



## RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database

Yuji Sawada<sup>a</sup>, Ryo Nakabayashi<sup>a</sup>, Yutaka Yamada<sup>a</sup>, Makoto Suzuki<sup>a</sup>, Muneo Sato<sup>a</sup>, Akane Sakata<sup>a</sup>,  
Kenji Akiyama<sup>a</sup>, Tetsuya Sakurai<sup>a</sup>, Fumio Matsuda<sup>a,b</sup>, Toshio Aoki<sup>c</sup>, Masami Yokota Hirai<sup>a,d</sup>,  
Kazuki Saito<sup>a,e,\*</sup>

<sup>a</sup>RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

<sup>b</sup>Organization of Advanced Sciences and Technology, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe 657-8501, Japan

<sup>c</sup>Department of Applied Biological Sciences, Nihon University, Fujisawa, Kanagawa 252-8510, Japan

<sup>d</sup>JST, CREST, 4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

<sup>e</sup>Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba 260-8675, Japan

### ARTICLE INFO

#### Article history:

Received 6 November 2011

Received in revised form 22 March 2012

Available online 4 August 2012

#### Keywords:

MS/MS

Metabolomics

Annotation

Database

Similarity search

MS/MS fragmentation association rule

### ABSTRACT

The fragment pattern analysis of tandem mass spectrometry (MS/MS) has long been used for the structural characterization of metabolites. The construction of a plant-specific MS/MS data resource and database will enable complex phytochemical structures to be narrowed down to candidate structures. Therefore, a web-based database of MS/MS data pertaining to phytochemicals was developed and named ReSpect (RIKEN tandem mass spectral database). Of the 3595 metabolites in ReSpect, 76% were derived from 163 literature reports, whereas the rest was obtained from authentic standards. As a main web application of ReSpect, a fragment search was established based on only the  $m/z$  values of query data and records. The confidence levels of the annotations were managed using the MS/MS fragmentation association rule, which is an algorithm for discovering common fragmentations in MS/MS data. Using this data resource and database, a case study was conducted for the annotation of untargeted MS/MS data that were selected after quantitative trait locus analysis of the accessions (Gifu and Miyakojima) of a model legume *Lotus japonicus*. In the case study, unknown metabolites were successfully narrowed down to putative structures in the website.

© 2012 Elsevier Ltd. All rights reserved.

### 1. Introduction

Metabolomics is an “omics” approach to the investigation of metabolites present in biological samples (Ferne et al., 2004; Weckwerth, 2003). Untargeted metabolomics methodologies that employ chromatographic separation and mass spectrometry (MS) are widely used for high throughput and high sensitivity/selectivity analysis of a broad range of phytochemicals (Fiehn et al., 2000; Guy et al., 2008; Hall et al., 2008; Saito and Matsuda, 2010; Schauer and Fernie, 2006; Sumner et al., 2003). These MS-based metabolomic approaches require reference data for authentic standard compounds to identify the detected metabolite candidates (Last et al., 2007). Phytochemicals are thought to be the lead chemicals involved in innovative bioactivities, and it is estimated that there are 0.2–1 million metabolites; however, the authentic compounds

available for qualitative or quantitative analysis are very limited. Thus, full identification of novel metabolites by isolation is the most important methodology in plant metabolomics. In this regard, a total of 37 metabolites have been successfully isolated and identified from spectroscopic data of the model plant *Arabidopsis*, but only two anthocyanins were elucidated as novel metabolites (Nakabayashi et al., 2009). An advanced tandem mass spectrometry (MS/MS) search will improve the ability to narrow down candidate complex phytochemical structures, thereby facilitating opportunities for unambiguous identification of the unknown metabolites.

MS/MS analysis facilitates the detection of fragment ions of a molecular weight-related ion (precursor ion), while using metabolite-specific fragment ions (product ions) for structural elucidation. Thus, the construction of an MS/MS data resource and database are required for metabolite identification in metabolomics. MS/MS data have recently been made available via web databases such as METLIN (Smith et al., 2005), BinBase (Fiehn et al., 2005, 2008), HMDB (Wishart et al., 2009, 2007), MMCD (Cui et al., 2008), and MassBank (Horai et al., 2010). Additionally, the XC-MS2 database

\* Corresponding author. Address: Metabolome Research Group, RIKEN Plant Science Center Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. Tel.: +81 45 503 9488; fax: +81 45 503 9489.

E-mail address: [ksaito@psc.riken.jp](mailto:ksaito@psc.riken.jp) (K. Saito).

has been used in metabolomics in conjunction with a similarity search algorithm from the MS/MS database to identify detected metabolites (Benton et al., 2008). These MS/MS databases have significantly enhanced the identification of detected metabolites (Hildebrandt et al., 2011).

Metabolomics data and MS/MS similarity search algorithms are useful for both identifying and assigning similar structures to detectable metabolites, a process termed annotation. Since April 2006, data was collected manually from the available literature for use as plant-specific reference MS/MS data (<http://spectra.psc.riken.jp/>). MS/MS data pertaining to approximately 500 metabolites can be collected from the literature each year using this method; therefore, thousands of phytochemicals have been successfully collected to date. Because the number of metabolites detected by MS-based metabolomics has recently been expanding, the MS/MS data derived from the literature are very valuable for help in the identification of detected phytochemicals.

In this work described herein, the ReSpect (RIKEN tandem mass spectral database) is a new web database and data resource composed of records from existing literature and the MS/MS data from our standard compounds. As a case study, the metabolite quantitative locus analysis of a model legume *Lotus japonicus* (Sato et al., 2008) was carried out by annotating untargeted MS/MS data using ReSpect as a first step towards metabolite identification. Various unknown metabolites were narrowed down to putative structures in the database, which is being provided as a Scientific Community Service.

