W.; Sidhu, H. S.; Sy, N. D.; Wang, X. R.; Gan, J., Conjugation of Di-n-butyl  
5 513 Phthalate Metabolites in *Arabidopsis thaliana* and Potential Deconjugation in Human Microsomes.  
6 514 *Environ Sci Technol* **2021**, *55*, (4), 2381-2391.  
7  
8 515 39. Tian, Z. Y.; Vila, J.; Yu, M. A.; Bodnar, W.; Aitken, M. D., Tracing the Biotransformation of  
9 516 Polycyclic Aromatic Hydrocarbons in Contaminated Soil Using Stable Isotope-Assisted  
10 517 Metabolomics. *Environmental Science & Technology Letters* **2018**, *5*, (2), 103-109.  
11  
12 518 40. Tripathi, A.; Vazquez-Baeza, Y.; Gauglitz, J. M.; Wang, M. X.; Duhrkop, K.; Nothias-Esposito, M.;  
13 519 Acharya, D. D.; Ernst, M.; van der Hooft, J. J. J.; Zhu, Q. Y.; McDonald, D.; Brejnrod, A. D.;  
14 520 Gonzalez, A.; Handelsman, J.; Fleischauer, M.; Ludwig, M.; Bocker, S.; Nothias, L. F.; Knight, R.;  
15 521 Dorrestein, P. C., Chemically informed analyses of metabolomics mass spectrometry data with  
16 522 Qemistree. *Nat Chem Biol* **2021**, *17*, (2), 146-+.  
17  
18 523 41. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray  
19 524 Ionization. *Anal Chem* **2013**, *85*, (1), 2-9.  
20  
21 525  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

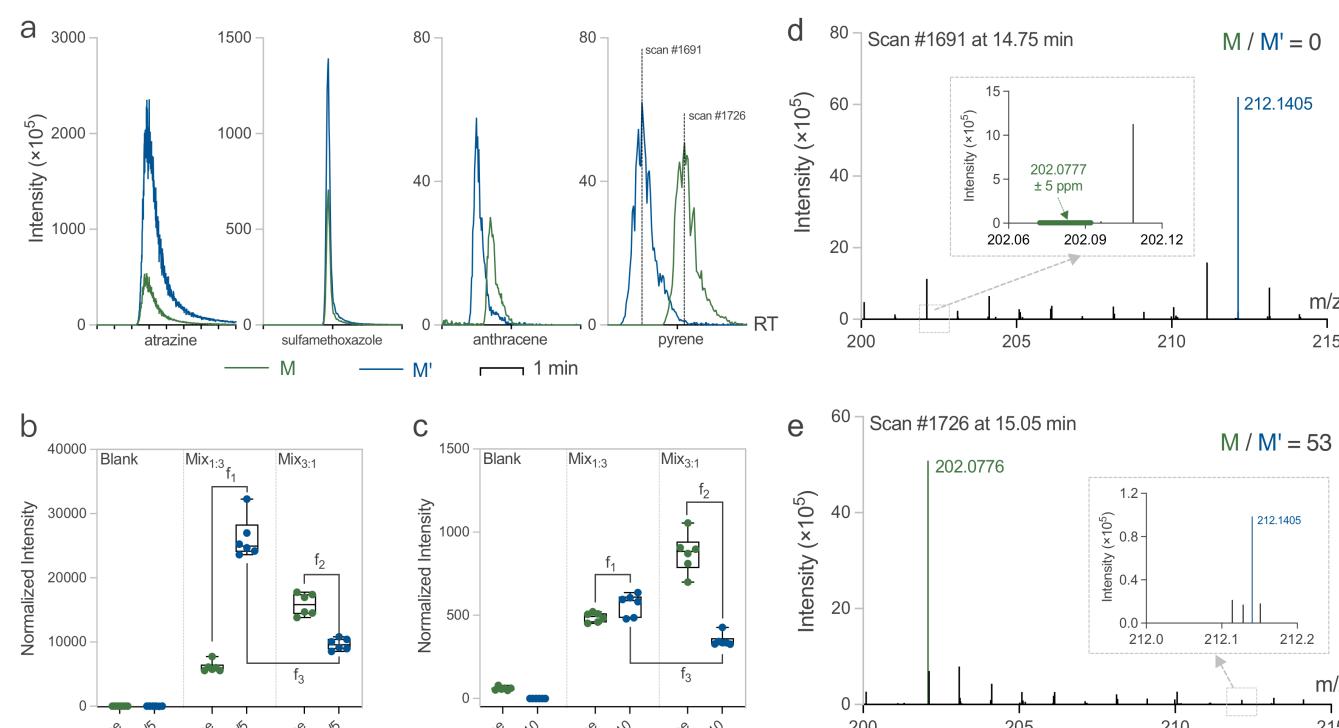
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

