Plant Biotechnology 31, 269–274 (2014)
DOI: 10.5511/plantbiotechnology.14.0609c

Note

ShiftedIonsFinder: A standalone Java tool for finding peaks with specified mass differences by comparing mass spectra of isotope-labeled and unlabeled data sets

Kota Kera[1], Yoshiyuki Ogata[2], Takeshi Ara[1], Yoshiki Nagashima[1,a], Norimoto Shimada[1,b], Nozomu Sakurai[1], Daisuke Shibata[1], Hideyuki Suzuki[1,*]

[1] Department of Research and Development, Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan;
[2] Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Osaka 599-8531, Japan
* E-mail: hsuzuki@kazusa.or.jp Tel: +81-438-52-3947 Fax: +81-438-52-3948

Received April 7, 2014; accepted June 9, 2014 (Edited by K. Yano)

Abstract In mass spectrometry (MS)-based metabolite annotation, the prediction of theoretical elemental compositions from accurate mass is effective to support database search results or annotation for unknown metabolites. However, there are deviations (error) from the accurate mass such that accurate estimation of elemental composition may not be possible. A technique to label the biological sample with a stable isotope is effective to solve this problem, but the software for handling labeled MS data is limited. In this paper, we present a standalone Java graphical user interface (GUI) program tool—ShiftedIonsFinder—that helps find peaks having specified mass differences by comparing the mass spectra between two data sets. Using this tool, it is possible to select candidate labeled peaks by comparing a sample labeled with a stable isotope against an unlabeled sample. This tool also detects peaks with chemical modifications, such as glycosylation and acylation. Since the search results can be generated in Excel format, the user can easily perform further analyses or edit the list as needed. ShiftedIonsFinder (http://www.kazusa.or.jp/komics/software/ShiftedIonsFinder) is available free of charge at KOMICS.

Key words: Metabolite annotation, elemental composition, stable isotope, mass spectrometry.

Liquid chromatography-mass spectrometry (LC-MS) is a powerful tool in plant metabolome analysis because of its wide range of target metabolites (Nakabayashi and Saito 2013; Okazaki and Saito 2012; Zhou et al. 2012). In non-targeted and large-scale metabolome analysis, automatic estimation of metabolites by using public compound databases and exact mass databases from accurate mass values is effective. However, when the intensity of a detected peak is not within an appropriate range, a deviation (error) from accurate mass occurs, and it may not be possible to accurately estimate the elemental composition. Furthermore, even if the deviation is quite low (≦1 ppm), multiple elemental compositions are sometimes predicted for peaks with high mass-to-charge ratio (m/z) values. In order to solve these problems, Giavalisco et al. proposed a technique to label the biological sample *Arabidopsis thaliana* with stable isotopes (^{13}C, ^{15}N, and ^{34}S) *in vivo* and to estimate the elemental composition with high reliability using the number of stable isotopes incorporated in the compound as an index (Giavalisco et al. 2011). Moreover, an *in vivo*-labeled sample provides information about elemental composition as well as metabolic fluxes, which is important for metabolite annotation and exploitation in any study. The crucial aspect of this strategy is the software to find stable isotope-labeled ions by comparing mass spectra between two data sets labeled with a stable isotope and an unlabeled sample, obtained from same analytical conditions.

To date, a few MS analysis tools have been proposed to handle labeled MS data, such as MetExtract (Bueschl et al. 2012), MAVEN (Melamud et al. 2010), and mzMatch-ISO (Chokkathukalam et al. 2013). In order to estimate elemental composition, stable isotope-labeled ions found from each kind of labeled sample are preferably all shown together. However, these tools were designed for individual analysis and do not have the function to summarize the results of multiple stable isotope

Abbreviations: LC-MS, liquid chromatography mass spectrometry; m/z, mass-to-charge ratio; RT, retention time; GUI, graphical user interface.
[a] Present address: Thermo Fisher Scientific, 3-9 Moriya-cho, Kanagawa-ku, Yokohama, Kanagawa 221-0022, Japan.
[b] Present address: Tokiwa Phytochemical Co., Ltd., 158 Kinoko, Sakura, Chiba 285-0801, Japan.
This article can be found at http://www.jspcmb.jp/
Published online August 23, 2014

