**BMC
Bioinformatics**

SOFTWARE

Open Access

AMDORAP: Non-targeted metabolic profiling based on high-resolution LC-MS

Hiroki Takahashi[1], Takuya Morimoto[1,2], Naotake Ogasawara[1] and Shigehiko Kanaya[1*]

Abstract

Background: Liquid chromatography-mass spectrometry (LC-MS) utilizing the high-resolution power of an orbitrap is an important analytical technique for both metabolomics and proteomics. Most important feature of the orbitrap is excellent mass accuracy. Thus, it is necessary to convert raw data to accurate and reliable *m/z* values for metabolic fingerprinting by high-resolution LC-MS.

Results: In the present study, we developed a novel, easy-to-use and straightforward *m/z* detection method, AMDORAP. For assessing the performance, we used real biological samples, *Bacillus subtilis* strains 168 and MGB874, in the positive mode by LC-orbitrap. For 14 identified compounds by measuring the authentic compounds, we compared obtained *m/z* values with other LC-MS processing tools. The errors by AMDORAP were distributed within ±3 ppm and showed the best performance in *m/z* value accuracy.

Conclusions: Our method can detect *m/z* values of biological samples much more accurately than other LC-MS analysis tools. AMDORAP allows us to address the relationships between biological effects and cellular metabolites based on accurate *m/z* values. Obtaining the accurate *m/z* values from raw data should be indispensable as a starting point for comparative LC-orbitrap analysis. AMDORAP is freely available under an open-source license at http://amdorap.sourceforge.net/.

Background

Metabolomics is defined as technology designed to give us the broadest, least biased insight into the richly diverse population of small molecules present in living things [1]. Understanding cells at the levels of the transcriptome and metabolome provides insight into the network of complex biological regulations [2-5]. Metabolites within cells have the diverse range of chemical and physical properties and the wide range of those concentrations [6]. To achieve metabolomics, two analytical platforms, i.e., mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR), have been widely used [7,8]. Chromatography-MS technologies play a central role in measuring the complex biological samples. Out of these, liquid chromatography-MS (LC-MS) is capable of detecting a broader range of metabolites than other MS technologies such as gas chromatography-MS and capillary electrophoresis-MS [9]. Therefore, LC-MS has become more widely used in metabolomics analysis. An orbitrap mass analyzer is the most recent addition to the set of tools that can be applied to identification, characterization and quantitation of components in biological systems since its commercial introduction in 2005 [10]. Orbitrap-based MSs have been proven to be a powerful tool in proteomics because they have ≈100 000 resolving power at a mass-to-charge ratio (*m/z*) 400 [11,12]. The most important feature of the orbitrap is that it can stably maintain excellent mass accuracy without re-calibration, and does not require the use of calibration standards [13]. Accurate *m/z* values can be used to define molecular formulae in putative identification of metabolites [7,14]. Consequently, in the field of non-targeted metabolomics, those instruments make it possible to identify candidate molecular formulae from mass differences in measured *m/z* values [15,16].

Public databases of chemical compounds such as ChEBI [17], HMDB [18], KEGG [19], KNApSAcK [20] and PubChem [21] provide suitable compounds for each molecular formula without measuring reference samples in advance. The species-metabolite relationship database

* Correspondence: skanaya@gtc.naist.jp
[1]Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
Full list of author information is available at the end of the article

KNApSAcK, for example, can easily narrow down candidates from accurate masses according to the species information or the type of ion adducts [22,23]. Several molecular ion adducts should be considered especially when the ionization of molecules in samples is performed by electrospray ionization [24,25]. Once given, the accurate m/z values can lead to the information of molecular formulae and candidate compounds by considering the mass differences, the appropriate ion adduct and the species together. However, it should be noted that structural isomers and stereoisomers with the same mass require the complicated chromatographic separation before mass analyzing [7].

Allen et al. [26] analyzed several "silent mutants" of yeasts (viable mutants with no obvious phenotype) by comparing extracellular metabolites using LC-MS data collected in a non-targeted approach. In preprocessing the LC-MS data, they skipped peak detection and annotation schemes typically used for such data; instead, they reduced data into a single aggregate MS vector and applied clustering and machine learning methods. Their study demonstrated the effectiveness of metabolic fingerprinting of extracellular extracts by non-complicated preprocessed data. Metabolic fingerprinting with the exclusion of m/z resolution, however, is impossible to get more insight from same data sets. The high-resolution of the orbitrap can be exploited in metabolic fingerprinting. In NMR or Fourier transform ion cyclotron resonance based MS (FT-ICR-MS), valid information about metabolic regulation in biological samples can be obtained by resolving power alone, even without any chromatographic separation [27].

An easy-to-use, flexible and automated tool is a key to success in metabolomics studies. This is particularly the case in high-resolution MS analyses mainly because of the data size. Our aim is to estimate more accurate m/z values and extract interesting m/z values from raw data in comparative LC-orbitrap analysis. In the present study, we describe a novel straightforward m/z detection method, "AMDORAP" (**A**ccurate **m**/z **d**etection method for LC-**o**rbit**rap**) for high-resolution MS (e.g., the orbitrap) by taking advantage of its stable mass accuracy.