526



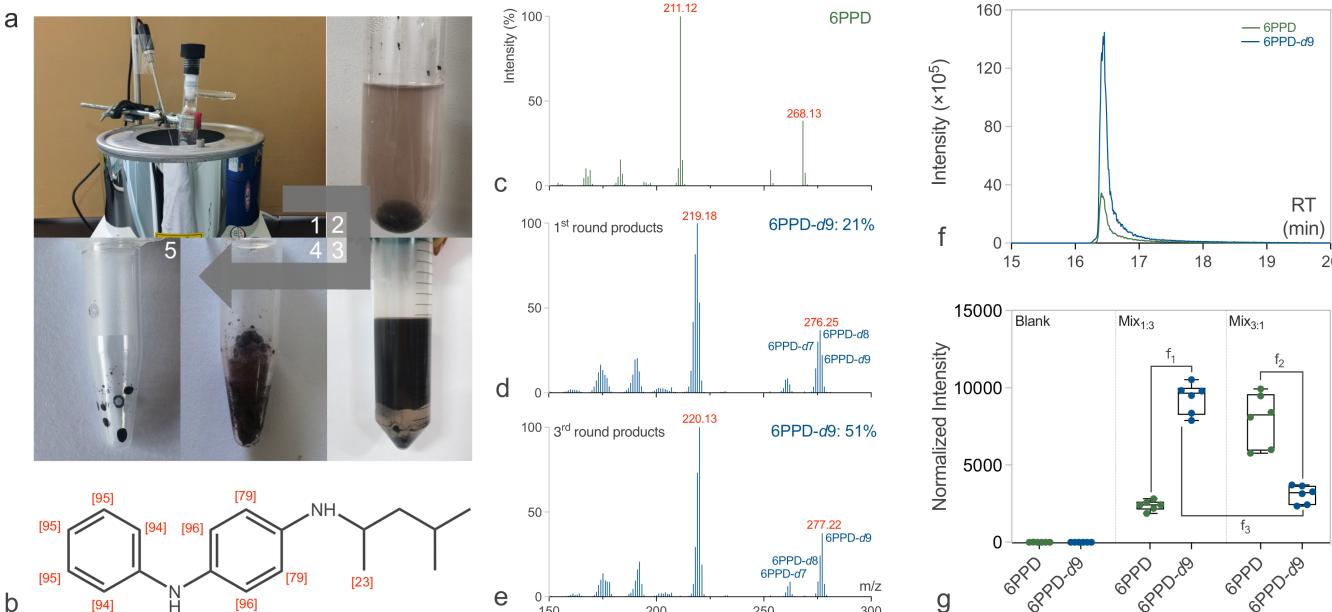
**Figure 1 Proposed pipeline for <sup>2</sup>H-labeled Stable Isotope Assisted Metabolomics (<sup>2</sup>H-SIAM).** Firstly, both natural and <sup>2</sup>H-labeled parent contaminants are individually incubated in the environment matrix for the indicated duration; they are subsequently extracted and mixed with the ratio of 1:3 and 3:1 (Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples); the mixtures are determined by LC-MS or GC-MS and thereafter obtained raw data are analyzed by MZmine2 to obtain feature list; then it is imported into the developed algorithm <sup>2</sup>H-SIAM(V1.0) to obtain possible features with isotopologues as TPs of the contaminants for further annotation and identification. Natural compounds are denoted as M, and the isotope-labeled compounds are denoted as M'.

