User guide for MetFamily 1.0

Hendrik Treutler

March 2016

Abstract

The MetFamily web application is designed for the identification of regulated metabolite families from untargeted LC-MS/MS metabolomics experiments. This is possible on the basis of metabolite profiles for a set of MS^1 features and the corresponding MS/MS spectra. Group-discriminating MS^1 features are identified using a principal component analysis (PCA) of MS^1 metabolite profiles and metabolite families are identified using a hierarchical cluster analysis (HCA) of MS/MS spectra. Regulated metabolite families are identified by considering group-discriminating MS^1 features from corporate metabolite families. The MetFamily web application is available at `http://msbi.ipb-halle.de/MetFamily/`.

1 General design

1.1 Approach

The data basis of MetFamily is a set of MS^1 mass / retention time features whose abundance was measured in a set of samples. Each MS^1 feature is associated with a CID-MS/MS spectrum. The software tool MS-DIAL is recommended to be used as a pre-analysis step [2]. MS-DIAL accepts several vendor-own raw data formats for MS^1 and MS/MS data. It performs peak picking from untargeted measurements, spectra deconvolution of LC-MS/MS data and associates precursor ions with MS/MS spectra from data independent CID-fragmentation. Preferably, MetFamily accepts metabolite profiles and MS/MS library outputs from MS-DIAL as data input.

In MetFamily, the user can analyse this data using two orthogonal approaches. First, the user is able to analyse the MS^1 abundances of different samples using principal component analysis (PCA) in order to identify group-discriminating MS^1 features. Second, the user is able to analyse the MS/MS spectra using a hierarchical cluster analysis (HCA) in order to group and annotate sets of MS^1 features which correspond to metabolite families. The iterative examination of both analyses aims at the identification of regulated metabolite families.

1.2 GUI design

The graphical user interface (GUI) of MetFamily is structured as follows. There is a tab 'Run' as well as a tab 'About'. The tab 'run' comprises the main functionality of MetFamily and the tab 'About' comprises information regarding the authors and a reference to the publication describing the methodology of MetFamily. The tab 'Run' is divided into a sidebar panel on the left for parameter settings and an analysis panel on the right for plots and annotation functionality.

In the sidebar panel, the user is able to feed input into MetFamily, apply global filters for MS/MS fragments, perform PCA and HCA, search (after PCA or HCA were performed) for MS^1 features or MS/MS fragments, and export different results. The analysis panel displays plots for HCA and PCA. In the lower section the user can annotate MS^1 features based on MS/MS information.

1.3 Additional filters specific to PCA and HCA

The user is able to filter sets of MS^1 features based on a threshold for the average MS^1 abundance in all samples. Additionally, a threshold for the $\log_2$-fold change of average MS^1 abundances in different sample groups can be set. This allows the user to specifically focus on MS^1 features which a predominantly present in a specific sample group. This filtering can be done at the HCA tab to perform a HCA on a subset of MS^1 features and at the PCA tab to perform a PCA on a subset of MS^1 features.

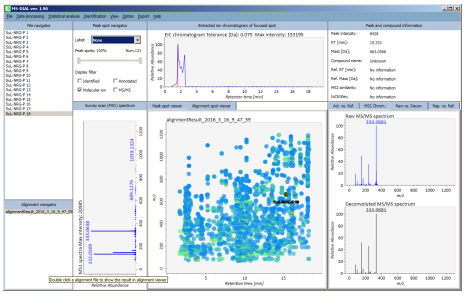
1.4 Selections

Once PCA or HCA have been performed, the user is able to select a set of MS^1 features using three different selection modes. A selected set of MS^1 features can be annotated in order to assign this set of MS^1 features to functional or structural annotations. Annotated MS^1 features are marked in HCA and PCA with a user-defined color and the iterative assignment of MS^1 features to annotations is a prerequisite for the identification of regulated metabolite families.

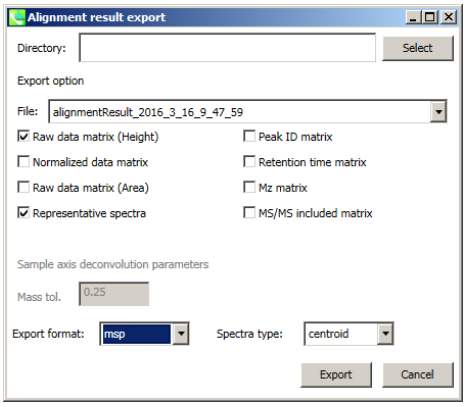
2 Input tab

The user can select between three different options for data input into MetFamily (see Figure 2).

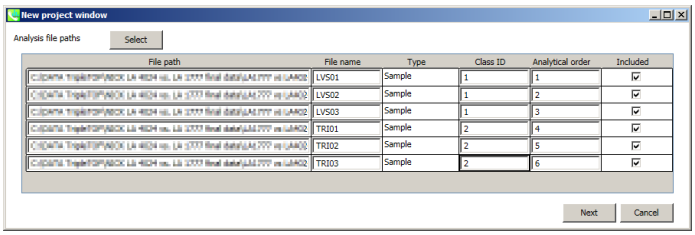
Using the option 'Import data', the user can import external data consisting of two files. The first file is a text document which comprises the MS^1 abundances of all MS^1 features and properties such as the retention time, m/z, and the ion species. The second file is a MS/MS library in the NIST MSP format. Please find an example for both files downloadable via the MetFamily web app in the tab 'Input' using option 'Example data'. The data import can be customized by a set of parameters as shown in table 1. In addition, the user is able to name the generated MetFamily project and add some comments as free text.