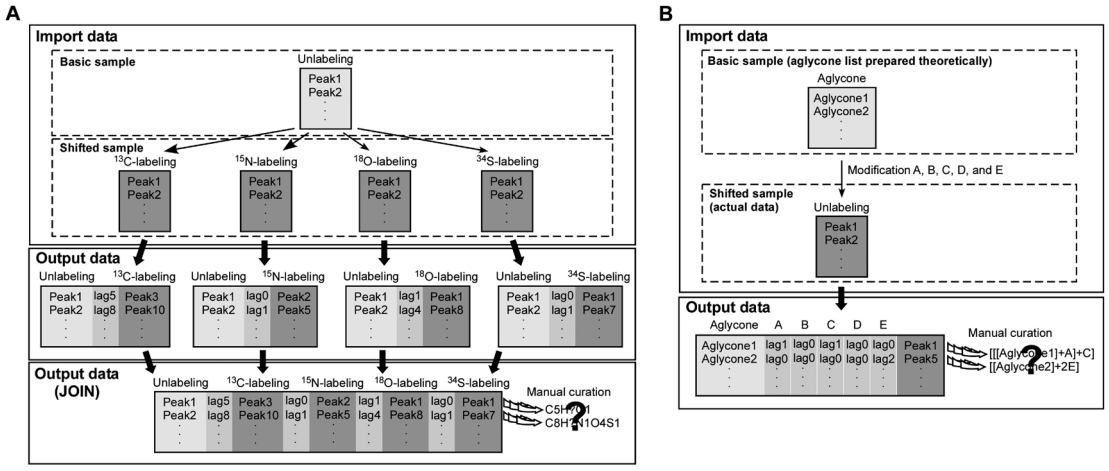


Figure 1. The scheme of data analysis by ShiftedIonsFinder. The overview of "Join" function in search for stable isotope labeled peaks is shown (A). The overview in search for peaks with chemical modifications is shown (B).

labeling analysis as described by Giavalisco et al. (2011). Moreover, these tools require the labeling efficiency of the sample or the elemental composition of interesting compounds to find stable isotope-labeled ions and reject false positives. These functions are good to analyze known metabolites or full-labeled samples, but are inconvenient in the analysis of unknown metabolites or partially labeled samples.

In this paper, we present a standalone Java graphical user interface (GUI) tool—ShiftedIonsFinder—that helps find peaks with specified mass differences attributable to stable isotopes or chemical modification by comparing the mass spectra between two data sets. As the search is performed using solely accurate mass values, it is also compatible with data generated using other (chromatography) mass spectrometers and is thus not limited to LC-MS. This tool is also designed for supporting manual annotation in metabolomics data analysis by experimental biologists without any programming techniques. For stable isotope labeling analysis, this tool can find not only fully labeled but also find partially labeled metabolites, and has a "Join" function that combines multiple search result lists into a single list (Figure 1A). Unlike other tools, ShiftedIonsFinder works as a standalone peak search tool. Therefore, it is possible to perform peak searches with this tool after peak picking, peak alignment, and database searches with other software. Since the search results can be displayed in Microsoft Excel format, the user can easily perform further analyses or edit the list as needed.

The import data for ShiftedIonsFinder is a set of "basic sample" and "shifted sample" data. The import data must be prepared as a tab-delimited text file corresponding to the output file of the PowerGet software (http://www.kazusa.or.jp/komics/software/PowerGet) (Iijima et al. 2008; Sakurai et al. 2014; Sano et al. 2012). The output files from other peak-picking software can be used in ShiftedIonsFinder if the format is converted to the PowerGet output format. This format is very simple and users can prepare the import data using PowerGet or Excel. The values "Ave. mass (detected)" and "Ave. R.T. (min)" (retention time) are used for the comparison of peaks. The average mass (detected) is the *m*/*z* value of the detected ions and the average retention time is the retention time for that case. In in-fusion analysis and the like, an arbitrary numeric value must be entered in the "Ave. R.T. (min)" column even when there is no retention time. The data that describes the specified mass difference is called "element data", which the user can edit as needed. The three parameters used in the search are the multiple (Max. Fold), the *m*/*z* value difference threshold (m/z Dif) and the retention time difference threshold (RT Dif). Figure 2 shows an overview of the search process flow used by ShiftedIonsFinder to search for peaks with the specified mass differences. The flow consists of the following four steps. First, the hypothetical *m*/*z* value ($Mass_{hyp}$) is calculated by adding the *m*/*z* value of the selected element data to those of the ions (M_{bn}) present in the basic sample. This $Mass_{hyp}$ is calculated for each value from zero to the value set by the multiple (F). Second, a search is performed to determine whether the calculated $Mass_{hyp}$ is present in the shifted sample (M_{sn}). Third, filters are applied using the set values for m/z Dif (m or m′) and RT Dif (t). Finally, the search results are displayed and exported in .xlsx (Excel) format. Peaks that match the search conditions are associated with the basic sample and the shifted sample, and are displayed in the search results table. The order of display in the table is as follows: 1) items of the basic sample, 2) associated information and 3) items of the shifted sample. Four calculated values—"Lag(times)", "RT Difference", "m/z Difference (u)", and "m/z Difference (ppm)"—are entered in the table as "Associated information". Lag(times) is the