537

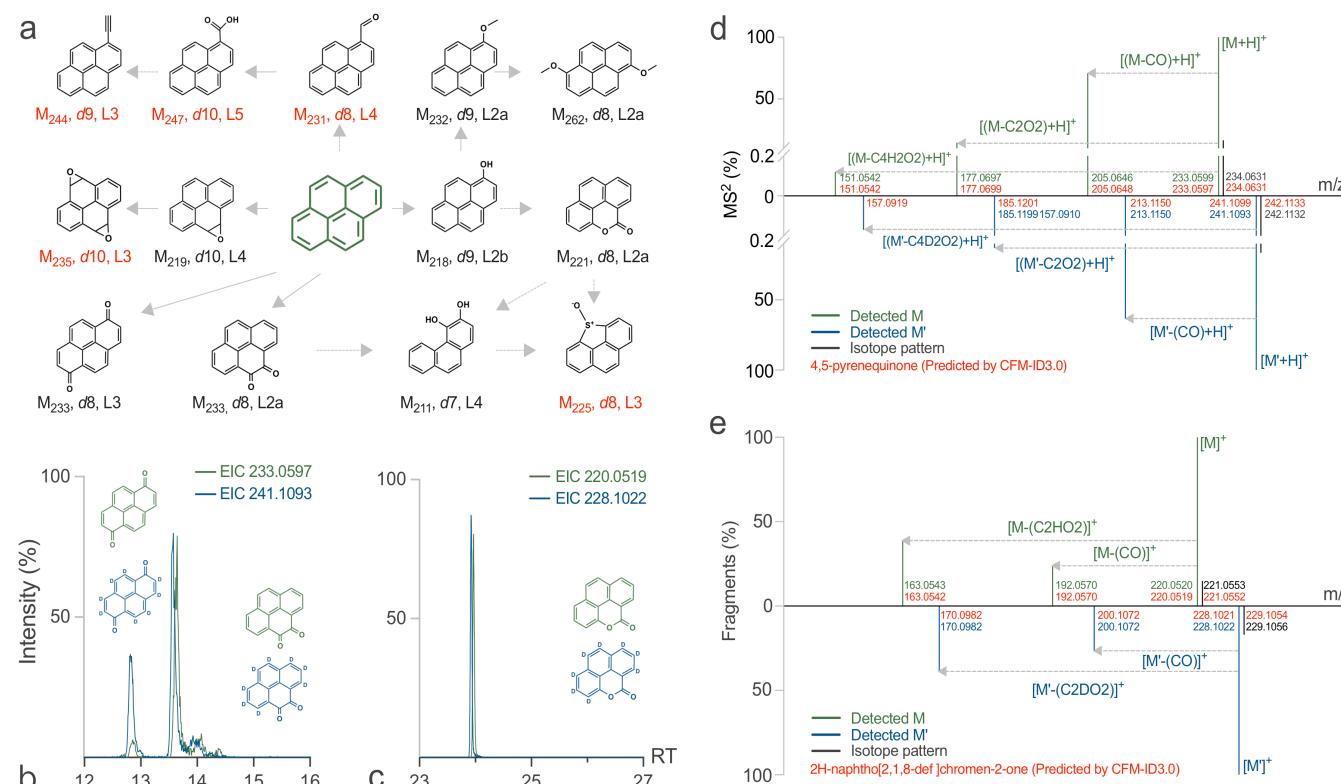


538

**Figure 2 Validation of  ${}^2\text{H}$ -SIAM pipeline by addition of 7 isotopologue pairs into soil extract.** 7 typical environmental contaminants with their  ${}^2\text{H}$ -labeled isotopologue were added into soil extract with the indicated concentration to obtain hypothetical Mix<sub>1:3</sub> and Mix<sub>3:1</sub> samples; then, they were analyzed by the UPLC-ESI-HRMS system, and obtained data were subjected to the proposed  ${}^2\text{H}$ -SIAM pipeline. (a) EIC of 4 typical isotopologue pairs from the Mix<sub>1:3</sub> sample which contain 10 ppm natural contaminants and 30 ppm  ${}^2\text{H}$ -labeled contaminants; (b, c) selected normalized height of isotopologue pairs of atrazine and pyrene for triple filters; (d, e) extracted mass spectra #1691 and #1726 at peaks of EIC of pyrene and pyrene-d10; natural compounds are denoted as M (dark green), and the isotope-labeled compounds are denoted as M' (dark blue).