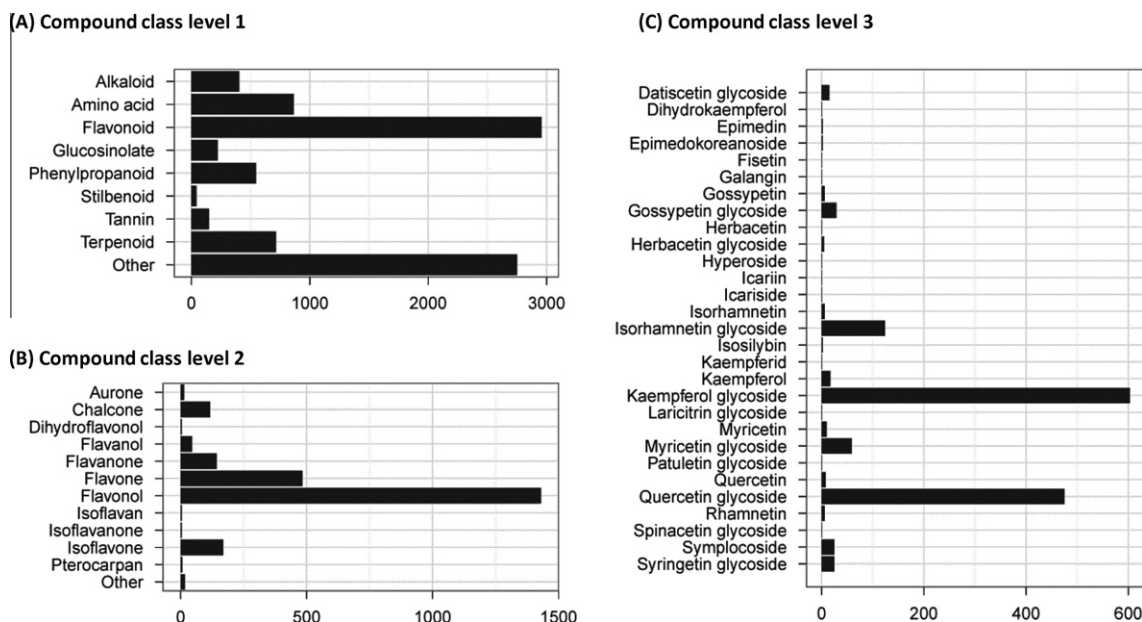
## 2. Results and discussion

### 2.1. ReSpect database

The ReSpect database has three major features that differ from other MS/MS databases; these features enhance the annotation efficiency of complex phytochemical structures. The first is a newly established fragment search system. ReSpect contains 3341 records derived from 163 literature reports (see Section 4). The available MS/MS literature is thus a valuable data resource for

phytochemicals, although only about half of the articles describe the  $m/z$  values of fragment ions, with no description of the intensities or MS spectra. In those cases, the relative intensities of all ions to 100% were adjusted. The adjusted intensity data might be overrated by the conventional spectrum search algorithm, which uses the cosine product method (Sokolow et al., 1978) in conjunction with  $m/z$  values and the intensity values of fragment ions for the MS/MS similarity search. Thus, a new search system (named fragment search) was established, which uses  $m/z$  values of fragment ions to search adjusted MS/MS data. The specific fragment pattern of MS/MS derived from a compound permitted narrowing down as candidate structures, which is quite important for the first steps towards unambiguous identification of unknown phytochemicals. Moreover, all records in the ReSpect database were manually categorized into a compound ontology system composed of three levels of compound classes that was essential for the interpretation of the fragment search; the records in the ReSpect database were categorized into the compound classes according to the book “Comprehensive Natural Products Chemistry,” Volumes 1, 2, 3, and 4 (Cane, 1999; Kelly, 1999; Pinto, 1999; Sankawa, 1999). In this study, the three levels were defined as follows: 1, general term; 2, main skeleton; 3, synonym. For the case study, major records in ReSpect were categorized as follows: level 1, flavonoids (2957 records); level 2, flavonols (1430 records); level 3, kaempferol glycosides (603 records) (Fig. 1). Multiple synonymous names have generally been used for phytochemical structures. As a result, evaluation of matched records in the MS/MS database has been a very difficult task for biologists. Using compound class level 1 (nine basic types of phytochemicals), the results of the fragment search can be easily understood by a wide range of users, while more detailed structural information is provided by compound class levels 2 and 3.

The second feature of ReSpect that differentiates it from other databases is the enumerated MS/MS fragmentation association rules among the ReSpect records. The association rules are an algorithm of frequent pattern mining for identification of interesting relationships between variables in the database (Agrawal et al., 1993). For example, the following association rule “milk  $\Rightarrow$  bread, eggs” in a selected dataset can be predicted as follows: if a cus-



**Fig. 1.** Distributions of major records with compound class levels. All ReSpect records were categorized in compound class level 1 (A). The records of “Flavonoids” in compound class 1 were categorized in compound class level 2 (B). The records of “Flavonols” in compound class level 2 were categorized in compound class level 3 (C).

**Table 1**

MS/MS fragmentation association rules in flavonoids (compound class level 1).

| Antecedent ( <i>m/z</i> ) | Consequent <sup>a</sup> ( <i>m/z</i> ) | Support <sup>b</sup> (%) | Confidence <sup>c</sup> (%) | Lift | Compound class (level 2/level 3)                                   |
|---------------------------|--|--------------------------|-----------------------------|------|--|
| 519                       | ⇒ 271                                  | 1.4                      | 73                          | 9.8  | Anthocyanin/pelargonidin glycoside, isoflavone/genistein glycoside |
| 465                       | ⇒ 303                                  | 2.5                      | 71                          | 4.7  | Anthocyanin/delphinidin glycoside, flavanol/quercetin glycoside    |
| 431                       | ⇒ 329                                  | 1.1                      | 77                          | 57.5 | Flavone/luteolin glycoside   |
| 397                       | ⇒ 313                                  | 1.2                      | 80                          | 44.4 | Flavone/apigenin glycoside   |

Transaction (a total of 1332 records of flavonoids in ReSpect) were selected by polarity (positive mode) and adduct type ( $[M+H]^+$ ), and MS/MS fragmentation association rules were selected with index values (Support > 1%, Confidence > 70%, Lift > 1).