Implementation
Several freely available frameworks for analyzing LC-MS data sets have been developed [28]. The typical MS data processing workflow comprises multiple stages, including filtering, feature detection, alignment and normalization. In MZmine 2 [29,30], peak alignment across samples, for example, follows peak detection for individual samples. The Bioconductor package XCMS [31,32] mainly consists of peak detection, peak matching and retention time alignment. A common concept shared by widely used methods, including MZmine 2 and XCMS, is that peak detection step for both m/z and retention time dimensions is executed for an individual sample, or scan, followed by an alignment (or merging) step across samples. The most important reason for using high-resolution MS is to obtain more accurate m/z values from biological samples. That makes it possible to identify correct candidate molecular formulae from mass differences alone. Since the orbitrap can determine m/z values extremely accurately, we assumed that m/z values derived from compounds with the same compositional formula, including structural isomers and stereoisomers, should be robust with respect to retention time and differences between samples.

In this study, we developed the preprocessing method, AMDORAP (**A**ccurate **m**/z **d**etection method for LC-**o**rbit**rap**) written in the R programming language [33] in order to attain the quick comparison of metabolic profiling by high-resolution MS. Figure 1 illustrates the AMDORAP procedure, which comprises three steps:
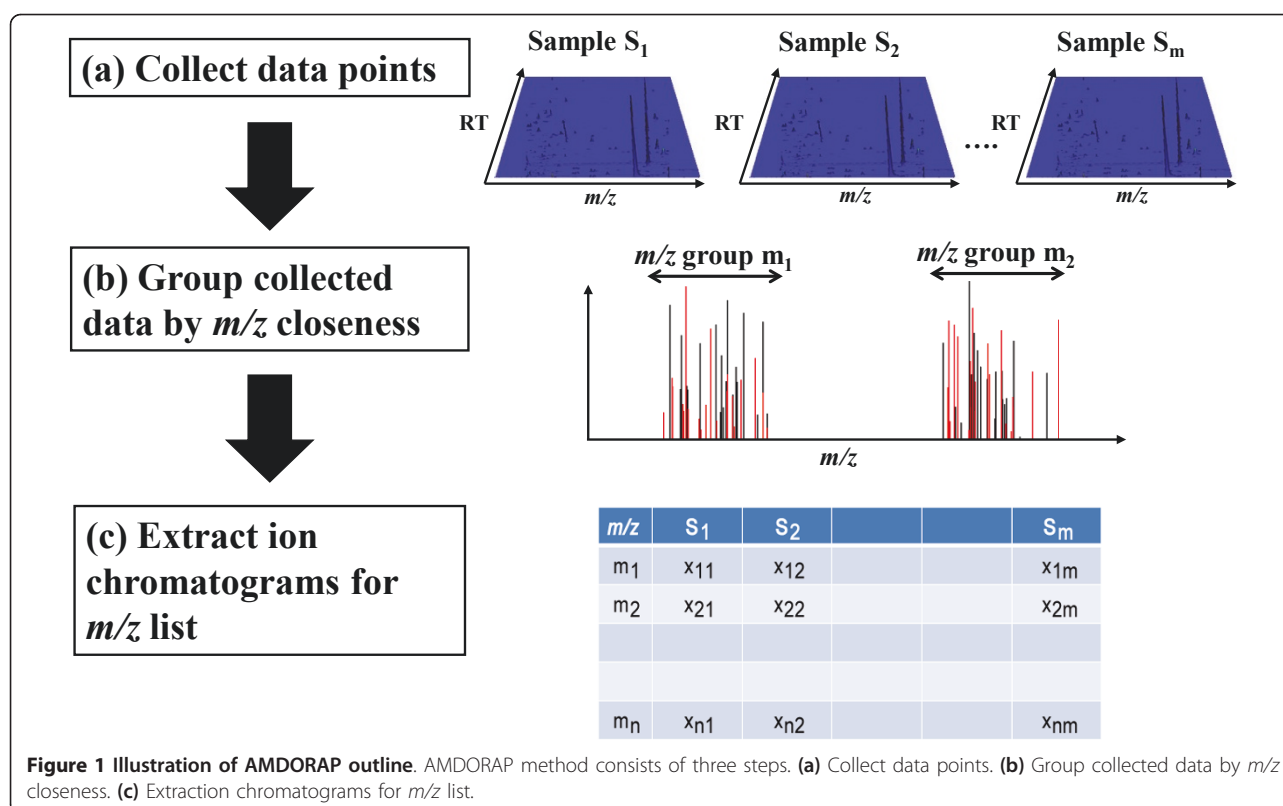
1. Collect data points with intensities larger than a threshold for all samples.
2. Group collected data points by m/z closeness, and estimate representative m/z values for individual m/z groups.
3. Extract ion chromatograms for the m/z list.

The main idea motivating this procedure is that peak picking and alignment steps of m/z values should be performed in a single step. In the following section, the AMDORAP performance was assessed using data sets in the positive mode from two *Bacillus subtilis* strains 168 and MGB874 [34].

Results and Discussion
Sample preparation and experimental conditions
In order to assess the AMDORAP performance, we performed the experiments and then prepared the biological data sets. Two *Bacillus subtilis* strains, wild-type 168 and the genome reduced strain MGB874 [34], were used for metabolome analysis. The cells were cultured at 37°C to an OD_{600} value of 4.0 in the early stationary phase of growth, in Spizizen's minimal medium (SMM) [35] supplemented with 0.5% glucose, 5 μg/ml tryptophan, 20 μg/ml methionine and trace elements [36]. Metabolite extraction was performed according to Takahashi et al. [23]. The culture media were passed through a 0.4 μm HTTP filter (Millipore). Residual cells on the filter were washed twice with HPLC grade water and then immersed in 2 ml of methanol. After incubation at 4°C overnight, the extracts were centrifuged at 9000 × g at 4°C for 10 min, filtered through 0.2 μm PTFE membrane (Advantec), evaporated

Figure 1 Illustration of AMDORAP outline. AMDORAP method consists of three steps. **(a)** Collect data points. **(b)** Group collected data by *m/z* closeness. **(c)** Extraction chromatograms for *m/z* list.

at room temperature and stored at -80°C. The extracts were dissolved in 200 μl of 80% methanol before analysis in the LC-orbitrap.