(a) Screenshot of the MS-DIAL program.



(b) Screenshot of the export panel.



(c) Screenshot of a import panel.

Figure 1: **The MS-DIAL program.** Figure 1a) The MS-DIAL program supports feature detection, ion species annotation, compound spectra extraction, and peak alignment between samples. Figure 1b) The export panel of the MS-DIAL program supports the export of the metabolite profile ('Raw data matrix (Height)') and a spectral library ('Representative spectra'). Figure 1c) An import panel of the MS-DIAL program supports the specification of, amongst others, Class ID to assign samples to sample groups.

The metabolite profile and the MS/MS library can be generated with MS-DIAL using the Export tab (see Figure 1b). Please note that the property 'Class ID' has to be specified for each sample during the import of data into MS-DIAL (see Figure 1c). Alternatively to SWATH data processed with MS-DIAL, MetFamily can be used with data from various other sources such as GC-MS, isMS/MS, and DDA. Please find the guide 'MetFamily Input Specification' enclosed with the MetFamily paper and downloadable from the MetFamily tool [1].

Using the option 'Load project', the user is able to load a project file. Project files can be generated and downloaded from MetFamily at different points. Project files are gzip-archives comprising a single spreadsheet with all data including MS[1] features and the fragment matrix. It is possible to uncompress

[1] `http://msbi.ipb-halle.de/MetFamily/`

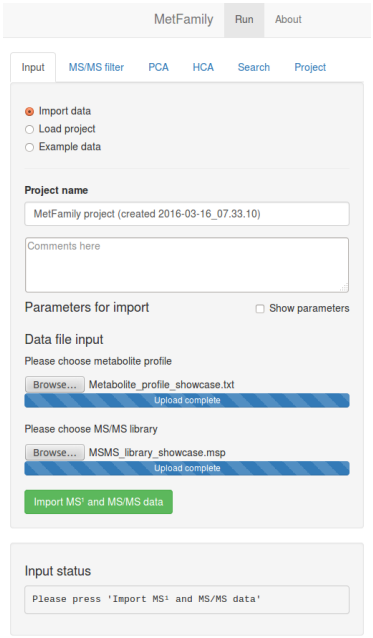


Figure 2: **MetFamily Input tab.** The user is able to import external data using the option 'Import data', to load an existing MetFamily project using the option 'Load project', and to download or load the example data from the MetFamily publication using the option 'Example data'.

project files and e.g. open and examine the data in a spreadsheet program.

Using the option 'Example data', the user is able to download and load example data which was published and analysed in the MetFamily publication [1]. Using this data, the user is able to reproduce the showcase presented there and to get a first impression of MetFamily without having to supply own data.

3 MS/MS filter tab

The user can filter the set of MS^1 features by a set of MS/MS fragments (see Figure 3). Subsequent filter operations in PCA and HCA are performed on the basis of this filtered set. This function can be used to narrow down the set of MS^1 features prior to further PCA or HCA analyses. This option is useful if the user intends to constrain the data analysis to certain metabolite families.

The user is able to filter using three lists of m/z values. The m/z values within each list are separated by comma or semicolon. Different values within each list are conjunct by 'and' and different lists are conjunct by 'or'. Positive entries refer to fragment ions. Negative entries refer to neutral losses. Thus, variable combinations for fragment ions and/or neutral losses to be filtered can be made. In addition, the user is able to specify the allowed m/z error for these MS/MS features in PPM (**p**arts **p**er **m**illion).

For assistance, the user is able to look up the m/z value of certain fragments or neutral losses in the 'Fragment overview' plot by checking 'Show frequent fragments'. Here, we plot on the abscissa the MS/MS feature m/z of fragments

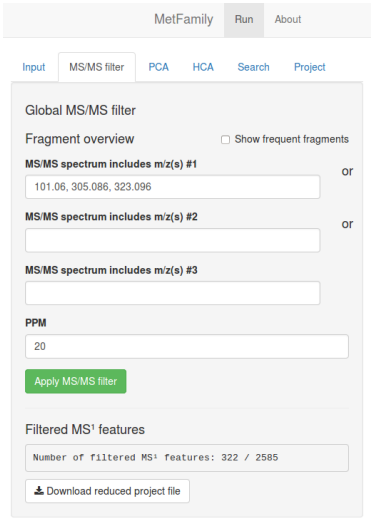


Figure 3: **MetFamily MS/MS filter tab.** The user is able to filter the full set of MS^1 features using three lists of fragment m/z values for a semi-targeted analysis approach. Subsequent PCA and HCA analyses will be performed on the filtered set of MS^1 features.

and neutral losses which are present in at least five MS/MS spectra and we plot on the ordinate the number of MS/MS spectra which comprise the MS/MS features. In this way, the user is able to look up abundant fragments and accurate m/z values for MS/MS features of interest.

4 PCA tab

The user can perform a principal component analysis (PCA) of MS^1 features on the basis of MS^1 abundances to identify group-discriminating sets of MS^1 features (see Figure 4).