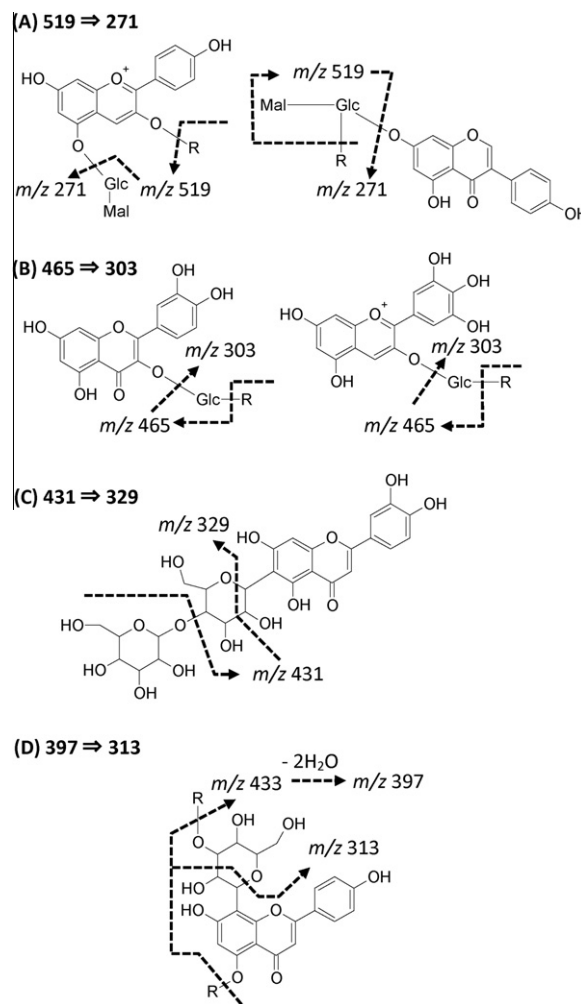
<sup>a</sup> Support value (0–100%) indicates frequency of antecedent in a dataset of transaction (See Supplementary Table 1).

<sup>b</sup> Confidence (0–100%) indicates frequency of a rule in all transactions including selected antecedent.

<sup>c</sup> Lift value (0–∞) indicates independence of frequency between the antecedent and the rule as follows: positive effect (<1), negative effect (>1) and no effect (≈1).

tomers bought milk, then they also bought bread and eggs (see Supplementary Fig. 1). Thus, the left- and right-hand side in a rule are named the antecedent (if) and consequent (then), and the datasets are named transactions, e.g., a transaction is a set of items in a basket. This approach was employed to estimate common MS/MS fragment patterns among the chemicals in compound class level 1 (described above). Among the association rules of the MS/MS data, the max *m/z* value of a product ion in a rule is defined as the antecedent, the other product ions are defined as the consequent, and invalid rules are excluded from analysis (see Section 4). As the transaction for the MS/MS fragmentation association rule, 30 datasets from the ReSpect records having the same polarity and adduct type are grouped as compounds in class level 1 (Supplementary Table 1). The index values of the association rules are calculated based on the number of transactions that include a specific combination of the antecedent and consequent, and the values (support, confidence and lift) are then used for evaluation of the frequency, strength, and independence (see Supplementary Fig. 1 and Section 4). Using the MS/MS fragmentation association rules with index values, the common MS/MS fragment patterns in flavonoids were successfully predicted in the ReSpect records (Table 1), and each flavonoid appears to have a common fragmentation pattern as a result of the positions of homolytic bond cleavage (Fig. 2). As shown in Fig. 2A and B, a common fragmentation scheme of glycosidic derivatives was estimated for different flavonoid skeletons. Moreover, very specific fragment ions derived from C-glycosidic flavonoids were estimated based on complex spectral interpretations (Fig. 2C and D) (Cavaliere et al., 2005; Rauter et al., 2005). In these cases, the results of a fragment search will be shown by the number of matches with a common fragment pattern. However, it is difficult to use such annotations to evaluate confidence levels, e.g., whether or not there is specific matching, based on only the number of matched records. Therefore, the MS/MS fragmentation association rules are also shown as the optional results of a fragment search of ReSpect, and the common fragment patterns among the compound classes of phytochemicals can easily be concurrently evaluated on the basis of the matched records with specific or common fragment patterns.