Mass analysis was performed using a Paradigm MS4 system (Michrome BioResources) coupled to an LTQ-orbitrap XL-HTC-PAL system (Thermo Fisher Scientific). All experimental events were controlled by Xcalibur software version 2.0.7 (Thermo Fisher Scientific). HPLC was performed under the conditions as described by Iijima et al. [37]. Samples were injected into to a TSKgel column ODS-100V (4.6 × 250 mm, 5 μm; TOSOH). Water (HPLC grade; solvent A) and acetonitrile (HPLC grade; solvent B) were used as the mobile phase with 0.1% v/v formic acid. The gradient program was as follows: 3% B to 97% B (45 min), 97% B (5 min) and 10% B (10 min). The flow rate was set to 0.5 ml/min. The ESI setting was as follows: spray voltage 4.5 kV and capillary temperature 350°C for the positive ionization mode. Nitrogen sheath gas and auxiliary gas were set at 60 and 20 arbitrary units, respectively. A full MS scan was performed in the *m/z* range 70-1500 at a resolution of 60 000. Simultaneously, top three MS^2 spectra within each full MS scan were gained by the linear ion trap at a collision energy of 35 eV. Thermo Fisher mass spectrometry RAW files were converted from profile mode into centroid mode using the ReAdW program [38].

AMDORAP performance
Collection of data points
Figure 2b shows the intensity distribution of a centroid data from *B. subtilis* strain 168. The total number of data points was 1 694 959 (1945 scans within 45 minutes). The top 1% of the data (represented by a red dot in Figure 2a) could explain 99.7% of the total variance of all data points. Thus, almost all data obtained by the LC-orbitrap can be considered as background noise. Here, we assumed that the top 1% of the data was detected ions, and the other 99% was noise for each sample in the collecting step. Figure 3 shows the total ion chromatograms and two dimensional map. The total ion chromatogram of top 1% of data highly correlates with that of all data (in Figure 3a) and then top 1% of data is extensively scattered in both dimensions (in Figure 3b), suggesting that top 1% of data can explain the characterization of all data with respect to intensities and dimensions.

Grouping collected data points and estimation of representative m/z values for individual groups
As the second step, all collected *m/z* values are grouped by closeness, i.e., if differences between the neighboring *m/z* values are within 5 ppm (default setting), they are grouped together. There is no limit of data points within one *m/z* group as long as this constrain is fulfilled. Out of the *m/z* alignment methods, Kazmi et al. [39]
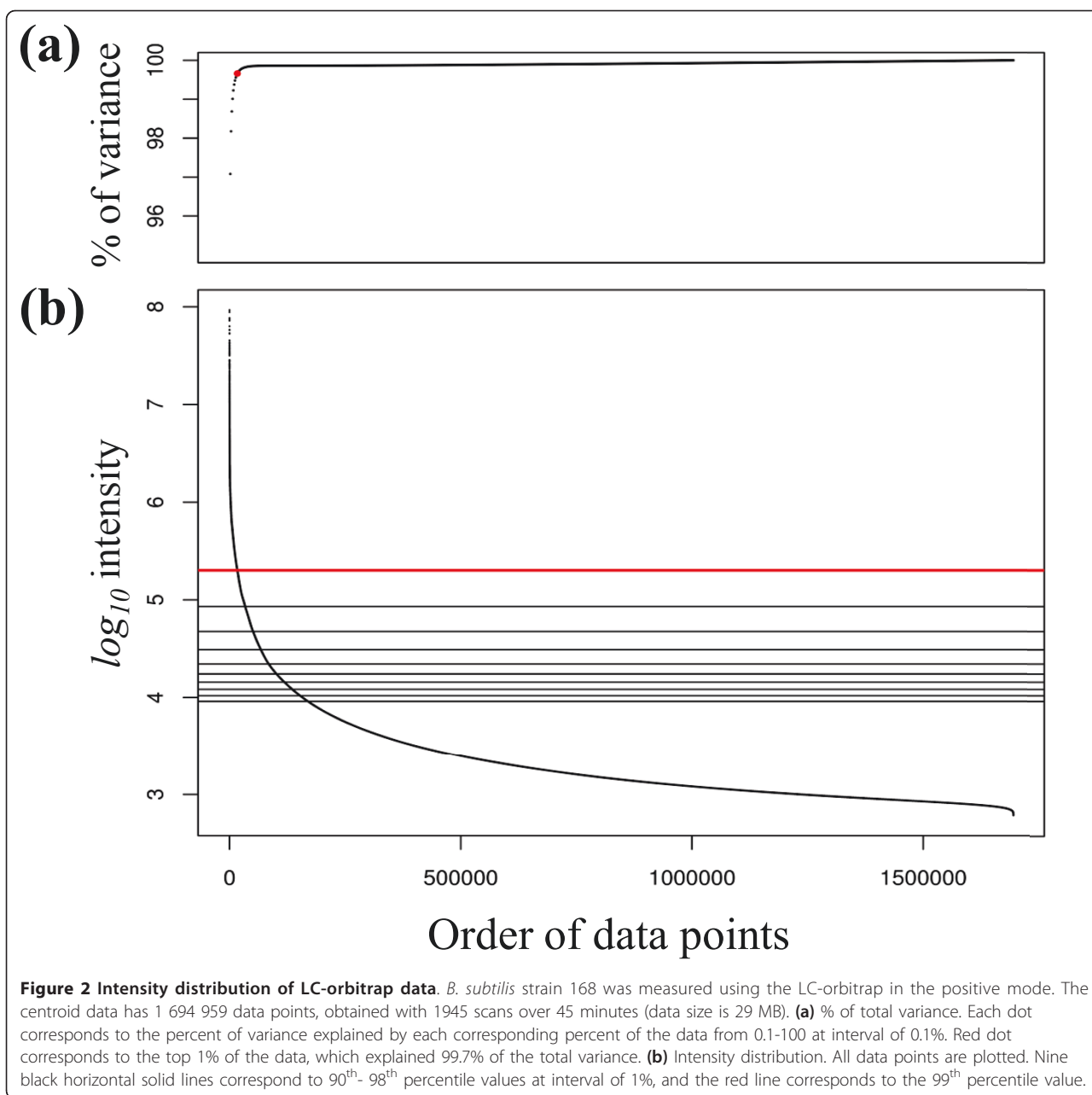
Figure 2 Intensity distribution of LC-orbitrap data. *B. subtilis* strain 168 was measured using the LC-orbitrap in the positive mode. The centroid data has 1 694 959 data points, obtained with 1945 scans over 45 minutes (data size is 29 MB). **(a)** % of total variance. Each dot corresponds to the percent of variance explained by each corresponding percent of the data from 0.1-100 at interval of 0.1%. Red dot corresponds to the top 1% of the data, which explained 99.7% of the total variance. **(b)** Intensity distribution. All data points are plotted. Nine black horizontal solid lines correspond to 90th- 98th percentile values at interval of 1%, and the red line corresponds to the 99th percentile value.

