

Methodology Report

MarVis-Filter: Ranking, Filtering, Adduct and Isotope Correction of Mass Spectrometry Data

Alexander Kaefer,¹ Manuel Landesfeind,¹ Mareike Possienke,² Kirstin Feussner,^{2,3}
Ivo Feussner,² and Peter Meinicke¹

¹ Department of Bioinformatics, Institute of Microbiology and Genetics, Georg-August-University Göttingen, 37077 Göttingen, Germany

² Department for Plant Biochemistry, Albrecht-von-Haller-Institute for Plant Sciences, Georg-August-University Göttingen, 37077 Göttingen, Germany

³ Department of Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August-University Göttingen, 37077 Göttingen, Germany

Correspondence should be addressed to Alexander Kaefer, alex@gobics.de

Received 28 July 2011; Revised 18 January 2012; Accepted 18 January 2012

Academic Editor: Brad Upham

Copyright © 2012 Alexander Kaefer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Statistical ranking, filtering, adduct detection, isotope correction, and molecular formula calculation are essential tasks in processing mass spectrometry data in metabolomics studies. In order to obtain high-quality data sets, a framework which incorporates all these methods is required. We present the MarVis-Filter software, which provides well-established and specialized methods for processing mass spectrometry data. For the task of ranking and filtering multivariate intensity profiles, MarVis-Filter provides the ANOVA and Kruskal-Wallis tests with adjustment for multiple hypothesis testing. Adduct and isotope correction are based on a novel algorithm which takes the similarity of intensity profiles into account and allows user-defined ionization rules. The molecular formula calculation utilizes the results of the adduct and isotope correction. For a comprehensive analysis, MarVis-Filter provides an interactive interface to combine data sets deriving from positive and negative ionization mode. The software is exemplarily applied in a metabolic case study, where octadecanoids could be identified as markers for wounding in plants.

1. Introduction

A central aim of untargeted Metabolomics and Metabolomics studies is the identification of marker metabolites which play a crucial role in the experimental context [1, 2]. Mass spectrometry combined with either gas chromatography (GC/MS) or liquid chromatography (LC/MS) has become a key technology for metabolome analysis under different experimental conditions [3, 4]. A typical data set after peak detection and sample alignment [5–7] consists of several thousand marker candidates which are characterized by a retention time (RT), a mass-to-charge value (m/z), and a multivariate intensity profile of abundance levels per condition, respectively [8]. The

experimental conditions are represented by replicate samples and may correspond to environmental disease or genetic perturbations [9–11]. In order to obtain a high-quality data set of experiment-related marker candidates, the raw data set is usually ranked and filtered using supervised machine learning techniques such as Random Forest classification [12, 13] or statistical analysis based on ANOVA or Kruskal-Wallis tests [14–16]. The filtered marker candidates are then annotated according to known metabolites from public biological and biomedical compound databases [17–21]. A central task of annotation is the calculation of actual molecular masses corresponding to each marker candidate by correcting the m/z ratios according to the ionization mode, potential adduct formation, and included natural

isotopes [22]. This problem can be addressed by applying the ionization rules $[xm + y]^{z[+/-]}$ [23], where x denotes the number of combined target molecules, y the mass of attached molecules (adduct formation), and z the degree of ionization (e.g., single or double). Additionally, the number of included isotopes has to be estimated in order to query databases which contain monoisotopic compound masses. Based on a potential ionization rule with parameters x , y , z and the number of included isotopes, the corresponding compound mass can be calculated.

For the corrected masses which cannot be assigned to particular compounds, the identification can be supported by calculating possible molecular formulas. The number of considered formulas can be significantly reduced by incorporating information from preprocessing as well as rules for heuristic filtering of molecular formulas [24], respectively. A major step in this process is the estimation of the number of included carbon atoms based on the intensity profiles of previously detected isotopologues.