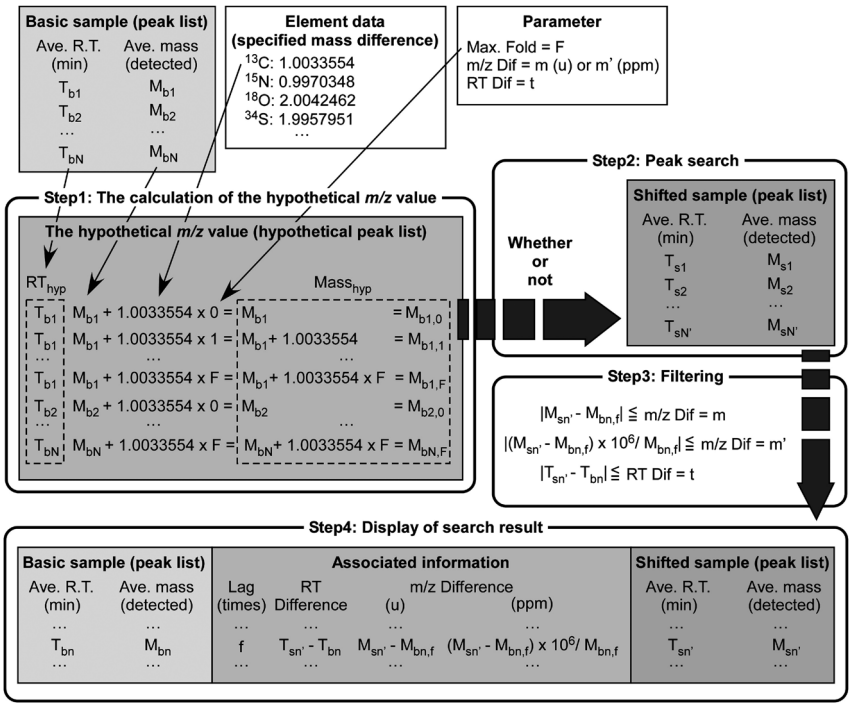


Figure 2. An overview of search process flow. The basic sample and shifted sample are applied as import data. The ShiftedIonsFinder finds candidate peaks with specified mass differences in four steps: 1) calculation of the hypothetical *m/z* value ($Mass_{hyp}$) based on basic sample, 2) peak search in shifted sample, 3) filtering using m/z value difference threshold (m/z Dif) and the retention time difference threshold (RT Dif), and 4) display of search result. Here, ^{13}C was selected as elemental data.

value of the multiple at the time of the search hit. The RT Difference and the m/z Difference represent, respectively, differences in elution time and the difference in the *m/z* value of the peak of the resulting basic sample and that of the shifted sample. Our tool does not require any other information, such as labeling efficiency.

We use ShiftedIonsFinder to find ^{13}C-, ^{15}N-, ^{18}O-, and ^{34}S-labeled metabolites from *Medicago truncatula* as a test case. After freeze-drying the samples, the metabolites are extracted with 80% methanol and analyzed by LC-Orbitrap-MS (Thermo Fisher Scientific Inc.). The details of the sample preparation methods and analytical conditions are available in the database Metabolonote (http://metabolonote.kazusa.or.jp/SE37:/) or in Supplemental Method, and the raw LC-MS chromatograms used in this study are available in the public metabolome database MassBase (http://webs2.kazusa.or.jp/massbase/). Their accession numbers are listed in Supplemental Table 1. The data were analyzed using PowerGet for ion feature detection. The resulting peak lists were exported as the basic sample from the unlabeled sample and as the shifted sample from the ^{13}C-labeled sample, ^{15}N-labeled sample, ^{18}O-labeled sample, and ^{34}S-labeled sample, respectively (Supplemental Data 1). After the labeled ions were searched with parameters (Max fold: C=100, N=50, O=50, S=50; Mass difference=3 ppm, RT difference=1), a summarized list from the four results is exported using the "Join" function (Supplemental Data 2). The search result for peak No. 1061 in the basic sample from the summarized list is shown in Table 1 as an example. Fourteen peaks in the ^{13}C-labeling shifted sample, four in the ^{15}N-labeling shifted sample, three in ^{18}O-labeling shifted sample and two peaks in the ^{34}S-labeling shifted sample were found. Here, the value of Lag(times) in associated information means the number of stable isotopes incorporated. For example, "C: Lag(times)=1" corresponding to the candidate ^{13}C-labeled peak [in shifted sample No. 14549] theoretically matched the metabolite that contains one ^{13}C atom in the molecule. Since ShiftedIonsFinder shows only the list of the candidate peaks to the user, the accuracy of this search result was checked using the full-mass spectrum (Supplemental Figure 1). In fact, three candidate ^{13}C-labeled peaks [in shifted samples No. 15065, No. 13133, and No. 15095] and one candidate ^{15}N-labeled peak [in shifted sample No. 13970] were false positives. However, the other peaks were accurate and covered all isotopic peaks that could be read from the full-mass spectrum. A small difference in ion mass between the results table and the mass spectrum was caused by the peak detection process in PowerGet. The elemental composition of peak No. 1061 was estimated (≦3 ppm) to be $C_{10}H_{16}N_2O_4S_1$ or $C_{18}H_{12}O_2$ using MFSearcher (Sakurai et al. 2013) (data not shown). Based on the number of stable isotopes incorporated, peak No. 1061 should contain at least 10 ^{13}C atoms, two ^{15}N atoms, two ^{18}O atoms, and one ^{34}S atom in the molecule.