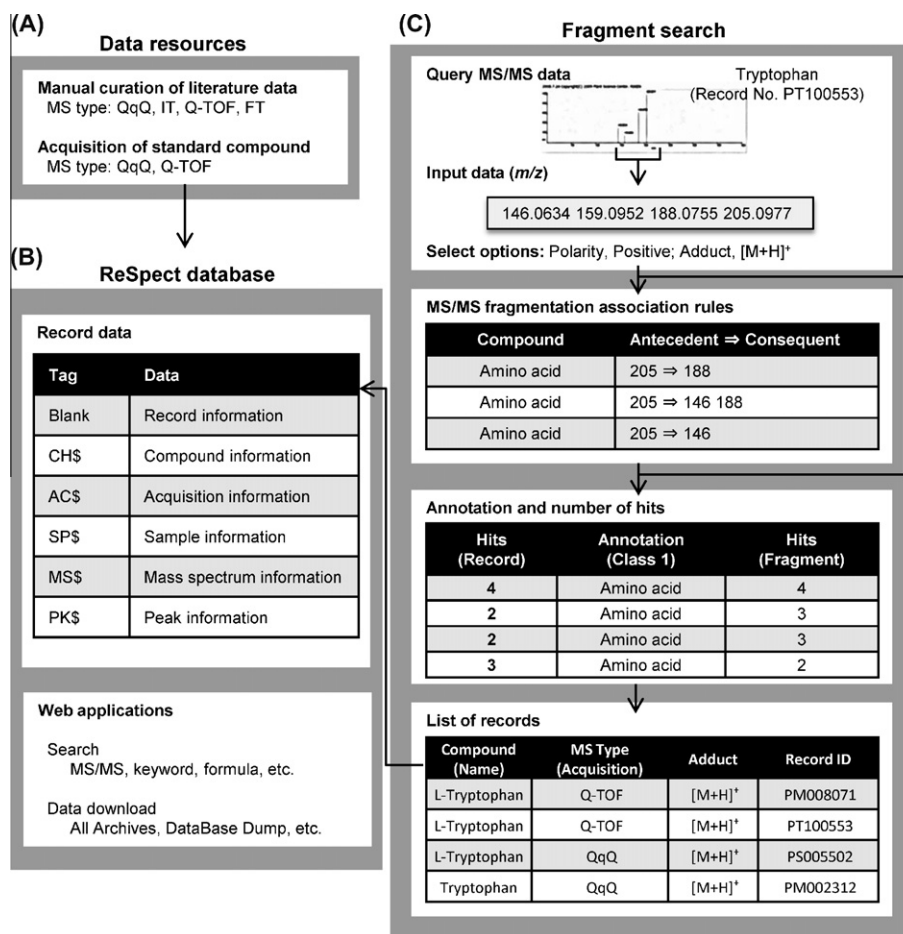
The third feature of ReSpect that differentiates it from other databases is that it is a plant-specific MS/MS data resource. Indeed, 100% of the records derived from the literature for the ReSpect database have been detected in plant species. Thousands of phytochemicals have been reported in the literature. The literature for MS/MS data of phytochemicals can be selected after running PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>) and Google Scholar (<http://scholar.google.com>) searches (Butler, 2004). The records derived from literature contained various types of MS instruments (e.g., tandem quadrupole, QqQ; ion trap, IT; quadrupole time of flight, Q-TOF; Fourier transfer, FT) under involuntary choice (Fig. 3A). In this study, the MS/MS data were manually digitalized, and all data were included in the ReSpect database. Because it is a plant-specific data resource, MS/MS data searches of ReSpect may lead to higher annotation success rates.



**Fig. 2.** Fragmentation scheme of MS/MS fragmentation association rules in flavonoids. Using predicted rules in Table 1, the common fragmentation sites corresponding to specific product ions were estimated for pelargonidin glycoside, genistein glycoside (A); delphinidin glycoside, quercetin glycoside (B); luteolin glycoside (C) and apigenin glycoside (D). The dashed arrow indicates the bond cleavage position in flavonoids.

## 2.2. Web applications of ReSpect

Web applications of the ReSpect database are already available for annotation of phytochemicals. The user can obtain a record for annotation of phytochemicals by simple handling. As the main web application of ReSpect (Fig. 3B), the query MS/MS data can be used for the newly established fragment search and existing algorithm of spectrum search using the cosine product. Additionally, all records are available for keyword searches, formula searches, exact mass searches, and index searches of the metabolite name.



**Fig. 3.** Workflow of database implementation and organization of Fragment search results. As data resources, we considered both the literature data and the data obtained using several types of MS instruments (A). All records were formatted using MassBank Record, and the web application supported several ways to search and download the records (B). Using Fragment search of ReSpect, the user can select the annotation of input data (*m/z* values), and the confidence levels can be managed by MS/MS fragmentation association rules (C).

Detailed instructions are displayed by help pop-ups on the web-site. All records can be downloaded from the website as text data and SQL databases under a Creative Commons Attribution – Non-commercial 2.1 Japan License (<http://creativecommons.org/licenses/by-nc/2.1/jp/deed.en>).

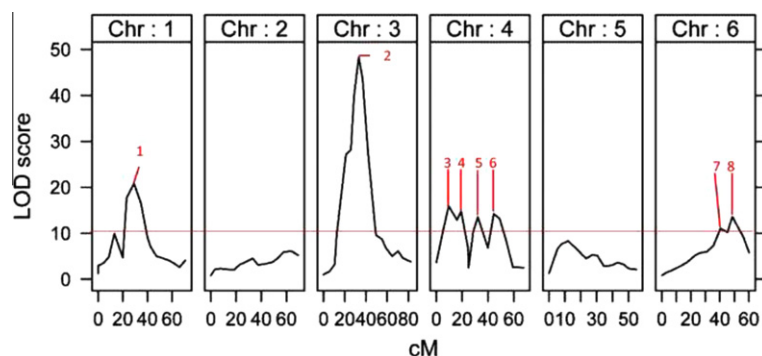
Using the fragment search of ReSpect for query data, users can select the polarity and adduct ion type (Fig. 3C). In addition, advanced search options can be used for more specific searches based on the user's threshold. The results of a fragment search conducted using a record of tryptophan (record No., PT100553; *m/z* tolerance,  $\pm 0.3$  Da; polarity, positive; adduct, [M+H]<sup>+</sup>; fragment filters: relative intensity  $\geq 0\%$ ; ranking, top 10) completely matched the fragment ions in four records derived from tryptophan (Fig. 3C and Supplementary Fig. 2). Among the optional results of the fragment search, the MS/MS fragmentation association rules corresponding to the precursor ion (*m/z* 205) of PT100553 were automatically summarized as a table on the webpage (Fig. 3C). In this case study, the top three rules were annotated as amino acids in compound class level 1. A total of 17 records corresponding to the rules (205  $\Rightarrow$  188; 205  $\Rightarrow$  146, 188; 205  $\Rightarrow$  146) were derived from tryptophan and tryptophan derivatives (Fig. 3C and Supplementary Fig. 3). These results suggest that the MS/MS fragmentation association rule of the fragment pattern is a powerful tool for prediction of fragmentation patterns in metabolites, as well as for managing the confidence level of annotations in fragment searches conducted using ReSpect.