There are a great number of software packages available, which provide tools for statistical analysis of multivariate experimental data [25, 26]. A number of tools for peak detection and sample alignment of mass spectrometry data, such as MetAlign or OpenMS, also support the deconvolution of isotopologues and statistical analysis [27, 28]. For the XCMS platform [7], a package for the annotation of LC/ESI-MS mass signals based on adduct rules has been implemented [23]. The calculation of possible ionization products and the rule-based heuristic filtering of molecular formulas is provided by several software packages [22, 24]. However, to the best of our knowledge, there is no software available which incorporates all of these methods in a single user-friendly tool as offered by MarVis-Filter.

2. Materials and Methods

In the following sections, the algorithm for adduct/isotope correction and the implementation of MarVis-Filter are described in detail.

2.1. Algorithm for Adduct and Isotope Correction. The algorithm is based on the input of the retention times, m/z ratios, and raw intensity profiles of all marker candidates in a data set and calculates as output the potential monoisotopic mass, ionization rule, and number of included ^{13}C -isotopes for every candidate. The approach is based on a greedy strategy which minimizes the number of potential molecular masses and simultaneously maximizes the similarity of intensity profiles between candidates with a similar retention time and actual mass. This concept follows the paradigm that in mass spectrometry analysis a metabolite is usually represented by several marker candidates with a similar retention time and intensity profile, but different m/z ratios according to the various possibilities of ionization and number of included isotopes. As parameters, the algorithm expects a list of ionization/adduct rules sorted according to their relevance, the assumed maximal number of ^{13}C -isotopes per marker candidate, a mass tolerance, an RT tolerance, and a minimal

cosine similarity of intensity profiles. The isotopologues correction is restricted to the detection of ^{13}C .

For storage of pairwise cosine similarities between candidate profiles, the algorithm utilizes a five-dimensional matrix M . Each entry $M_{(m,a_1,i_1,a_2,i_2)}$ corresponds to the maximal cosine similarity between the intensity profile of candidate m , assuming ionization rule a_1 and i_1 ^{13}C -isotopes, and another candidate, which has a similar retention time (within tolerance) and corrected mass (within tolerance) assuming ionization rule a_2 and i_2 ^{13}C -isotopes. For each candidate m , the algorithm then chooses the ionization rule and number of ^{13}C -isotopes which is supported by the highest sum of cosine similarities. In the following, the algorithm is described in detail.

- (1) Initialize M with zeros.
- (2) Calculate all possible masses by applying all ionization rules and number of ^{13}C -isotopes to all candidate m/z ratios.
- (3) Consider all pairs of potential masses under the following constraints and fill M with pairwise cosine similarities of corresponding candidate profiles.
 - (i) Consider only pairs of different marker candidates.
 - (ii) Consider only pairs within the mass and RT tolerance.
 - (iii) Consider only pairs with at least the requested cosine similarity.
 - (iv) Consider only pairs with different combinations of adduct rules and number of isotopes.
 - (v) For each entry in M hold only the maximum cosine similarity.
- (4) Calculate the reduced three-dimensional matrix M^{red} with summed entries:

$$M_{(m,a_1,i_1)}^{\text{red}} = \sum_{a_2,i_2} M_{(m,a_1,i_1,a_2,i_2)}. \quad (1)$$
- (5) Choose for each candidate m : the adduct rule and isotope number with the maximal sum of similarities $c_{\text{max}} = \max_{a_1,i_1} (M_{(m,a_1,i_1)}^{\text{red}})$. If $c_{\text{max}} = 0$, use the first ionization rule and zero ^{13}C -isotopes as default.
- (6) Calculate the masses according to chosen rules and isotope numbers.

In order to avoid apparently false associations between marker candidates, negative cosine similarities are disregarded. If for a given candidate different selections of the ionization rule and the number of isotopes maximize the sum of cosine similarities, the ionization rules with the highest relevance and the minimal number of ^{13}C -isotopes are selected.

Following the annotation of the ionization rules and ^{13}C -isotopes, the number of carbon atoms per candidate is estimated by comparing the raw intensities of marker candidates with zero predicted ^{13}C -isotopes (I_M) and the

respective marker candidates including one ^{13}C -isotope (I_{M+1}) according to the following formula:

$$n_C = \frac{98.9 I_{M+1}}{1.1 I_M}, \quad (2)$$

corresponding to the natural abundances of carbon isotopes. Given a pair of candidates, annotated as isotopologues (M and $M + 1$) and with the same ionization rule, a robust estimation of the number of carbon atoms is obtained by calculating the median n_C over all samples included in both intensity profiles.