developed the method to create bins and then combine consecutive bins together according to the constrains, similar to complete linkage hierarchical clustering. While they must consider the origins of *m/z* values, our method is to collect all data points with relatively higher intensities and then deal with collected data as one spectrum. Consequently, the grouping of *m/z* values is feasible in one step.

Median *m/z* values of individual *m/z* groups are defined as the peak values among all samples. Figure 4 shows the relationship between closeness and the number of *m/z* groups by simultaneously using two data sets.

In case of closeness 5 ppm (default setting) for the top 1 and 5% of data points, 624 (black dots in Figure 4) and 2821 (red dots) *m/z* groups were obtained, respectively. According to Werf et al. [40], the *in silico* metabolome of *B. subtilis* is covered by 537 compounds. Of those, 282 compounds are commercially available. Other compounds can not be identified by the method of measuring authentic compounds. Additionally, Pluskal et al. [41] and Iijima et al. [37], for example, identified 123 metabolites from approximately 1900 peaks in yeast and at most 29 metabolites identified by comparison with authentic compounds (they called grade A) from ~4700 peaks in
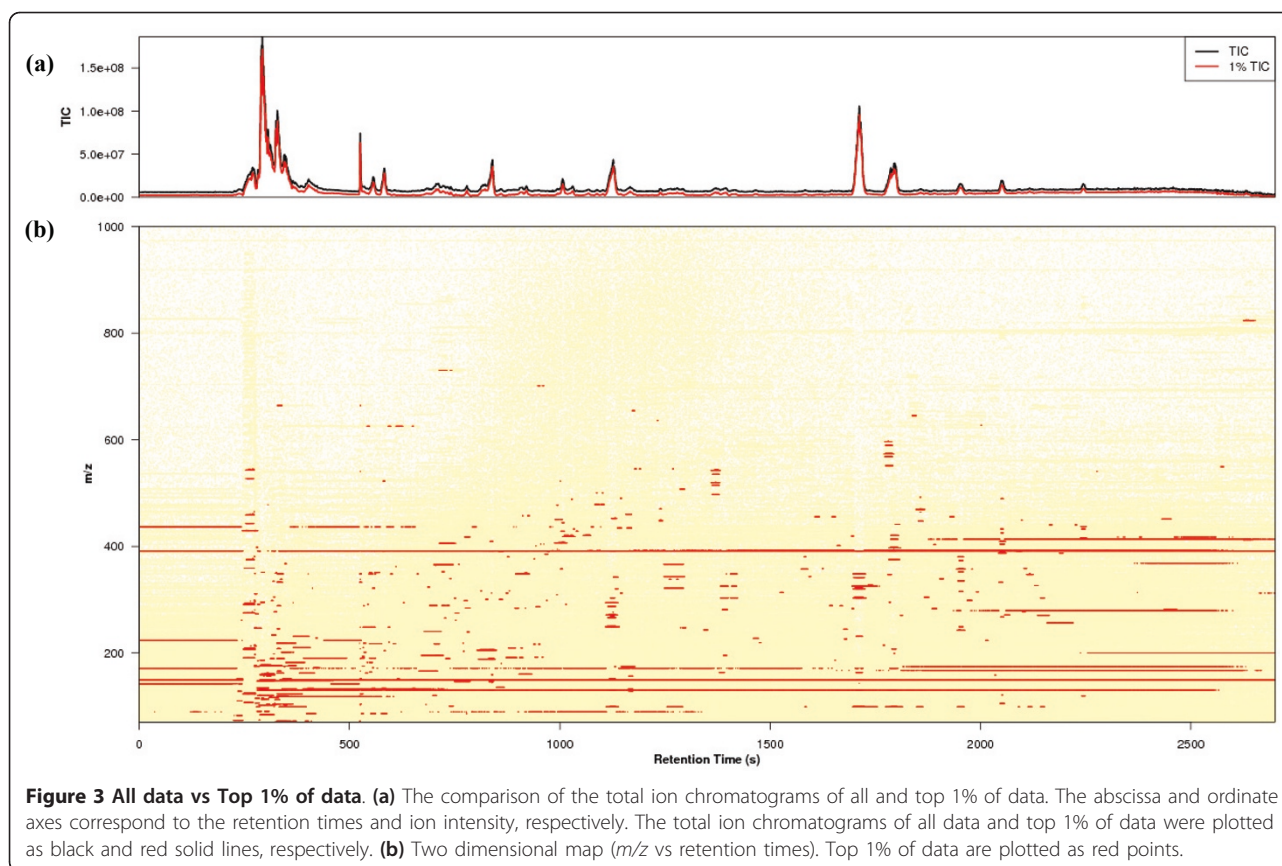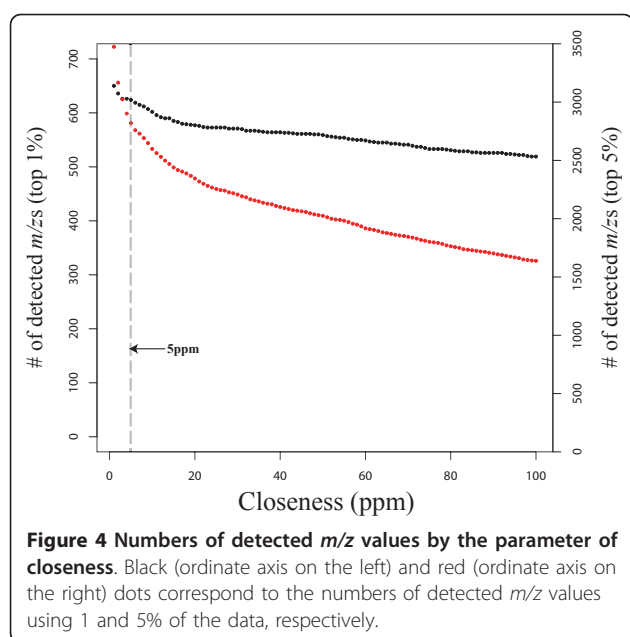
Figure 3 All data vs Top 1% of data. **(a)** The comparison of the total ion chromatograms of all and top 1% of data. The abscissa and ordinate axes correspond to the retention times and ion intensity, respectively. The total ion chromatograms of all data and top 1% of data were plotted as black and red solid lines, respectively. **(b)** Two dimensional map (*m/z* vs retention times). Top 1% of data are plotted as red points.