In section 'MS^1 abundance filter for PCA', the user can apply filters on the basis of the MS^1 signal abundance and the fold change of the average MS^1 abundance between two selected sample groups. Here, the user is able to select a set of replicate groups using the check boxes of the parameter 'Groups'. To specify a threshold for the average abundance of the MS^1 features in the selected replicate groups in the field 'Average MS^1 abundance'. A threshold for the $\log_2$-fold change of the average MS^1 abundance between the selected replicate groups is set using the parameter 'MS^1 log2-fold change' (only in case of exactly two selected replicate groups). This parameter should only be used if group-specific PCA loading shall be analysed. Furthermore, it is possible to include or exclude MS^1 features which have been annotated with annotation 'Ignore' (see section MS^1 feature annotation) by checking/ unchecking the box 'Include ignored MS^1 features'. In section 'Filtered MS^1 features', the number of MS^1 features which meet the specified filter criteria is shown and a project file reduced to this set of MS^1 features can be downloaded.

In section 'PCA properties', the scaling and $\log_2$ transformation of MS^1 abundances can be selected as advanced options. See table 2 for the supported

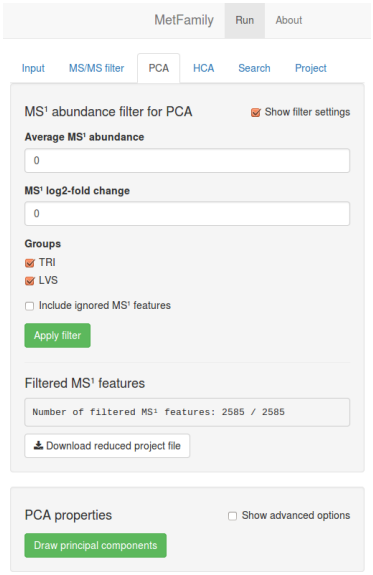


Figure 4: **MetFamily PCA tab.** The user is able to filter the set of MS^1 features (possible pre-filtered in tab 'MS/MS filter tab' as described in section MS/MS filter tab) and perform a PCA using different scaling functions.

scaling functions. In addition, the user is able to select two principal components which are plotted subsequently. After clicking the button 'Draw principal components', the scores and loadings of the PCA are plotted on the right side of the screen. For further details on the analysis of the scores and the loadings please see section PCA scores and PCA loadings.

The scores plot displays the set of samples in MS^1 two-dimensionally for two selected principal components. The loadings plot shows the same principal components as the scores plot and displays the set of MS^1 features projected to these principal components. Group-discriminating MS^1 features are displayed off the point of origin and usually these are considered as MS^1 features of interest.

5 HCA tab

The user is able to perform hierarchical cluster analysis (HCA) of MS^1 features on the basis of the spectral similarity in the corresponding MS/MS spectra in order to identify and annotate sets of MS^1 features as metabolite families (see Figure 5).

As described for the PCA tab, in section 'MS^1 abundance filter for HCA' the user is able to set filters for the MS^1 abundance (precursor ion abundance) and fold change of the MS^1 abundance between selected sample groups. Here, the user can select two replicate groups using the radio buttons of the parameters 'Group 1' and 'Group 2'. As for the PCA tab ignored MS^1 features can be included or excluded using the checkbox 'Include ignored MS^1 features'. Please note, that these filters can be set for PCA and HCA independently. In section 'Filtered MS^1 features', the number of MS^1 features that meet the filter criteria

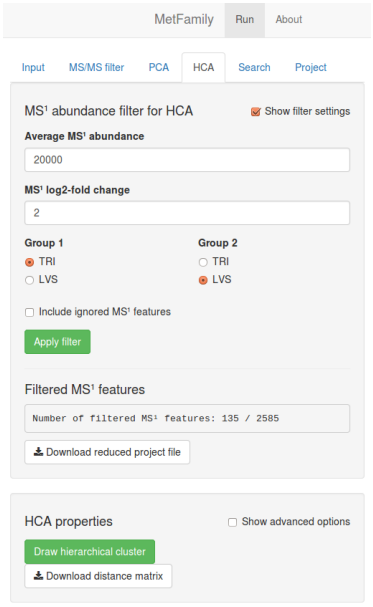


Figure 5: **MetFamily HCA tab.** The user is able to filter the set of MS^1 features (possible pre-filtered in tab 'MS/MS filter tab' as described in section MS/MS filter tab) and perform a HCA using different distance functions.

is shown and a project file reduced to this set of MS^1 features can be downloaded.

In section "HCA properties", the user is able to choose between different distance functions for the computation of MS/MS spectral similarities as well as the cluster method for the assembly of the hierarchical cluster dendrogram as advanced options. See table 3 for the supported distance functions. Click the button 'Draw hierarchical cluster' to compute and display the hierarchical cluster dendrogram. For further details about the analysis of HCA dendrograms please see section Hierarchical cluster dendrogram.