Table 1. Search result of peak No. 1061 in the basic sample for stable isotope labeling.

From basic sample							^{13}C-Labeling							^{15}N-Labeling						
							Associated information			From shifted sample				Associated information			From shifted sample			
No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. mass (actual)	Ave. intensity of peak	Prefioniz. mode	Ave. MS/MS exists	C: lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. intensity of peak	Ave. MS/MS exists	N: lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. intensity of peak	Ave. MS/MS exists
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	0	−0.02	0.0001	0.5	14560	14.61	261.0908	1028		0	−0.21	−0.0001	−0.2	13574	14.42	261.0906	8057	
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	1	−0.03	0.0001	0.4	14549	14.60	262.0941	1519		1	−0.21	0.0000	−0.2	13573	14.42	262.0877	18253	
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	2	−0.02	0.0001	0.5	14558	14.61	263.0975	4072		2	−0.22	0.0000	0.0	13571	14.42	263.0847	20542	
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	3	−0.04	−0.0001	−0.4	14535	14.59	264.1006	11190	yes	2	0.10	−0.0002	−0.8	13970	14.73	263.0845	841	
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	4	−0.04	0.0001	0.2	14539	14.60	265.1041	3358										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	5	−0.03	0.0001	0.5	14545	14.60	266.1076	1834										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	6	−0.03	0.0001	0.2	14552	14.60	267.1109	2574										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	7	−0.04	−0.0001	−0.3	14528	14.59	268.1141	4589										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	8	−0.04	0.0001	0.4	14532	14.59	269.1176	6057										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	9	−0.04	0.0002	0.6	14533	14.59	270.1210	6324										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	10	−0.03	0.0001	0.3	14542	14.60	271.1243	3011										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	27	0.41	−0.0002	−0.8	15065	15.04	288.1810	1923										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	53	−0.92	0.0009	3.0	13133	13.72	314.2694	2355										
1061	14.63	261.0907	260.0834	44562	[Actual+H]$^+$	yes	53	0.44	0.0009	2.9	15095	15.07	314.2694	362										

^{18}O-Labeling									^{34}S-Labeling								
	Associated information				From shifted sample				Associated information				From shifted sample				
	O: lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. intensity of peak	Ave. MS/MS exists	S: lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. intensity of peak	Ave. MS/MS exists
	0	−0.50	0.0000	0.0	13339	14.13	261.0907	15169		0	−0.60	−0.0001	−0.2	13241	14.03	261.0906	14525	
	1	−0.50	−0.0002	−0.8	13350	14.13	263.0947	18493		1	−0.60	−0.0001	−0.2	13246	14.03	263.0864	30662	yes
	2	−0.50	0.0000	−0.1	13347	14.13	265.0991	8222										

Therefore, the elemental composition of peak No. 1061 was estimated to be $C_{10}H_{16}N_2O_4S_1$ by manual annotation process, not an automatic process by ShiftedIonsFinder. These results indicate that ShiftedIonsFinder can find labeled peaks by comparing two data sets and provide beneficial information to estimate elemental composition.