### 2.3. A case study of annotation for untargeted MS/MS data

Metabolomics is developing toward an integral component of functional genomic approaches. The integrated analysis of metabolomics and transcriptomics is a powerful tool for the discovery of genes related to novel plant metabolisms (Saito et al., 2008). The integrated approaches may also be effective for searching key metabolites in plant metabolism. Among these approaches, metabolite quantitative locus (mQTL) analyses have shown that the gene marker-assisted selection of valuable metabolites has great potential for discovering the gene to metabolite networks in plants (Brotman et al., 2011; Lisec et al., 2008). Elucidation of the natural variation of metabolites in plant is an important topic in Phytochemistry; however, it is difficult to cover a broad and a minor target with the isolation technique of metabolites. Thus, the case study of untargeted mQTL analysis of *L. japonicus* accessions (MG-20 and B-129) is a rather important technique for narrowing down the key metabolites of natural variation. mQTL and the annotation process of unknown metabolites with ReSpect are very effective to reach a broad target audience in Phytochemistry.

More than 80,000 untargeted MS/MS tags (MS2T) derived from a mixture of seed extracts from MG-20 and B-129 are available on our website (Matsuda et al., 2010, 2009) (MS2T viewer, <http://prime.psc.riken.jp/>). Based on the biological and analytical thresholds (see Section 4), 88 MS2Ts were effectively selected for mQTL analysis. The results yielded eight gene marker positions that were suc-





**Fig. 4.** mQTL of *L. japonicus* B-129 × MG20 RILs. The maximum logarithm of odds (LOD) score values obtained from triplicate experiments were plotted in each chromosome (Chr: 1–6). The major mQTLs were estimated based on the LOD scores (>10). The genetic markers of mQTLs were indicated by red color as follows: 1, TM0199; 2, TM0110; 3, TM0832; 4, TM0194; 5, TM0030; 6, TM0429; 7, TM0880; 8, TM0756. The information regarding the genetic markers was provided by KDRI (<http://www.kazusa.or.jp/lotus/>).

**Table 2**  
ReSpect fragment search with MS2Ts of *Lotus japonicus*.

| MS2T id <sup>a</sup> | mQTL <sup>b</sup> | RT <sup>c</sup> | Query data <sup>d</sup> |  | Annotation with compound classes <sup>e</sup> |                       |
|----------------------|-------------------|-----------------|-------------------------|--|---|-----------------------|
|                      |                   |                 | Precursor ion           | Product ions                             | Level 2                                       | Level 3               |
| LJA01p24687          | 1                 | 2.8             | 465                     | 303                                      | Anthocyanin                                   | Delphinidin glycoside |
| LJA01p06814          | 1                 | 3.2             | 449                     | 287                                      | Anthocyanin                                   | Cyanidin glycoside    |
| LJA01p41308          | 1                 | 4.8             | 435                     | 303                                      | Flavonol                                      | Kaempferol glycoside  |
| LJA01n36824          | 1                 | 4.8             | 433                     | 301, 300, 271, 255, 243, 227             | Flavonol                                      | Quercetin glycoside   |
| LJA01n04957          | 1                 | 4.9             | 433                     | 301, 300, 271, 255, 243                  | Flavonol                                      | Quercetin glycoside   |
| LJA01n20490          | 3,4               | 5.4             | 431                     | 285, 284, 255, 229, 227                  | Flavonol                                      | Kaempferol glycoside  |
| LJA01p38469          | 3,4               | 5.5             | 287                     | 258, 213, 157, 137, 121, 107, 93, 77, 69 | Flavonol                                      | Kaempferol glycoside  |
| LJA01n04969          | 3,4               | 5.5             | 431                     | 285, 284, 255, 227                       | Flavonol                                      | Kaempferol glycoside  |
| LJA01n09025          | 7,8               | 4.0             | 625                     | 301, 300, 271, 255                       | Flavonol                                      | Quercetin glycoside   |

<sup>a</sup> MS2Ts were named as follows: plant species (LJA, *Lotus japonicus*), sequential project id (01), polarity (p, positive; n, negative); sequential id.

<sup>b</sup> The number of mQTL corresponds to Fig. 4.

<sup>c</sup> RT, Retention time (minutes) of MS2Ts.

<sup>d</sup> Query data for the ReSpect fragment search were selected based on the intensity and *m/z* values of MS2Ts (see Section 4).

<sup>e</sup> All annotated MS2Ts with ReSpect records were categorized in flavonoids of compound class level 1, and detailed annotations were assigned with compound class levels 2 and 3.

cessfully estimated as major mQTLs (Fig. 4). A total of 18 MS2Ts were associated with the major mQTLs (Supplementary Table 2). A fragment search of metabolite candidates with ReSpect successfully annotated nine MS2Ts as flavonoid glycosides (Table 2). These metabolites have been reported in *L. japonicus* (Suzuki et al., 2008). In MS2T of LJA01p24687, the association rules of the fragment search revealed abundant matching of 34 records dependent on a common fragmentation pattern (465 ⇒ 303) of a flavonoid, as described in Table 1. These results suggest that the use of ReSpect with integrated metabolomics approaches has the potential for effective elucidation of the putative structure of phytochemicals.

### 3. Concluding remarks

The first plant-specific MS/MS data resource and web database has been established. The newly established fragment search of ReSpect enables prediction of the MS/MS-based annotation of detectable phytochemicals, and this search system is suitable to search literature data for which only the *m/z* values of MS/MS are available. Using a fragment search of ReSpect with MS/MS query data, the user can obtain annotations based on the available records, and the confidence levels of the annotations can be evaluated based on the number of matched records and MS/MS fragmentation association rules. As a case study, the annotation of flavonoid glycosides was demonstrated in two accessions of *L. japonicus*, for which there were significant differences in the accumulation patterns. In this case study, MS/MS fragmentation association rules

clearly demonstrated common fragmentation patterns among the flavonoids.