The generated dendrogram contains the set of filtered MS^1 features specified in the HCA tab as leaves and each branch represents a cluster of MS^1 features. The spectral similarity amongst the hierarchical clusters is inversely proportional to the height of the corresponding branch, i.e. the shorter the branch, the more similar are the connected clusters. The cluster-discriminating power of each cluster is visualized by the size of the node at each branch (see legend 'Cluster-discriminating power'). Each branch is labeled with the number of characteristic fragments which are present in more than 50% of comprised MS/MS spectra (see option 'Show labels' in section 'HCA dendrogram' on the right). Branches and leaves can be selected using three selections modes as described in section MS^1 feature selection modes (see legend 'Selection marks'). The m/z and retention time of each MS^1 feature is displayed below the leaves. In addition, the average MS^1 abundance of the MS^1 features in both selected groups is colour-coded in a heatmap below as well as the color-coded $\log_2$-fold change (abbreviated LFC) between the two selected groups (see legends 'log2(MS^1 fold change)' and 'log10(MS^1 abundance)' on the right).

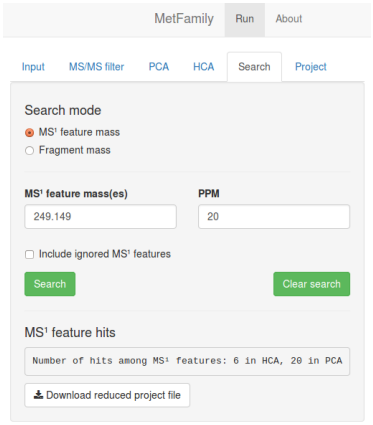


Figure 6: **MetFamily Search tab.** The user is able to search for MS^1 features in PCA and HCA using MS^1 feature m/z or lists of fragments m/z values.

6 Search tab

Post analysis, the user is able to select a set of MS^1 features in the PCA as well as in the HCA. Here, there are two options for the selection of MS^1 features (see Figure 6). First, the user can select MS^1 features using the 'MS^1 feature m/z' and the allowed m/z error in PPM (**p**arts **p**er **m**illion). Second, the user is able to select MS^1 features based on a set of MS/MS fragments or neutral losses in analogy to the global MS/MS filter described in section MS/MS filter. In contrast to these filters, the "Search" tab allows to highlight a selection in the previously processed data, while the former is a global filter, which will reduce the data set before analysis. MS^1 feature hits are marked in HCA and PCA by red circles and can be examined and annotated in selection mode 'Selection by search' as described in section MS^1 feature annotation.

7 Project tab

In the 'Project' tab, the user is able to view and edit properties of the MetFamily project (see Figure 7). In section 'MetFamily project', the name of the MetFamily project is shown and the user is able to view and edit the project description. Please note that an edited project description does only apply with a click on button 'Update project desciption'. In section 'Data import parameters', the user is able to view the set of import parameters which have been used for the initial data import in the course of the MetFamily project creation.

In section 'Export', the user is able to download different exports from MetFamily. First, the user can export the full project as gzipped project file. This file can be shared with other researchers and used later as input for MetFamily. Second, the user is able to export a parameter file containing the set of parameters which have been used for the initial data import. This parameter file can be used for data import of a new project or shared with other researchers to facilitate reproducible research. Third, the user is able to export the PCA and

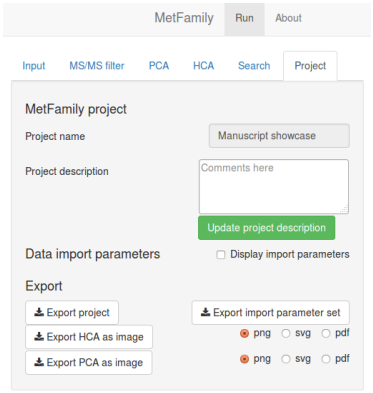


Figure 7: **MetFamily Project tab.** The user is able to examine and edit different properties of the present MetFamily project. In addition, the user is able to download different results.

HCA plots in the current state as Portable Network Graphics (*.png) image file (resolution: 600 DPI), Scalable Vector Graphics (*.svg) image file, or Portable Document Format (*.pdf) file.