ShiftedIonsFinder can also be used to find the shift in the derivatization of metabolites (Figure 1B). Secondary metabolites, such as glycosylation (Gachon et al. 2005) and acylation (D'Auria 2006), are accumulated with modifications in more complex plants. Further, the variety of modification patterns among species makes it difficult to identify metabolites in biological samples. As an example, flavonoid derivatives caused by glycosylation and acylation in the unlabeled sample were searched using ShiftedIonsFinder. The peak list having typical flavonoid information was prepared theoretically as the basic sample (Supplemental Data 1). We entered 1 as the arbitrary numeric value of Ave. R.T. (min), whereas a theoretical *m*/*z* value taking into account the ionization mode was entered for Ave. mass (detected). The peak list of the unlabeled sample shown above was used as the shifted sample. Glucosylation Glc ($C_6H_{10}O_5$), *m*/*z* 162.05282, rhamnosylation Rha ($C_6H_{10}O_4$) *m*/*z* 146.05791, glucronidation GlcUA ($C_6H_8O_6$) *m*/*z* 176.03209, malonylation Malonyl ($C_3H_2O_3$), *m*/*z* 89.0003939, and succinylation, Succinyl ($C_4H_4O_3$) with a *m*/*z* value of 100.016044 were applied to the element data as possible modification groups. Thus, potential peaks were searched with parameters (Max fold: Glc=3, Rha=3, GlcUA=3, Malonyl=3, Succinyl=3; Mass difference=3 ppm, RT difference=60). If the user is interested in the apigenin derivative, the user can choose "Apigenin ($C_{15}H_{10}O_5$)" (Table 2) using the filter function in Excel from the search result (output file, Supplemental Data 3). As a result, nine potential peaks were found regardless of the presence of tandem mass spectrometry (MS^2) information. Here, the value of Lag(times) in associated information means the number of each possible modification group. For example, "GlcUA ($C_6H_{10}O_5$): Lag(times)=2" corresponding to peak No. 2127 theoretically matched [[apigenin]+2 GlcUA] (Table 2, Supplemental Figure 2). As is commonly misunderstood, these nine peaks are not always apigenin derivatives but might be something non-flavonoid. For example, peak No. 2127 in the shifted sample could represent at least 10 kinds of flavonoid derivatives in the basic sample (Table 3). In order to narrow down the candidate peaks, other information, such as MS^n, ultra-violet (UV) absorbance or nuclear magnetic resonance (NMR) data, is necessary.

In conclusion, ShiftedIonsFinder automatically finds peaks having the specified mass differences by comparing the mass spectra of two data sets and making a list. Further, it is easy to apply to the user's environment. For

Table 2. Search result for possible apigenin derivatives.

From basic sample							Associated information								From shifted sample						
No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. mass (actual)	Name	Class	ioniz.mode	Glc ($C_6H_{10}O_5$): lag (times)	Rha ($C_6H_{10}O_4$): lag (times)	GlcUA ($C_6H_8O_6$): lag (times)	Malonyl ($C_3H_2O_3$): lag (times)	Succinyl ($C_4H_4O_3$): lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. mass (actual)	Ave. intensity of peak	pref.ioniz. mode	Ave. MS/MS exists
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	2	0	0	0	0	18.61	0.0010	1.7	1639	19.61	595.1667	594.1595	46999	$[Actual+H]^+$	yes
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	1	0	1	0	20.51	0.0010	2.1	1889	21.51	503.1194	502.1122	13424	$[Actual+H]^+$	
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	0	2	0	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	1	0	1	0	22.92	0.0009	1.8	2316	23.92	503.1193	502.1120	60051	$[Actual+H]^+$	yes
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	0	1	0	0	24.03	0.0007	1.5	2498	25.03	447.0929	446.0856	63687	$[Actual+H]^+$	
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	0	1	0	0	24.34	0.0007	1.5	2531	25.34	447.0929	446.0856	41603	$[Actual+H]^+$	
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	1	0	0	1	27.03	0.0011	2.1	2794	28.03	517.1351	516.1279	48147	$[Actual+H]^+$	yes
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	1	0	0	0	1	28.00	0.0011	2.1	2942	29.00	533.1301	532.1228	16517	$[Actual+H]^+$	
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	0	0	0	0	31.41	0.0005	1.7	3201	32.41	271.0606	270.0533	70514	$[Actual+H]^+$	yes

Table 3. Search result of peak No. 2127 in shifted sample for flavonoid derivatives.