Currently, ReSpect does not support multiple stage mass spectrometries (MS<sub>n</sub>, *n* > 2) derived from IT-MS because the focus was mainly on the MS/MS data and its search algorithm. One of the key challenges in the next database is the development of a novel search algorithm for MS<sub>n</sub>. The relational search for interpretation of MS<sub>n</sub> data in the literature will be ultimately effective in helping towards the structural characterization of unknown metabolites, and it also is a very valuable data resource for the next database. In addition, the accurate mass measurement technique using high resolution MS and MS/MS instruments, Q-TOF-MS and FT-MS, also allowed us to search by narrowing down the *m/z* tolerance. These improved search systems depended on not only the MS acquisition and instrument type, but also the metadata of sample in a literature; these metadata include plant species and growth conditions, which can effectively narrow down the candidate structure of an unknown metabolite. These detailed data collection and curation activities should be organized as a common data deposit system, which is supported by each scientific field. With the increase of MS/MS database activities, the annotation process requires a more detailed compound ontology construction. However, the current simple ontology system in ReSpect is certainly biased towards secondary metabolites, which is our area of expertise. Hence, the major metabolites in the primary metabolites were categorized into “Others” in compound level 1, and this will be improved in the next database.

At present, a few hundred unknown phytochemicals in representative plant species can be predicted using MS2Ts (MS2T viewer, <http://prime.psc.riken.jp/>), and a framework including ReSpect has been reported (Matsuda et al., 2011). However, computational prediction of the structure of metabolites based on MS/MS data is still very challenging without reference data (standard compounds and biological data) owing to the limitations of the available algorithms and model data. MS/MS data resources developed in previous studies cannot be downloaded and are therefore difficult to reuse for the development of novel methodology. ReSpect is the first tool for annotation of phytochemicals based on downloadable MS/MS data resources and a database. It is expected that users of bioinformatics and metabolomics will be able to develop novel algorithms and methodologies using our database and data resources.

## 4. Experimental

### 4.1. MS/MS data in ReSpect

The literature used for the MS/MS data were obtained from PubMed and Google Scholar, and the MS/MS data were manually digitized. All literature used are listed in the information section in ReSpect (<http://spectra.psc.riken.jp/>). For studies in which the relative intensity value of the MS fragment spectra was not described digitally, the heights of the MS peaks in the figures were manually measured and calculated. For literature data that have only *m/z* values without intensity data, the intensities of all ions to 100% were adjusted. In addition, the MS/MS data of the authentic standard compounds were acquired using Q-TOF/MS and QqQ/MS (Matsuda et al., 2010, 2009; Sawada et al., 2009a,b,c). All records used for ReSpect were categorized into compound classes according to the book (Cane, 1999; Kelly, 1999; Pinto, 1999; Sankawa, 1999). The records were formatted using MassBank Record (Horai et al., 2010), which can accept MS/MS data as well as its metadata (summary, chemical, analytical, spectral and outer database links). As of September 2011, ReSpect contained 8649 records corresponding to 3595 metabolites. Specifically, the database contains 163 selected literature reports (3341 records corresponding to 2741 metabolites), Q-TOF/MS data (1050 records corresponding to 575 standard compounds) (Matsuda et al., 2010, 2009) and QqQ/MS data (4258 records corresponding to 861 standard compounds) (Sawada et al., 2009a,b,c). The ratios of the metabolite ontology classes are biased toward the available data and referenced classification of metabolites. In the data section of ReSpect, the example analyses (statistics and evaluation) were automatically executed based on the updated records.

### 4.2. MS/MS fragment association rules in ReSpect

In this study, association rules and index values (transaction, support, confidence, lift) of MS/MS data in the ReSpect records were enumerated using a module of Perl (Data Mining Association-Rules Version 0.10). To mimic the MS/MS data structure, the maximum value of the product ion and other product ions are defined as the antecedent and consequent, respectively, an antecedent is defined as a single ion, and the difference in *m/z* values between the antecedent and consequent is greater than 10. Moreover, a rare rule (only once in the transaction) was also excluded as invalid rules. On the basis of these definitions, the following invalid rules were excluded from the association rules: multiple antecedents in a rule, antecedent < consequent, and the estimated ion of the consequent corresponding to the isotopic ion of the antecedent. The index values of MS/MS fragmentation association rules with an antecedent (X), consequent (Y) and transaction (Z) are calculated

as follows: support (X) = count (X)/count (Z), support (Y) = count (Y)/count (Z), support (X  $\Rightarrow$  Y) = count (X  $\cup$  Y)/count (Z), confidence (X  $\Rightarrow$  Y) = count (X  $\cup$  Y)/count (X), lift (X  $\Rightarrow$  Y) = confidence (X  $\Rightarrow$  Y)/support (Y). The MS/MS fragmentation association rules can be downloaded from the data section of ReSpect.

### 4.3. Construction of web applications

The web applications of ReSpect were implemented using HTML5.0, JavaScript, Perl, and MySQL and were hosted on CentOS. ReSpect has been successfully tested on Mozilla Firefox 3.6+, Safari 5.0+, Google Chrome 5.0+, Opera 10.0+, and Microsoft Internet Explorer 8.0+. The user can define the threshold for query and data resources with the following search options: polarity (positive, negative), spectra filters (similarity score, relative intensity, precursor ion and *m/z* tolerance), data type selection (ignores all single fragments and one major fragment spectrum) and data set selection (literature, Q-TOF/MS and QqQ/MS).

### 4.4. mQTL analysis of *L. japonicus*

#### 4.4.1. Plant materials

The *L. japonicus* accessions (MG-20 and B-129) and mapping population of 129 recombinant inbred lines (RILs) have been described in detail elsewhere (Gondo et al., 2007; Hayashi et al., 2001; Kawaguchi et al., 2001; Klein and Grusak, 2009).