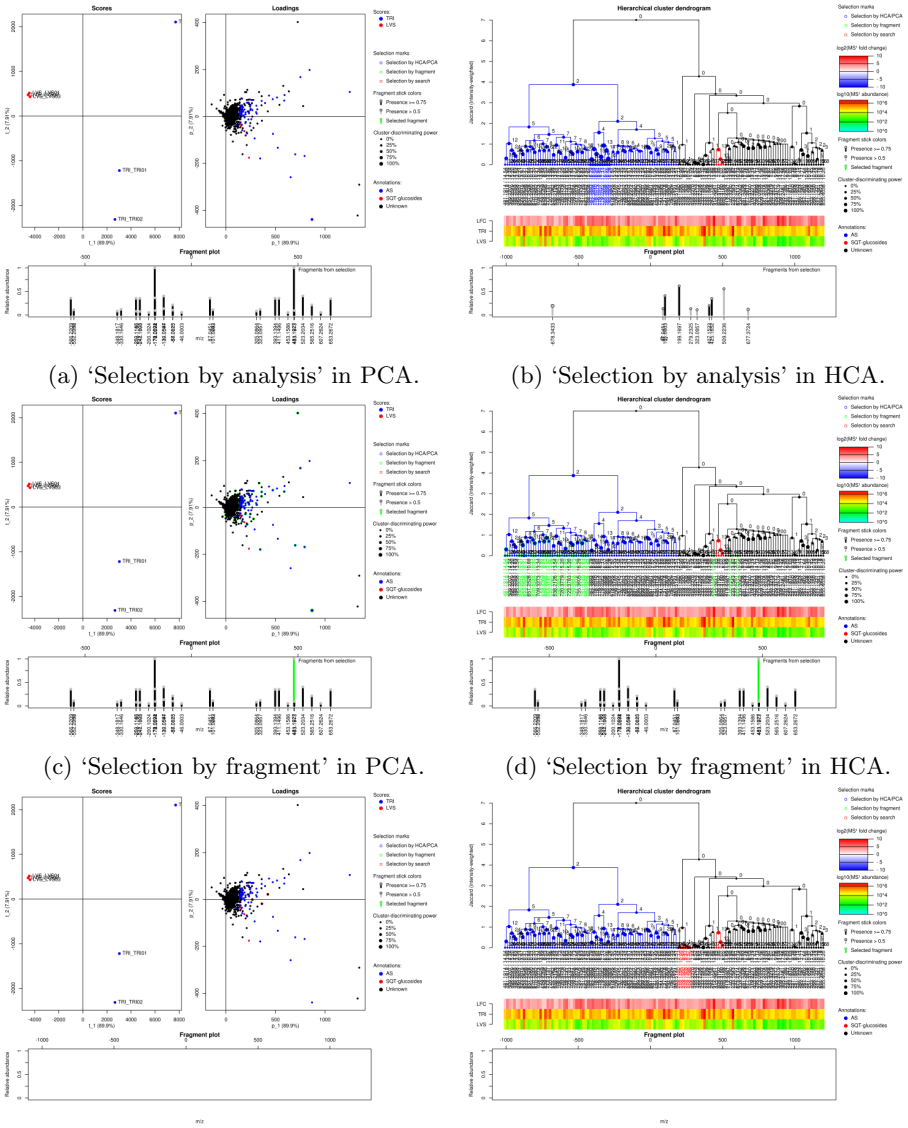
8 MS^1 feature selection modes

Post PCA or HCA analysis, the user is able to select sets of MS^1 features in three different selection modes (see Figure 8). Selected MS^1 features are marked in the hierarchical cluster dendrogram and the PCA loadings plot by mode-specific colors (see legend 'Selection marks' on the upper right). The three selection modes are as follows.

The selection mode 'Selection by analysis' enables the user to select a set of MS^1 features in the hierarchical cluster dendrogram or in the PCA loadings plot. In the hierarchical cluster dendrogram, the user can select a single MS^1 feature by selecting a dendrogram leaf and the user can select a set of MS^1 features by selecting a dendrogram cluster (see section Hierarchical cluster dendrogram). In the PCA loadings plot, the user can select a single MS^1 feature by selecting a loading (see section PCA scores and PCA loadings). The selection color for this selection mode is blue.

The selection mode 'Selection by fragment' enables the user to select a set of MS^1 features by selecting a fragment in the fragment plot below the hierarchical cluster dendrogram or the PCA plots respectively (see section Fragment plot). All MS^1 features in the hierarchical cluster dendrogram and in the PCA loadings plot, which MS/MS spectrum comprises the selected fragment, are selected. The selection color for this selection mode is green.

The selection mode 'Selection by search' enables the user to select a set of MS^1 features using a m/z mass in MS^1 or using a set of fragment m/z values (see section Search tab).



(a) ‘Selection by analysis’ in PCA.

(b) ‘Selection by analysis’ in HCA.

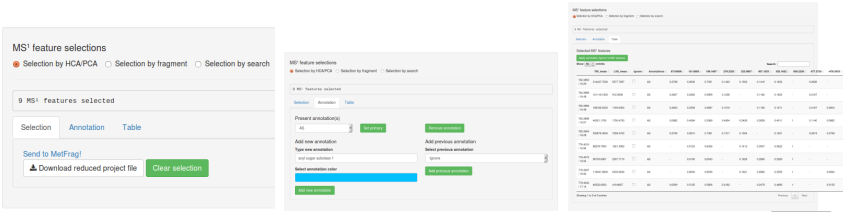
(c) ‘Selection by fragment’ in PCA.

(d) ‘Selection by fragment’ in HCA.

(e) ‘Selection by search’ in PCA.

(f) ‘Selection by search’ in HCA.

Figure 8: **Selected sets of MS^1 features are visualized both in PCA and HCA.** Figures 8a,8b) The selection mode ‘Selection by analysis’ is visualized in blue. Figures 8c,8d) The selection mode ‘Selection by fragment’ is visualized in green. Figures 8e,8f) The selection mode ‘Selection by search’ is visualized in red.



(a) Tab 'Selection'. (b) Tab 'Annotation'. (c) Tab 'Table'.

Figure 9: **The panel 'MS^1 feature selections' provides options for selected sets of MS^1 features**. Figure 9a) The tab 'Selection' enables the download of the selected set of MS^1 features. Figure 9b) The tab 'Annotation' enables the annotation of the selected set of MS^1 features. Figure 9c) The tab 'Table' enables the examination of the selected set of MS^1 features.

9 MS^1 feature annotation

For the purpose of examining, extracting, and annotating selected MS^1 features, we created the panel 'MS^1 feature selections' in the lower part of the PCA and HCA plots, respectively (see Figure 9). This panel consists of three tabs as follows.