2.2. Implementation. MarVis-Filter is implemented in the Matlab and C programming language and has been compiled together with the MarVis-Cluster tool [29] for Microsoft Windows XP/Vista/7. Execution of the software requires installation of the Matlab Compiler Runtime, which is provided with the software. The installation packages, the documentation, and example data sets can be downloaded from the project home page <http://marvis.gobics.de/>.

For data import and export MarVis-Filter uses the CSV (Comma Separated Values) file format, which can easily be processed by statistical analysis software and spreadsheet applications. MarVis-Filter also supports the direct import of aligned mass spectrometry samples from MarkerLynx Application Manager of MassLynx (Waters Corporation, Milford). For interactive analysis, ranking and filtering of multivariate intensity profiles MarVis-Filter provides the well-known one-way ANOVA and Kruskal-Wallis tests [14] combined with methods for P value adjustment for multiple-hypothesis testing [30, 31]. Based on customizable lists of ionization rules, the adduct/isotope correction can be performed on raw or filtered data sets. The ionization rules are imported as text files and can easily be adapted or extended.

Figure 1 shows the main window of MarVis-Filter after import and ranking. The “Ranking plot” (1) displays the adjusted P values (y -axis) of all candidate intensity profiles in the current data set sorted in ascending order. The data set can interactively be filtered according to a user-defined significance level by selecting a marker, sliding the red separator line or jumping to a predefined level. The “Profile plot” (2) shows the raw intensity profile of the currently selected marker candidate. Intensity values of replicated samples belonging to the same experimental conditions are marked in the same color. The “Marker information box” (3) displays information about all marker candidates of the data set arranged according to the P values and characterized by the m/z ratio, RT and additional user-defined scores, which can be imported along with the data set. After adduct and isotope correction, the additional annotations are displayed in this listbox as well. The “Data set clipboard listbox” (4) shows data sets which are currently held in the MarVis clipboard. The current (filtered or unfiltered) data set can simply be added or removed to/from this list. The data set clipboard supports an adduct and isotope correction of selected data sets in a batch mode. Data sets which were corrected based on different sets of ionization rules (e.g.,

positive and negative ionization) may be combined into one single data set.

For selected candidate profiles, bar plots, standard error plots, and boxplots can easily be inspected and exported in various image formats. For detailed analysis, the user can zoom into all plots. Additionally, MarVis-Filter provides a convenient interface for quick candidate search based on the ID, RT, m/z , or mass value.

MarVis-Filter also provides a molecular formula calculator, which is based on the Seven Golden Rules [24] and utilizes the estimated number of carbon atoms per marker candidate obtained after adduct and isotope correction.

MarVis-Filter and MarVis-Cluster [29] are combined in the MarVis-Suite which features the direct data exchange between preprocessing in MarVis-Filter and convenient visualization of multivariate intensity profiles and high-level cluster analysis in MarVis-Cluster.

3. Results and Discussion

The functionality of MarVis-Filter is demonstrated using two data sets of a metabolomic case study for plant wounding experiments [8]. The data sets are available on the project homepage <http://marvis.gobics.de/> together with a detailed description of the extraction and UPLC-TOF method. Additionally, the data sets are available for import in MarVis-Filter after installation of the MarVis-Suite (wound_neg_raw.csv and wound_pos_raw.csv in the examples directory).

3.1. Case Study and Data Sets. The case study reflects a wounding time course of *Arabidopsis thaliana* wild-type (WT) plants as well as of mutant plants (dde 2-2), which are deficient in the biosynthesis of the plant wound hormone jasmonic acid and its derivatives [32]. The wounding time course represents eight experimental conditions. The first four conditions reflect the metabolic situation within a wounding time course of wild-type (WT) plants, starting with the unwounded control plants (abbreviation wt_0) followed by the plants harvested 0.5 (wt_30), 2 (wt_2), and 5 hours past wounding (wt_5). The conditions 5 to 8 represent the analogous time course for the jasmonate deficient mutant plant dde 2-2 (aos_0, aos_30, aos_2, aos_5). Each condition contains nine replicate samples.