tomato, respectively. Those studies indicate that most of obtained peaks from LC-MS data would remain unknown even after peak detection. We concluded that 624 *m/z* groups could be sufficient to express the cell state as starting point for LC-orbitrap analysis.

For identification of the ions by MS^2 data, we made an in-house database for *B. subtilis* compounds by using KEGG database. All reactions associated with *B. subtilis* were extracted and 890 compounds were set to be as the database (Additional file 1). After database search ([M +H]$^+$) within ±5 ppm for MS^2 precursor *m/zs* in two *B. subtilis* data, 20 available authentic compounds (Additional file 2) were measured under the same conditions for *B. subtilis* strains. Out of limited MS^2 spectra in *B. subtilis* samples, 14 compounds were manually identified by measuring the authentic compounds. Next, we performed a comparison study for *m/z* accuracy between AMDORAP, MZmine 2 and XCMS. The steps including Chromatogram builder (*m/z* tolerance = 0.01), RANSAC aligner and Peak finder were performed by MZmine 2. In XCMS parameters for UPLC-orbitrap data, Dunn et al. [6] showed that two parameters, *snthresh* and *bw*, significantly affected the processed data, e.g., the number of peaks detected and the peak area reproducibility. For XCMS, the parameters were set to be "centWave", *bw* =

60, *snthresh* = 2, *ppm* = 3 and *mzwid* = 0.02 with all other default settings. Table 1 summarizes the comparisons of observed *m/z* values associated with 14 identified compounds. Seven *m/z* values obtained by AMDORAP were closest to the theoretical masses. While all errors of observed *m/z* values in AMDORAP were distributed within ±3 ppm, some errors in MZmine 2 and XCMS were over ±100 ppm, e.g., tryptophan, uridine and glutamine, suggesting that our procedure can detect more accurate *m/z* values than others. In the case of other parameter settings for XCMS, a few compounds were not detected (data not shown). In compound searches using mass differences alone, *m/z* values with errors over ±100 ppm could be no longer correctly annotated by leveraging the high-resolution power of the orbitrap. This comparison shows that our method has the best performance in detecting accurate *m/z* values and can allow us to identify correct candidate compounds by mass differences alone. According to Goerlach et al. [25], 30 and 14 different types of molecular ion adducts exist in the positive and negative modes, respectively. Furthermore, structural isomers and stereoisomers have the same mass. Therefore, it should be noted that putative identification of metabolites based on the accurate *m/z* values is carefully performed to avoid the misleading results.

Figure 4 Numbers of detected *m/z* values by the parameter of closeness. Black (ordinate axis on the left) and red (ordinate axis on the right) dots correspond to the numbers of detected *m/z* values using 1 and 5% of the data, respectively.