In the tab 'Selection', the user is able to download a project file reduced to the selected set of MS^1 features as well as to clear the selection. In case of exactly one selected MS^1 feature, the user is able to send the selected MS^1 feature to MetFrag for identification.

In the tab 'Annotation', the user is able to add and remove annotations for the selected set of MS^1 features. In section 'Present annotation(s)', the user can examine and remove present annotations. In case of multiple present annotations, the user is also able to define one of these annotations as 'primary' which means that this annotation will be used preferentially for coloring the cluster dendrogram and the PCA loadings (see legend 'Annotations' for a summary of the displayed annotations and the corresponding colors). In section 'Add new annotation', the user is able to add an annotation the selected set of MS^1 features by typing the name and choosing a color for the new annotation. In section 'Add previous annotation', the user can assign an annotation which is already assigned to other MS^1 features.

In the tab 'Table', the user is able to examine m/z, retention time, MS^1 abundance, present annotations, and common fragments of the selected set of MS^1 features. In addition, the user can annotate specific MS^1 features with the annotation 'Ignore' to mark artifacts. MS^1 features with the annotation 'Ignore' can be excluded in subsequent PCA and HCA analyses.

10 Hierarchical cluster dendrogram

The hierarchical cluster dendrogram represents a set of MS^1 features as leaves which are clustered hierarchically according to the pairwise similarity of the corresponding MS/MS spectra (see Figure 10). Both leaves and clusters are marked by nodes. The size of the cluster nodes represents the maximal cluster-discriminating power of all characteristic fragments of this cluster. The cluster-discriminating power is a measure for a fragment to which degree this fragment is specific to the cluster (see MetFamily publication for details [1]).

The user can zoom in by dragging the left mouse button horizontally and double-clicking the selected region. A double-click with the left mouse-button resets the zoomed state to the full range. The user can select a leaf or cluster using a click with the left mouse-button on the corresponding node. Selected **p**arts **p**er **m**illion of the hierarchical cluster dendrogram are colored blue and can be annotated via the selection mode 'Selection by analysis' as described in section MS^1 feature annotation. When a leaf is selected, the corresponding MS/MS spectrum is displayed in the Fragment plot. In case of a cluster selection, a set of fragments which are frequently present in the MS/MS spectra of the corresponding MS^1 features denoted characteristic fragments is displayed in the fragment plot (see section Fragment plot). Specifically, characteristic fragments which are present in at least 75% of the corresponding MS^1 features are coloured black and fragments which are present in more than 50% of the corresponding MS^1 features are colored grey (see legend 'Fragment stick colors' in the right). In addition, the user is able to hover leaves and cluster nodes with the mouse to compare the corresponding fragments in 'head to tail' manner to the fragments which are already present in the fragment plot from the previous selection.

11 PCA scores and PCA loadings

The PCA scores plot and the PCA loadings plot displayes two principal components of the scores and the loadings of the PCA (see Figure 11). The scores represent the individual samples in the calculated coordinate system and the loadings represent the individual MS^1 features in the calculated coordinate system.

The user can zoom in by dragging the left mouse button and double-clicking the selected region. A double-click with the left mouse-button resets the zoomed state to the full range. The user can select a loading in the PCA loadings plot by clicking with the left mouse button on the corresponding node. Selected loadings are marked in blue and can be annotated in selection mode 'Selection by analysis' (see section MS^1 feature annotation). The MS/MS spectrum of the selected MS^1 feature is displayed in the fragment plot (see section Fragment plot). In addition, the user is able to hover loadings to compare the corresponding MS/MS spectra in the fragment plot to the fragments of the MS/MS spectrum which was selected previously.

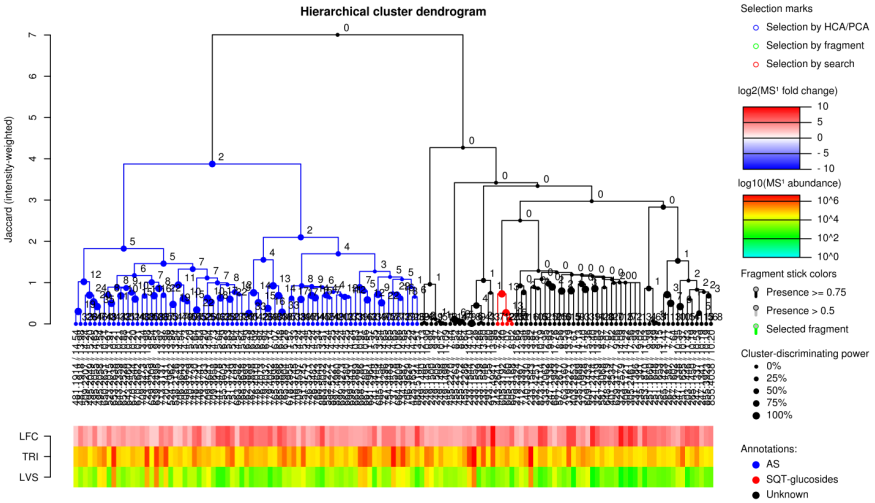


Figure 10: **The Hierarchical cluster dendrogram.** The HCA dendrogram represents a hierarchy of putative metabolite families on different levels of granularity. Leaves represent MS^1 features and are labeled with the number of fragments in the corresponding MS/MS spectrum. Branches represent clusters of MS^1 features and are labeled with the number of fragments which are common in the MS/MS spectra of at least 75% of the MS^1 features in the cluster. See legends on the right.