From basic sample							Associated information								From shifted sample						
No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. mass (actual)	Name	Class	ioniz.mode	Glc ($C_6H_{10}O_5$): lag (times)	Rha ($C_6H_{10}O_4$): lag (times)	GlcUA ($C_6H_8O_6$): lag (times)	Malonyl ($C_3H_2O_3$): lag (times)	Succinyl ($C_4H_4O_3$): lag (times)	RT difference	m/z difference (u)	m/z difference (ppm)	No.	Ave. R.T. (min)	Ave. mass (detected)	Ave. mass (actual)	Ave. intensity of peak	pref.ioniz. mode	Ave. MS/MS exists
1	1.00	289.0707	288.0634	Pentahydroxychalcone ($C_{15}H_{12}O_6$)	Chalcone	$[M+H]^+$	1	0	0	2	0	21.90	0.0011	1.8	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
2	1.00	275.0914	274.0841	Phloretin ($C_{15}H_{14}O_5$)	Dihydrochalcone	$[M+H]^+$	0	0	1	2	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
4	1.00	289.0707	288.0634	Eriodictyol ($C_{15}H_{12}O_6$)	Flavanone	$[M+H]^+$	1	0	0	2	0	21.90	0.0011	1.8	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
6	1.00	271.0601	270.0528	Sulphuretin ($C_{15}H_{10}O_5$)	Aurone	$[M+H]^+$	0	0	2	0	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
9	1.00	271.0601	270.0528	Apigenin ($C_{15}H_{10}O_5$)	Flavone	$[M+H]^+$	0	0	2	0	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
15	1.00	271.0601	270.0528	Genistein ($C_{15}H_{10}O_5$)	Isoflavone	$[M+H]^+$	0	0	2	0	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
18	1.00	289.0707	288.0634	Dihydrokaempferol ($C_{15}H_{12}O_6$)	Flavanonol	$[M+H]^+$	1	0	0	2	0	21.90	0.0011	1.8	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
19	1.00	305.0656	304.0583	Dihydroquercetin ($C_{15}H_{12}O_7$)	Flavanonol	$[M+H]^+$	0	1	0	2	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
26	1.00	323.0761	322.0689	Leucodelphinidin ($C_{15}H_{14}O_8$)	Leucoanthocyanidin	$[M+H]^+$	0	0	0	0	3	21.90	0.0011	1.8	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes
29	1.00	271.0601	271.0606	Pelargonidin ($C_{15}H_{11}O_5$)	Anthocyanidin	$[M]^+$	0	0	2	0	0	21.90	0.0011	1.7	2127	22.90	623.1254	622.1181	253795	$[Actual+H]^+$	yes

stable isotope labeling analysis, this tool can find not only fully labeled but also partially labeled metabolites and supplies a "Join" function, which is useful for the estimation of elemental composition. This tool shows the user the list of the candidate peaks and increases the efficiency of the narrowing-down process by screening first, especially in large-scale datasets.

Acknowledgements

This work was supported by a grant from the Strategic International Research Cooperative Program of the Japanese Science and Technology Agency (Metabolomics for a Low Carbon Society, JST-NSF). We would like to thank Dr. Lloyd W. Sumner (The Samuel Roberts Noble Foundation) for providing seeds of *M. truncatula* Jemalong A17. We would like to thank Dr. Toru Nakayama (Tohoku University) for advising on the flavonoid list. We would also like to thank Mr. Hiroshi Chiba (Tohoku Chemical Co., Ltd.) and Ms. Noriko Tamayama (Tohoku Chemical Co., Ltd.) for helping write the program.

References

Bueschl C, Kluger B, Berthiller F, Lirk G, Winkler S, Krska R, Schuhmacher R (2012) MetExtract: A new software tool for the automated comprehensive extraction of metabolite-derived LC/MS signals in metabolomics research. *Bioinformatics* 28: 736–738

Chokkathukalam A, Jankevics A, Creek DJ, Achcar F, Barrett MP, Breitling R (2013) mzMatch-ISO: An R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics* 29: 281–283

D'Auria JC (2006) Acyltransferases in plants: A good time to be BAHD. *Curr Opin Plant Biol* 9: 331–340

Gachon CMM, Langlois-Meurinne M, Saindrenan P (2005) Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends Plant Sci* 10: 542–549

Giavalisco P, Li Y, Matthes A, Eckhardt A, Hubberten HM, Hesse H, Segu S, Hummel J, Kohl K, Willmitzer L (2011) Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J* 68: 364–376

Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K, Suzuki T, Suzuki H, Okazaki K, Kitayama M, et al. (2008) Metabolite annotations based on the integration of mass spectral information. *Plant J* 54: 949–962

Melamud E, Vastag L, Rabinowitz JD (2010) Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* 82: 9818–9826

Nakabayashi R, Saito K (2013) Metabolomics for unknown plant metabolites. *Anal Bioanal Chem* 405: 5005–5011

Okazaki Y, Saito K (2012) Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol Rep* 6: 1–15

Sakurai N, Ara T, Enomoto M, Motegi T, Morishita Y, Kurabayashi A, Iijima Y, Ogata Y, Nakajima D, Suzuki H, et al. (2014) Tools and databases of the KOMICS web portal for preprocessing, mining, and dissemination of metabolomics data. *Biomed Res Int* 2014: 194812