Extraction of ion chromatograms for the m/z list

The final step is to extract ion chromatograms for the *m/z* list within ±5 ppm (default setting). AMDORAP provides two types of representative values for detected *m/z* values. One is the sum of total ion chromatogram and another is the sum of selected peak area by a signal-to-noise ratio cutoff for Gaussian filtered chromatogram [42]. Of 624 *m/z* values, 603 reliable chromatograms were extracted by manually checking. We judged the chromatograms with noisy baseline or stretched across the experimental time, i.e., 45 min, as unreliable chromatograms in this study.

Additional file 3 shows 21 extracted ion chromatograms judged to be unreliable chromatograms. The numbers of chromatograms with only one peak through 45 minutes, were 471 (79%) and 453 (75%) in *B. subtilis* strains 168 and MGB874, respectively; the numbers of chromatograms with two peaks were 86 (14%) and 113 (18%). As showing in Figure 5, two peaks were seen in a chromatogram of phenylalanine; this phenomenon was confirmed under our experimental conditions by measuring the authentic phenylalanine, indicating that some of the chromatograms with two peaks originate from unique compounds. Those results suggest that to separate the peaks by the retention time could mislead the identification of the ions and clues about the chemical structures corresponding to those peaks could be obtained without separating chromatograms by the retention time. Hence, almost all chromatograms based on AMDORAP could be identified as unique compounds even without separation of identical *m/z* peaks by the retention time. Taken together, the reliable *m/z* grouping process is sufficient for comparison of metabolic fingerprinting based on high-resolution LC-MSs.

Conclusions

In metabolic profiling by the high-resolution mass technologies, it is important to convert raw data to reliable *m/z* values in order to quickly get the information of correct candidate metabolites in biological samples. With respect to obtained *m/z* accuracy, comparison study was performed using only 14 identified compounds. Clearly, the *m/z* errors by AMDORAP are smallest, although the number of compared compounds might be not enough. In the range of tested parameters,

Table 1 Comparison of detected *m/z* values for fourteen compounds by AMDORAP, MZmine 2 and XCMS

metabolite	formula	theoretical [M + H]$^+$	AMDORAP observed m/z	error ppm	MZmine 2 observed m/z	error ppm	XCMS Observed m/z	error ppm
serine	$C_3H_7NO_3$	106.04987	**106.04960**	**-2.55**	106.04955	-3.02	106.04958	-2.68
valine	$C_5H_{11}NO_2$	118.08626	**118.08604**	**-1.78**	118.08602	-2.01	118.08422	-17.26
glutamine	$C_5H_{10}N_2O_3$	147.07642	147.07666	1.64	**147.07642**	**0.03**	147.14403	459.68
lysine	$C_6H_{14}N_2O_2$	147.11280	**147.11277**	**-0.23**	147.11067	-14.49	147.10871	-27.82
glutamic acid	$C_5H_9NO_4$	148.06043	148.06055	0.76	**148.06042**	**-0.09**	148.05154	-60.10
methionine	$C_5H_{11}NO_2S$	150.05833	150.05817	-1.06	**150.05818**	**-0.96**	150.04993	-55.93
D-alanyl-D-alanine	$C_6H_{12}N_2O_3$	161.09207	161.09181	-1.59	**161.09197**	**-0.60**	161.08698	-31.61
phenylalanine	$C_9H_{11}NO_2$	166.08626	166.08617	-0.53	**166.08618**	**-0.44**	166.08285	-20.50
citrulline	$C_6H_{13}N_3O_3$	176.10297	176.10280	-0.96	**176.10281**	**-0.87**	176.11662	77.51
tyrosine	$C_9H_{11}NO_3$	182.08117	182.08113	-0.21	182.08110	-0.41	**182.08119**	**0.09**
tryptophan	$C_{11}H_{12}N_2O_2$	205.09715	**205.09732**	**0.81**	205.06284	-167.29	205.09010	-34.38
pantothenate	$C_9H_{17}NO_5$	220.11795	**220.11798**	**0.14**	220.11414	-17.29	220.12327	24.18
uridine	$C_9H_{12}N_2O_6$	245.07681	**245.07692**	**0.44**	245.10884	130.69	245.08308	25.58
methylthioadenosine	$C_{11}H_{15}N_5O_3S$	298.09684	**298.09708**	**0.80**	298.08231	-48.72	298.09719	1.18

Comparison of detected *m/z* values between AMDORAP, MZmine 2 and XCMS. Seven *m/z* values obtained by AMDORAP, i.e., serine, valine, lysine, tryptophan, pantothenate, uridine and methylthioadenosine, were closest to the theoretical masses.
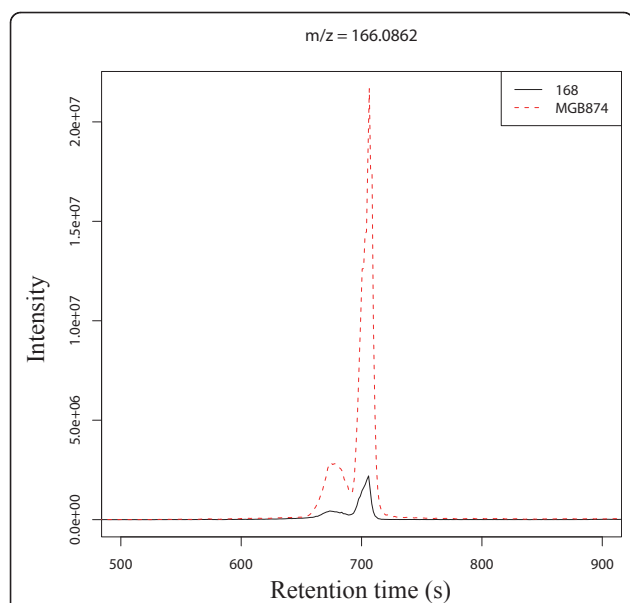
Figure 5 A chromatogram trace of phenylalanine. Chromatogram trace of *m/z* slice 166.08534-166.08700 is presented. Black solid and red dashed lines correspond to *B. subtilis* strains 168 and MGB874, respectively. Two peak areas were also observed for an authentic compound under our conditions.