3.2. Data Import and Analysis in MarVis-Filter. The two data sets are imported sequentially in MarVis-Filter using the “Import raw CSV data” entry in the “File” menu with the following options: Delimiter: “;”; Start row: 5, Start column: 3; ID label: “id”; Generate IDs: activated; x column: 2; x label: “rt”; y column: 3; y label: “ m/z ”; Condition identifiers: “wt_0, wt_30, wt_2, wt_5, aos_0, aos_30, aos_2, aos_5”.

After data import, the marker candidates are sorted and ranked according to the P values of a Kruskal-Wallis test and the Bonferroni-Holm adjustment for multiple hypothesis testing [30] by selecting the corresponding checkboxes in the “Filter dialog” and the “Adjustment for multiple testing” dialog.

Adduct and isotope correction are performed on the full data sets separately using predefined sets of adduct

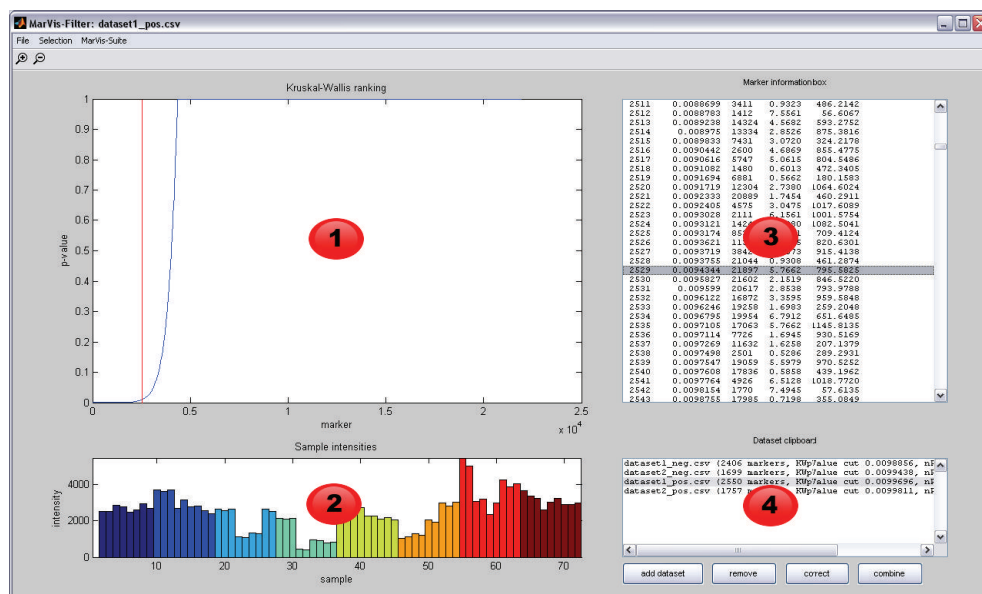


FIGURE 1: The main window of MarVis-Filter after data import and ranking. The “Ranking plot” (1) displays the adjusted P values (y -axis) of all candidate intensity profiles. The “Profile plot” (2) shows the raw intensity profile of the currently selected marker candidate. The “Marker listbox” (3) displays information about the current marker candidate. The “Data set clipboard listbox” (4) shows data sets which are currently held in the MarVis clipboard.

TABLE 1: Overview on data sets from the metabolomic case study for plant wounding experiments. The columns “Candidates” and “Filtered candidates” contain the number of marker candidates in the raw data set and the number of significant candidates in the filtered data set, respectively.