12 Fragment plot

The fragment plot displays the set of MS/MS features of a selected MS/MS spectrum or the characteristic fragments of a cluster of MS^1 features, where the fragment m/z is plotted on the abscissa and the fragment intensity is plotted on the ordinate (see Figure 12). Additionally, it is possible to compare a spectrum from a selected dendrogram node or a selected PCA loadings node with a second spectrum from a dendrogram node or a PCA loadings node with a mouse-hover. Each fragment has a selectable node. In case of characteristic fragments from a cluster of MS^1 features, the node size represents the cluster-discriminating power, which is a measure to which degree the fragment is specific to the cluster (see MetFamily publication for details [1]).

The user can zoom in by dragging the left mouse button horizontally and double-clicking the selected region. A double-click with the left mouse-button resets the zoomed state to the full range. The user is able to select a fragment using a click with the left mouse-button on the corresponding node. Selected fragments are colored green (see legend 'Fragment stick colors' on the upper right) and all MS^1 features in the HCA and PCA loadings, which MS/MS spectrum comprises the selected fragment are selected and colored green as well. The user is able to annotate the selected MS^1 features in selection mode

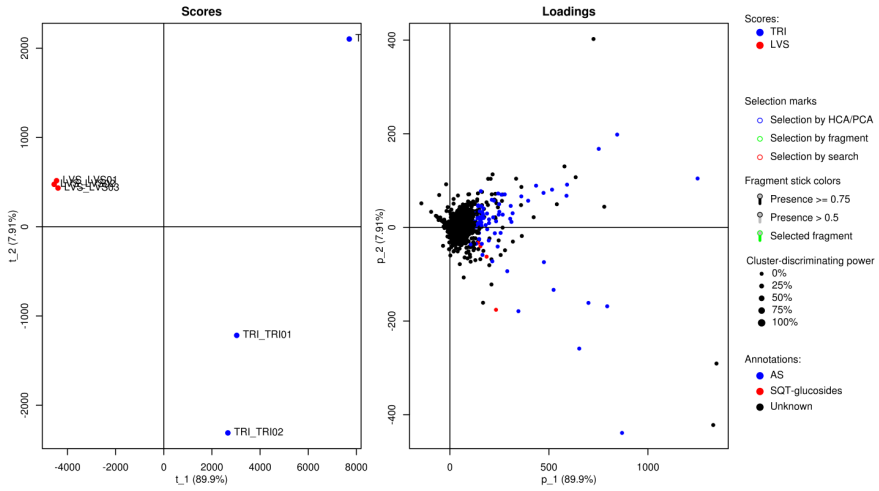


Figure 11: **The PCA scores plot and the PCA loadings plot.** The PCA scores plot on the left represents the set of samples and the PCA loadings plot right of the PCA scores plot represents the set of MS[1] features. Here, both plots display the first two principal components. See legends on the right.

'Selection by fragment' as described in section MS[1] feature annotation.

Parameter	**Description**
	Fragment filter
Min. spectrum intensity	A MS/MS spectrum is considered if the MS/MS feature intensity is greater or equal than this value
Fragment proportion	A MS/MS feature is considered if its intensity relative to the base peak in the associated CID spectrum is greater or equal than this value
	Neutral losses
Fragment vs. precursor	Include neutral losses relative to the precursor ion, i.e. the m/z difference between the m/z of the precursor ion and the m/z of each fragment ion of the corresponding MS/MS spectrum
Fragment vs. fragment	Include neutral losses amongst fragment ions, i.e. the m/z difference between the m/z of all pairs of fragment ions within each MS/MS spectrum; this involves the incorporation of potentially nonexistent neutral losses and needs more time for processing
	Fragment grouping
m/z deviation (abs.)	A MS/MS feature is added to a fragment group if the absolute m/z difference is smaller or equal than this value
m/z deviation (PPM)	A MS/MS feature is added to a fragment group if the absolute m/z difference is smaller or equal than the m/z times this value divided by 1,000,000 (**p**arts **p**er **m**illion)
	Precursor deisotoping
m/z deviation (abs.)	A MS^1 feature is considered a +1 isotopic peak if the absolute of the m/z difference to the (putative) monoisotopic peak minus 1.0033548378 ($=^{13}C$ - ^{12}C) is smaller or equal than this value (analog for the +2 isotopic peak)
m/z deviation (PPM)	A MS^1 feature is considered a +1 isotopic peak if the absolute of the m/z difference to the (putative) monoisotopic peak minus 1.0033548378 ($=^{13}C$ - ^{12}C) is smaller or equal than the m/z times this value divided by 1,000,000 (**p**arts **p**er **m**illion, analog for the +2 isotopic peak)
Retention time difference	A MS^1 feature is considered an isotopic peak if the absolute of the retention time difference to the (putative) monoisotopic peak is smaller or equal than this value (in minutes)
	Fragment deisotoping
m/z deviation (abs.)	A MS/MS feature is considered a +1 isotopic peak if the absolute of the m/z difference to the (putative) monoisotopic peak minus 1.0033548378 ($=^{13}C$ - ^{12}C) is smaller or equal than this value
m/z deviation (PPM)	A MS/MS feature is considered a +1 isotopic peak if the absolute of the m/z difference to the (putative) monoisotopic peak minus 1.0033548378 ($=^{13}C$ - ^{12}C) is smaller or equal than the m/z times this value divided by 1,000,000 (**p**arts **p**er **m**illion)