we couldn't get better results for 14 compounds by MZmine 2 and XCMS. This suggests that parameter optimization of those tools is time consuming process and difficult to find out best settings for both dimensions, i.e., *m/z* and retention time. Furthermore, it would suggest that both mass and retention time alignment processes introduce the larger errors for obtained *m/z* values, while AMDORAP uses only the ions with relatively higher intensities for estimating the *m/z* values. In addition, a signal-to-noise ratio cutoff by Gaussian filtering could allow us to achieve a reliable comparison of the ion abundances between samples, even when there are peaks with noisy baseline. Thus, AMDORAP can detect more accurate *m/z* values from raw data and provide the platform for metabolic fingerprinting. Information of MS^n, retention time and behaviors of the authentic compounds has the essential roles to finally verify the ions as particular metabolites. However, the extraction of interesting accurate *m/z* values by AMDORAP should be indispensable as a starting point for comparative LC-orbitrap analysis, because of the limitations of available authentic compounds and simultaneously obtained MS^2 spectra with a full MS scan per sample.

Availability and requirements
Project name: AMDORAP
  Project home page: http://amdorap.sourceforge.net/
  Operating systems: Platform independent

Programming language: R
License: GPL v2
Any restrictions to use by non-academics: No

Additional material

Additional file 1: A list of 890 compounds. This list contains 890 compounds associated with all reactions in *B. subtilis* of KEGG.

Additional file 2: A list of 20 authentic compounds. These compounds were measured by LC-orbitrap. Obtained information (MS^2 and retention time) were used to identify the compounds.

Additional file 3: 21 unreliable chromatograms. 21 extracted ion chromatograms judged to be unreliable chromatograms are shown. The abscissa and ordinate axes correspond to the retention times and ion intensity, respectively.

Author details
[1]Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan. [2]Biological Science Laboratories, Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi 321-3497, Japan.

Authors' contributions
HT performed experimental parts, implemented the methods, analyzed data sets and wrote the manuscript. TM performed extraction of metabolites. NO and SK supervised this project. All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Hall RD: **Plant metabolomics: from holistic hope, to hype, to hot topic.** *New Phytol* 2006, **169(3)**:453-468.
2. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L: **Metabolite profiling for plant functional genomics.** *Nat Biotechnol* 2000, **18(11)**:1157-1161.
3. Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K: **Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana.** *Proc Natl Acad Sci USA* 2004, **101(27)**:10205-10210.
4. Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K: **Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics.** *J Biol Chem* 2005, **280(27)**:25590-25595.
5. Takahashi H, Morioka R, Ito R, Oshima T, Altaf-Ul-Amin M, Ogasawara N, Kanaya S: **Dynamics of Time-Lagged Gene-to-Metabolite Networks of Escherichia coli Elucidated by Integrative Omics Approach.** *OMICS* 2011, **15(1-2)**:15-23.

6. Dunn WB, Broadhurst D, Brown M, Baker PN, Redman CW, Kenny LC, Kell DB: **Metabolic profiling of serum using Ultra Performance Liquid Chromatography and the LTQ-Orbitrap mass spectrometry system.** *J Chromatogr B Analyt Technol Biomed Life Sci* 2008, **871(2)**:288-298.

7. Brown M, Wedge DC, Goodacre R, Kell DB, Baker PN, Kenny LC, Mamas MA, Neyses L, Dunn WB: **Automated work flows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets.** *Bioinformatics* 2011, **27(8)**:1108-1112.

8. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL: **Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy.** *Chem Soc Rev* 2011, **40**:387-426.

9. Verhoeven HA, de Vos CHR, Bino RJ, Hall RD: **Plant metabolomics strategies based upon Quadruple Time of Flight Mass Spectrometry (QTOF-MS).** In *Plant Metabolomics. Biotechnology in agriculture and forestry. Volume 57.* Springer; 2006:33-48.

10. Scigelova M, Makarov A: **Orbitrap mass analyzer-overview and applications in proteomics.** *Proteomics* 2006, **6(Suppl 2)**:16-21.

11. Han X, Aslanian A, Yates JRr: **Mass spectrometry for proteomics.** *Curr Opin Chem Biol* 2008, **12(5)**:483-490.

12. Perry RH, Cooks RG, Noll RJ: **Orbitrap mass spectrometry: instrumentation, ion motion and applications.** *Mass Spectrom Rev* 2008, **27(6)**:661-699.

13. Zhang NR, Yu S, Tiller P, Yeh S, Mahan E, Emary WB: **Quantitation of small molecules using high-resolution accurate mass spectrometers - a different approach for analysis of biological samples.** *Rapid Commun Mass Spectrom* 2009, **23(7)**:1085-1094.

14. Marshall AG, Hendrickson CL: **High-resolution mass spectrometers.** *Annu Rev Anal Chem (Palo Alto Calif)* 2008, **1**:579-599.