Data set	Ionization	Candidates	Filtered candidates	Samples per condition
1	Negative	24796	1719	9
2	Positive	23325	1785	9

rules for the negative (Table 2) and positive ionization mode (Table 3), an RT tolerance of 0.04 minutes, a mass tolerance of 0.005 Da, a minimal cosine similarity of 0.75, and a maximum number of two ¹³C-isotopes per candidate. The adduct rules had been determined in previous targeted UPLC-TOF-MS experiments. After correction, the data sets are filtered according to a significance level for adjusted P values of 0.01 (“Goto level” entry in “Selection” menu) and added to the MarVis data set clipboard. Table 1 shows the initial number of imported marker candidates and the number of high-quality marker candidates after filtering. Finally, the two data sets in the MarVis clipboard are concatenated using the “combine” button. The combined data set can be sorted according to a user-defined method once again and is then presented in a new MarVis-Filter window. After selecting the whole data set, the combined subset of 3504 high-quality marker candidates can be exported as a CSV file, and clustered as well as visualized using MarVis-Cluster (“Goto MarVis-Cluster” entry in the “MarVis-Suite” menu). Figure 2 shows the results from clustering of the filtered and combined data in MarVis-Cluster.

TABLE 2: List of adduct rules for correction of data measured in negative ionization mode.

Rule	Description	Rule
1	Deprotonation	$[m - H]^-$
2	Formate adduct	$[m + CH_2O_2 - H]^-$
3	Formate adduct with sodium	$[m + CH_2O_2 - 2H + Na]^-$

TABLE 3: List of adduct rules for correction of data measured in positive ionization mode.

Rule	Description	Rule
1	Protonation	$[m + H]^+$
2	Ammonium adduct	$[m + NH_4]^+$
3	Sodium adduct	$[m + Na]^+$

3.3. Identification of Metabolites. The corrected, filtered, and combined data sets were used to identify metabolites which show a significant change of abundance in the wound time course in WT and/or jasmonate deficient mutant plants. First, the corrected masses of marker candidates were matched to molecular masses of all compounds recorded in the KEGG [17] and AraCyc [18] database or literature [33] based on a tolerance of 0.005 Da. The identity of marker candidates was confirmed based on the isotopic pattern and coelution with identical standards or MS/MS fragmentation [34]. Thus, a number of oxylipins could be identified as wound-induced metabolite markers (see Table 4). Oxylipins are metabolites deriving from lipid peroxidation and are involved in regulating developmental processes as well as environmental responses, like the inflammatory or wound response, in nearly every organism. Among these bioactive

TABLE 4: Identified metabolites in the combined and filtered data set. The retention time is measured in minutes and the exact compound mass is stated in Dalton. The columns “Negative” and “Positive” contain the number of associated marker candidates/ions obtained in the negative or positive ionization mode. The column “Ions” contains the sum of associated marker candidates/ions per compound. The column “*P* value” contains the minimal adjusted *P* value of the Kruskal-Wallis test over all associated marker candidates, respectively. The column “Mass error” contains the absolute difference between the corrected mass of the marker candidate with the minimal adjusted *P* value and the exact compound mass in Dalton.

RT	Exact mass	Mass error	Name	Formula	Ions	Negative	Positive	<i>P</i> value
0.73	210.1256	0.0015	Jasmonic acid	C12H18O3	5	5	0	6.67e-8
2.08	292.2038	0.0021	OPDA	C18H28O3	8	4	4	3.41e-8
1.85	310.2144	0.0016	13-HPOT	C18H30O4	1	1	0	2.04e-4
2.49	292.2038	0.0027	13-KOT	C18H28O3	4	4	0	1.70e-7
1.33	264.1725	0.0037	dn-OPDA	C16H24O3	5	3	2	7.65e-8
0.5	226.1205	0.0009	11/12-Hydroxy jasmonic acid	C12H18O4	4	4	0	2.87e-8
0.51	339.2046	0.0008	12-Hydroxy jasmonoyl isoleucine	C18H29NO5	1	1	0	3.44e-5
0.51	353.1838	0.001	12-Carboxy jasmonoyl isoleucine	C18H27NO6	1	1	0	2.20e-5
4.02	760.4762	0.005	18:3/dn-OPDA-MGDG	C43H68O11	4	0	4	1.96e-6
2.85	774.4554	0.0022	OPDA/dn-OPDA-MGDG	C43H66O12	8	0	8	1.93e-7
3.26	802.4867	0.0034	OPDA/OPDA-MGDG	C45H70O12	7	0	7	1.79e-7
4.59	1048.6487	0.0033	OPDA/dn-OPDA-MGDG-OPDA	C61H92O14	9	0	9	1.48e-6
4.89	1076.68	0.002	OPDA/OPDA-MGDG-OPDA	C63H96O14	8	0	8	2.13e-6
2.36	936.5083	0.0023	OPDA/dn-OPDA-DGDG	C49H76O17	4	0	4	1.86e-6
2.76	964.5396	0.0021	OPDA/OPDA-DGDG	C51H80O17	6	0	6	1.98e-7