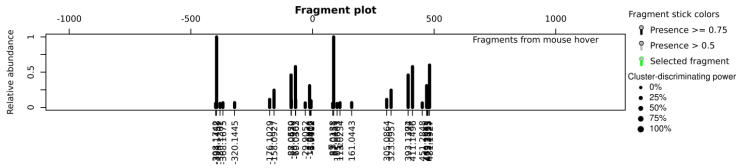
Table 1: Overview of parameters for the import of external data into MetFamily

Scaling function	Description
None	No data scaling
Mean center	Data minus the mean*
Autoscaling (unit variance)	Data minus the mean* is divided by the standard deviation**
Pareto	Data minus the mean* is divided by the square root of the standard deviation**

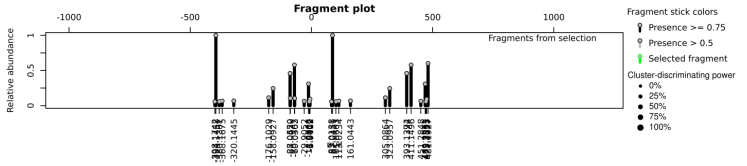
Table 2: Overview of scaling functions for the computation of the PCA. * The mean of the selected sample groups ** The standard deviation of the selected sample groups

Distance function	Description
Jaccard	The number of intersecting fragments divided by the number of distinct fragments
Jaccard (intensity-weighted)	The number of intersecting fragments divided by the number of distinct fragments weighted by fragment intensity
Jaccard (fragment-count-weighted)	The number of intersecting fragments divided by the number of distinct fragments weighted by fragment frequency in all MS/MS spectra
NDP (Normalized dot product)	Spectral distance measure as used by Lia et al

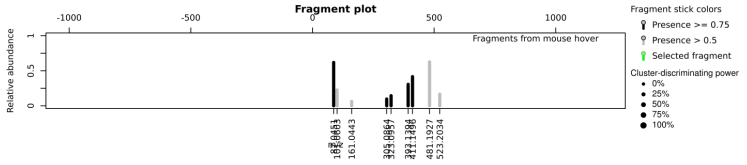
Table 3: Overview of distance functions for hierarchical clustering of MS/MS spectra.



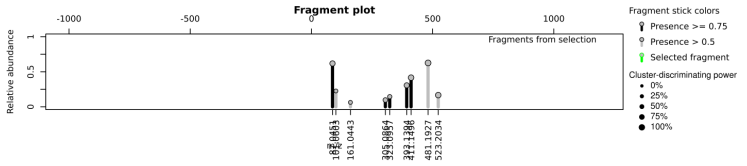
(a) MS/MS spectrum by mouse hover of dendrogram leaf or PCA loading.



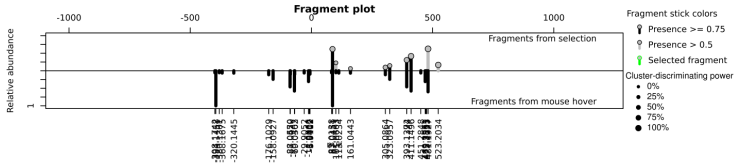
(b) MS/MS spectrum by selection of dendrogram leaf or PCA loading.



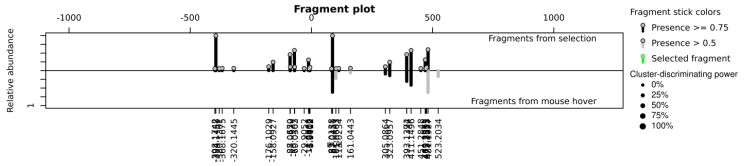
(c) Characteristic fragments by mouse hover of a dendrogram cluster.



(d) Characteristic fragments by selection of a dendrogram cluster.



(e) Characteristic fragments by selection of a dendrogram cluster versus MS/MS spectrum by mouse hover of a dendrogram leaf or PCA loading.



(f) MS/MS spectrum by selection of a dendrogram leaf or PCA loading versus characteristic fragments by mouse hover of a dendrogram cluster.

Figure 12: **The fragment plot.** Figure 12a,12b) MS/MS spectra of MS^1 features are displayed with a stick diagram representing the fragment m/z and fragment intensity normalized to one. Figure 12c,12d) Characteristic fragments corresponding to clusters of MS^1 features are displayed with a stick diagram representing the fragment m/z and mean fragment intensity. Figure 12e,12f) It is possible to display two spectra in 'head to tail' manner.

References

[1] Hendrik Treutler, Hiroshi Tsugawa, Andrea Porzel, Alain Tissier, Steffen Neumann, and Gerd Balcke, *Discovering regulated metabolite families in untargeted metabolomics studies*, Submitted (2016).

[2] Hiroshi Tsugawa, Tomas Cajka, Tobias Kind, Yan Ma, Brendan Higgins, Kazutaka Ikeda, Mitsuhiro Kanazawa, Jean VanderGheynst, Oliver Fiehn, and Masanori Arita, *MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis*, Nat Meth **12** (2015), no. 6, 523–526.