15. Breitling R, Pitt AR, Barrett MP: **Precision mapping of the metabolome.** *Trends Biotechnol* 2006, **24(12)**:543-548.

16. Lu W, Clasquin MF, Melamud E, Amador-Noguez D, Caudy AA, Rabinowitz JD: **Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer.** *Anal Chem* 2010, **82(8)**:3212-3221.

17. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcantara R, Darsow M, Guedj M, Ashburner M: **ChEBI: a database and ontology for chemical entities of biological interest.** *Nucleic Acids Res* 2008, , **36 Database**:D344-D350.

18. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L: **HMDB: the Human Metabolome Database.** *Nucleic Acids Res* 2007, , **35 Database**:D521-D526.

19. KEGG. [http://www.genome.jp/KEGG/].

20. Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, Asahi H, Kurokawa K, Arita M, Saito K, Ohta D, Shibata D, Kanaya S: **KNApSAcK: A Comprehensive Species-Metabolite Relationship Database.** In *Plant Metabolomics. Biotechnology in agriculture and forestry. Volume 57.* Springer; 2006:165-181.

21. PubChem. [http://pubchem.ncbi.nlm.nih.gov/].

22. Oikawa A, Nakamura Y, Ogura T, Kimura A, Suzuki H, Sakurai N, Shinbo Y, Shibata D, Kanaya S, Ohta D: **Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies.** *Plant Physiol* 2006, **142(2)**:398-413.

23. Takahashi H, Kai K, Shinbo Y, Tanaka K, Ohta D, Oshima T, Altaf-Ul-Amin M, Kurokawa K, Ogasawara N, Kanaya S: **Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry.** *Anal Bioanal Chem* 2008, **391(8)**:2769-2782.

24. Huanga N, Siegela MM, Kruppab HG, Laukienb HF: **Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data.** *J Am Soc Mass Spectr* 1999, **10(11)**:1166-1173.

25. Goerlach E, Richmond R: **Discovery of Quasi-Molecular Ions in Electrospray Spectra by Automated Searching for Simultaneous Adduct Mass Differences.** *Anal Chem* 1999, **71(24)**:5557-5562.

26. Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB: **High-throughput classification of yeast mutants for functional genomics using metabolic footprinting.** *Nat Biotechnol* 2003, **21(6)**:692-696.

27. Fiehn O: **Metabolomics-the link between genotypes and phenotypes.** *Plant Mol Biol* 2002, **48(1-2)**:155-171.

28. Katajamaa M, Oresic M: **Data processing for mass spectrometry-based metabolomics.** *J Chromatogr A* 2007, **1158(1-2)**:318-328.

29. Katajamaa M, Oresic M: **Processing methods for differential analysis of LC/MS profile data.** *BMC Bioinformatics* 2005, **6**:179.

30. Pluskal T, Castillo S, Villar-Briones A, Oresic M: **MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data.** *BMC Bioinformatics* 2010, **11**:395.

31. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G: **XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification.** *Anal Chem* 2006, **78(3)**:779-787.

32. Tautenhahn R, Bottcher C, Neumann S: **Highly sensitive feature detection for high resolution LC/MS.** *BMC Bioinformatics* 2008, **9**:504.

33. R Development Core Team: *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing, Vienna, Austria; 2008 [http://www.R-project.org], [ISBN 3-900051-07-0].

34. Morimoto T, Kadoya R, Endo K, Tohata M, Sawada K, Liu S, Ozawa T, Kodama T, Kakeshita H, Kageyama Y, Manabe K, Kanaya S, Ara K, Ozaki K, Ogasawara N: **Enhanced recombinant protein productivity by genome reduction in Bacillus subtilis.** *DNA Res* 2008, **15(2)**:73-81.

35. Anagnostopoulos C, Spizizen J: **Requirements for Transformation In Bacillus Subtilis.** *J Bacteriol* 1961, **81(5)**:741-746.

36. Harwood CR, Archibald AR: **Growth, maintenance and general techniques.** In *Molecular Biological Methods for Bacillus.* Edited by: Harwood CR, Cutting SM. John Wiley 1990:549.

37. Iijima Y, Nakamura Y, Ogata Y, Tanaka K, Sakurai N, Suda K, Suzuki T, Suzuki H, Okazaki K, Kitayama M, Kanaya S, Aoki K, Shibata D: **Metabolite annotations based on the integration of mass spectral information.** *Plant J* 2008, **54(5)**:949-962.

38. Keller A, Eng J, Zhang N, Li XJ, Aebersold R: **A uniform proteomics MS/MS analysis platform utilizing open XML file formats.** *Mol Syst Biol* 2005, **1**, 2005.0017.

39. Kazmi AS, Ghosh S, Shin DG, Hill WD, Grant FD: **Alignment of high resolution mass spectra: development of a heuristic approach for metabolomics.** *Metabolomics* 2006, **2(2)**:75-83.

40. van der Werf MJ, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T: **Microbial metabolomics: toward a platform with full metabolome coverage.** *Anal Biochem* 2007, **370**:17-25.

41. Pluskal T, Nakamura T, Villar-Briones A, Yanagida M: **Metabolic profiling of the fission yeast S. pombe: quantification of compounds under different temperatures and genetic perturbation.** *Mol Biosyst* 2010, **6**:182-198.

42. Danielsson R, Bylund D, Markides KE: **Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography-mass spectrometry.** *Analytica Chimica Acta* 2002, **454(2)**:167-184.