The identified oxylipins are found in literature under the following synonyms: Jasmonic acid (3-Oxo-2*R*-(2*Z*)-penten-1*R*-yl-cyclopentaneacetic acid), OPDA (12-Oxo-10,15(*Z*)-phytyldienoic acid or 4-Oxo-5 α -(2(*Z*)-pentenyl)-2-cyclopentene-1 α -octanoic acid), 13-HPOT (13-Hydroperoxy-octadeca-9(*Z*),11(*Z*),15(*Z*)-trienoic acid), 13-KOT (13-Keto-octadeca-9(*Z*),11(*Z*),15(*Z*)-trienoic acid), dn-OPDA (4-Oxo-5*S*-(2*Z*)-2-penten-1-yl-2-cyclopentene-1*S*-hexanoic acid), 18:3/dn-OPDA-MGDG (Arabidopsis F, Monogalactosyldiacylglycerol), OPDA/dn-OPDA-MGDG (Arabidopsis A, Monogalactosyldiacylglycerol), OPDA/OPDA-MGDG (Arabidopsis B, Monogalactosyldiacylglycerol), OPDA/dn-OPDA-MGDG-OPDA (Arabidopsis E, Acylated Monogalactosyldiacylglycerol), OPDA/OPDA-MGDG-OPDA (Arabidopsis G, Acylated Monogalactosyldiacylglycerol), OPDA/dn-OPDA-DGDG (Arabidopsis C, Digalactosyldiacylglycerol), and OPDA/OPDA-DGDG (Arabidopsis D, Digalactosyldiacylglycerol).

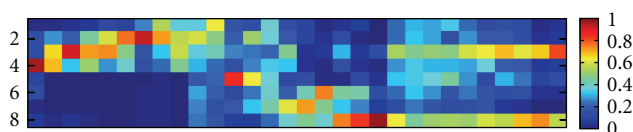


FIGURE 2: Prototype plot of the filtered and combined data in MarVis-Cluster using 30 prototypes for clustering. Every column represents the average intensity profile (prototype) of associated marker candidates. The prototypes are ordered according to similarity based on a one-dimensional self-organizing map. The first prototypes represent marker candidates in a WT-specific wound time course (high intensities in the first four conditions and almost no intensities in the last four conditions).

lipids, the mammalian and plant oxylipins are the best characterized ones. Mammals use predominantly C20 fatty acids (eicosanoids), while in plants C18 fatty acids are most abundantly used for the biosynthesis of oxylipins or so-called octadecanoids [35]. The identified oxylipins (see Table 4) are part of the α -linolenic acid metabolism or members of the compound class of mono- and digalactosyldiacylglycerols. They are described in the context of plant wounding [33, 34, 36]. Thirteen of the fifteen identified oxylipins could only be detected in either the negative or the positive ionization mode. On average, five ions/marker candidates could be

assigned per compound. The findings are supported by very low adjusted *P* values from the Kruskal-Wallis test of the intensity profiles (see previous section and Table 4).