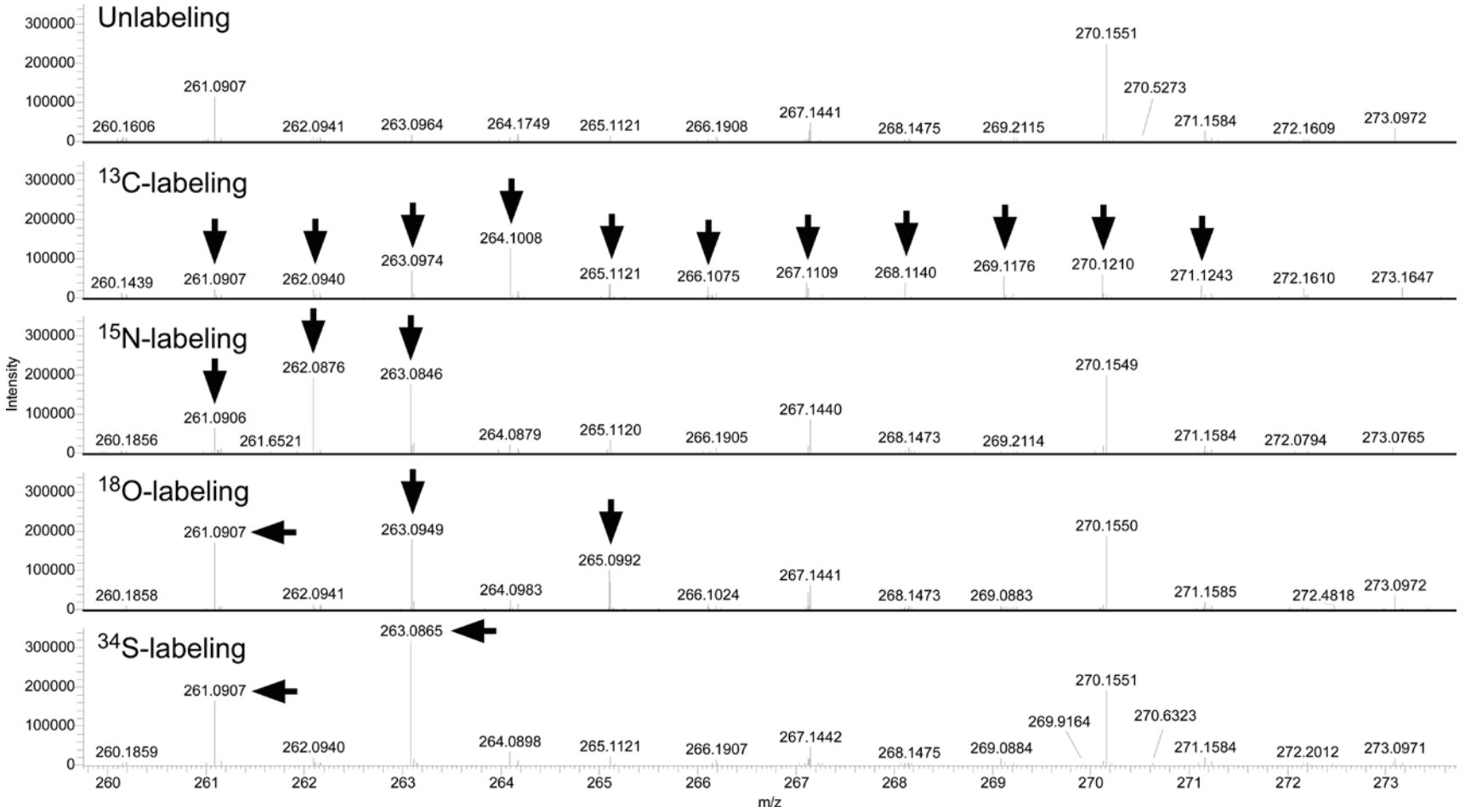
Sakurai N, Ara T, Kanaya S, Nakamura Y, Iijima Y, Enomoto M, Motegi T, Aoki K, Suzuki H, Shibata D (2013) An application of a relational database system for high-throughput prediction of elemental compositions from accurate mass values. *Bioinformatics* 29: 290–291

Sano R, Ara T, Akimoto N, Sakurai N, Suzuki H, Fukuzawa Y, Kawamitsu Y, Ueno M, Shibata D (2012) Dynamic metabolic changes during fruit maturation in Jatropha curcas L. *Plant Biotechnol* 29: 175–178