546  
547 **Figure 3 Synthesis of 6PPD-d9 by direct H/D exchange reactions and its performance in  $^2\text{H}$ -SIAM.** The reaction was carried out in a  
548 glass tube with  $\text{D}_2\text{O}$  as deuterium source,  $\text{Pd/C}$  and  $\text{Pt/C}$  as catalysts, and  $\text{H}_2$  atmosphere at  $80^\circ\text{C}$  for 24 hours to obtain 1<sup>st</sup> round 6PPD-d9  
549 products (a, 1-4); the reaction was then repeated for twice, and a grease-like brown product was obtained (a, 5); (b) 600M NMR determined  
550 the structure of the 3<sup>rd</sup> round 6PPD-d9 products; GC-MS spectrum of 6PPD (c),  $^2\text{H}$ -labeled 6PPD from 1<sup>st</sup> round reaction (d), and  $^2\text{H}$ -  
551 labeled 6PPD from the 3<sup>rd</sup> round reaction (e); (f) EIC of 6PPD and 6PPD-d9 from UPLC-ESI-HRMS data; (g) normalized height of 6PPD  
552 and 6PPD-d9 for  $^2\text{H}$ -SIAM quantitative triple filters; natural compounds are marked as dark green, and the isotope-labeled compounds are  
553 marked as dark blue.



557 **Figure 4 Performance of <sup>2</sup>H-SIAM pipeline in a soil pyrene degradation study.** Pyrene and pyrene-*d*10 were added into the soil and  
558 incubated for 2 months; they were subsequently subjected to the proposed <sup>2</sup>H-SIAM pipeline study, which finally selected 52 features as  
559 isotopologue pairs, and 13 features were annotated as TPs of pyrene. (a) proposed TPs of pyrene in this study, *dn* indicates numbers of <sup>2</sup>H-  
560 labels, Ln indicates identification confidence levels (see Methods); 5 of them (red) were annotated as the TPs of pyrene for the first time;  
561 (b) selected EIC of M<sub>233</sub> and M'<sub>233-d8</sub> from UPLC-ESI-HRMS data; (c) selected EIC of M<sub>220</sub> and M'<sub>220-d8</sub> from GC-EI-HRMS data; (d)  
562 comparison of MS<sup>2</sup> spectrum (NCE 40, UPLC-ESI-Q-HRMS) between M<sub>233</sub>, M'<sub>233-d8</sub> at 13.6 min and CFM-ID 3.0 predicted spectrum of  
563 4,5-pyrenequinone; (e) comparison of fragments (70eV, GC-EI-HRMS) of M<sub>220</sub>, M'<sub>220-d8</sub> and CFM-ID 3.0 predicted spectrum of 2H-  
564 naphtho[2,1,8-def]chromen-2-one. Isotope pattern was extracted from MS<sup>1</sup>. Natural compounds are denoted as M (dark green), and the  
565 isotope-labeled compounds are denoted as M' (dark blue).

**Table 1 Performance of  $^2\text{H}$ -SIAM(1.0) in this study.**

Features	Experiments		
	7 Contaminants		Pyrene Degradation
	UPLC-ESI-HRMS	GC-EI-HRMS	UPLC-ESI-HRMS
Total	7704	21955	4376
Paired	1720	2548	658
Filter 1	891	1122	318
Filter 2	249	301	81
Filter 3	130	162	52
Annotated	7	5	14

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

1  
2  
3  
4  
5 569 For Table of Contents only  
6