4. Conclusions

MarVis-Filter combines essential preprocessing tools for mass spectrometry data analysis within a single user-friendly tool. Large data sets from the negative and positive ionization mode can easily be imported, corrected, filtered, and combined. Lists of ionization rules for adduct correction can be customized, extended, and commented in a convenient way using a standard text editor. Within the MarVis-Suite filtered and combined data sets can directly be clustered, visualized, and analyzed in detail using the MarVis-Cluster tool. In a case study 75 high-quality marker candidates could be clearly assigned to fifteen compounds of the oxylipin class based on the adduct and isotope correction in MarVis-Filter. The combination of data sets deriving from the negative and positive ionization mode is an important step for further data analysis. In the case study, most of the identified metabolites could only be detected in either the negative or the positive mode. The significance of the selected wound markers is supported by a high number of annotated and assigned ions/marker candidates and by very low adjusted *P* values from the Kruskal-Wallis test. The statistical filtering of

marker candidates reduced the complexity of the data sets from about 48000 to 3500 significant candidates (about 7 percent).

Acknowledgments

This work was partially supported by the DFG FOR-546-FE 446/2-3 to I. Feussner and by the Federal Ministry of Education and Research (BMBF 0315595A) to P. Meinicke and I. Feussner. A. Kaever and M. Landesfeind were supported by the Biomolecules program of the Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences (GGNB). The authors are grateful to Lars Söder for support in software development and testing, to Farina Schill for critical reading the manuscript and to Pia Meyer for expert technical assistance.