#### 4.4.2. Metabolic profiling

All raw data for mQTL analyses carried out in this study are available in our webpage (DropMet, <http://prime.psc.riken.jp/>). Metabolic profiling data was collected by analyzing seed extract mixtures of *L. japonicus* (accessions: MG-20 and B-129), which were prepared as follows: a seed was disrupted using a multi beads shocker (Shake Master NEO, BMS, Tokyo, Japan), and the seed powder was extracted with 1 mL of extraction buffer (0.1% HCO<sub>2</sub>H, MeOH–H<sub>2</sub>O (4:1 v/v), and 33.6 nM lidocaine and 840 nM camphor sulfonic acid as internal standards of positive mode and negative mode, respectively) using a multi beads shocker. After centrifugation (4 °C, 10,000 rpm, 5 min), the sample tubes were subjected to sample preparation (buffer transfer, 250  $\mu$ L of seed extract; dried up, resolution, 250  $\mu$ L of LC-MS grade H<sub>2</sub>O; filtration, 384 well formatted filter (0.45  $\mu$ m PVDF, Whatman, NJ, USA) with a liquid handling system (Microlab Star Plus, Hamilton, Ontario, Canada) (Sawada et al., 2009a). The metabolites were analyzed using a liquid chromatography quadrupole time of flight mass spectrometer and tandem quadrupole mass spectrometer system (UPLC-QTOF Premier and UPLC-TQS, Waters Co., MA, USA) (Matsuda et al., 2009; Sawada et al., 2009a).

After cutting off the low intensity data of MS2Ts, 702 MS2Ts were used for a selected reaction monitoring (SRM) assay of UPLC-TQS. To determine the optimal collision energy for MS2Ts, collision induced dissociation fragmentation analyses at six energy steps (10–60 eV) were conducted using the same seed extracts. As a result, 4212 SRMs were used for analysis. The SRM condition that gave the highest signal intensity among the six energy steps was defined as the optimal SRM. The optimal SRM conditions were further selected based on the following criterion: the relative standard deviation of UPLC-TQS peak area values must be less than 10% (analytical replicates = 3). Finally, 342 SRM conditions for MS2Ts were successfully assigned to the seed extract of *L. japonicus*. Based on the natural variations in the metabolite accumulation patterns among the accessions (MG-20 and B-129), 88 SRM conditions were used for mQTL analysis.

#### 4.4.3. mQTL analysis

Metabolic profiling data of 129 RILs were processed as follows: the peak area values of 88 SRMs were divided by the peak area values of the internal standards, which were lidocaine and camphor sulfonic acid for the positive and negative ion modes, respectively. The divided values were then converted into z-scores of binary log-arithms after missing values were replaced with 0.1. The genotype data of each RIL were obtained from KDRI (<http://www.kazu-sa.or.jp/lotus/>). Using processed metabolic profiling data and gene marker data of RILs, mQTL analysis was carried out using R/qtl (Broman et al., 2003) (Supplementary Data).

#### Acknowledgements

**Funding:** This work was supported by the RIKEN Plant Science Center, Japan Science and Technology Agency (CREST Grant), Japan Advanced Plant Science Research Network, and Fuji Foundation for Protein Research. This research was supported, in part, by JST, Strategic International Collaborative Research Program, SICORP. The RILs of *L. japonicus* were constructed by the Forage Legume Breeding Laboratory, National Agricultural Research Center for Hokkaido Region, and KDRI. The RILs were distributed by the National BioResource Project (Legume Base).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.phytochem.2012.07.007>.