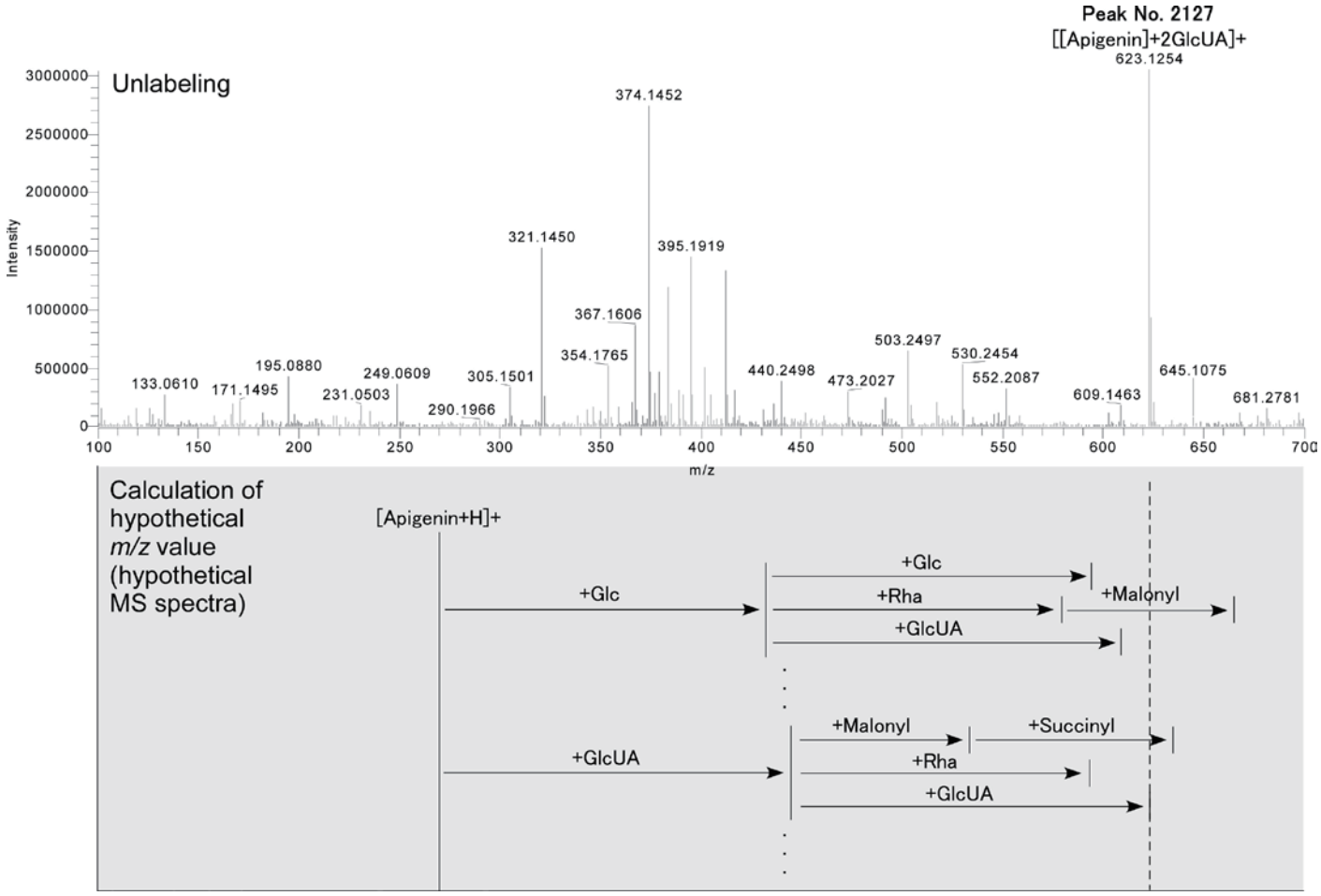
Zhou B, Xiao JF, Tuli L, Ressom HW (2012) LC-MS-based metabolomics. *Mol Biosyst* 8: 470–481

Supplemental Table 1. The accession numbers of Metabolonote and MassBase

Raw file name	MassBase_ID	Import file name	Metabolonote_ID
Blank_1	MDLC1_43484	Shoot_unlabeling	SE37_S17_M1_D1
Blank_2	MDLC1_43485		
Blank_3	MDLC1_43486		
shoot_control_1	MDLC1_43514		
shoot_control_2	MDLC1_43515		
shoot_control_3	MDLC1_43516		
shoot_13C_2	MDLC1_43503	Shoot_13C_labeling	SE37_S05_M1_D1
shoot_15N_1	MDLC1_43505	Shoot_15N_labeling	SE37_S07_M1_D1
shoot_18O_1	MDLC1_43508	Shoot_18O_labeling	SE37_S10_M1_D1
shoot_34S_1	MDLC1_43511	Shoot_34S_labeling	SE37_S13_M1_D1



Supplemental Figure 1. The mass spectra corresponding to peak No. 1061 in basic sample for stable isotope labeling. Raw data (shoot_control_2, shoot_13C_2, shoot_15N_1, shoot_18O_1, and shoot_34S_1) were used. The black allow indicates the mono-isotopic and isotopic peaks in each labeled sample.



Supplemental Figure 2. The mass spectra corresponding to peak No. 2127 in shifted sample for possible apigenin derivatives. Here, modification groups were hypothetically added to aglycone, apigenin, like a building block and the resulting hypothetical *m/z* values (MS spectra) were compared with that of peaks in Shifted sample. Raw data (shoot_control_2) was used.