References

- [1] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. N. Trethewey, and L. Willmitzer, "Metabolite profiling for plant functional genomics," *Nature Biotechnology*, vol. 18, no. 11, pp. 1157–1161, 2000.
- [2] J. K. Nicholson, J. C. Lindon, and E. Holmes, "Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [3] J. Lisec, N. Schauer, J. Kopka, L. Willmitzer, and A. R. Fernie, "Gas chromatography mass spectrometry-based metabolite profiling in plants," *Nature Protocols*, vol. 1, no. 1, pp. 387–396, 2006.
- [4] R. C. H. De Vos, S. Moco, A. Lommen, J. J. B. Keurentjes, R. J. Bino, and R. D. Hall, "Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry," *Nature Protocols*, vol. 2, no. 4, pp. 778–791, 2007.
- [5] K. Dettmer, P. A. Aronov, and B. D. Hammock, "Mass spectrometry-based metabolomics," *Mass Spectrometry Reviews*, vol. 26, no. 1, pp. 51–78, 2007.
- [6] M. Katajamaa and M. Orešič, "Data processing for mass spectrometry-based metabolomics," *Journal of Chromatography A*, vol. 1158, no. 1–2, pp. 318–328, 2007.
- [7] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, and G. Siuzdak, "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification," *Analytical Chemistry*, vol. 78, no. 3, pp. 779–787, 2006.
- [8] P. Meinicke, T. Lingner, A. Kaever et al., "Metabolite-based clustering and visualization of mass spectrometry data using one-dimensional self-organizing maps," *Algorithms for Molecular Biology*, vol. 3, no. 1, article 9, 2008.
- [9] V. Shulaev, D. Cortes, G. Miller, and R. Mittler, "Metabolomics for plant stress response," *Physiologia Plantarum*, vol. 132, no. 2, pp. 199–208, 2008.
- [10] L. Tarpley, A. L. Duran, T. H. Kebrom, and L. W. Sumner, "Biomarker metabolites capturing the metabolite variance present in a rice plant developmental period," *BMC Plant Biology*, vol. 5, article 8, 2005.
- [11] K. Nahlik, M. Dumkow, Ö Bayram et al., "The COP9 signalosome mediates transcriptional and metabolic response to hormones, oxidative stress protection and cell wall rearrangement during fungal development," *Molecular Microbiology*, vol. 78, no. 4, pp. 964–979, 2010.
- [12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] M. Beckmann, D. P. Enot, D. P. Overy, and J. Draper, "Representation, comparison, and interpretation of metabolome fingerprint data for total composition analysis and quality trait investigation in potato cultivars," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 9, pp. 3444–3451, 2007.
- [14] J. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, CRC Press, 2003.
- [15] A. Koulman, B. A. Tapper, K. Fraser, M. Cao, G. A. Lane, and S. Rasmussen, "High-throughput direct-infusion ion trap mass spectrometry: a new method for metabolomics," *Rapid Communications in Mass Spectrometry*, vol. 21, no. 3, pp. 421–428, 2007.
- [16] D. A. MacKenzie, M. Defernez, W. B. Dunn et al., "Relatedness of medically important strains of *Saccharomyces cerevisiae* as revealed by phylogenetics and metabolomics," *Yeast*, vol. 25, no. 7, pp. 501–512, 2008.
- [17] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [18] P. Zhang, H. Foerster, C. P. Tissier et al., "MetaCyc and AraCyc. Metabolic pathway databases for plant research," *Plant Physiology*, vol. 138, no. 1, pp. 27–37, 2005.
- [19] M. Sud, E. Fahy, D. Cotter et al., "LMSD: LIPID MAPS structure database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D527–D532, 2007.
- [20] D. S. Wishart, C. Knox, A. C. Guo et al., "HMDB: a knowledgebase for the human metabolome," *Nucleic Acids Research*, vol. 37, supplement 1, pp. D603–D610, 2009.
- [21] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, "PubChem: a public information system for analyzing bioactivities of small molecules," *Nucleic Acids Research*, vol. 37, no. 2, pp. W623–W633, 2009.
- [22] J. Draper, D. P. Enot, D. Parker et al., "Metabolite signal identification in accurate mass metabolomics data with MZedDB, an interactive m/z annotation tool utilising predicted ionisation behaviour 'rules'," *BMC Bioinformatics*, vol. 10, article 227, 2009.
- [23] R. Tautenhahn, C. Böttcher, and S. Neumann, "Annotation of LC/ESI-MS mass signals," in *1st International Conference on Bioinformatics Research and Development (BIRD '07)*, vol. 4414 of *Lecture Notes in Computer Science*, pp. 371–380, Berlin, Germany, March 2007.
- [24] T. Kind and O. Fiehn, "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry," *BMC Bioinformatics*, vol. 8, article 105, 2007.
- [25] R. Gentleman, R. Ihaka et al., <http://www.r-project.org/>.
- [26] B. Jones, *MATLAB: Statistics Toolbox User's Guide*, MathWorks, 1993.
- [27] A. Lommen, "Metalign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing," *Analytical Chemistry*, vol. 81, no. 8, pp. 3079–3086, 2009.
- [28] M. Sturm, A. Bertsch, C. Gröpl et al., "OpenMS—an open-source software framework for mass spectrometry," *BMC Bioinformatics*, vol. 9, article 163, 2008.
- [29] A. Kaever, T. Lingner, K. Feussner, C. Göbel, I. Feussner, and P. Meinicke, "MarVis: a tool for clustering and visualization of metabolic biomarkers," *BMC Bioinformatics*, vol. 10, article 92, 2009.

- [30] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.
- [31] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B*, vol. 57, pp. 289–300, 1995.
- [32] B. Von Malek, E. Van Der Graaff, K. Schneitz, and B. Keller, "The Arabidopsis male-sterile mutant dde2-2 is defective in the ALLENE OXIDE SYNTHASE gene encoding one of the key enzymes of the jasmonic acid biosynthesis pathway," *Planta*, vol. 216, no. 1, pp. 187–192, 2002.
- [33] C. Göbel and I. Feussner, "Methods for the analysis of oxylipins in plants," *Phytochemistry*, vol. 70, no. 13-14, pp. 1485–1503, 2009.
- [34] A. Ibrahim, A. Schütz, J. Galano et al., "The alphabet of galactolipids in Arabidopsis thaliana," *Frontiers in Plant Physiology*, vol. 2, article 95, 2011.
- [35] A. Andreou, F. Brodhun, and I. Feussner, "Biosynthesis of oxylipins in non-mammals," *Progress in Lipid Research*, vol. 48, no. 3-4, pp. 148–170, 2009.
- [36] G. A. Howe and G. Jander, "Plant immunity to insect herbivores," *Annual Review of Plant Biology*, vol. 59, pp. 41–66, 2008.