#### References

- Agrawal, R., Imielinski, T., Swami, A., 1993. Mining association rules between sets of items in large databases. SIGMOD Conference, 207–216.
- Benton, H.P., Wong, D.M., Trauger, S.A., Siuzdak, G., 2008. XCMS2: processing tandem mass spectrometry data for metabolite identification and structural characterization. Anal. Chem. 80, 6382–6389.
- Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19, 889–890.
- Brotman, Y., Riewe, D., Lise, J., Meyer, R.C., Willmitzer, L., Altmann, T., 2011. Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in Arabidopsis. J. Plant Physiol. 168, 1387–1394.
- Butler, D., 2004. Science searches shift up a gear as Google starts Scholar engine. Nature 432, 423.
- Cane, D.E., 1999. Comprehensive Natural Products Chemistry: Isoprenoids Including Carotenoids and Steroids. Pergamon, Oxford.
- Cavaliere, C., Foglia, P., Pastorini, E., Samperi, R., Lagana, A., 2005. Identification and mass spectrometric characterization of glycosylated flavonoids in *Triticum durum* plants by high-performance liquid chromatography with tandem mass spectrometry. Rapid Commun. Mass Spectrom. 19, 3143–3158.
- Cui, Q., Lewis, I.A., Hegeman, A.D., Anderson, M.E., Li, J., Schulte, C.F., Westler, W.M., Eghbalian, H.R., Sussman, M.R., Markley, J.L., 2008. Metabolite identification via the Madison Metabolomics Consortium Database. Nat. Biotechnol. 26, 162–164.
- Fernie, A.R., Trethewey, R.N., Krotzky, A.J., Willmitzer, L., 2004. Metabolite profiling: from diagnostics to systems biology. Nat. Rev. Mol. Cell Biol. 5, 763–769.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N., Willmitzer, L., 2000. Metabolite profiling for plant functional genomics. Nat. Biotechnol. 18, 1157–1161.
- Fiehn, O., Wohlgenuth, G., Scholz, M., 2005. Setup and annotation of metabolomic experiments by integrating biological and mass spectrometric metadata. Proc. Lect. Notes Bioinformatics 3615, 224–239.
- Fiehn, O., Wohlgenuth, G., Scholz, M., Kind, T., Lee Do, Y., Lu, Y., Moon, S., Nikolau, B., 2008. Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J. 53, 691–704.
- Gondo, T., Sato, S., Okumura, K., Tabata, S., Akashi, R., Isobe, S., 2007. Quantitative trait locus analysis of multiple agronomic traits in the model legume *Lotus japonicus*. Genome 50, 627–637.
- Guy, C., Kopka, J., Moritz, T., 2008. Plant metabolomics coming of age. Physiol. Plant. 132, 113–116.
- Hall, R.D., Brouwer, I.D., Fitzgerald, M.A., 2008. Plant metabolomics and its potential application for human nutrition. Physiol. Plant. 132, 162–175.
- Hayashi, M., Miyahara, A., Sato, S., Kato, T., Yoshikawa, M., Taketa, M., Pedrosa, A., Onda, R., Imaizumi-Anraku, H., Bachmair, A., Sandal, N., Stougaard, J., Murooka, Y., Tabata, S., Kawasaki, S., Kawaguchi, M., Harada, K., 2001. Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. DNA Res. 8, 301–310.
- Hildebrandt, C., Wolf, S., Neumann, S., 2011. Database supported candidate search for metabolite identification. J. Integr. Bioinform. 8, 157.
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. 45, 703–714.
- Kawaguchi, M., Motomura, T., Imaizumi-Anraku, H., Akao, S., Kawasaki, S., 2001. Providing the basis for genomics in *Lotus japonicus*: the accessions Miyakojima and Gifu are appropriate crossing partners for genetic analyses. Mol. Genet. Genomics 266, 157–166.
- Kelly, J.W., 1999. Comprehensive Natural Products Chemistry: Amino acids, Peptides, Porphyrins, and Alkaloids. Pergamon, Oxford.
- Klein, M.A., Grusak, M.A., 2009. Identification of nutrient and physical seed trait QTL in the model legume *Lotus japonicus*. Genome 52, 677–691.
- Last, R.L., Jones, A.D., Shachar-Hill, Y., 2007. Towards the plant metabolome and beyond. Nat. Rev. Mol. Cell Biol. 8, 167–174.
- Lise, J., Meyer, R.C., Steinfath, M., et al., 2008. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. Plant J. 53, 960–972.
- Matsuda, F., Hirai, M.Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N.J., Sakurai, T., Shimada, Y., Saito, K., 2010. AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol. 152, 566–578.
- Matsuda, F., Nakabayashi, R., Sawada, Y., Suzuki, M., Hirai, M.Y., Kanaya, S., Saito, K., 2011. Mass spectra-based framework for automated structural elucidation of metabolome data to explore phytochemical diversity. Front. Plant Physiol. 2, 1–10.
- Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K., Saito, K., 2009. MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. Plant J. 57, 555–577.
- Nakabayashi, R., Kusano, M., Kobayashi, M., Tohge, T., Yonekura-Sakakibara, K., Kogure, N., Yamazaki, M., Kitajima, M., Saito, K., Takayama, H., 2009. Metabolomics-oriented isolation and structure elucidation of 37 compounds including two anthocyanins from *Arabidopsis thaliana*. Phytochemistry 70, 1017–1029.
- Pinto, B.M., 1999. Comprehensive Natural Products Chemistry: Carbohydrates and Their Derivatives Including Tannins, Cellulose, and related Lignins. Pergamon, Oxford.
- Rauter, A.P., Martins, A., Borges, C., Ferreira, J., Jorge, J., Bronze, M.R., Coelho, A.V., Choi, Y.H., Verpoorte, R., 2005. Liquid chromatography-diode array detection-electrospray ionisation mass spectrometry/nuclear magnetic resonance analyses of the anti-hyperglycemic flavonoid extract of *Genista tenera*. Structure elucidation of a flavonoid-C-glycoside. J. Chromatogr. A 1089, 59–64.
- Saito, K., Hirai, M.Y., Yonekura-Sakakibara, K., 2008. Decoding genes with coexpression networks and metabolomics – ‘majority report by precogs’. Trends Plant Sci. 13, 36–43.
- Saito, K., Matsuda, F., 2010. Metabolomics for functional genomics, systems biology, and biotechnology. Annu. Rev. Plant Biol. 61, 463–489.
- Sankawa, U., 1999. Comprehensive Natural Products Chemistry: Polyketides and Other Secondary Metabolites Including Fatty Acids and Their Derivatives. Pergamon, Oxford.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K., Fujishiro, T., Katoh, M., Kohara, M., Kishida, Y., Minami, C., Nakayama, S., Nakazaki, N., Shimizu, Y., Shinpo, S., Takahashi, C., Wada, T., Yamada, M., Ohmido, N., Hayashi, M., Fukui, K., Baba, T., Nakamichi, T., Mori, H., Tabata, S., 2008. Genome structure of the legume, *Lotus japonicus*. DNA Res. 15, 227–239.
- Sawada, Y., Akiyama, K., Sakata, A., Kuwahara, A., Otsuki, H., Sakurai, T., Saito, K., Hirai, M.Y., 2009a. Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. Plant Cell Physiol. 50, 37–47.
- Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., Saito, K., Hirai, M.Y., 2009b. Omics-based approaches to methionine side chain elongation in Arabidopsis: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. Plant Cell Physiol. 50, 1181–1190.
- Sawada, Y., Toyooka, K., Kuwahara, A., Sakata, A., Nagano, M., Saito, K., Hirai, M.Y., 2009c. Arabidopsis bile acid:sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. Plant Cell Physiol. 50, 1579–1586.
- Schauer, N., Fernie, A.R., 2006. Plant metabolomics: towards biological function and mechanism. Trends Plant Sci. 11, 508–516.
- Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G., 2005. METLIN: a metabolite mass spectral database. Ther. Drug Monit. 27, 747–751.
- Sokolow, S., Karnofsky, J., Gustafson, P., 1978. The Finnigan Library Search Program Finnigan Application, Report 2.
- Sumner, L.W., Mendes, P., Dixon, R.A., 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62, 817–836.
- Suzuki, H., Sasaki, R., Ogata, Y., Nakamura, Y., Sakurai, N., Kitajima, M., Takayama, H., Kanaya, S., Aoki, K., Shibata, D., Saito, K., 2008. Metabolic profiling of flavonoids in *Lotus japonicus* using liquid chromatography Fourier transform ion cyclotron resonance mass spectrometry. Phytochemistry 69, 99–111.
- Weckwerth, W., 2003. Metabolomics in systems biology. Annu. Rev. Plant Biol. 54, 669–689.

- Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J.A., Lim, E., Sobsey, C.A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazzyrova, A., Shaykhtudinov, R., Li, L., Vogel, H.J., Forsythe, I., 2009. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 37, D603–D610.
- Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D.D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G.E., Macinnis, G.D., Weljie, A.M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B.D., Vogel, H.J., Querengesser, L., 2007. HMDB: the Human Metabolome Database. *Nucleic Acids Res.* 35, D521–D526.